Asymptotic Learning with Ambiguous Information*

Pëllumb Reshidi † João Thereze ‡ Mu Zhang §

October 30, 2024

Abstract

We study asymptotic learning when the decision maker faces ambiguity in the precision of her information sources. She aims to estimate a state and evaluates outcomes according to the worst-case scenario. Under prior-by-prior updating, we characterize the set of asymptotic posteriors the decision maker entertains, which consists of a continuum of degenerate distributions over an interval. Moreover, her asymptotic estimate of the state is generically incorrect. We show that even a small amount of ambiguity may lead to large estimation errors and illustrate how an econometrician who learns from observing others' actions may over- or underreact to information.

*We thank Roland Benabou, Simon Board, Sylvain Chassang, Alessandro Lizzeri, Pietro Ortoleva, Wolfgang Pesendorfer, Can Urgun, and Leeat Yariv for very helpful suggestions and feedback. We also thank audiences at Princeton University, BEAT Conference (Tsinghua University), and The 33rd Stony Brook International Conference on Game Theory, for their helpful comments and discussions. An earlier version of this paper was circulated under the title "Information Aggregation under Ambiguity."

†Department of Economics, Duke University, Durham, NC — pellumb.reshidi@duke.edu

‡Fuqua School of Business, Duke University, Durham, NC — joao.thereze@duke.edu

§Department of Economics, University of Michigan, Ann Arbor, MI — muzhang@umich.edu

1 Introduction

Consider agents who rely on multiple information sources to learn about a payoff-relevant state. A voter turns to poll results and advertising to gauge a politician's competence and agenda, while an investor leverages the reports of various analysts to project the future returns of a stock. A common assumption in the literature is that the decision maker (DM) has beliefs about the quality of her information sources and that these beliefs are correctly specified. In such cases, asymptotic learning is successful. Often, however, forming beliefs is not straightforward. For example, a customer consulting online reviews for a product may not have particular beliefs about the quality of reviewers because they are being consulted for the first time. In such settings, little is known about learning. We aim to close this gap.

We analyze asymptotic learning when the DM lacks particular beliefs about her information sources. In our baseline model, she observes unbiased signals and aims to estimate a state by minimizing a loss function. The state and the signals are jointly normally distributed, but the DM does not know the signals' precisions—the inverse of their variances. The DM is not probabilistically sophisticated; instead, she faces ambiguity in the precision of each information source and perceives them to lie in a bounded interval. Each assignment of precisions to information sources pins down a belief of the DM, a joint distribution over signals and the state. Thus, an interval of perceived precisions induces a set of beliefs. We assume the DM updates her beliefs prior by prior. Concretely, upon observing information, she updates each belief in her belief set according to Bayes' rule. In doing so, the DM obtains multiple posterior distributions for the state. Thus, ambiguity about precisions induces ambiguity about the state. Finally, she takes a robust approach and evaluates the expected loss according to the worst case across all posteriors.

Our first result characterizes the DM's set of posteriors as the number of signals grows large. We show that the induced ambiguity concerning the state does not vanish asymptotically. That is, the posterior beliefs of the DM do not converge to a single distribution as the number of information sources grows. As in standard Bayesian learning, the variance of each posterior converges to zero. However, different beliefs about precisions lead to different weighting of signals and, consequently, to different posterior means. For example, for any realization of signals, the DM's belief set contains a belief that assigns higher precisions to signals with high realizations and lower precisions otherwise. In this case, the posterior mean converges to a relatively high value. Similarly, there exist beliefs that lead to a relatively low posterior mean. Applying this process to all beliefs over precisions generates an interval of posterior means. We show that the set of asymptotic posteriors is the set of Dirac measures over values in that interval. Importantly, this set is independent of the objective of the DM and includes a degenerate belief over the true state. We then extend this characterization of the limit belief set to settings in which unbiased signals are not directly observable, but the DM has access to certain functions of the signals. This setup encompasses a range of environments in which the DM monitors the actions of other agents.

Our second result characterizes the DM's asymptotic estimate of the state. Her decision problem can be interpreted as a zero-sum game against nature. Initially, the DM receives information and chooses the estimate that minimizes her expected loss. Subsequently, nature chooses the precision of each source to maximize the DM's loss. In doing so, nature affects the DM's posterior distribution. We show that, asymptotically, this is equivalent to nature choosing posterior means in the interval described in the previous paragraph. If the DM chooses a relatively low estimate within the interval, nature will maximize her loss by selecting the highest value possible, and vice versa. To minimize the maximal loss, the DM's estimate makes nature indifferent between choosing the highest or lowest value from the interval of asymptotic posterior means. We show that, in our setting, asymptotic learning typically fails. That is, the DM's estimate is not consistent. To the best of our knowledge, this is the first paper to show how ambiguity aversion disrupts classical inference from large samples.

These results have several implications. First, we show that when the true signal precisions are relatively low, even a small amount of ambiguity can lead to arbitrarily large losses and estimation errors. Therefore, the magnitude of ambiguity and estimation error need not be proportional—a minor degree of ambiguity can have a major impact on the quality of asymptotic learning.

Second, we show the DM can be worse off even if she perceives all her information sources to be more informative. Consider two decision problems, a and b, in which the DM directly observes unbiased signals but has different intervals of perceived precision. We show that even if the lowest precision in decision problem a is higher than the highest precision in decision problem b, the DM may be better off in decision problem b. To carry out this comparison, we study how the initial ambiguity about precisions maps into induced ambiguity about the state. In particular, we show that the asymptotic interval of posterior means is determined by the ratio between the highest and the lowest possible perceived precisions. Because the length of this interval pins down the DM's loss, her welfare is monotonic in that ratio, regardless of the level of perceived precision.

Lastly, we analyze the problem of an ambiguity-averse econometrician who observes choices made by Bayesian agents. The agents estimate a payoff-relevant state given their private information. The econometrician aims to estimate the same state but does not know the precision of the private signals observed by the Bayesian agents. For example, consider a healthcare official assessing the prevalence of a disease in a region. She relies on hospital reports to do so but is not sure about the quality of their data collection protocols. Generically, the econometrician fails to aggregate information. We characterize how she may over- or underreact to the information contained in the observed actions as a function of her initial belief and the true precisions of signals observed by the Bayesian agents.

Related Literature Our paper studies learning under ambiguity, in which the DM follows the maxmin expected utility model (Gilboa and Schmeidler, 1989) and prior-by-prior updating (full Bayesian updating).[1] Under the same assumptions, Epstein and Schneider (2008) studies a financial market where the representative agent observes one signal with ambiguous precision. They show how this ambiguity affects reactions to information and the asset price. Follow-up papers extend these results to other environments (Illeditsch, 2011; Gollier, 2011; Condie and Ganguli, 2017). We consider a similar setup as Epstein and Schneider (2008) but focus on asymptotic learning and consistency of the DM's estimate.[2] Another closely related paper is Chen (2023), which shows that under prior-by-prior updating, ambiguous signals generically lead to herding behavior and information cascades in a social learning setting. By comparison, we focus on the asymptotic learning of a single DM. The two papers are complementary in facilitating our understanding of the implications of prior-by-prior updating on learning.

Prior-by-prior updating has also been widely adopted in many applications. Recent examples include mechanism design (Bose and Daripa, 2009; Bose and Renou, 2014), auctions (Ghosh and Liu, 2021; Auster and Kellner, 2022), persuasion and cheap talk (Kellner and Le Quement, 2018; Beauchêne et al., 2019), and optimal stopping (Auster and Kellner, 2023; Auster et al., 2024). In Section 6, we discuss alternative updating rules under ambiguity, such as the maximum likelihood updating rule (Gilboa and Schmeidler, 1993) and its generalization in Epstein and Schneider (2007) and Cheng (2022).

Of relevance is also the misspecification literature, which provides an alternative foundation for the failure of asymptotic learning. In this literature, a misspecified agent typically has an initial belief that assigns probability 0 to (a neighborhood of) the true model. Berk (1966) and Shalizi (2009) show that with exogenous information, under mild conditions, the

[1] See Pires (2002) for an axiomatization.

[2] Al-Najjar (2009) show that individuals who use frequentist models might compensate for the scarcity of data by limiting inference to a statistically simple family of events, which leads to statistically ambiguous beliefs. In their setting, such ambiguity vanishes in standard continuous outcome spaces as data increases without bounds.

agent's beliefs converge, although not to the true state. Other works focused on settings where the signals can be affected by the actions of the agent and are, hence, endogenous (Frick et al., 2023; Esponda et al., 2021; Fudenberg et al., 2021; Heidhues et al., 2021).[3] Our paper differs from the existing work in three ways. First, the agent in the misspecified learning literature is a Bayesian learner, whereas, in our setup, the DM holds multiple beliefs and adopts prior-by-prior updating. Second, the DM in our model can be interpreted as correctly specified as we will explain in Footnote 6. Third, we show that in our setting, even when information is exogenous, as in Berk (1966) and Shalizi (2009), the belief set diverges almost surely.

Our paper also relates to the robust statistics literature (Huber, 2004). Roughly speaking, robust statistics are statistics that produce good performance even with deviations from assumptions on the data-generating process. Cerreia-Vioglio et al. (2013) highlight the close relation between decision making under ambiguity, akin to the approach in this paper, and robust statistics, and characterize conditions under which the two approaches are equivalent. However, the problems studied in the robust statistics literature typically differ from the ones studied in this paper. For instance, Giacomini and Kitagawa (2020) and Giacomini et al. (2019) propose new tools for Bayesian inference in set-identified models to reconcile the asymptotic disagreement between Bayesian and frequentist inferences. By contrast, our focus is on characterizing learning outcomes as the number of signals grows large. Finally, this result is in contrast to Marinacci (2002), where ambiguity vanishes because all observations are drawn from the same ambiguous distribution.

2 Setup

A DM aims to learn the state of the world, $\theta \in \Theta := \mathbb{R}$, and has access to N information sources, $I =: \{1, ..., N\}$. It is common knowledge that the state

[3]See Nyarko (1991) and Fudenberg et al. (2017) for examples in which the convergence of beliefs fails.

θ is normally distributed with mean μ and variance $\frac{1}{\rho_\mu}$. Each information source $i \in I$ produces a signal $s_i = \theta + \varepsilon_i$, where the noise ε_i is independent of the state and is normally distributed with mean 0 and variance $\frac{1}{\rho_i}$. We call $\rho_\mu, \rho_i > 0$ the *precision* of the initial belief and information source i, respectively.[4] We assume that the state and signal noise are independent of each other.

The actual precisions of the information sources are drawn i.i.d. from some distribution function G on $[\underline{\rho}, \overline{\rho}]$ with $\overline{\rho} > \underline{\rho} > 0$. However, the DM faces ambiguity in these precisions. In particular, she knows that the precision of each information source lies in $[\underline{\rho}, \overline{\rho}]$, but she cannot form a probabilistic belief about it. Rather, the DM considers a range of conjectures for the precision of each information source. We denote the DM's conjectured precision for information source i by $\hat{\rho}_i \in [\underline{\rho}, \overline{\rho}]$.

Conditional on the realized state θ, we assume the signals are i.i.d. according to the cumulative distribution function F, where

$$F(s) = \int_{[\underline{\rho}, \overline{\rho}]} F_\rho(s) dG(\rho),$$

with F_ρ being the CDF of a normal distribution with mean θ and variance $\frac{1}{\rho}$ for each $\rho \in [\underline{\rho}, \overline{\rho}]$.[5] For now, we assume that the DM directly observes the realized signals. In Section 3.4, we extend the framework to consider more general assumptions on observability.

Belief Updating For each $N \geq 1$, denote the profile of precisions by $\rho^N := (\rho_1, ..., \rho_N)$, the profile of conjectured precisions by $\hat{\rho}^N := (\hat{\rho}_1, ..., \hat{\rho}_N)$, and the profile of signals by $s^N := (s_1, ..., s_N)$. Let $\Delta(\mathbb{R}^N)$ be the set of distributions over $\mathbb{R}^N$. Following Epstein and Schneider (2007) and Epstein and

[4]Our framework is suitable for analyzing biased signals as well. While our main insights remain unchanged in that case, the comparison with the Bayesian benchmark in Section 3.1 is less clear.

[5]We use this assumption of the true distribution of signals as a benchmark for our analysis. A more general interpretation of F is that it represents the limit of the empirical distribution of signals. Based on this interpretation, the characterization in Theorem 1 and Theorem 2 does not rely on whether and how we model the true distribution of signals.

Schneider (2008), we define $L^{s}(\hat{\rho}^N,\theta) \in \Delta(\mathbb{R}^N)$ as the likelihood function for the profile of signals, which is the conditional distribution for signals given conjectured precisions $\hat{\rho}^N$ and the realized state θ. Then the set of likelihood functions of the DM can be represented by $\mathcal{L}_N^{s}$, where

$$\mathcal{L}_N^{s} = \{L^{s}(\hat{\rho}^N,\theta) \in \Delta(\mathbb{R}^N) : \hat{\rho}^N \in [\underline{\rho},\overline{\rho}]^N,\ \theta \in \mathbb{R}\}.$$

We assume the DM adopts prior-by-prior updating (Pires, 2002) to derive posteriors using the initial belief and the set of likelihood functions $\mathcal{L}_N^{s}$. In other words, given the realized profile of signals s^N and a vector of conjectured precisions $\hat{\rho}^N$, the posterior over the state $P_N^{s}(s^N,\hat{\rho}^N) \in \Delta(\mathbb{R})$ is obtained by applying Bayes' rule.[6] Then, the posteriors of the DM can be represented by the following set:

$$\mathbb{P}^{s}(s^N) = \left\{P_N^{s}(s^N,\hat{\rho}^N) \in \Delta(\mathbb{R}) : \hat{\rho}^N \in [\underline{\rho},\overline{\rho}]^N\right\}.$$

We want to emphasize that prior-by-prior updating does not select or rule out any of the available conjectures.

After observing the profile of signals, the DM chooses a potentially random estimate g, with distribution $\Gamma \in \Delta(\mathbb{R})$. Given the set of beliefs $\mathbb{P}^{s}(s^N)$, the DM is a maxmin expected utility (MEU) maximizer (Gilboa and Schmeidler, 1989), and she evaluates her estimate based on the worst possible belief. That is, for a given loss function u, the DM's objective is to minimize the maximal expected loss across all distributions in the set of posteriors. This preference might be a result of the DM being ambiguity-averse or the DM's intention to derive a robust upper bound for the expected loss. Formally, she picks an estimate $g^*(s^N)$, with distribution $\Gamma^*(s^N)$, to solve the

[6] Expanding the DM's belief set by allowing her to consider non-degenerate distributions over precisions does not affect our results. In particular, the belief set could contain the true distribution of precisions, G, and, in this sense, we interpret the DM's model to be correctly specified.

following min-max problem:[7]

$$\min_{\Gamma \in \Delta(\mathbb{R})} \max_{p \in \mathbb{P}^s(s^N)} \mathbb{E}_p \left[\int_g u(g-\theta) d\Gamma \right],$$

where we assume that $u : \mathbb{R} \to \mathbb{R}_+$ is strictly convex, minimized at 0, and smooth with bounded second-order derivatives.

In Section 6, we discuss how our findings depend on the several assumptions made above. There, we highlight that results are primarily driven by our belief updating rule. In the rest of the paper, we are interested in the asymptotic properties of the DM's learning behavior as the number of signals N goes to infinity.

3 Asymptotic Beliefs

In this section, we characterize the DM's posterior set as the number of signals grows large. Recall that for any N, the DM observes s^N. Given a profile of conjectured precisions $\hat{\rho}^N$, the DM's posterior belief is $P_N^{\boldsymbol{s}}(s^N, \hat{\rho}^N)$. We are interested in the asymptotic behavior of the DM's posterior set, $\mathbb{P}^s(s^N)$. Thus, we define the limit set of posteriors as follows:

$$\mathbb{P}_\infty^{\boldsymbol{s}}(s) = \{P : \exists \hat{\rho} \in [\underline{\rho}, \overline{\rho}]^\infty \text{ s.t. } P = \lim_{N\to\infty} P_N^{\boldsymbol{s}}(s^N, \hat{\rho}^N)\},$$

where the limit above is in the sense of convergence in distribution. Note $\mathbb{P}_\infty(s)$ is defined as the set of limits of posteriors that can be generated by some profile of precisions. This definition is silent about which posterior beliefs converge. In fact, many non-convergent sequences of posterior beliefs exist, but, as we clarify in Section 4, these sequences are immaterial for our discussion of the asymptotic estimate.

It is worth differentiating the asymptotic behavior of the DM's posterior set from the DM's optimal estimate. We say the DM **learns successfully**

[7] Lemma 2 in the Appendix shows that the problem below is well-defined. That is, the minimum and maximum can be achieved.

if, for each state θ, $\mathbb{P}^{s}_{\infty}(s) = \{\delta_\theta\}$ for almost all signal realizations. Further, we will say that the DM's **estimate is consistent** if, for each θ, her optimal estimate converges to the true state for almost all signal realizations.

We first establish a useful benchmark in which the DM is Bayesian and minimizes expected loss. Then we characterize the limit set of posteriors for our ambiguity-averse DM, and we finally extend the characterization result to the case where signals cannot be directly observed.

3.1 A Bayesian Decision Maker

In this section, we consider a Bayesian DM. That is, a DM who entertains a joint distribution $B \in \Delta[\underline{\rho}, \overline{\rho}]^\infty$ over the space of infinite sequences of precisions. We assume that the sequence of precisions is independent of the state θ. Upon observing signal realizations s^N, the Bayesian DM updates her belief about the state using Bayes rule. Note that while the DM is Bayesian, her belief may be misspecified in the sense that the distribution of precision sequences she considers, B, may be different from the true distribution.[8]

The DM seeks to minimize the expected loss conditional on the signal:

$$\min_{\Gamma \in \Delta(\mathbb{R})} \mathbb{E}\left[\int_g u(g-\theta) d\Gamma \;\middle|\; s^N \right],$$

where the expectation is taken with respect to the DM's subjective belief. This problem has a solution, which is unique and deterministic by strict convexity of u. We denote this solution by $g^B(s^N) \in \mathbb{R}$. Let $\mathbb{P}^{B}_{\infty}(s)$ be the limit set of posteriors for the Bayesian DM. The next result shows that this set contains exclusively the Dirac measure on the true state, and the DM's estimate is consistent regardless of the possible misspecification or the specific shape of the loss function. That is, the Bayesian DM learns successfully and estimates consistently.

[8] Given the definition of the true data-generating process in Section 2, the true distribution over a sequence of precisions is the product measure over $[\underline{\rho}, \overline{\rho}]^\infty$ produced by the CDF G on each dimension.

Proposition 1. *For all states $\theta \in \mathbb{R}$ and for almost all sequences of signals s, we have $\mathbb{P}^B_\infty(s) = \{\delta_\theta\}$ and $g^B(s^N) \xrightarrow{a.s.} \theta$.*

By Proposition 1, a Bayesian DM's estimate almost surely converges to the true state in our setting. Thus, any deviation from this benchmark is a consequence of ambiguity aversion and the adopted belief updating rule. This finding can be attributed to two factors. First, because each of the signals is at least minimally informative—$\underline{\rho} > 0$—despite the potential misspecification, the posterior belief of the Bayesian DM always becomes degenerate. The Bayesian DM correctly believes that the realizations of the signals are uncorrelated with their precisions, making the posterior mean a consistent estimate of the true state. Second, employing a common econometrics approach, the asymptotic estimate is equal to the optimal estimate at the limiting set of beliefs—which, we argued, is degenerate at the true state. Because the loss function is uniquely minimized at zero, this optimal estimate is the true state itself, even if the loss function is asymmetric.

The key difference between the Bayesian DM and our ambiguity-averse DM is the limit set of posterior beliefs they entertain. Next, we characterize this set for our DM.

3.2 The Game against Nature

It is often useful to interpret the ambiguity-averse DM's decision problem as a zero-sum game between the DM and nature. Under this interpretation, after signals are realized, the DM chooses an estimate for the state to minimize her expected loss function. Subsequently, with knowledge of the estimate, nature is free to choose, for each signal, any precision within the uncertainty set of the DM. The DM's objective is then to guarantee the lowest expected loss conditional on the fact that nature acts after her and to her detriment. Note that, for each possible sequence of signals, s^N, and conjectured precisions, $\hat{\rho}^N$, the corresponding posterior distribution on the state is normal. Denote the posterior mean by $\mathbb{E}[\theta|s^N, \hat{\rho}^N] = \frac{\hat{\rho}^N \cdot s^N + \rho_\mu \mu}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}$, and the posterior variance by $\mathbb{V}[\theta|s^N,\hat{\rho}^N] = \left(1 - \frac{\hat{\rho}^N \cdot \mathbb{1}^N}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}\right)\frac{1}{\rho_\mu}$.

By changing the precision of each signal, nature affects both the posterior mean and the posterior variance. It determines variance by choosing the sum of precisions across signals, and, importantly, it affects the posterior mean by assigning different precisions to different signal realizations. Given an estimate of the DM, Γ, how does nature act to increase her loss? A simple heuristic is to notice that the loss function can be approximated by:

$$\mathbb{E}[\int_g u\,(g-\theta)\,d\Gamma|s^N,\hat{\rho}^N] \approx \int_g u\left(g - \mathbb{E}\left[\theta|s^N,\hat{\rho}^N\right]\right)d\Gamma + \lambda\mathbb{V}[\theta|s^N,\hat{\rho}^N],$$

for some constant $\lambda > 0$. Thus, nature can increase the DM's loss by making the posterior variance large or by pushing the posterior mean far from the DM's estimate. The latter can be interpreted as increasing the posterior bias. From the definition of the posterior variance, we see that as long as each signal is somewhat informative ($\underline{\rho} > 0$), the posterior variance converges to 0 as the number of available signals N increases. Hence, the more signals the DM receives, the less able nature is to use the posterior variance against her. In the extreme case of $N \to \infty$, for any choice of precisions that nature may consider, the posterior variance is equal to 0, and nature utilizes the posterior mean as its only instrument. We now characterize nature's behavior when $N \to \infty$, where nature's choice of what precision to attribute to which signal exclusively affects the posterior mean. For a fixed sequence of realized signals s, nature's strategy is a precision assignment $\hat{\rho} \in [\underline{\rho},\overline{\rho}]^\infty$, an infinite sequence of precisions, one for each information source.

Definition 1. *For any sequence of realized signals s, an assignment $\hat{\rho}$ is a threshold assignment if there exists $x \in \mathbb{R}$ such that either (i) $\hat{\rho}_i = \overline{\rho}$ for all $s_i > x$ and $\hat{\rho}_i = \underline{\rho}$ for all $s_i \leq x$, or (ii) $\hat{\rho}_i = \underline{\rho}$ for all $s_i > x$ and $\hat{\rho}_i = \overline{\rho}$ for all $s_i \leq x$.*

Lemma 1. *For any sequence of realized signals s, if $\hat{\rho}^*$ solve $\max_{\hat{\rho}\in[\underline{\rho},\overline{\rho}]^\infty} u\Big(g - \mathbb{E}[\theta|s,\hat{\rho}]\Big)$ for some $g \in \mathbb{R}$, then $\hat{\rho}^*$ is a threshold assignment.*

Thus, nature chooses a threshold in the space of signal realizations and assigns the highest precision to signals above the threshold and the lowest precision to those below it, or vice-versa. Nature aims to maximize the posterior bias, which entails either maximizing or minimizing the posterior mean. The intuition for Lemma 1 follows from the expression of the posterior mean. First, we show that it is without loss of optimality for nature to assign the same precision to identical signal realizations. We slightly abuse notation and denote by $\hat{\rho}(x)$ the precision assigned to signal realization x. Given an observed empirical distribution of signals, F, this assignment induces posterior mean $\mathbb{E}[\theta|s, \hat{\rho}] = \frac{\int x\hat{\rho}(x)dF(x)}{\int \hat{\rho}(x)dF(x)}$. Consider nature's choice to maximize this expression while keeping the same expected value of conjectured precisions $\int \hat{\rho}(x)dF(x) = c$. Because c pins down the denominator of the expression for the posterior mean, nature chooses an assignment to maximize the numerator. To do so, nature assigns high-valued signals high precisions and low-valued signals low precisions, thereby moving the posterior mean towards higher signal realizations. Using the extreme precisions $\overline{\rho}$ and $\underline{\rho}$ is the best way to achieve this, therefore justifying the optimality of threshold strategies. An analogous strategy is optimal to minimize the posterior mean.

3.3 Characterization of Asymptotic Beliefs

We now characterize the DM's set of asymptotic beliefs. As discussed in the previous section, each one of nature's strategies corresponds to a plausible belief in the DM's belief set. As the number of signals goes to infinity, nature loses the ability to influence the posterior variance, as aggregate information becomes infinitely precise. However, through the assignment of precisions to signals, nature can still affect the DM's posterior bias. This affects the DM's limit set of beliefs: Her posteriors converge to an interval of degenerate distributions. The next result characterizes this interval. Recall that F is the actual distribution over signals given state θ.

Theorem 1. *Define:*

$$\overline{m} = \frac{\underline{\rho}\int_{-\infty}^{\overline{m}} x dF(x) + \overline{\rho}\int_{\overline{m}}^{\infty} x dF(x)}{\underline{\rho} F(\overline{m}) + \overline{\rho}(1 - F(\overline{m}))}, \quad \underline{m} = \frac{\overline{\rho}\int_{-\infty}^{\underline{m}} x dF(x) + \underline{\rho}\int_{\underline{m}}^{\infty} x dF(x)}{\overline{\rho} F(\underline{m}) + \underline{\rho}(1 - F(\underline{m}))}.$$

Then, for almost all sequences of realized signals s,

1. *for all sequences $\hat{\rho} \in [\underline{\rho}, \overline{\rho}]^{\infty}$,*

$$\underline{m} \leq \lim_{N\to\infty} \inf \mathbb{E}_{P_N^s(s^N, \hat{\rho}^N)}[\theta | s^N, \hat{\rho}^N] \leq \lim_{N\to\infty} \sup \mathbb{E}_{P_N^s(s^N, \hat{\rho}^N)}[\theta | s^N, \hat{\rho}^N] \leq \overline{m};$$

2. *the limit set of posteriors is a set of degenerate distributions independent of s:*

$$\mathbb{P}_{\infty}^{s}(s) = \{\delta_b : \underline{m} \leq b \leq \overline{m}\}.$$

Theorem 1 starts by establishing that for any precision assignment, posterior means are bounded by two real numbers: $\underline{m}$, $\overline{m}$. These numbers formalize the notion of maximal and minimal posterior means that nature can achieve asymptotically. The second part of the theorem shows that any converging posterior approximates a degenerate distribution and that distribution may have any mean between the boundaries $\underline{m}$ and $\overline{m}$. Notably, Theorem 1 characterizes the values of these boundaries. For example, $\overline{m}$ is generated by the following threshold assignment: give the highest precision to signals higher than $\overline{m}$ and the lowest precision to values below it. By giving more weight to high signals, nature moves the posterior mean up. The fixed point $\overline{m}$ expresses the highest achievable posterior mean. For a concrete example, see Figure 1 below. It is worthwhile to note that the true state θ always lies between the two bounds of posterior means, $\underline{m}$ and $\overline{m}$, and hence, the DM never rules out θ as the number of signals grows.

For an intuition of the definition of $\overline{m}$ in Theorem 1, suppose we start with a threshold m that is lower than the posterior mean. Then, increasing the threshold to $m' > m$ has two effects. First, by assigning all values in (m, m') to low precisions, this precision-weighted sum of signals is reduced.

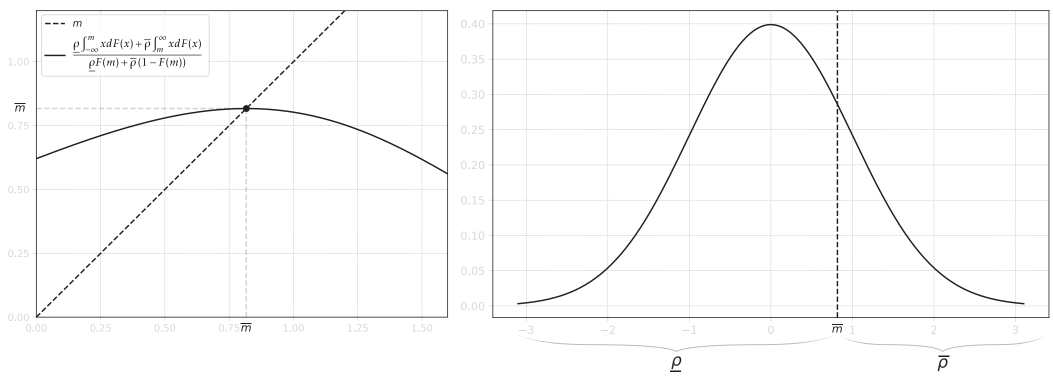
When m' is close to m, this effect is roughly proportional to the marginal signal, m. Second, since more signals are assigned the lowest precision, the precision-weighted mass of signals is reduced. This effectively increases the value of all the inframarginal signals, so it is proportional to the precision-weighted posterior mean. Because this value was higher than the threshold to begin with, the second effect dominates the first, and the expected value of signals increases. This process can be repeated until the marginal signal equals the posterior mean.[9]

To see why the ambiguity-averse DM does not learn successfully while a Bayesian DM does, consider the following example. Assume the Bayesian DM believes that signals are drawn according to precision sequence $\rho = (\rho_1, \rho_2, \dots)$. Although Proposition 1 shows that the Bayesian DM's learns the truth almost surely, there exist signal realizations such that her posterior beliefs converge away from the truth. Because our ambiguity-averse DM considers many of these conjectures, it is possible to find, for almost any signal realizations, one of such conjectures that leads to a sequence of posteriors converging far away from the truth.

Theorem 1 implies that induced ambiguity about the state does not vanish asymptotically. Rather, the DM still entertains a wide range of values for the state θ even when she has access to an arbitrarily large number of informative signals. This finding is in stark contrast to quantifiable risk. In fact, a secondary consequence of the result above is that quantifiable risk completely disappears even in our setting: all the limit posteriors are degenerate. In Section 4, we show how the presence of ambiguity in the limit set of posteriors affects the optimal estimate of the DM. Before that, we prove that a similar characterization applies even when signals are not directly observable.

[9] This parallels the argument that the average cost curve is minimized when it intersects the marginal cost curve.

Figure 1: Identifying the Upper Bound



The graphs above depict the procedure for identifying the upper bound with observable signals. In this example, the initial ambiguity set is fixed at $\underline{\rho} = 0.5$ and $\overline{\rho} = 4$, while the actual signals are drawn from a standard normal distribution. The solid curve on the left figure is the mean of the DM's posterior for different thresholds, which vary in the horizontal axis. As Theorem 1 establishes, the maximum of this function coincides with the optimal threshold, $\overline{m}$. The picture on the right illustrates the optimal threshold: signal realizations above (below) $\overline{m}$ are assigned $\overline{\rho}$ ($\underline{\rho}$), leading to the highest attainable posterior mean.

3.4 General Observables

In this section, we allow the possibility that the DM cannot observe signals directly.[10] Suppose that for each information source i with precision ρ_i, there is a one-to-one mapping between its realized signal s_i and the observable of the DM, a_i. Denote $s_i = \boldsymbol{s^a}(a_i, \rho_i)$, in which the mapping $\boldsymbol{s^a}$ reflects the way the DM backtracks the unobserved signal from the observed a_i. When $\boldsymbol{s^a}(a_i, \rho_i) \equiv a_i$, we are back to the case of observable signals.

We can use $\boldsymbol{s^a}$ to define the following notation, in line with those in Section 2. We slightly abuse notation and still use F to denote the distribution function of the observables. That is, $F(a) = \int_{[\underline{\rho},\overline{\rho}]} F_\rho\Big(\boldsymbol{s^a}(a,\rho)\Big)dG(\rho)$ in which $F_\rho \sim \mathcal{N}\left(\theta, \frac{1}{\rho}\right)$ for each $\rho \in [\underline{\rho}, \overline{\rho}]$. We assume that for any $\rho \in [\underline{\rho}, \overline{\rho}]$, the integral $\int \boldsymbol{s^a}(a,\rho)dF(a)$ is finite. Denote the profile of observables by $a^N := (a_1, ..., a_N)$. Let $L^{\boldsymbol{a}}(\hat{\rho}^N, \theta)$ be the likelihood function for observables and $\mathcal{L}_N^{\boldsymbol{a}}$ be the set of all likelihood functions.[11] Let $P_N^{\boldsymbol{a}}(a^N, \hat{\rho}^N)$ be the poste-

[10] As shown in Theorem 1, our main results do not rely on the unobservability of signals, which is in contrast to Battigalli et al. (2019).

[11] To calculate the likelihood function of observables, one can first calculate the likelihood function of signals, which is just a multivariate normal distribution with independent marginals, and then make use of the one-to-one mapping between signals and observables

rior given realized observables a^N and conjectured precisions $\hat{\rho}^N$, $\mathbb{P}^{\boldsymbol{a}}(a^N)$ be the set of all posteriors, and $\mathbb{P}^{\boldsymbol{a}}_\infty(a)$ be the limit set of posteriors. To guarantee the optimality of threshold assignments as in Lemma 3, we impose the following assumption on $\boldsymbol{s^a}$.

Assumption 1. *There exists a strictly increasing and surjective function* $\beta : \mathbb{R} \to \mathbb{R}$ *and a weakly increasing function* $\gamma : \mathbb{R} \to \mathbb{R}$ *such that* $\boldsymbol{s^a}(a,\rho) = \beta(a) + \frac{\gamma(a)}{\rho}$ *for all* $a \in \mathbb{R}$ *and* $\rho \in [\underline{\rho}, \overline{\rho}]$.

This assumption allows for a broad range of observables relevant to several economic applications. Consider the following three examples:

(i) With observable signals, $\boldsymbol{s^a}(a,\rho) = a$, corresponding to $\beta(a) = a$ and $\gamma(a) = 0$ for all a.

(ii) When the DM observes estimates of Bayesian agents based on their common initial belief and private signals, which is $a_i = \frac{\rho_i s_i + \rho_\mu \mu}{\rho_i + \rho_\mu}$, we can set $\boldsymbol{s^a}(a,\rho) = a + \frac{\rho_\mu(a-\mu)}{\rho}$ with $\beta(a) = a$ and $\gamma(a) = \rho_\mu(a-\mu)$ for all a.

(iii) Let θ be the value of a financial asset. In many market microstructure models (Kyle, 1985; Lambert et al., 2018), risk-neutral investors may submit a market order to be executed by an uninformed market maker. In equilibrium, investors face an affine pricing function: $p = \mu + \alpha + \lambda x$, $\lambda > 0$. Then their demand is: $a_i = \frac{1}{2\lambda}\left(\frac{\rho(s_i-\mu)}{\rho+\rho_\mu} - \alpha\right)$. We can set $\gamma(a) = (2\lambda a + \alpha)\rho_\mu$, and $\beta(a) = \mu + \alpha + 2\lambda a$, since $\boldsymbol{s^a}(a,\rho) = \mu + \alpha + 2\lambda a + \frac{(2\lambda a+\alpha)\rho_\mu}{\rho}$.

Under Assumption 1, we can generalize Theorem 1. Note that the true state θ still lies between the two bounds of posterior means.

Theorem 2. *Let Assumption 1 hold. Define:*

$$\overline{m} = \frac{\underline{\rho}\int_{-\infty}^{\beta^{-1}(\overline{m})} \boldsymbol{s^a}(x,\underline{\rho})dF(x) + \overline{\rho}\int_{\beta^{-1}(\overline{m})}^{\infty} \boldsymbol{s^a}(x,\overline{\rho})dF(x)}{\underline{\rho}F(\beta^{-1}(\overline{m})) + \overline{\rho}(1 - F(\beta^{-1}(\overline{m})))},$$

$$\underline{m} = \frac{\overline{\rho}\int_{-\infty}^{\beta^{-1}(\underline{m})} \boldsymbol{s^a}(x,\overline{\rho})dF(x) + \underline{\rho}\int_{\beta^{-1}(\underline{m})}^{\infty} \boldsymbol{s^a}(x,\underline{\rho})dF(x)}{\overline{\rho}F(\beta^{-1}(\underline{m})) + \underline{\rho}(1 - F(\beta^{-1}(\underline{m})))}.$$

given the profile of conjectured precisions.

Then, for almost all sequences of realized observables a,

1. *for all sequences $\hat{\rho} \in [\underline{\rho}, \overline{\rho}]^\infty$,*

$$\underline{m} \leq \lim_{N\to\infty} \inf \mathbb{E}_{P_N(a^N,\hat{\rho}^N)}[\theta | a^N, \hat{\rho}^N] \leq \lim_{N\to\infty} \sup \mathbb{E}_{P_N(a^N,\hat{\rho}^N)}[\theta | a^N, \hat{\rho}^N] \leq \overline{m};$$

2. *the limit set of posteriors is a set of degenerate distributions independent of a:*

$$\mathbb{P}^{\boldsymbol{a}}_\infty(a) = \{\delta_b : \underline{m} \leq b \leq \overline{m}\}.$$

4 Asymptotic Estimate

In this section, we characterize the DM's asymptotic estimate under general observables. Recall that, for each realization of observables a^N, the DM's estimate, $g^*(a^N)$, which has distribution $\Gamma^*(a^N)$, minimizes her loss function, considering the worst-case posterior in $\mathbb{P}^{\boldsymbol{a}}(a^N)$. Because observables and the loss function are arbitrary, obtaining an explicit solution to $g^*(a^N)$ for finite N is intractable. However, as N goes to infinity, we can make use of Theorem 2. The main result of this section characterizes the asymptotic estimate by demonstrating that the following limit exchange holds:

$$\lim_{N\to\infty} \arg \min_{\Gamma\in\Delta(\mathbb{R})} \max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p\Big[\int_g u(g-\theta) d\Gamma\Big] = \arg \min_{\Gamma\in\Delta(\mathbb{R})} \max_{p\in\mathbb{P}^{\boldsymbol{a}}_\infty(a)} \mathbb{E}_p\Big[\int_g u(g-\theta) d\Gamma\Big]. \tag{1}$$

Theorem 2 states that $\mathbb{P}^{\boldsymbol{a}}_\infty(a) = \{\delta_m : m \in [\underline{m}, \overline{m}]\}$ for almost all realizations of observables. The limit swap above implies that, as N grows, the optimal estimate converges to the estimate of a DM who does not know the mean of θ but seeks to minimize loss within the interval $[\underline{m}, \overline{m}]$. This observation greatly simplifies the characterization: The asymptotic behavior of the estimate is pinned down by an extremely simple optimization problem. In this problem, the DM only cares about how biased her estimate is in the worst-case scenario. Recall that the DM's loss increases as her estimate diverges from the true state. If her estimate is too far from $\underline{m}$, she has a large utility loss in the worst case, in which the state is actually $\underline{m}$. A symmetric argument holds for $\overline{m}$. Therefore, she guarantees minimal loss by being indifferent between these two extreme possible values of the state. This intuition is formalized in the next result.

Theorem 3. *For all N, $g^*(a^N)$ is deterministic. Moreover, $g^*(a^N) \xrightarrow{a.s.} g^*$, where g^* is the unique solution to $u(g^* - \underline{m}) = u(g^* - \overline{m})$.*

Although intuitive, this result depends on the non-trivial exchange of limits in equation 1. Ex ante, it is not clear that this limit swap works. First, the limits of optimizers of a sequence of optimization problems are not guaranteed to coincide with the optimizers of the limit problem. Second, not all distributions in the set $\mathbb{P}^{\boldsymbol{a}}(a^N)$ converge. Indeed, there always exist sequences of precisions such that posterior beliefs diverge. Still, Theorem 3 confirms the limit swap is valid, and the heuristic argument we presented goes through formally. We make this argument in two steps, addressing the two concerns highlighted above.

The first step is to show the DM's optimization can be approximated by an optimization that only considers the means of posterior distributions as N grows large. For any finite N, the DM's loss is clearly affected by higher moments of the posteriors, but because quantifiable risk vanishes as the number of observable information sources grows, the mean progressively becomes the only relevant moment. The second step relies on an extension of the Glivenko-Cantelli theorem. It provides the important result that the sequence $g^*(a^N)$ is bounded. Recall, from part 1 of Theorem 2, that non-converging posteriors are bounded. Thus, intuitively, for each N, the payoff obtained by a non-convergent sequence can be bounded by the payoff of two convergent sequences. Consequently, restricting attention to convergent sequences turns out to be without loss of generality. We prove these two steps are sufficient to guarantee the convergence of g^*.

Theorem 3 shows the asymptotic estimate is typically inconsistent. To illustrate, recall that the interval $[\underline{m}, \overline{m}]$ in Theorem 2 is independent of the particular choice of the loss function. The interval is determined by the initial ambiguity and the mapping from observables to signals $\boldsymbol{s^a}$. By contrast, the asymptotic estimate is a consequence of the shape of the loss function, u, on this interval. This demonstrates that the DM's estimate is consistent only in knife-edge cases. Moreover, even in such cases, slightly perturbing either u or $\boldsymbol{s^a}$ would again lead to an inconsistent asymptotic estimate.

Finally, it is worth noting that the DM does not benefit from randomization. This is in contrast with most of the work on decision-making under maxmin expected utility. In our setting, the optimality of deterministic estimates follows from the richness of the action space and the convexity of the loss function. In the literature, either randomized actions are directly excluded (Tillio et al., 2017) or assuming deterministic actions is proved to be with loss of generality (Bose and Renou, 2014; Gershkov et al., 2023). One exception is Tang and Zhang (2021).

4.1 Observable signals: Loss Symmetry

We now focus on the case in which signals are directly observable to illustrate how the shape of the loss function affects consistency of the estimate. A loss function is symmetric if $u(x) = u(-x)$ for any real number x. We show that when signals are directly observable, the symmetry of the loss function plays a prominent role in the asymptotic estimate. Indeed, by Theorem 3, we have that, under symmetric losses, $g^* = \frac{\overline{m}+\underline{m}}{2}$. In turn, normality implies the true distribution of signals F is symmetric around the true state θ. In this case, the DM's estimate of the state is consistent. The next result formalizes this relationship between the symmetry of the loss function and the consistency of the estimate and proves a partial converse.

Corollary 1. *Fix a state θ and assume signals are observable. If u is symmetric, then $g^*(s^N) \xrightarrow{a.s.} \theta$ for any perceived precision set $[\underline{\rho}, \overline{\rho}]$. If u is not symmetric, then there exists a perceived precision set $[\underline{\rho}, \overline{\rho}]$ such that $g^*(s^N) \xrightarrow{a.s.} g^* \neq \theta$.*[12]

[12]Note that the DM does not learn successfully, even if her estimate is consistent under symmetric losses and observable signals. Moreover, the consistency of the DM's estimate

The following example illustrates Corollary 1.

Example: Asymmetric Quadratic Loss Let the loss function be given by

$$u(g-\theta)=\begin{cases}(g-\theta)^2 & \text{if } g\geq\theta\\ \lambda(g-\theta)^2 & \text{if } g<\theta\end{cases}.$$

with $\lambda>0$. That is, the DM might perceive losses differently based on whether the state θ is over- or under-estimated. For example, if $\lambda>1$, she is less concerned about overestimating than underestimating the true state. This difference may arise from various factors. For instance, a health official estimating disease prevalence might face greater consequences for underestimating transmission rates, while a product developer could incur significant losses from overestimating market demand and developing a costly product that fails to be marketed. Following Theorem 3, the optimal estimate satisfies $g^*=\frac{\underline{m}+\sqrt{\lambda}\overline{m}}{1+\sqrt{\lambda}}$. However, as argued in the previous section, observable signals imply $\frac{\underline{m}+\overline{m}}{2}=\theta$, which does not equal g^* whenever $\lambda\neq 1$.

The above example highlights that the loss function directly affects the DM's estimate, even asymptotically. As outlined in Section 3.1, a Bayesian DM's posterior belief converges to a Dirac measure on the true state. In a setting with multiple Bayesian DMs, as the available information grows, regardless of their loss functions, Bayesian DMs agree on the optimal estimate of the state. By contrast, our DM's asymptotic estimate continues to depend on the particular form of the loss function. Thus, ambiguity about the precision of information sources, coupled with potentially different loss functions, can rationalize disagreement between well-informed experts who aim to find out the truth, such as scientists with access to the same large dataset.

depends on the symmetry of the loss function and the symmetry of the normal distribution, even under observable signals.

5 Implications

We now turn to various implications of our main result.

5.1 Comparative Statics of Ambiguity

First, we illustrate that contrary to intuition, making all signals more precise is not necessarily beneficial to the DM. For simplicity, we consider the case of symmetric loss functions and observable signals. Recall that by Theorem 1, the limit set of posteriors is a set of degenerate distributions δ_b with $\underline{m} \leq b \leq \overline{m}$. Note that $\overline{m}$ and $\underline{m}$ only depend on the fraction of the highest and the lowest possible precisions and not on their absolute values because they can be rewritten as

$$\overline{m} = \frac{\int_{-\infty}^{\overline{m}} x dF(x) + \eta \int_{\overline{m}}^{\infty} x dF(x)}{F(\overline{m}) + \eta\,(1 - F(\overline{m}))}, \quad \underline{m} = \frac{\eta \int_{-\infty}^{\underline{m}} x dF(x) + \int_{\underline{m}}^{\infty} x dF(x)}{\eta F(\underline{m}) + (1 - F(\underline{m}))},$$

where $\eta = \frac{\overline{\rho}}{\underline{\rho}}$. The following proposition shows that both $\overline{m}$ and $\underline{m}$ change with η monotonically.

Proposition 2. *Let $\eta = \frac{\overline{\rho}}{\underline{\rho}} \in (1, +\infty)$. Under observable signals, $\overline{m}$ is monotonically increasing in η and $\underline{m}$ is monotonically decreasing in η. Moreover, when $\eta \to +\infty$, we have $\overline{m} \to \infty$ and $\underline{m} \to -\infty$; when $\eta \to 1$, we have $\overline{m} - \underline{m} \to 0$.*

In Proposition 2, $\eta = \frac{\overline{\rho}}{\underline{\rho}}$ can be interpreted as the degree of ambiguity in the set of perceived precisions $[\underline{\rho}, \overline{\rho}]$. When more ambiguity exists regarding precisions of signals ex-ante, the limit set of posteriors expands, and hence, ambiguity regarding states is greater ex-post. We now examine the welfare implications of these comparative statics. By Corollary 1, under observable signals, a DM with a symmetric loss function estimates the state correctly at the limit. Thus, utility depends solely on the size of the limit set of posterior means, that is, $\overline{m} - \underline{m}$. Corollary 2 directly follows from Proposition 2.

Corollary 2. *Let u be symmetric. Under observable signals, as η increases, the DM is strictly worse off asymptotically.*

Corollary 2 reveals a counterintuitive implication. Increasing $\underline{\rho}$ and/or $\overline{\rho}$ in a manner that leads to a higher η makes the DM strictly worse off. Therefore, even making all signals more precise is not necessarily beneficial to the DM and might, in fact, be detrimental to her.

5.2 Comparative Statics of the True Distribution

In this section, we study how asymptotic ambiguity and the DM's estimate change as we vary the true distribution of precisions. Once again, for ease of exposition, assume signals are directly observable. We fix the perceived precisions $[\underline{\rho}, \overline{\rho}]$, and we let G and H be distributions of true precisions generating asymptotic belief boundaries $\{\underline{m}_G, \overline{m}_G\}$ and $\{\underline{m}_H, \overline{m}_H\}$, respectively.

Proposition 3. *If G first-order stochastically dominates H, then the asymptotic belief set is larger for H. That is, for any state $\theta \in \mathbb{R}$,*

$$H \preceq_{FOSD} G \implies \underline{m}_H \leq \underline{m}_G \leq \overline{m}_G \leq \overline{m}_H.$$

Intuitively, if the true precision of information sources is lower, the signals they produce become more scattered. Consequently, this widens the range of potential posterior probabilities by thickening the tails of the signal distribution. This result has implications for asymptotic estimates. As detailed in Section 3.1, while the asymptotic loss for a Bayesian DM remains zero, irrespective of her knowledge about precision distributions, this distribution becomes pivotal for an ambiguity-averse DM. Certain precision distributions can result in the ambiguity-averse DM making significantly large estimation errors. Corollary 3 offers sufficient conditions to generate such errors.

Corollary 3. *Assume signals are observable and u is the asymmetric quadratic loss function in Section 4.1 with any $\lambda \neq 1$. For any $\eta > 1$ and any constant $C > 0$, true distributions of precisions G exist such that $|g^* - \theta| > C$.*

5.3 Aggregating Estimates

Finally, we study the problem of an ambiguity-averse econometrician who aims to estimate the state by aggregating estimates from many Bayesian agents. The agents share the same initial belief about the state, $\theta \sim \mathcal{N}\left(\mu, \frac{1}{\rho_\mu}\right)$, but have access to different private information sources. The econometrician knows the initial belief but does not know the precision of the individual sources. This setup finds practical relevance across various scenarios. For instance, consider a healthcare official who estimates disease prevalence in a region using hospital reports but is uncertain about their data-collection protocols.

Conditional on the realization of θ, agent i receives a private signal $s_i = \theta + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}\left(0, \frac{1}{\rho_i}\right)$ is independent noise. We consider the case in which each agent and the econometrician attempt to estimate the realized value of θ to minimize the mean-squared error. Given the initial belief and the private signal, the optimal Bayesian estimate for agent i is $\mathbb{E}[\theta|\rho_i, s_i] = \frac{\rho_\mu \mu + \rho_i s_i}{\rho_\mu + \rho_i}$. These actions are observed by the econometrician.

Although each agent knows the precision of their private signal, the econometrician does not and considers a set of possible precisions $\left[\underline{\rho}, \overline{\rho}\right]$. Because each action is a convex combination of the private signal s_i and the mean of the initial belief μ, an econometrician who intends to estimate θ will first have to transform the actions back to signals. For a conjectured precision $\hat{\rho}_i$, the recovered signal will be $\mathbf{s^a}(a_i, \hat{\rho}_i) = a_i + \frac{\rho_\mu}{\hat{\rho}_i}(a_i - \mu)$.

Applying Theorem 1, the boundaries of the limit set of posteriors are defined by a fixed point similar to the example with observable signals. However, here, the econometrician backtracks realized signals from observed estimates, leading to a bias in recovered signals. Because of this bias, and as a corollary of Theorem 3, the econometrician's estimation converges away from the truth almost surely.

To give an example where we can clearly characterize the optimal estimate and analyze comparative statics, we continue with the following assumption.

Assumption 2. *For some* $\rho^* \in [\underline{\rho}, \overline{\rho}]$, $G = \delta_{\rho^*}$.

Although the econometrician might consider different precisions for each signal, under Assumption 2, in reality, all signals share the same precision. We say the econometrician *overreacts* if $|g^* - \mu| > |\theta - \mu|$ and *underreacts* if the inequality is reversed. Put simply, an estimator overreacts to information when it deviates further from the mean of initial beliefs than the true state does.

Proposition 4 (Guess Characterization)**.** *Let* $\mathbf{s^a}(a_i, \hat{\rho}_i) = a_i + \frac{\rho_\mu}{\hat{\rho}_i}(a_i - \mu)$ *and the loss function be quadratic. Under* Assumption 2, $g(A^n) \xrightarrow{a.s.} g^*$, *and*

1. *If* $\mu = \theta$, *then* $g^* = \theta$;
2. *If* $\mu \neq \theta$, *then there exists* $\tilde{\rho} < \tilde{\tilde{\rho}}$ *such that*
   - *If* $\rho^* \leq \tilde{\rho}$, *then the DM underreacts;*
   - *If* $\rho^* \geq \tilde{\tilde{\rho}}$, *then the DM overreacts;*
   - *If* $\tilde{\rho} < \rho^* < \tilde{\tilde{\rho}}$, *then the DM underreacts when* $|\theta - \mu|$ *is small and overreacts when* $|\theta - \mu|$ *is large,*

where:
$$\tilde{\rho} = \frac{2\underline{\rho}\overline{\rho}}{\underline{\rho} + \overline{\rho}} \qquad \text{and} \qquad \tilde{\tilde{\rho}} = \underline{\rho} F\left(\overline{m}(\tilde{\tilde{\rho}}, \mu)\right) + \overline{\rho}\left(1 - F\left(\overline{m}(\tilde{\tilde{\rho}}, \mu)\right)\right).$$

Proposition 4 reveals that whether the DM over- or underreacts depends on the true precision of the signals. Essentially, the estimation involves the DM attempting to infer the mean of the unobserved signals from the mean of observed actions. Because signals are unbiased, their unobserved mean is effectively θ. When ρ^* is high, θ must be relatively close to the mean of actions because agents place a high weight on their signals, and vice versa if ρ^* is low. However, the econometrician does not know the true precision, so she backtracks signals from actions using roughly the same method regardless of ρ^*, leading to over or under-reaction.

Lastly, when the initial belief is extremely imprecise, $\rho_\mu \approx 0$, actions remain unaffected by the initial belief, resulting in straightforward backtracking of signals and correct estimation by the econometrician. Conversely, when the precision of the initial belief grows to infinity, $\rho_\mu \to \infty$, the econometrician estimates correctly by simply following the initial belief. For intermediate values, however, the estimate is wrong almost surely. In other words, the accuracy of the estimate is not monotonic with the precision of the initial belief: Better information ex-ante does not guarantee a more correct estimate asymptotically.

6 Discussion

To maintain tractability and clarity, our analysis has relied on four main assumptions: (i) The DM adopts prior-by-prior updating; (ii) the DM only knows the highest possible and lowest possible precisions of each information source and nothing else; (iii) both the state and signals follow normal distributions; and (iv) the DM is a MEU maximizer regarding ambiguity. In this section, we briefly argue our main result that ambiguity does not vanish asymptotically remains valid when we relax the last three assumptions. Hence, the essential assumption is the updating rule under ambiguity.

Updating Rule First, we note our result *does* rely on the updating rule under ambiguity. Since we assume an unambiguous initial belief and ambiguous likelihood functions, the existence of ambiguity in posterior beliefs results from the dilation property (Seidenfeld and Wasserman, 1993; Shishkin and Ortoleva, 2023) of the prior-by-prior updating rule.[13] Another alternative to the prior-by-prior updating rule is the maximum likelihood updating rule, which also satisfies the dilation property. Unlike prior-by-prior updating, where the DM applies Bayes' rule to the entire set of initial beliefs, the DM with the maximum likelihood updating rule would discard initial beliefs that do not ascribe the maximal probability to the observed signals and update the remaining initial beliefs according to Bayes' rule. We conjecture that it will generically generate a single posterior (a unique maximizer of

[13] Our results may not hold under updating rules that restrict or rule out dilation, such as the proxy updating rule of Gul and Pesendorfer (2021) and the contraction updating rule of Tang (2024).

the likelihood problem) as the number of signals goes to infinity, so ambiguity should vanish.[14]

There are also intermediate cases between full Bayesian and maximum likelihood updating. Under the likelihood-ratio updating rule in Epstein and Schneider (2007) and the relative maximum likelihood updating rule in Cheng (2022), the set of asymptotic beliefs would be either an intermediate set between or a convex combination of those under prior-by-prior and maximum likelihood updating, and hence our qualitative results will be maintained. More recently, Cheng (2024) studies when data can improve robust decisions in terms of the expected payoff under the true distribution. He shows that it is necessary and sometimes sufficient for the revision rule to accommodate the true distribution. In our setting, one such revision rule is to retain a state if and only if it is close enough to the sample mean. He shows that ambiguity vanishes asymptotically under this revision rule (Proposition 2 in Cheng (2024)).

Belief Set In the main analysis we have assumed that the DM's belief set consists of degenerate distributions over all possible sequences of precisions in $[\underline{\rho}, \overline{\rho}]$. This can be interpreted as a fully ambiguous belief set. Our main results remain qualitatively valid for other specifications of the belief set, so long as nature has enough flexibility to condition the (distribution of) the sequence of precisions on realized signals.[15] Consider the following two examples and note that, in both cases, nature can only assign two different (distributions of) precisions to each signal.

First, suppose the DM believes that there are two groups of information sources. Group 1 consists of a fraction $\alpha \in (0,1)$ of information sources with shared high precision $\overline{\rho}$, and Group 2 consists of fraction $1-\alpha$ with shared

[14] Under maximum likelihood updating, the DM maximizes the expected likelihood function of signals under the initial belief over states. As a result, the optimal precision for a particular information source depends on the entire sequence of realized signals and it is difficult to leverage usual asymptotic results.

[15] As an example where this condition fails, suppose the DM's belief set is the set of degenerate distributions over sequences of identical precisions. In this case, she will learn the true state eventually. We thank an anonymous referee for this example.

low precision $\underline{\rho}$. That is, given any sequence of realized signals, the DM belief sets consists of sequences of precisions in $\{\overline{\rho}, \underline{\rho}\}^{\infty}$ with a fraction α taking the value $\overline{\rho}$ and a fraction $1-\alpha$ taking the value $\underline{\rho}$. Although the belief set is much more restrictive than the fully ambiguous one we have focused on, nature can still induce the DM to have a relatively high (low) posterior mean of the state by assigning high signals to Group 1 (Group 2) subject to the new constraint. Hence, the DM does not learn successfully.

Second, we consider an example where the DM holds a non-degenerate belief over the sequence of precisions in the belief set. For any fixed $\rho^1 > \rho^2 > \rho^3 > \rho^4 > 0$ and $\alpha \in (0,1)$, we assume that the precision of each information source is independently drawn from some distribution that can be either $\overline{Q}$ or $\underline{Q}$, where $\overline{Q}$ assigns probability α to ρ^1 and $1-\alpha$ to ρ^2, and $\underline{Q}$ assigns probability α to ρ^3 and $1-\alpha$ to ρ^4. In other words, the belief set can be interpreted as the set of all sequences of independent distributions $\{\overline{Q}, \underline{Q}\}^{\infty}$. Fix a sequence of signals s. Inspired by the notion of $\hat{\rho}$, define $\hat{Q} \in \{\overline{Q}, \underline{Q}\}^{\infty}$ as the conjectured sequence of precision distributions. To show that ambiguity does not vanish, it suffices to show that the limit set of posterior means contains at least two elements. Consider the following two simple threshold assignments $\hat{Q}^*$ and $\hat{Q}^{**}$ of nature: $\hat{Q}^*_i = \overline{Q}$ if $s_i \geq 0$ and $\hat{Q}^*_i = \underline{Q}$ if $s_i < 0$; $\hat{Q}^{**}_i = \underline{Q}$ if $s_i \geq 0$ and $\hat{Q}^{**}_i = \overline{Q}$ if $s_i < 0$. We can show that posterior means resulting from these two assignments do not converge to the same limit and hence learning is not successful.

Distributions We have assumed that the state and signals are normally distributed. For general distributions, the precision of each signal is no longer fully captured by the reciprocal of its variance. To extend our model to other distributions, we can assume the DM considers a set of likelihood functions for each information source. As in the main model, each allocation of likelihoods to information sources defines a belief for the DM. Under prior-by-prior updating, for each belief, the DM forms a posterior on the state. The analysis would be less tractable since higher moments of the posterior no longer necessarily vanish asymptotically, but we conjecture that as long as two different beliefs result in two different posterior means, our results that ambiguity does not vanish hold. We leave the detailed analysis for future research.

Ambiguity Preferences Finally, we can consider more general preferences under ambiguity. As long as the ambiguity the DM faces takes the form of a set of beliefs over the state and signals, and she adopts the prior-by-prior updating rule upon receiving signals, Theorem 2 still holds. Indeed, our analysis of asymptotic beliefs does not rely on the specification of the DM's ambiguity preferences. For instance, the DM might use the α-maxmin expected utility (α-MEU) criterion (Hurwicz, 1951), where she considers the weighted average of each act's worst-case and best-case expected utility. With this preference, the DM might not be ambiguity-averse.

Appendix: Proofs

Proof of Proposition 1

We first show that $\mathbb{P}^B_\infty(s) = \{\delta_\theta\}$. Fix a realization of the state, $\theta \in \mathbb{R}$. We let θ^N represent the random variable that is distributed according to the subjective Bayesian posterior of the DM after observing signal s^N. Conditioning on any realized sequence of precisions, $\hat{\rho}^N$, we know θ^N is normally distributed with mean $\mu(s^N, \hat{\rho}^N) = \frac{\hat{\rho}^N \cdot s^N + \rho_\mu \mu}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}$, and variance $\sigma^2(\hat{\rho}^N) = \left(1 - \frac{\hat{\rho}^N \cdot \mathbb{1}^N}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}\right)\frac{1}{\rho_\mu}$. We can then write the characteristic function of θ^N, for any N and $t \in \mathbb{R}$:

$$\Psi_N(t) \equiv \mathbb{E}\left[e^{it\theta^N}\right] = \mathbb{E}_B\left[\mathbb{E}\left[e^{it\theta^N} | \hat{\rho}^N\right]\right] = \mathbb{E}_B\left[e^{it\mu(s^N, \hat{\rho}^N) - \frac{t^2}{2}\sigma^2(\hat{\rho}^N)}\right],$$

where the first equality follows from the law of iterated expectations, and the second equality holds because of the normality of θ^N conditional on $\hat{\rho}^N$.

We now take limits on this function to obtain that:

$$\lim_{N\to\infty} \Psi_N(t) \equiv \lim_{N\to\infty} \mathbb{E}_B\left[e^{it\mu(s^N,\hat{\rho}^N)-\frac{t^2}{2}\sigma^2(\hat{\rho}^N)}\right] = \mathbb{E}_B\left[\lim_{N\to\infty} e^{it\mu(s^N,\hat{\rho}^N)-\frac{t^2}{2}\sigma^2(\hat{\rho}^N)}\right] = e^{it\theta}.$$

The exchange of the limit operator and the expectation operator above follows from the dominated convergence theorem, since $|e^{it\mu(s^N,\hat{\rho}^N)-\frac{t^2}{2}\sigma^2(\hat{\rho}^N)}| \leq |e^{it\mu(s^N,\hat{\rho}^N)}| = 1$, by Euler's formula. For the last equality, note that $\sigma(\hat{\rho}^N) \to 0$ for all possible sequences of precision because $\underline{\rho} > 0$. Moreover,

$$\mu(s^N, \hat{\rho}^N) = \frac{\hat{\rho}^N \cdot s^N + \rho_\mu \mu}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu} = \frac{\hat{\rho}^N \cdot \mathbb{1}^N}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}\theta + \frac{\frac{\hat{\rho}^N \cdot \hat{\varepsilon}^N + \rho_\mu \mu}{N}}{\frac{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}{N}} \xrightarrow{a.s.} \theta,$$

by the strong law of large numbers for independent random variables with mean zero. Thus, $e^{it\mu(s^N,\hat{\rho}^N)-\frac{t^2}{2}\sigma^2(\hat{\rho}^N)} \xrightarrow{a.s.} e^{it\theta}$ by the continuous mapping theorem. We have now proved $\Phi_N(t)$ converges pointwise to $e^{it\theta}$. Levy's continuity theorem then implies $\theta^N \xrightarrow{p} \theta$ and hence $\mathbb{P}^B_\infty = \{\delta_\theta\}$.

To prove the estimate's consistency, denote the sample mean of signals by $\mu(s^N) = \frac{1}{N}\left(s^N \cdot \mathbb{1}^N\right)$. Because each ε_i is independent from one another, and their variances are uniformly bounded, a strong law of large number holds and implies $\mu(s^N) \xrightarrow{a.s.} \theta$. Given that $\mu(s^N)$ is a feasible estimate:

$$\mathbb{E}\left[u\left(g^*(s^N) - \theta\right)\middle|s^N\right] \leq \mathbb{E}\left[u\left(\mu(s^N) - \theta\right)\middle|s^N\right] \xrightarrow{a.s.} 0$$

Because ε is arbitrary, we have proved $\mathbb{E}[u(g^*(s^N) - \theta^N)] \xrightarrow{a.s.} 0$. Since u is strictly convex and has a unique minimizer at 0, we obtain $|g^*(s^N) - \theta^N| \xrightarrow{a.s.} 0$, which implies $g^*(s^N) \xrightarrow{a.s.} \theta$. $\square$

We prove an auxiliary result, which establishes that the supremum in Nature's problem and the infimum in the DM's problem are both achieved.

Lemma 2. *For each realization a^N of observables, there exist $g^*(a^N) \in \mathbb{R}$ and $p^*(a^N) \in \mathbb{P}^{\boldsymbol{a}}(a^N)$ such that:*

$$\inf_{\Gamma\in\Delta(\mathbb{R})} \sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p\left[\int_g u(g-\theta)d\Gamma\right] = \mathbb{E}_{p^*(a^N)}\left[u\left(g^*(a^N)-\theta\right)\right].$$

Proof of Lemma 2

We start by showing that it is without loss to focus on a deterministic estimates. For that, for any fixed a^N, consider a sequence Γ^n of potentially stochastic estimates such that:

$$\sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p\left[\int_g u\left(g-\theta\right)d\Gamma^n\right] \xrightarrow{n\to\infty} \inf_{\Gamma\in\Delta(\mathbb{R})} \sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p\left[\int_g u\left(g-\theta\right)d\Gamma^n\right].$$

Fix any posterior p. We then have, by convexity of u:

$$\mathbb{E}_p\left[u\left(\int_g g d\Gamma^n-\theta\right)\right] \leq \mathbb{E}_p\left[\int_g u\left(g-\theta\right)d\Gamma^n\right].$$

By taking supremum on both sides, we obtain:

$$\begin{aligned}\sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p\left[u\left(\int_g g d\Gamma^n-\theta\right)\right] &\leq \sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p\left[\int_g u\left(g-\theta\right)d\Gamma^n\right]\\ &\xrightarrow{n\to\infty} \inf_{\Gamma\in\Delta(\mathbb{R})} \sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p\left[\int_g u\left(g-\theta\right)d\Gamma\right],\end{aligned}$$

which implies that the left-hand side also converges to the infimum. We henceforth focus attention on deterministic estimates. Our next step is to show that the sup of Nature's problem is achieved. For that, notice that any $p\in\mathbb{P}^{\boldsymbol{a}}(a^N)$ is distributed normally according to:

$$\mathcal{N}\left(\frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s}^{\boldsymbol{a}}(a_i,\hat{\rho}_i)+\rho_\mu\mu}{\sum_{i=1}^N \hat{\rho}_i+\rho_\mu}, \left(1-\frac{\sum_{i=1}^N \hat{\rho}_i}{\sum_{i=1}^N \hat{\rho}_i+\rho_\mu}\right)\frac{1}{\rho_\mu}\right).$$

Consider a sequence of such posteriors, which are parameterized by the conjecture $\hat{\rho}^n=\left(\hat{\rho}_1^n,\hat{\rho}_2^n,\ldots,\hat{\rho}_N^n\right)$. We know that, because $\hat{\rho}^n$ is a finite sequence of numbers bounded between $[\underline{\rho}, \overline{\rho}]$, this sequence of precision conjectures has a converging subsequence:

$$\rho^{n_k} \to \hat{\rho},$$

for some $\hat{\rho} \in [\underline{\rho}, \overline{\rho}]^N$. Let p be the posterior obtained by conjecture $\hat{\rho}$. Now, it is clear that:

$$\mathbb{E}_{p^{n_k}}[\theta] = \frac{\sum_{i=1}^N \hat{\rho}_i^{n_k} \boldsymbol{s}^{\boldsymbol{a}}(a_i, \hat{\rho}_i^{n_k}) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i^{n_k} + \rho_\mu} \to \frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s}^{\boldsymbol{a}}(a_i, \hat{\rho}_i) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu} = \mathbb{E}_p[\theta],$$

and

$$\mathbb{V}_{p^{n_k}}[\theta] = \left(1 - \frac{\sum_{i=1}^N \hat{\rho}_i^{n_k}}{\sum_{i=1}^N \hat{\rho}_i^{n_k} + \rho_\mu}\right)\frac{1}{\rho_\mu} \to \left(1 - \frac{\sum_{i=1}^N \hat{\rho}_i}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}\right)\frac{1}{\rho_\mu} = \mathbb{V}_p[\theta].$$

It is well known that the convergence in mean and variance above implies $p^{n_k} \xrightarrow{d} p$. We have then shown that any sequence in $\mathbb{P}^{\boldsymbol{a}}(a^N)$ has a weakly converging subsequence with limit in $\mathbb{P}^{\boldsymbol{a}}(a^N)$. Thus, $\mathbb{P}^{\boldsymbol{a}}(a^N)$ is weakly compact.

Fix any $g \in \mathbb{R}$. Because all distributions $p \in \mathbb{P}^{\boldsymbol{a}}(a^N)$ are Gaussian with an uniformly bounded mean and variance, and because u has a bounded second-derivative, and is thus bounded above by a polynomial, for each $\varepsilon > 0$, there must exist a value $K > 0$ such that:

$$\sup_{p \in \mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p[u(g-\theta)\mathbb{1}_{|g-\theta|>K}] \leq \frac{\varepsilon}{2}.$$

Thus, take a sequence p^n such that $\mathbb{E}_{p^n}[u(g-\theta)]$ converges to the supremum. By our previous argument on sequential compactness, it is without loss of generality, up to a subsequence, to let $p^n \xrightarrow{d} p^* \in \mathbb{P}^{\boldsymbol{a}}(a^N)$. Then:

$$
\begin{aligned}
&\frac{\varepsilon}{2} \geq \mathbb{E}_{p^n}[u(g-\theta)] - \mathbb{E}_{p^n}[u(g-\theta)\mathbb{1}_{|g-\theta|\leq K}] \\
\xrightarrow{n\to\infty} &\sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p[u(g-\theta)] - \mathbb{E}_{p^*}[u(g-\theta)\mathbb{1}_{|g-\theta|\leq K}] \\
&\geq \sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p[u(g-\theta)] - \mathbb{E}_{p^*}[u(g-\theta)],
\end{aligned}
$$

where the first inequality comes from the definition of K; the first component after the limit follows from p^n being a sequence that takes the objective function to the sup; the second component after the limit follows from p^n weakly converging to p^* and $u(g-\theta)\mathbb{1}_{|g-\theta|\leq K}$ being a bounded function. Finally, the last inequality follows from u being non-negative. Because ε is arbitrary, we obtain that $\mathbb{E}^*_p[u(g-\theta)] = \sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p[u(g-\theta)]$, so the maximum is achieved.

To conclude, we prove that the inf of the DM's problem is achieved. Take a sequence g^n of nonrandom estimates such that

$$\max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p[u(g^n-\theta)] \to \inf_{g\in\mathbb{R}} \max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p[u(g-\theta)].$$

Notice that g^n cannot be unbounded. Indeed, in that case:

$$\max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p[u(g^n-\theta)] \geq \max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} u(g^n-\mathbb{E}_p[\theta]) \to \infty,$$

whereas for any fixed g the DM can obtain a finite loss, since the max is achieved. Therefore, g^n is bounded and, thus, we can assume it to be convergent, up to a subsequence. Let $g^n \to g^*$. Notice that, because the sequence g^n is bounded, we can choose a compact set G such that $g^n \in G$ for all n. Then, we can again argue that for all $\varepsilon > 0$, there exists a large number K with:

$$\sup_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N), g\in G} \mathbb{E}_p[u(g-\theta)\mathbb{1}_{|g-\theta|>K}] \leq \frac{\varepsilon}{2}.$$

Take the sequence p^n such that p^n solves the maximization problem for each g^n. The previous argument ensures the existence, up to a subsequence, of p^* such that $p^n \xrightarrow{d} p^*$. Therefore:

$$0 \leq \mathbb{E}_{p^n}[u(g^n-\theta)]-\mathbb{E}_{p^*}[u(g^n-\theta)] \to \inf_{g\in\mathbb{R}} \max_{p\in\mathbb{P}^{a}(a^N)} \mathbb{E}_p[u(g-\theta)]-\mathbb{E}_{p^*}[u(g^*-\theta)],$$

where the inequality follows from p^n maximizing losses for g^n, the first expression after the limit follows from the definition of g^n, and the second expression follows from the dominated convergence theorem and the fact that g^n is bounded and u is bounded above by a polynomial. We have thus proved that the inf is also achieved. $\square$

Below we prove a more general version of Lemma 1 by allowing for general observables introduced in Section 3.4.

Lemma 3. *For any sequence of realized observables a, if Assumption 1 holds and $\hat{\rho}^*$ solve $\max_{\hat{\rho}\in[\underline{\rho},\overline{\rho}]^\infty} u\big(g-\mathbb{E}[\theta|a,\hat{\rho}]\big)$ for some $g \in \mathbb{R}$, then $\hat{\rho}^*$ is a threshold assignment.*

Proof of Lemma 3

Since u is strictly convex and minimized at 0, it is easy to see that $\hat{\rho}^*$ solves either $\max_{\hat{\rho}\in[\underline{\rho},\overline{\rho}]^\infty} \mathbb{E}[\theta|a,\hat{\rho}]$ or $\min_{\hat{\rho}\in[\underline{\rho},\overline{\rho}]^\infty} \mathbb{E}[\theta|a,\hat{\rho}]$.

Given an infinite sequence of observables, a, we have that:

$$\mathbb{E}[\theta|a,\hat{\rho}] = \frac{\sum_i \hat{\rho}_i s^{a}(a_i,\hat{\rho}_i)}{\sum_i \hat{\rho}_i}.$$

First, we show that it is without loss of generality to assume that if two observables are identical, the precision assigned to them is also identical: $a_i = a_j \implies \hat{\rho}_i = \hat{\rho}_j$. Fix $\hat{\rho}$. For any $A \in \mathbb{R}$, let $\mathcal{A} = \{j : a_j = A\}$, and consider $\hat{\rho}_i' = \frac{\sum_{j\in\mathcal{A}}\hat{\rho}_j}{|\mathcal{A}|}$, for all $i \in \mathcal{A}$. Let $\hat{\rho}_i' = \hat{\rho}_i$, for $i \notin \mathcal{A}$. Then:

$$\begin{aligned}
\mathbb{E}[\theta|a,\hat{\rho}] &= \frac{\sum_{i\in\mathcal{A}}\hat{\rho}_i \boldsymbol{s^a}(a_i,\hat{\rho}_i) + \sum_{i\notin\mathcal{A}}\hat{\rho}_i \boldsymbol{s^a}(a_i,\hat{\rho}_i)}{\sum_{i\in\mathcal{A}}\hat{\rho}_i + \sum_{i\notin\mathcal{A}}\hat{\rho}_i} \\
&= \frac{\sum_{i\in\mathcal{A}}(\hat{\rho}_i\beta(A)+\gamma(A)) + \sum_{i\notin\mathcal{A}}\hat{\rho}_i \boldsymbol{s^a}(a_i,\hat{\rho}_i)}{|\mathcal{A}|\frac{\sum_{i\in\mathcal{A}}}{|\mathcal{A}|}\hat{\rho}_i + \sum_{i\notin\mathcal{A}}\hat{\rho}_i} \\
&= \frac{|\mathcal{A}|\left(\frac{\sum_{i\in\mathcal{A}}\hat{\rho}_i}{|\mathcal{A}|}\beta(A)+\gamma(A)\right) + \sum_{i\notin\mathcal{A}}\hat{\rho}_i \boldsymbol{s^a}(a_i,\hat{\rho}_i)}{|\mathcal{A}|\frac{\sum_{i\in\mathcal{A}}}{|\mathcal{A}|}\hat{\rho}_i + \sum_{i\notin\mathcal{A}}\hat{\rho}_i} \\
&= \frac{\sum_{i\in\mathcal{A}}\hat{\rho}'_i \boldsymbol{s^a}(a_i,\hat{\rho}'_i) + \sum_{i\notin\mathcal{A}}\hat{\rho}_i \boldsymbol{s^a}(a_i,\hat{\rho}_i)}{\sum_{i\in\mathcal{A}}\hat{\rho}'_i + \sum_{i\notin\mathcal{A}}\hat{\rho}_i} = \mathbb{E}[\theta|a,\hat{\rho}'],
\end{aligned}$$

where the second equality follows from Assumption 1 and the definition of $\mathcal{A}$. Because A was arbitrary, we have shown that we can restrict attention to assignments that give the same precision to identical observables. That allows us to define assignments as allocations of precisions to realized observables. We henceforth consider empirical distributions of observables $F \in \Delta(\mathbb{R})$ generated by the sequence of observables a, and we abuse notation to write $\hat{\rho} : \mathbb{R} \to [\underline{\rho}, \overline{\rho}]$ as a precision assignment. In the space of distributions over observables, we can rewrite:

$$v(\hat{\rho}) \equiv \mathbb{E}[\theta|a,\hat{\rho}] = \int \frac{\hat{\rho}(x)\boldsymbol{s^a}(x,\hat{\rho}(x))dF(x)}{\int \hat{\rho}(x)dF(x)}.$$

Fix a value $M \in [\underline{\rho}, \overline{\rho}]$ and consider the problem:

$$\begin{aligned}
&\max_{\hat{\rho}}\left\{v(\hat{\rho}) : \int \hat{\rho}(x)dF(x) = M\right\} \\
=&\frac{1}{M}\max_{\hat{\rho}}\left\{\int \hat{\rho}(x)\boldsymbol{s^a}(x,\hat{\rho}(x))dF(x) : \int \hat{\rho}(x)dF(x) = M\right\}
\end{aligned}$$

where the last equality is justified because we are equating the denominator of v to M. By the Lagrange multiplier theorem in Banach spaces, we obtain that there is $\lambda \in \mathbb{R}$ such that, for almos all $x \in \operatorname{supp} F$:

$$\hat{\rho}(x) \in \arg\max_{\rho \in [\underline{\rho}, \overline{\rho}]} \{\rho \boldsymbol{s^a}(x, \rho) - \lambda(\rho - M)\}.$$

By Assumption 1, this problem can be rewritten as:

$$\hat{\rho}(x) \in \arg\max_{\rho \in [\underline{\rho}, \overline{\rho}]} \{\rho\beta(x) + \gamma(x) - \lambda(\rho - M)\}.$$

Because β is increasing, the objective function of each of these optimizations is supermodular in (ρ, x), so $\hat{\rho}$ is increasing with x, according to Topkis' lemma. Because each of these programs is linear, the solution can be assumed to be an extreme point of the interval $[\underline{\rho}, \overline{\rho}]$. Therefore, for each M, the solution is a threshold assignment. Thus, maximizing over M's implies that the solution must also be a threshold assignment. Clearly, the symmetric result holds for minimization problem. This completes the proof.

Lemma 1 is a special case of Lemma 3 and hence is also proved. □

Proof of Theorem 1 and Theorem 2

Theorem 1 is a special case of Theorem 2, so we provide a proof for the more general result.

For any realization of observables a^N, let $F^N \in \Delta(\mathbb{R})$ be the empirical distribution of observables. We abuse notation to write $s^a(a^N, \hat{\rho}^N)$ as the vector in which the i-th entry is $s^a(a_i^N, \hat{\rho}_i^N)$. Given a conjecture $\hat{\rho}^N$, we know the backtracked signals $s^a(a_i^N, \hat{\rho}_i^N)$ are jointly normal with the state, allowing us to calculate the posterior mean as:

$$\mathbb{E}[\theta | a^N, \hat{\rho}^N] = \frac{\hat{\rho}^N \cdot s^a(a^N, \hat{\rho}^N) + \rho_\mu \mu}{\hat{\rho}^N \cdot \mathbb{1} + \rho_\mu}.$$

Define:

$$\underline{m}^N \equiv \min_{\hat{\rho} \in [\underline{\rho}, \overline{\rho}]^N} \mathbb{E}[\theta | a^N, \hat{\rho}^N] \ , \ \overline{m}^N \equiv \max_{\hat{\rho} \in [\underline{\rho}, \overline{\rho}]^N} \mathbb{E}[\theta | s^N, \hat{\rho}^N].$$

The above $\underline{m}^N$ and $\overline{m}^N$ are (random) bounds on posterior means. Assume that $\underline{\rho}^N$ and $\overline{\rho}^N$ are the minimizer and maximizer, respectively.

By the proof of Lemma 3, let $\hat{\rho} : \mathbb{R} \to [\underline{\rho}, \overline{\rho}]$ be a precision assignment. Let F be the true distribution of observables. Again, given a precision assignment, the posterior mean can be written as:

$$\mathbb{E}[\theta|\hat{\rho}] = \int \frac{\hat{\rho}(x) s^a(x, \hat{\rho}(x)) dF(x)}{\int \hat{\rho}(x) dF(x)}.$$

Finally, let:

$$\underline{m} = \min_{\hat{\rho}:\mathbb{R}\to[\underline{\rho},\overline{\rho}]} \mathbb{E}[\theta|\hat{\rho}] \text{ and } \overline{m} = \min_{\hat{\rho}:\mathbb{R}\to[\underline{\rho},\overline{\rho}]} \mathbb{E}[\theta|\hat{\rho}].$$

We start the proof by showing, in **Step 1**, that the random bounds on posterior means converge to $\underline{m}$ and $\overline{m}$ asymptotically. Then, we show that the $\underline{m}$ and $\overline{m}$ are indeed asymptotic bounds of posterior means, proving part 1 of the Theorem in **Step 2**.

Step 1. $\underline{m}^N \xrightarrow{a.s.} \underline{m}$ **and** $\overline{m}^N \xrightarrow{a.s.} \overline{m}$**.**

Step 1.1. $\overline{m} = \frac{\underline{\rho}\int_{-\infty}^{\beta^{-1}(\overline{m})} \boldsymbol{s^a}(x,\underline{\rho})dF(x) + \overline{\rho}\int_{\beta^{-1}(\overline{m})}^{\infty} \boldsymbol{s^a}(x,\overline{\rho})dF(x)}{\underline{\rho}F(\beta^{-1}(\overline{m})) + \overline{\rho}\left(1-F(\beta^{-1}(\overline{m}))\right)}$.

By the proof of Lemma 3, $\overline{m}$ is achieved by a threshold assignment. We can then write the corresponding optimization problem by

$$\overline{m} = \max_{a\in\mathbb{R}} \{v(a)\},$$

where $v(a) = \frac{\underline{\rho}\int_{-\infty}^{a} \boldsymbol{s^a}(x,\underline{\rho})dF(x) + \overline{\rho}\int_{a}^{\infty} \boldsymbol{s^a}(x,\overline{\rho})dF(x)}{\underline{\rho}F(a) + \overline{\rho}(1-F(a))}$.

Using Assumption 1, the first order condition leads to:

$$\beta(\overline{a}) = \frac{\underline{\rho}\int_{-\infty}^{\overline{a}} \boldsymbol{s^a}(x,\underline{\rho})dF(x) + \overline{\rho}\int_{\overline{a}}^{\infty} \boldsymbol{s^a}(x,\overline{\rho})dF(x)}{\underline{\rho}F(\overline{a}) + \overline{\rho}\left(1-F(\overline{a})\right)},$$

which implicitly defines the threshold value $\overline{a}$, in the space of observables, that solves that maximization. If $\overline{a}$ is indeed the optimal threshold, then by noticing that the right-hand side of the equation above is $v(\overline{a}) = \overline{m}$, we obtain $\overline{m} = \beta(\overline{a})$, which proves the expression for $\overline{m}$ in the statement of the theorem. In what follows, we prove that $\overline{a}$ as defined above exists and is the optimal threshold.

We start by showing that the objective function v is quasiconcave, so that the first order condition is sufficient. The first derivative of v can be written as:

$$v'(a) = (v(a) - \beta(a))\frac{(\overline{\rho} - \underline{\rho})f(a)}{\underline{\rho}F(a) + \overline{\rho}(1 - F(a))}.$$

First, notice that because the second term is positive for all $a \in \mathbb{R}$, the sign of v' is determined by $v(a) - \beta(a)$. This implies that v is quasiconcave: If there is $\overline{a}$ such that $v'(\overline{a}) < 0$, then $v'(a) < 0$ for all $a \geq \overline{a}$. To see this, assume there is $\overline{a}$ such that $v'(\overline{a}) < 0$ and, to obtain a contradiction, let there be $a > \overline{a}$ with $v'(a) \geq 0$. Let $b = \min\{x \in [\overline{a}, a] : v'(x) \geq 0\}$. By continuity of v, b is well-defined and $v'(b) = 0$—that is, $v(b) = \beta(b)$. We then have:

$$0 = v(b) - \beta(b) < v(b) - \beta(\overline{a}) = v(\overline{a}) - \beta(\overline{a}) + \int_{\overline{a}}^{b} v'(x)dx < 0.$$

Where the first inequality above comes from monotonicity of β, and the second one holds because $v'(\overline{a}) < 0$ implies $v(\overline{a}) < \beta(\overline{a})$. We have thus obtained a contradiction.

We now show that the solution to the first order condition above, $\overline{a}$, exists and is unique.

As $a \to -\infty$, $v(a) \to \int \boldsymbol{s}^{\boldsymbol{a}}(x, \underline{\rho})dF(x)$, which is finite by assumption. Because β is surjective by Assumption 1 and strictly increasing, that implies that we can find a sufficiently small number $\underline{a}$ such that $v(\underline{a}) - \beta(\underline{a}) > 0$, implying $v'(\underline{a}) > 0$. Notice that the same should be true for all $a \leq \underline{a}$, so that v is a strictly increasing function in $(-\infty, \underline{a}]$.

On the other hand, as $a \to \infty$, again we have $v(a) \to \int \boldsymbol{s}^{\boldsymbol{a}}(x, \overline{\rho})dF(x)$, which is also finite by assumption. Then, there is a sufficiently high number $\overline{a}$ with $v(a) - \beta(a) < 0$, so $v'(a) < 0$ for all $a \geq \overline{a}$.

Because v' is continuous, there is $a^* \in [\underline{a}, \overline{a}]$ with $v'(a^*) = 0$, so the solution exists. We now prove uniqueness. Let a' satisfy $v'(a') = 0$, and let $a' > a^*$ without loss of generality. By the quasiconcavity argument above, $v'(x) = 0$ for all $x \in [a^*, a']$. Then

$$0 = v(a') - \beta(a') < v(a') - \beta(a^*) = v(a^*) - \beta(a^*) + \int_{a^*}^{a'} v'(x)dx = 0,$$

again, yielding a contradiction. Therefore a^* is unique. This concludes Step 1.1. By symmetry, we have the definition of $\underline{m}$.

Step 1.2. Approximating $\overline{m}^N$ using a threshold. In this step we show how to approximate $\overline{m}^N$ using the expectation generated by a threshold strategy as N grows large. For any realized sequence of observables, a^N, let F^N be the associated empirical distribution of observables. We then define:

$$\tilde{m}^N = \max_{a \in \mathbb{R}} \frac{\underline{\rho} \int_{-\infty}^{a} \boldsymbol{s}^{\boldsymbol{a}}(x, \underline{\rho}) dF^N(x) + \overline{\rho} \int_{a}^{\infty} \boldsymbol{s}^{\boldsymbol{a}}(x, \overline{\rho}) dF^N(x)}{\underline{\rho} F^N(a) + \overline{\rho}(1 - F^N(a))}$$

We call the objective function of the problem above $\Psi^N(a)$.

At the same time, using the proof of Lemma 3 without assuming the distribution of observables is non-atomic, we obtain that $\overline{m}^N$ can be obtained by an assignment that is a threshold except for possibly one of the observables receiving an intermediate precision. Thus, we can find $\overline{m}^N$ through the alternative optimization:

$$\overline{m}^N = \max_{a, \rho \in [\underline{\rho}, \overline{\rho}]} \left\{ \frac{\underline{\rho} \int_{-\infty}^{a-} \boldsymbol{s}^{\boldsymbol{a}}(x, \underline{\rho}) dF^N(x) + \rho \boldsymbol{s}^{\boldsymbol{a}}(a, \rho)(F^N(a) - F^N(a-)) + \overline{\rho} \int_{a}^{\infty} \boldsymbol{s}^{\boldsymbol{a}}(x, \overline{\rho}) dF^N(x) + \frac{\rho_\mu}{N}\mu}{\underline{\rho} F^N(a-) + \rho(F^N(a) - F^N(a-)) + \overline{\rho}(1 - F^N(a)) + \frac{\rho_\mu}{N}} \right\}$$

We call the objective function above $\tilde{\Psi}^N(a,\rho)$. We next prove

$$\sup_{a\in\mathbb{R},\rho\in[\underline{\rho},\overline{\rho}]} |\tilde{\Psi}^N(a,\rho)-\Psi^N(a)| \xrightarrow{a.s} 0.$$

To see this, notice that for almost any sequences of realized observables a^N, it must be that $\sup_a \left\{F^N(a)-F^N(a-)\right\} \le \frac{1}{N}$. This observation directly leads to the uniform convergence result.

Denote

$$\Psi(a) = \frac{\underline{\rho}\int_{-\infty}^{a} \boldsymbol{s^a}(x,\underline{\rho})dF(x)+\overline{\rho}\int_{a}^{\infty} \boldsymbol{s^a}(x,\overline{\rho})dF(x)}{\underline{\rho}F(a)+\overline{\rho}(1-F(a))}.$$

where F is, again, the true distribution of observables.

Step 1.3. $\sup_{\mathbf{a}\in\mathbb{R}} |\boldsymbol{\Psi}^{\mathbf{N}}(\mathbf{a})-\boldsymbol{\Psi}(\mathbf{a})| \xrightarrow{\mathbf{a.s.}} \mathbf{0}.$ Given the Glivenko–Cantelli theorem, we know that the empirical distribution function converges to the true cumulative distribution function uniformly over x, that is,

$$\|F^N - F\| := \sup_{x\in\mathbb{R}} |F^N(x)-F(x)| \xrightarrow{a.s.} 0.$$

For each real-valued function v, denote

$$F^N(v) = \int v dF^N \ \ , \ \ F(v) = \int v dF.$$

A class of real-valued functions $\mathcal{V}$ is defined to be a *P-Glivenko-Cantelli class of functions* if

$$\|F^N - F\|_{\mathcal{V}} := \sup_{v\in\mathcal{V}} |F^N(v)-F(v)| \xrightarrow{a.s.} 0.$$

Recall that the $L_1(F)$ norm is defined for real-valued functions such that

$$\|v\|_{L_1(F)} = \int |v| dF.$$

Given two real-valued functions l and u and $\epsilon > 0$, a *ε-bracket* $[l,u]$ is the set of all functions f such that $l \leq f \leq u$ and $\|u-l\|_{L_1(F)} \leq \varepsilon$. The *bracketing number* $N(\varepsilon, \mathcal{V}, \|\cdot\|_{L_1(F)})$ is the minimum number of ε-brackets needed to cover $\mathcal{V}$. The following theorem provides a sufficient condition for a P-Glivenko-Cantelli class.

Theorem 4. *((Blum, 1955; DeHardt, 1971)) If $N(\varepsilon, \mathcal{V}, \|\cdot\|_{L_1(F)}) < \infty$ for any $\varepsilon > 0$, then $\mathcal{V}$ is a P-Glivenko-Cantelli class.*

Denote

$$\mathcal{V}_1 = \left\{ v_1^a : v_1^a(x) = \underline{\rho} \mathbb{1}_{\{x \leq a\}} + \overline{\rho} \mathbb{1}_{\{x > a\}}, \forall x \in \mathbb{R}, \text{ for some } a \in \mathbb{R} \right\},$$

$$\mathcal{V}_2 = \left\{ v_2^a : v_2^a(x) = \underline{\rho} x \mathbb{1}_{\{x \leq a\}} + \overline{\rho} x \mathbb{1}_{\{x > a\}}, \forall x \in \mathbb{R}, \text{ for some } a \in \mathbb{R} \right\}.$$

It is easy to see that

$$\Psi^N(a) = \frac{F^N(v_2^a)}{F^N(v_1^a)} \text{ and } \Psi(a) = \frac{F(v_2^a)}{F(v_1^a)}.$$

We want to show that both $\mathcal{V}_1$ and $\mathcal{V}_2$ are P-Glivenko-Cantelli classes. Note that F is a continuous distribution whose expectation is well-defined, that is, $\int |x| dF < \infty$.

Fix $\varepsilon > 0$. For any $a > b$, the $L_1(F)$-distance between v_1^a and v_1^b is

$$\|v_1^a - v_1^b\|_{L_1(F)} = (\overline{\rho} - \underline{\rho}) \int_b^a dF(x).$$

Since $\int_{-\infty}^{\infty} dF(x) = 1$, for M large enough, we can find a finite increasing sequence $\{a_1, \ldots, a_M\}$ on the extended real line such that $a_1 = -\infty$, $a_M = \infty$ and

$$\int_{a_i}^{a_{i+1}} dF(x) = \frac{1}{M-1} \leq \frac{\varepsilon}{\overline{\rho} - \underline{\rho}}, \forall i = 1, \ldots, M-1$$

This is feasible as F is a continuous distribution. Then it is easy to show that the set of ε-brackets $\{[v_1^{a_i}, v_1^{a_{i+1}}] : i = 1, \ldots, M-1\}$ covers $\mathcal{V}_1$ and $N(\varepsilon, \mathcal{V}_1, \|\cdot\|_{L_1(F)}) \leq M-1 < \infty$. Hence $\mathcal{V}_1$ is a P-Glivenko-Cantelli class.

Similarly, for any $a > b$, the $L_1(F)$-distance between v_2^a and v_2^b is

$$\|v_2^a - v_2^b\|_{L_1(P)} = (\overline{\rho} - \underline{\rho}) \int_b^a |x| dF(x).$$

Since $\int |x| dF < \infty$ and F is continuous, for M' large enough, again we can fine a finite increasing sequence $\{b_1, ..., b_{M'}\}$ on extended real line such that $b_1 = -\infty$, $b_{M'} = \infty$ and

$$\int_{b_i}^{b_{i+1}} |x| dF(x) = \frac{\int |x| dF}{M' - 1} \leq \frac{\varepsilon}{\overline{\rho} - \underline{\rho}}, \forall i = 1, ..., M' - 1.$$

Then it is easy to show that the set of ε-brackets $\{[v_2^{b_i}, v_2^{b_{i+1}}] : i = 1, ..., M' - 1\}$ covers $\mathcal{F}_2$ and $N(\varepsilon, \mathcal{V}_2, \|\cdot\|_{L_1(F)}) \leq M' - 1 < \infty$. Hence $\mathcal{V}_2$ is a P-Glivenko-Cantelli class.

The definition of the P-Glivenko-Cantelli class implies that

$$\|F^N - F\|_{\mathcal{V}_1} = \sup_{v \in \mathcal{V}_1} |F^N(v) - F(v)| = \sup_{a \in \mathbb{R}} |F^N(v_1^a) - F(v_1^a)| \xrightarrow{a.s.} 0. \tag{2}$$

$$\|F^N - F\|_{\mathcal{V}_1} = \sup_{v \in \mathcal{V}_1} |F^N(v) - F(v)| = \sup_{a \in \mathbb{R}} |F^N(v_1^a) - F(v_1^a)| \xrightarrow{a.s.} 0. \tag{3}$$

Now we can show the convergence of Ψ^N:

$$\begin{aligned}
&\sup_{a\in\mathbb{R}}|\Psi^N(a)-\Psi(a)| \\
=&\sup_{a\in\mathbb{R}}|\frac{F^N(v_2^a)}{F^N(v_1^a)}-\frac{F(v_2^a)}{F(v_1^a)}| \\
\leq&\sup_{a\in\mathbb{R}}|\frac{F^N(v_2^a)}{F^N(v_1^a)}-\frac{F^N(v_2^a)}{F(v_1^a)}|+\sup_{a\in\mathbb{R}}|\frac{F^N(v_2^a)}{F(v_1^a)}-\frac{F(v_2^a)}{F(v_1^a)}| \\
\leq&\sup_{a\in\mathbb{R}}|\frac{F^N(v_2^a)}{F(v_1^a)F^N(v_1^a)}||F^N(v_1^a)-F(v_1^a)|+\sup_{a\in\mathbb{R}}\frac{1}{|F(v_1^a)|}|F^N(v_2^a)-F(v_2^a)| \\
\leq&\sup_{a\in\mathbb{R}}|\frac{F^N(v_2^a)}{F(v_1^a)F^N(v_1^a)}|\sup_{a\in\mathbb{R}}|F^N(v_1^a)-F(v_1^a)|+\sup_{a\in\mathbb{R}}\frac{1}{|F(v_1^a)|}\sup_{a\in\mathbb{R}}|F^N(v_2^a)-F(v_2^a)|.
\end{aligned}$$

Notice that $0<\underline{\rho}\leq F(v_1^a)\leq\overline{\rho}<\infty$ and $0<\underline{\rho}\leq F^N(v_1^a)\leq\overline{\rho}<\infty$ for each N. That is, $F(v_1^a)$ and $F^N(v_1^a)$ are uniformly bounded away from 0 and ∞. Also, by applying strong law of large numbers,

$$\sup_{a\in\mathbb{R}}|F^N(v_2^a)|\leq(\underline{\rho}+\overline{\rho})\int|x|dF^N\xrightarrow{a.s.}(\underline{\rho}+\overline{\rho})\int|x|dF<+\infty.$$

By equations 2 and 3, we know

$$\sup_{a\in\mathbb{R}}|\Psi^N(a)-\Psi(a)|\xrightarrow{a.s.}0.$$

Step 1.4. $\overline{m}^N\xrightarrow{a.s.}\overline{m}$**.** This result follows directly from the following standard results about consistency of M–estimators. We include the proof for completeness.

Lemma 4. *Suppose that*

1. $\sup_{a\in\mathbb{R},\rho\in[\underline{\rho},\overline{\rho}]}|\tilde{\Psi}^N(a,\rho)-\Psi(a)|\xrightarrow{a.s.}0,$
2. $\overline{m}^N\in\arg\max_{a\in\mathbb{R},\rho\in[\underline{\rho},\overline{\rho}]}\tilde{\Psi}^N(a,\rho)$ *for each* N*,*
3. $\overline{m}=\arg\max_{a\in\mathbb{R}}\Psi(a)$ *is the unique maximum of* Ψ*,*

Then $\overline{m}^N\xrightarrow{a.s.}\overline{m}.$

Proof of Lemma 4. We ignore the argument ρ throughout the proof without loss of generality. By conditions (2) and (3), we know $\tilde{\Psi}^N(\overline{m}^N) \geq \tilde{\Psi}^N(\overline{m})$ and $\Psi(\overline{m}) \geq \Psi(\overline{m}^N)$ for each N. Using these inequalities we have

$$\tilde{\Psi}^N(\overline{m}^N) - \Psi(\overline{m}^N) \geq \tilde{\Psi}^N(\overline{m}^N) - \Psi(\overline{m}) \geq \tilde{\Psi}^N(\overline{m}) - \Psi(\overline{m})$$

Therefore from the above we have

$$|\tilde{\Psi}^N(\overline{m}^N) - \Psi(\overline{m})| \leq \max\left\{|\tilde{\Psi}^N(\overline{m}^N) - \Psi(\overline{m}^N)|, |\tilde{\Psi}^N(\overline{m}) - \Psi(\overline{m})|\right\} \leq \sup_{a \in \mathbb{R}} |\tilde{\Psi}^N(a) - \Psi(a)|$$

Hence by condition (1), we know $|\tilde{\Psi}^N(\overline{m}^N) - \Psi(\overline{m})| \xrightarrow{a.s.} 0$. Finally, consider any ω in set M such that $|\tilde{\Psi}^N(\overline{m}^N(\omega)) - \Psi(\overline{m}(\omega))| \to 0$. For that ω:

$$\Psi(\overline{m}^N) - \Psi(\overline{m}) = \left(\Psi(\overline{m}) - \tilde{\Psi}^N(\overline{m}^N)\right) + \left(\tilde{\Psi}^N(\overline{m}^N) - \Psi(\overline{m}^N)\right).$$

Each term in parentheses in the right hand side converges to zero because of the definition of the set M and condition (1), respectively. Thus, the left hand side must converge to zero. Since $\overline{m}$ is the unique minimizer of Ψ, we get $\overline{m}^N \to 0$ for any $\omega \in M$. Because M set with probability one, this convergence is almost sure. $\square$

Now it suffices to show that the conditions in Lemma 4 hold in our case. Condition (1) is shown in Step 1.2 and 1.3. Explicitly, $\sup_{a,\rho} |\tilde{\Psi}^N(a,\rho) - \Psi(a)| \xrightarrow{a.s} 0$ and $\sup_a |\Psi^N(a) - \Psi(a)| \xrightarrow{a.s.} 0$ imply that condition. Condition (2) holds by the definition of $\overline{m}^N$. Condition (3) is shown in the proof of Step 1.1. This completes the proof for $\overline{m}^N \xrightarrow{a.s.} \overline{m}$. The same arguments apply for showing $\underline{m}^N \xrightarrow{a.s.} \underline{m}$.

Step 2. Part 1 of Theorem: Boundedness of Posterior Means. For any N, with observables a^N and conjectured precisions $\hat{\rho}^N$, recall we have:

$$\theta | s^N, \hat{\rho}^N \sim \mathcal{N}\left(\frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s}^{\boldsymbol{a}}(a_i, \hat{\rho}_i) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}, \left(1 - \frac{\sum_{i=1}^N \hat{\rho}_i}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}\right)\frac{1}{\rho_\mu}\right). \tag{4}$$

Since $\hat{\rho}_i \geq \underline{\rho} > 0$, it is clear that $\lim_{N\to\infty} \frac{\sum_{i=1}^N \hat{\rho}_i}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu} = 1$, so the variance converges to zero for all sequences of signal realizations.

As for the posterior mean, notice that, by definition of $\underline{m}^N, \overline{m}^N$:

$$\underline{m}^N \leq \frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s}^{\boldsymbol{a}}(a_i, \hat{\rho}_i) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu} \leq \overline{m}^N.$$

By taking limit inferior in the first inequality above and limit superior in the second, we obtain, using the result in Step 2, that for almost all sequences of signal realizations, the asymptotic bounds on expected values hold.

Step 3. Part 2 of Theorem: Limit Set of Posteriors. Fix a sequence of realizations a. We want to characterize the set of distributions the posterior beliefs of the DM converge to, $\mathbb{P}_\infty(a)$. By (4), it is clear that a necessary condition for weak convergence is that the posterior mean $\frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s}^{\boldsymbol{a}}(a_i, \hat{\rho}_i) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}$ converges. We can then focus on sequences with convergent means. Define $b = \lim_{N\to\infty} \frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s}^{\boldsymbol{a}}(a_i, \hat{\rho}_i) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}$.

We can write the characteristic function of $P_N(s^N, \hat{\rho}^N)$ as:

$$\varphi^N(t) = e^{it\left\{\frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s}^{\boldsymbol{a}}(a_i, \hat{\rho}_i) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu} - \frac{1}{2}\left(1 - \frac{\sum_{i=1}^N \hat{\rho}_i}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}\right)\frac{1}{\rho_\mu}\right\}}$$

By Step 2, the variance converges to zero. We then have, for all t:

$$\varphi^N(t) \to e^{itb}$$

which is the characteristic function of δ_b. Then, by Levy's continuity theorem: $P_N(s^N, \hat{\rho}^N) \xrightarrow{w} \delta_b$.

We finally show that any $b \in [\underline{m}, \overline{m}]$ can be achieved. For that, fix a threshold assignment $\rho : \mathbb{R} \to \{\underline{\rho}, \overline{\rho}\}$. Then $\{\rho(a_i)\boldsymbol{s}^{\boldsymbol{a}}(a_i, \rho(a_i))\}_{i\geq 1}$ is a sequence of independent signals with uniformly bounded variance. Then, by the strong law of large numbers:

$$\frac{\sum_{i=1}^{N}\rho(s_i)\boldsymbol{s^a}(a_i,\rho(a_i))+\rho_\mu\mu}{\sum_{i=1}^{N}\rho(s_i)+\rho_\mu}=\frac{N\int\rho(x)\boldsymbol{s^a}(x,\rho(x))dF^N(x)+\rho_\mu\mu}{N\int\rho(x)dF^N(x)+\rho_\mu}$$
$$\xrightarrow{a.s.}\frac{\int\rho(x)\boldsymbol{s^a}(x,\rho(x))dF(x)}{\int\rho(x)dF(x)}.$$

We finish this step by showing that by appropriately choosing the assignment ρ, the function $\frac{\int\rho(x)\boldsymbol{s^a}(x,\rho(x))dF(x)}{\int\rho(x)dF(x)}$ can achieve any point between $\underline{m}$ and $\overline{m}$. To see that, recall that $\overline{m}=\max_a\Psi(a)$. We can show that $\min_a\Psi(a)=\min\{\int\boldsymbol{s^a}(x,\overline{\rho})dF(x),\int\boldsymbol{s^a}(x,\underline{\rho})dF(x)\}$. Since Ψ is continuous, by choosing different a's, any number in $[\min_a\Psi(a),\overline{m}]$ can be achieved. Similarly, for the function that defines $\underline{m}$, we can show that it can achieve any value in $[\underline{m},\max\{\int\boldsymbol{s^a}(x,\overline{\rho})dF(x),\int\boldsymbol{s^a}(x,\underline{\rho})dF(x)\}]$. Therefore, any value between $[\underline{m},\overline{m}]$ can be achieved by $\frac{\int\rho(x)\boldsymbol{s^a}(x,\rho(x))dF(x)}{\int\rho(x)dF(x)}$. This completes the proof. □

Proof of Theorem 3

We start by proving that the optimal estimate is deterministic. For any fixed N, and observable realization a^N, recall that the DM's problem is to choose a distribution $\Gamma^*(a^N)$ to solve:

$$\min_{\Gamma\in\Delta(\mathbb{R})}\max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)}\mathbb{E}_p\Big[\int_g u(g-\theta)d\Gamma\Big].$$

Fix any non-degenerate distribution of estimates, Γ, and let $\tilde{g}=\int g d\Gamma$ be its expected value. Note that, for any $p\in\mathbb{P}^{\boldsymbol{a}}(a^N)$:

$$\int u(\tilde{g}-\theta)dp<\int\int u(g-\theta)d\Gamma dp,$$

by Jensen's inequality and strict convexity of u. By taking supremum with respect to the compact set $\mathbb{P}^{\boldsymbol{a}}(a^N)$ and noticing that the objective function on the left-hand-side is continuous, we obtain:

$$\max_{p \in \mathbb{P}^{a}(a^N)} \int u(\tilde{g}-\theta) dp < \sup_{p \in \mathbb{P}^{a}(a^N)} \int \int u(g-\theta) d\Gamma dp.$$

Therefore, any distribution over estimates is strictly outperformed by its expected value, and the optimal estimate is deterministic. Henceforth, we take as given that the optimal estimate is deterministic to prove the rest of the result.

Define

$$\chi^N(g) \equiv \max_{p \in \mathbb{P}^{a}(a^N)} \mathbb{E}_p \left[u(g-\theta) \right].$$

By definition, assuming that the limits exist, we have:

$$\lim_{N \to \infty} g^*(s^N) = \lim_{N \to \infty} \arg\min_{g} \chi^N(g).$$

Also denote

$$\chi(g) = \max\{u(g-\overline{m}), u(g-\underline{m})\}$$

where $\underline{m}$ and $\overline{m}$ are defined in Theorem 2.

We start with introducing an auxiliary problem with finitely many signals by ignoring the effect of any moment of the posterior distribution that is not the mean. Explicitly:

$$\tilde{\chi}^N(g) \equiv \max_{p \in \mathbb{P}^{a}(a^N)} u(g-\mathbb{E}_p[\theta]) = \max\left\{u(g-\overline{m}^N), u(g-\underline{m}^N)\right\}$$

where $\overline{m}^N$ and $\underline{m}^N$ are defined in the proof of Theorem 2 and the equality follows from the fact that u is convex.

Lemma 5. *Let f^N be a sequence of real-valued random mappings such that $x^N \in \arg\min_{x \in \mathbb{R}} f^N(x)$, for all $N \in \mathbb{N}$. Assume there is another random mapping f and that the following are satisfied:*

1. $\sup_{x \in C} |f(x) - f^N(x)| \xrightarrow{a.s} 0$, *as $N \to \infty$, for all compact sets $C \subset \mathbb{R}$.*

2. *$x^* \in \arg\min_{x\in\mathbb{R}} f(x)$ is the unique minimum of f.*

3. *The sequence x^N is bounded almost everywhere.*

Then $x^N \xrightarrow{a.s} x^*$.

Proof of Lemma 5. By condition (3), there exists an event M with $\mathbb{P}(M) = 1$ such that for all $\omega \in M$, there is a compact set $C(\omega) \subseteq \mathbb{R}$ with $\{x^N(\omega)\}_{N\geq 1} \cup \{x^*(\omega)\} \subseteq C(\omega)$. By condition (1), we can find $M' \subseteq M$ with $\mathbb{P}(M') = 1$ such that for all $\omega \in M'$, $\sup_{x\in C(\omega)} |f(x) - f^N(x)| \to 0$. Easy to see that x^* is the unique minimum of f on $C(\omega)$ and x^N is a minimum of f^N on $C(\omega)$. Following the same proof of Lemma 4, we know for all $\omega \in M'$, $x^N(\omega) \to x^*(\omega)$, which implies $x^N \xrightarrow{a.s} x^*$. □

In the remainder of this proof, we aim to show that χ^N, χ, $g^N \equiv g^*(s^N)$ and g^* solving $u(g^* - \overline{m}) = u(g^* - \underline{m})$ satisfy the conditions of Lemma 5. We do so in three steps, one for each condition in the lemma. This allows us to obtain that $g^*(s^N) \xrightarrow{a.s} g^*$.

Step 1. $\sup_{\mathbf{g}\in\mathbf{C}} |\chi(\mathbf{g}) - \chi^{\mathbf{N}}(\mathbf{g})| \xrightarrow{\mathbf{a.s.}} \mathbf{0}$**, as** $n \to \infty$**, for all compact sets** $C \subset \mathbb{R}$**.**

Step 1.1. $\sup_{\mathbf{g}\in\mathbf{C}} |\chi^{\mathbf{N}}(\mathbf{g}) - \tilde{\chi}^{\mathbf{N}}(\mathbf{g})| \xrightarrow{\mathbf{a.s.}} \mathbf{0}$**, as** $n \to \infty$**, for all compact sets** $C \subset \mathbb{R}$**.**

We first consider the auxiliary function $\tilde{\chi}^N$. As N grows to infinity, the gap between χ^N and $\tilde{\chi}^N$ shrinks uniformly. To prove this claim, note that for any a^N and $p \in \mathbb{P}^a(a^N)$:

$$\chi^N(g) \geq \mathbb{E}_p[u(g-\theta)] \geq u\Big(g - \mathbb{E}_p[\theta]\Big),$$

where we use convexity of u for the second inequality. Then, by taking max over $p \in \mathbb{P}^a(a^N)$ we obtain $\tilde{\chi}^N(g) \leq \chi^N(g)$.

Now, for each g, $\tilde{\theta}$, and $p \in \mathbb{P}^{\boldsymbol{a}}(a^N)$, convexity of u implies:

$$u(g-\tilde{\theta}) \leq u(g-\mathbb{E}_p[\theta]) + u'(g-\tilde{\theta})(\mathbb{E}_p[\theta]-\tilde{\theta}) \leq u(g-\mathbb{E}_p[\theta]) + \Big|u'(g-\tilde{\theta})\Big|\Big|\mathbb{E}_p[\theta] - \tilde{\theta}\Big|.$$

Because u'' is bounded, and u is smooth and minimized at zero by assumption, we have, for some $L > 0$,

$$|u'(g-\tilde{\theta})| \leq L|g-\tilde{\theta}|.$$

Then, by taking expectations on the previous expression and applying this bound, we have

$$\begin{aligned}\mathbb{E}_p\left[u(g-\tilde{\theta})\right] &\leq \mathbb{E}_p\left[u(g-\mathbb{E}_p[\theta])\right] + L\,\mathbb{E}_p\left[\left|g-\tilde{\theta}\right|\left|\mathbb{E}_p[\theta]-\tilde{\theta}\right|\right] \\ &\leq \mathbb{E}_p\left[u(g-\mathbb{E}_p[\theta])\right] + L\,\sqrt{\mathbb{E}_p\left[\left(g-\tilde{\theta}\right)^2\right]}\sqrt{\mathbb{E}_p\left[\left(\mathbb{E}_p[\theta]-\tilde{\theta}\right)^2\right]},\end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality. Fix an arbitrary compact set $C \subset \mathbb{R}$ and define $v = \max_{g\in C, p\in\mathbb{P}^a(a^N)} \sqrt{\mathbb{E}_p\left[(g-\tilde{\theta})^2\right]}$ — which is finite by normality and uniformly bounded variance of the distributions in $\mathbb{P}^a(a^N)$. We can then maximize the expression above over $p \in \mathbb{P}^a(a^N)$ and use subadditivity of the max operator to obtain that for any $g \in C$,

$$\chi^N(g) \leq \tilde{\chi}^N(g) + Lv \max_{p\in\mathbb{P}^a(a^N)} \sqrt{\mathrm{Var(p)}},$$

and therefore:

$$0 \leq \chi^N(g) - \tilde{\chi}^N(g) \leq Lv \max_{p\in\mathbb{P}^a(a^N)} \sqrt{\mathrm{Var}(p)}.$$

Notice that neither bound depends on g within this compact set C. Moreover, the upper bound converges to zero, as we have proved that all variances converge to zero and are uniformly bounded above by the variance of assigning $\underline{\rho}$ for all signals. Thus:

$$\sup_{g\in C} |\chi^N(g) - \tilde{\chi}^N(g)| \xrightarrow{a.s.} 0.$$

Step 1.2. $\sup_{\mathbf{g}\in\mathbf{C}} |\chi(\mathbf{g}) - \tilde{\chi}^{\mathbf{N}}(\mathbf{g})| \xrightarrow{\mathbf{a.s.}} \mathbf{0}$**, as** $n \to \infty$**, for all compact sets** $C \subset \mathbb{R}$.

Recall that we can write $\tilde{\chi}^N(g) = \max\{u(g-\overline{m}^N), u(g-\underline{m}^N)\}$. Also by

Theorem 2, $\overline{m}^N \xrightarrow{a.s.} \overline{m}$ and $\underline{m}^N \xrightarrow{a.s.} \underline{m}$.

We use the following lemma:

Lemma 6. *Let f^N, g^N, f, g for $N \in \mathbb{N}$ be functions from $D \subset \mathbb{R}$ into the reals, and let $h^N = \max\{f^N, g^N\}$ and $h = \max\{f, g\}$. If $\sup_x |f^N - f| \to 0$ and $\sup_x |g^N - g| \to 0$ then, $\sup_x |h^N - h| \to 0$.*

Proof. For any fixed ε there exist N_f and N_g such that, for all $x \in D$:

$$|f^N(x) - f(x)| < \varepsilon \;\text{ if }\; N \geq N_f,$$

$$|g^N(x) - g(x)| < \varepsilon \;\text{ if }\; N \geq N_g.$$

Take $N \geq \tilde{N} = \max\{N_f, N_g\}$. We then have:

$$h(x) \leq (f^N(x) + \varepsilon)\mathbb{1}_{f(x) \geq g(x)} + (g^N(x) + \varepsilon)\mathbb{1}_{g(x) \geq f(x)} \leq h^N(x) + \varepsilon.$$

where the second inequality comes from the definition of h^N. By the same logic, inverting the roles of h and h^N:

$$h^N(x) \leq (f(x) + \varepsilon)\mathbb{1}_{f^N(x) \geq g^N(x)} + (g(x) + \varepsilon)\mathbb{1}_{g^N(x) \geq f^N(x)} \leq h(x) + \varepsilon.$$

By joining the two inequalities above: $|h(x) - h^N(x)| \leq \varepsilon$ for all $N \geq \tilde{N}$. Because x is arbitrary, we have our result. $\square$

In order to apply the result above, notice that $\sup_{g \in C} |u(g - x) - u(g - y)|$ is a continuous function of x and, thus, converges to 0 as $x \to y$. Thus, $\sup_{g \in C} |u(g - \overline{m}^N) - u(g - \overline{m})| \xrightarrow{a.s.} 0$ and similarly $\sup_{g \in C} |u(g - \underline{m}^N) - u(g - \underline{m})| \xrightarrow{a.s.} 0$. Therefore, applying the above lemma by defining $f^N(x) = u(x - \overline{m}^N)$ and $g^N(x) = u(x - \underline{m}^N)$ gives us our result.

Step 1.3. $\sup_{\mathbf{g} \in \mathbf{C}} |\chi(\mathbf{g}) - \chi^{\mathbf{N}}(\mathbf{g})| \xrightarrow{\text{a.s.}} \mathbf{0}$**, as** $n \to \infty$**, for all compact sets** $C \subset \mathbb{R}$.

This is directly implied by the previous two steps.

Step 2. g^* **such that** $u(g^* - \underline{m}) = u(g^* - \overline{m})$ **is the unique minimum of** χ.

Recall that $\chi(g) = \max\{u(g - \underline{m}), u(g - \overline{m})\}$. First, notice that g^* that minimizes χ must be in $[\underline{m}, \overline{m}]$. Assume, for a contradiction, that $\min \chi(g) = u(g^* - \underline{m}) > u(g^* - \overline{m})$. By continuity of u, we can choose $\underline{m} < g' < g^*$ such that $u(g' - \underline{m}) > u(g' - \overline{m})$, that is, $\chi(g') = u(g' - \underline{m})$. Because u is strictly convex and minimized at 0, it must be that $u(g^* - \underline{m}) > u(g' - \underline{m})$. But then, $\chi(g') < \chi(g^*)$, which is a contradiction. A similar contradiction is found if we assume $\min \chi(g) = u(g^* - \overline{m}) < u(g^* - \underline{m})$. Thus, the equality must hold. The uniqueness of g^* is guaranteed by u being strictly convex and minimized at 0.

Step 3. The sequence g^N **is bounded almost everywhere.** For a sequence of realized observables a^N, recall that $\underline{m}^N = \min_{p \in \mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}[\theta]$ and, symmetrically, $\overline{m}^N = \max_{p \in \mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}[\theta]$. Assume, for a contradiction, that there is an event M with positive probability, such that g^N is unbounded. If that's the case, up to a subsequence, we have: $g^N > N$. Then, by strict convexity of u we have:

$$\chi(g^N) = \max_{p \in \mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p[u(g^N - \theta)] \geq \max_{p \in \mathbb{P}^{\boldsymbol{a}}(a^N)} u(g^N - \mathbb{E}_p[\theta]) \geq u(g^N - \overline{m}^N).$$

Now, because $\overline{m}^N \xrightarrow{a.s.} \overline{m}$, we can choose an event $M' \subset M$, with the same probability of M, in which $\overline{m}^N \to \overline{m}$. That implies, with the unboundedness of g^N and strict convexity of u, that the lower bound above diverges, so $\chi(g^N)$ is unbounded. To show that g^N cannot be optimal, it suffices to show that there is a sequence x^N such that $\chi(x^N)$ is bounded in this event. For any real number a, take the sequence $x^N = a$ for all N. Because $\chi^N \xrightarrow{a.s.} \chi$ uniformly in any compact set, we have that, for a further event $M'' \subset M'$, with the same probability of M', that for any ε, for sufficiently large N,

$$\chi^N(a) < \chi(a) + \varepsilon.$$

Thus, $\chi^N(a)$ is a bounded sequence, proving that, for sufficiently large

N:

$$\chi^N(a) < \chi^N(g^N),$$

which is the contradiction that we were seeking. $\square$

Proof of Corollary 1

Symmetry implies Consistency. Define

$$\overline{\zeta}(m) = \frac{\underline{\rho}\int_{-\infty}^{m} x dF(x) + \overline{\rho}\int_{m}^{\infty} x dF(x)}{\underline{\rho}F(m) + \overline{\rho}(1-F(m))}, \quad \underline{\zeta}(m) = \frac{\overline{\rho}\int_{-\infty}^{m} x dF(x) + \underline{\rho}\int_{m}^{\infty} x dF(x)}{\overline{\rho}F(m) + \underline{\rho}(1-F(m))}$$

Clearly, $\overline{\zeta}(\overline{m}) = \overline{m}$ and $\underline{\zeta}(\underline{m}) = \underline{m}$. Because F is symmetric around θ, for $m \in \mathbb{R}$:

$$\begin{aligned}\overline{\zeta}(2\theta - m) &= \frac{\underline{\rho}\int_{-\infty}^{2\theta-m} x dF(x) + \overline{\rho}\int_{2\theta-m}^{\infty} x dF(x)}{\underline{\rho}F(2\theta-m) + \overline{\rho}(1-F(2\theta-m))} \\ &= 2\theta - \frac{\overline{\rho}\int_{-\infty}^{m} x dF(x) + \underline{\rho}\int_{m}^{\infty} x dF(x)}{\overline{\rho}F(m) + \underline{\rho}(1-F(m))} = 2\theta - \underline{\zeta}(m).\end{aligned}$$

Then, $2\theta - \underline{m} = 2\theta - \underline{\zeta}(\underline{m}) = \overline{\zeta}(2\theta - \underline{m})$. But because $\overline{m}$ is the unique fixed point of $\overline{\zeta}$:[16] $\overline{m} = 2\theta - \underline{m}$, and we are done.

Asymmetry implies non-consistency for some sets. Let x^* be such that $u(x^*) \neq u(-x^*)$. Define $\eta = \frac{\overline{\rho}}{\underline{\rho}}$. Notice, from the proof of Proposition 2 that, that $\frac{\overline{m}-\underline{m}}{2}$ is an function of η onto the real line. Then, choose η^* such that $\frac{\overline{m}-\underline{m}}{2} = x^*$. Finally, recall that, by observable signals, $\theta = \frac{\overline{m}+\underline{m}}{2}$. Then:

$$u(\theta - \underline{m}) = u(x^*) \neq u(-x^*) = u(\theta - \overline{m})$$

Thus, by Theorem 3, $g^* \neq \theta$. $\square$

[16] See the Proof of Theorem 2.

Proof of Proposition 2

$\overline{m}(\underline{m})$ **monotonically increases(decreases) in** η We go through the proof for $\overline{m}$, a symmetric argument holds for $\underline{m}$. Define $k_\eta(a)$ as

$$k_\eta(a) \equiv \mathbb{E}[\theta](a) = \frac{\underline{\rho}\int_{-\infty}^{a} xf(x)dx + \overline{\rho}\int_{a}^{\infty} xf(x)dx}{\underline{\rho}F(a) + \overline{\rho}(1-F(a))} = \frac{\int_{-\infty}^{a} xf(x)dx + \eta\int_{a}^{\infty} xf(x)dx}{F(a) + \eta(1-F(a))}$$

For convenience we can rewrite $k_\eta(a)$ as

$$k_\eta(a) = \frac{F(a)\mathbb{E}[x|x<a] + \eta(1-F(a))\mathbb{E}[x|x \geq a]}{F(a) + \eta(1-F(a))}$$

We know that

$$\overline{m} = \arg\max_{a\in\mathbb{R}} k_\eta(a) \qquad \textit{and} \qquad \overline{m} = \max_{a\in\mathbb{R}} k_\eta(a)$$

Then, via the envelope theorem we have

$$\frac{d\overline{m}}{d\eta} = \frac{dk_\eta(\overline{m})}{d\eta} = \frac{F(\overline{m})(1-F(\overline{m}))}{(F(\overline{m}) + \eta(1-F(\overline{m})))^2}\left(\mathbb{E}[x|x \geq \overline{m}] - \mathbb{E}[x|x < \overline{m}]\right) > 0$$

Step 1. As $\eta \to +\infty(-\infty)$, $\overline{m} \to \infty(\underline{m} \to -\infty)$.   First note that

$$\lim_{\eta\to\infty} k_\eta(a) = \mathbb{E}[x|x \geq a] > a.$$

The above inequality follows from the full support of the distribution. For any $z \in \mathbb{R}$ we want to show that $\exists\,\tilde{\eta}$ such that $k_{\tilde{\eta}}(\overline{m}) \geq z$. From the above limit, we know that $\exists\tilde{\eta}$ such that $k_{\tilde{\eta}}(z) > z$. Because $\overline{m} = \arg\max_{a\in\mathbb{R}} k_\eta(a)$ we know that $k_{\tilde{\eta}}(\overline{m}) \geq k_{\tilde{\eta}}(z) > z$.

Step 2. As $\eta \to 1$, $\overline{m} - \underline{m} \to 0$.   When $\eta \to 1$, $k_\eta(a)$ reduces to the unconditional expected value for any a. Similarly, the optimization problem that determines $\underline{m}$ reduces to the unconditional expected value, completely unaffected by a. Thus, as $\eta \to 1$ both $\overline{m}$ and $\underline{m}$ converge to the unconditional expectation. $\square$

Proof of Proposition 3

Let $H \preceq_{\text{FOSD}} G$ be precisions distributions, and let the true value of the state be θ. Assume these distributions of precision generate signal distributions F_H and F_G respectively. We first prove F_H is a mean-preserving spread of F_G — denoted $F_G \preceq_{\text{MPS}} F_H$.

Signals are more disperse under H. Indeed, notice that:

$$F_H(x) = \int_\rho F_\rho(x) dH(\rho) \qquad\qquad F_G(x) = \int_\rho F_\rho(x) dG(\rho),$$

where F_ρ is the CDF of the normal distribution with mean θ and precision ρ. Now, because all F_ρ's have the same mean, and a higher precision means lower dispersion of signals, we have $F_\rho \preceq_{m.p.s.} F_{\rho'}$, for $\rho < \rho'$. Therefore, for any z, $\int_{-\infty}^{z} F_\rho(x)dx$ is decreasing in ρ. Thus:

$$\int_{-\infty}^{z} F_H(x)dx = \int_\rho \int_{-\infty}^{z} F_\rho(x)dxdH(\rho) \geq \int_\rho \int_{-\infty}^{z} F_\rho(x)dxdG(\rho) = \int_{-\infty}^{z} F_G(x)dx,$$

where the change in the integration order is a consequence of Tonelli's theorem, and the inequality is justified because $\int_{-\infty}^{z} F_\rho(x)dx$ is decreasing in ρ, and G first-order stochastically dominates H. This inequality implies F_H second-order stochastically dominates F_G. But it is clear F_H and F_G have the same mean, θ. So we proved F_H is a mean-preserving spread of F_G. We now use this result to conclude the proof.

Asymptotic belief set is larger under H. We prove the result for the upper bound. The result holds for the lower bound by symmetry. For a signal distribution P, define:

$$k_P(a) = \frac{\int_{-\infty}^{a} x dP(x)dx + \eta \int_{a}^{\infty} x dP(x)dx}{P(a) + \eta\,(1-P(a))} = \frac{P(a)\mathbb{E}_P[x|x \le a] + \eta\,(1-P(a))\,\mathbb{E}_P[x|x \ge a]}{P(a) + \eta\,(1-P(a))}.$$

We know:

$$\overline{m}_G = \max_a k_{F_G}(a) \quad \text{and} \quad \overline{m}_H = \max_a k_{F_H}(a).$$

First, we implement a change of variables. For each a, there exists a quantile $q \in [0,1]$ such that $P(a) = q$. We can then write:

$$k_P(a) = \hat{k}_P(q) \equiv \frac{q\mathbb{E}_P[x|P(x) \le q] + \eta(1-q)\mathbb{E}_P[x|P(x) \ge q]}{q + \eta(1-q)}$$

Because F_H is a mean-preserving spread of F_G,

$$\mathbb{E}_{F_H}[x|F_H(x) \ge q] \ge \mathbb{E}_{F_G}[x|F_G(x) \ge q] \;\; \text{and} \;\; \mathbb{E}_{F_G}[x|F_G(x) \le q] \ge \mathbb{E}_{F_H}[x|F_H(x) \le q].$$

Moreover, because F_G and F_H have the same mean:

$$\begin{aligned} &\mathbb{E}_{F_H}[x] = q\mathbb{E}_{F_H}[x|F_H(x) \le q] + (1-q)\mathbb{E}_{F_H}[x|F_H(x) \ge q] \\ =&\mathbb{E}_{F_G}[x] = q\mathbb{E}_{F_G}[x|F_G(x) \le q] + (1-q)\mathbb{E}_{F_G}[x|F_G(x) \ge q]. \end{aligned}$$

Because $\eta > 1$, the expressions above above imply $\hat{k}_{F_G}(q) \le \hat{k}_{F_H}(q)$. To conclude the argument, we note:

$$\overline{m}_G = \max_a k_{F_G}(a) = \max_{q \in [0,1]} \hat{k}_{F_G}(q) \le \max_{q \in [0,1]} \hat{k}_{F_H}(q) = \max_a k_{F_H}(a) = \overline{m}_H.$$

$\square$

Proof of Proposition 4

By setting $\mathbf{s^a}(a_i, \hat{\rho}_i) = a_i + \frac{\rho_\mu}{\hat{\rho}_i}(a_i - \mu)$ and $\beta(a) = a$ in Theorem 3, the bounds of the limiting posterior set are given by

$$\overline{m}_a = \frac{\underline{\rho}\int_{-\infty}^{\overline{m}_a} x dF(x) + \overline{\rho}\int_{\overline{m}_a}^{\infty} x dF(x) + c}{\underline{\rho}F(\overline{m}_a) + \overline{\rho}(1 - F(\overline{m}_a))}, \qquad \underline{m}_a = \frac{\overline{\rho}\int_{-\infty}^{\underline{m}_a} x dF(x) + \underline{\rho}\int_{\underline{m}_a}^{\infty} x dF(x) + c}{\overline{\rho}F(\underline{m}_a) + \underline{\rho}(1 - F(\underline{m}_a))} \tag{5}$$

where $c = \frac{\rho\rho_\mu}{\rho_\mu + \rho}(\theta - \mu)$.

The optimal estimate is $m_a = \frac{\overline{m}_a + \underline{m}_a}{2}$. When $\theta = \mu$, $c = 0$ and by **Corollary 1** $m_a = \theta = \mu$ and the observer's estimate is consistent. From now on, we first focus on the case where $\theta > \mu$.

Denote $\overline{G}(z) = \underline{\rho}F(z) + \overline{\rho}(1 - F(z))$ and $\underline{G}(z) = \overline{\rho}F(z) + \underline{\rho}(1 - F(z))$. Rearranging the first equation and using integration by parts, we get

$$\begin{aligned}
\overline{m}_a\overline{G}(\overline{m}_a) &= \underline{\rho}\left(xF(x)\Big|_{-\infty}^{\overline{m}_a} - \int_{-\infty}^{\overline{m}_a} F(x)dx\right) + \overline{\rho}\left(-x(1 - F(x))\Big|_{\overline{m}_a}^{\infty} + \int_{\overline{m}_a}^{\infty} (1 - F(x))\, dx\right) + c \\
&= \underline{\rho}\left(\overline{m}_a F(\overline{m}_a) - \int_{-\infty}^{\overline{m}_a} F(x)dx\right) + \overline{\rho}\left(\overline{m}_a(1 - F(\overline{m}_a)) + \int_{\overline{m}_a}^{\infty} (1 - F(x))\, dx\right) + c \\
&= \overline{m}_a\overline{G}(\overline{m}_a) - \left(\underline{\rho}\int_{-\infty}^{\overline{m}_a} F(x)dx - \overline{\rho}\int_{\overline{m}_a}^{\infty} (1 - F(x))\, dx\right) + c.
\end{aligned}$$

This implies

$$\underline{\rho}\int_{-\infty}^{\overline{m}_a} F(x)dx - \overline{\rho}\int_{\overline{m}_a}^{\infty} (1 - F(x))\, dx = c. \tag{6}$$

A symmetric argument for $\underline{m}_a$ shows that

$$\overline{\rho}\int_{-\infty}^{\underline{m}_a} F(x)dx - \underline{\rho}\int_{\underline{m}_a}^{\infty} (1 - F(x))\, dx = c. \tag{7}$$

Taking the derivative with respect to the state θ on both sides of equation 6 and equation 7, we get

$$\frac{d\overline{m}_a}{d\theta} = \frac{\rho}{\rho_\mu + \rho} + \frac{\rho_\mu}{\rho_\mu + \rho}\frac{\rho}{\overline{G}(\overline{m}_a)} \qquad \text{and} \qquad \frac{d\underline{m}_a}{d\theta} = \frac{\rho}{\rho_\mu + \rho} + \frac{\rho_\mu}{\rho_\mu + \rho}\frac{\rho}{\underline{G}(\underline{m}_a)}.$$

The derivative of the optimal guess $m_a = \frac{\overline{m}_a + \underline{m}_a}{2}$ with respect to θ is then:

$$\frac{dm_a}{d\theta} = \frac{\rho}{\rho_\mu + \rho} + \frac{\rho_\mu}{\rho_\mu + \rho}\frac{\rho}{2}\left(\frac{1}{\overline{G}(\overline{m}_a)} + \frac{1}{\underline{G}(\underline{m}_a)}\right). \tag{8}$$

Recall that H is normally distributed and denote its density function as h. Then, we can use the derivative of the optimal bounds obtained above to calculate:

$$\begin{aligned}
\frac{dF(\overline{m}_a)}{d\theta} &= \frac{\partial F(\overline{m}_a)}{\partial \overline{m}_a}\frac{d\overline{m}_a}{d\theta} + \frac{\partial F(\overline{m}_a)}{\partial \theta} = -\frac{\rho_\mu}{\rho_\mu + \rho}\frac{\rho}{\overline{G}(\overline{m}_a)} f(\overline{m}_a),\\
\frac{dF(\underline{m}_a)}{d\theta} &= \frac{\partial F(\underline{m}_a)}{\partial \underline{m}_a}\frac{d\underline{m}_a}{d\theta} + \frac{\partial F(\underline{m}_a)}{\partial \theta} = -\frac{\rho_\mu}{\rho_\mu + \rho}\frac{\rho}{\underline{G}(\underline{m}_a)} f(\underline{m}_a).
\end{aligned}$$

It then follows that

$$\begin{aligned}
\frac{d^2 m_a}{d\theta^2} &= \frac{\overline{\rho} - \underline{\rho}}{2}\left(\frac{\rho\rho_\mu}{\rho_\mu + \rho}\right)^2 \left(\frac{f(\overline{m}_a)}{\overline{G}^3(\overline{m}_a)} - \frac{f(\underline{m}_a)}{\underline{G}^3(\underline{m}_a)}\right)\\
&= \frac{\overline{\rho} - \underline{\rho}}{2}\left(\frac{\rho\rho_\mu}{\rho_\mu + \rho}\right)^2 \left(\left(\frac{f(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right)\frac{1}{\underline{G}^2(\underline{m}_a)} + \frac{f(\overline{m}_a)}{\overline{G}(\overline{m}_a)}\left(\frac{1}{\overline{G}^2(\overline{m}_a)} - \frac{1}{\underline{G}^2(\underline{m}_a)}\right)\right).
\end{aligned}$$

Lemma 7. $\left(\frac{1}{\overline{G}^2(\overline{m}_a)} - \frac{1}{\underline{G}^2(\underline{m}_a)}\right) > 0$ *whenever* $\theta > \mu$.

Proof. The statement is equivalent to $\underline{G}(\underline{m}_a) > \overline{G}(\overline{m}_a)$, which is also equivalent to $F(\overline{m}_a) + F(\underline{m}_a) > 1$. Since H is symmetric around $\frac{\rho\theta + \rho_\mu\mu}{\rho + \rho_\mu}$, the latter is true if and only if $m_a > \frac{\rho\theta+\rho_\mu\mu}{\rho+\rho_\mu}$. We show that this is the case. Define

$$\overline{\zeta}(z,u) = \frac{\underline{\rho}\int_{-\infty}^{z} xdF(x) + \overline{\rho}\int_{z}^{\infty} xdF(x) + u}{\underline{\rho}F(z) + \overline{\rho}(1-F(z))},$$
$$\underline{\zeta}(z,u) = \frac{\overline{\rho}\int_{-\infty}^{z} xdF(x) + \underline{\rho}\int_{z}^{\infty} xdF(x) + u}{\overline{\rho}F(z) + \underline{\rho}(1-F(z))}$$

We know $\overline{m}_a = \overline{\zeta}(\overline{m}_a, c)$, and it was previously proved that $\overline{m}_a$ maximizes $\overline{\zeta}(\overline{m}_a, c)$. By the envelope theorem we have:

$$\frac{d\overline{\zeta}(\overline{m}_a,c)}{du} = \frac{\partial\overline{\zeta}(\overline{m}_a,c)}{\partial u} = \frac{1}{\underline{\rho}F(\overline{m}_a) + \overline{\rho}(1-F(\overline{m}_a))} > 0$$

A similar argument implies that $\frac{\underline{\zeta}(\underline{m}_a,u)}{du} > 0$, for all $u \in \mathbb{R}$. Finally, by an equivalent argument to the proof of Corollary 1, we have $\frac{\overline{\zeta}(\overline{m}_a,0)+\underline{\zeta}(\underline{m}_a,0)}{2} = \int xdH = \frac{\rho\theta+\rho_\mu\mu}{\rho+\rho_m u}$. Then, if $\theta > \mu$ - which implies $c > 0$:

$$m_a = \frac{\overline{m}_a + \underline{m}_a}{2} = \frac{\overline{\zeta}(\overline{m}_a, c) + \underline{\zeta}(\underline{m}_a, c)}{2} > \frac{\overline{\zeta}(\overline{m}_a, 0) + \underline{\zeta}(\underline{m}_a, 0)}{2}.$$

This concludes the proof of the lemma. □

Therefore,

$$\left(\frac{f(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) \geq 0 \implies \frac{d^2 m_a}{d\theta^2} > 0. \tag{9}$$

We next consider the partial derivative of the optimal guess with respect to ρ. We start with an alternative implicit function of $\overline{m}_a$ and $\underline{m}_a$. Notice that if f as the density function of a normal distribution with mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$, then $\frac{\partial f(x)}{\partial x} = -\frac{x-\tilde{\mu}}{\tilde{\sigma}^2} f(x)$. This implies $xf(x) = \tilde{\mu}f(x) - \tilde{\sigma}^2 \frac{\partial f(x)}{\partial x}$.

Plugging this into the initial implicit functions 5, we get

$$\overline{m}_a = \frac{\rho_\mu \mu + \rho\theta}{\rho_\mu + \rho} + \frac{c}{\overline{G}(\overline{m}_a)} + (\overline{\rho} - \underline{\rho})\frac{\rho}{(\rho_\mu + \rho)^2}\frac{f(\overline{m}_a)}{\overline{G}(\overline{m}_a)},$$
$$\underline{m}_a = \frac{\rho_\mu \mu + \rho\theta}{\rho_\mu + \rho} + \frac{c}{\underline{G}(\underline{m}_a)} - (\overline{\rho} - \underline{\rho})\frac{\rho}{(\rho_\mu + \rho)^2}\frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}.$$

By definition of m_a, we have

$$m_a = \theta + (\theta - \mu)\left(\frac{dm_a}{d\theta} - 1\right) + \frac{(\overline{\rho} - \underline{\rho})\rho}{2(\rho_\mu + \rho)^2}\left(\frac{f(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right). \tag{10}$$

Based on the implicit function theorem, we can calculate the following derivative:

$$\frac{dm_a}{d\rho} = \frac{\rho_\mu(m_a - \mu) + \rho(\theta - m_a)}{2\rho^2 + 2\rho_\mu\rho} + \frac{c}{2}\frac{\rho_\mu + (\rho_\mu + \rho)\rho}{(\rho_\mu + \rho)\rho}\left(\frac{1}{\overline{G}(\overline{m}_a)} + \frac{1}{\underline{G}(\underline{m}_a)}\right).$$

As $\theta > \mu$, it is easy to show that $m_a > \mu$ and $c > 0$. This leads to the following result.

$$\theta > \mu \quad \text{and} \quad m_a \le \theta \quad \Longrightarrow \quad \frac{dm_a}{d\rho} > 0. \tag{11}$$

Note that the last term of $\frac{dm_a}{d\rho}$ can be rewritten as $\left(\frac{dm_a}{d\theta} - \frac{\rho}{\rho_\mu + \rho}\right)\frac{\rho_\mu + \rho}{\rho_\mu}\frac{2}{\rho}$. Let $\kappa_1 = \frac{1}{2\rho^2 + 2\rho_\mu\rho}$ and $\kappa_2 = \frac{\rho_\mu + (\rho_\mu + \rho)\rho}{\rho_\mu\rho^2}$, then:

$$\frac{d^2 m_a}{d\rho\, d\theta} = \rho_\mu\kappa_1\frac{dm_a}{d\theta} - \rho\kappa_1\left(\frac{dm_a}{d\theta} - 1\right) + \frac{\rho\rho_\mu}{\rho_\mu + \rho}\kappa_2\left(\frac{dm_a}{d\theta} - \frac{\rho}{\rho_\mu + \rho}\right) + c\kappa_2\frac{d^2 m_a}{d\theta^2}. \tag{12}$$

We know that $\frac{dm_a}{d\theta} > \frac{\rho}{\rho_\mu + \rho} > 0$ and when $\theta = \mu$, $\frac{d^2 m_a}{d\theta^2} = 0$. This leads to the following result:

$$\theta = \mu \quad \text{and} \quad \frac{dm_a}{d\theta} \leq 1 \quad \Longrightarrow \quad \frac{d^2 m_a}{d\rho\, d\theta} > 0. \tag{13}$$

To make it clear that the optimal guess depends on θ and ρ, we sometimes denote $\underline{m}_a$, $\overline{m}_a$ and m_a as $\underline{m}_a(\rho,\theta)$, $\overline{m}_a(\rho,\theta)$ and $m_a(\rho,\theta)$. Notice that $\tilde{\rho}$ is determined by forcing $\frac{dm_a}{d\theta}$ to approach 1 when θ goes to infinity, while at $\tilde{\tilde{\rho}}$ we have $\frac{dm_a}{d\theta}(\tilde{\tilde{\rho}},\mu) = 1$.

The rest of the proof will be divided by the following lemmas. We will fix μ and consider the case with $\theta \geq \mu$.

Lemma 8. *For any given ρ, if $m_a(\rho,\hat{\theta}) > \hat{\theta}$ and $\frac{dm_a}{d\theta}(\rho,\hat{\theta}) > 1$, then $m_a(\rho,\theta) > \theta$ for all $\theta > \hat{\theta}$.*

Proof. Fix ρ. Assume that there exists $\hat{\theta}$, $m_a(\rho,\hat{\theta}) > \hat{\theta}$ and $\frac{dm_a}{d\theta}(\rho,\hat{\theta}) > 1$. Suppose by contradiction that there exists some $\overline{\theta} > \hat{\theta}$ such that $m_a(\rho,\overline{\theta}) = \overline{\theta}$. By continuity of $\frac{dm_a}{d\theta}$, there exists $\theta' < \theta'' \in (\hat{\theta},\overline{\theta}]$ where $\frac{dm_a}{d\theta}(\rho,\theta') = 1$ and $\frac{dm_a}{d\theta}(\rho,\theta'') < 1$. By continuity of m_a, $m_a(\rho,\theta') > \theta'$.

At θ', equation (10) implies $\left(\frac{f(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) > 0$, which guarantees $\frac{d^2 m_a}{d\theta^2}(\rho,\theta') > 0$. This implies that for a neighborhood to the right of θ', $\frac{dm_a}{d\theta} > 1$. Notice that this holds for any $\theta \in [\hat{\theta},\overline{\theta}]$ with $\frac{dm_a}{d\theta}(\rho,\theta) = 1$. Thus $\frac{dm_a}{d\theta}(\rho,\theta) \geq 1$ for all $\theta \in [\hat{\theta},\overline{\theta}]$, which contradicts the assumption that $m_a(\rho,\overline{\theta}) = \overline{\theta}$. As a result, we know $m_a(\rho,\theta) > \theta$ for $\theta > \hat{\theta}$. This concludes the proof of the lemma. □

Lemma 9. *For any given ρ, if there exists $\theta^* > \mu$ such that $m_a(\rho,\theta^*) = \theta^*$ and $m_a(\rho,\theta) < \theta$ for all $\mu < \theta < \theta^*$, then $m_a(\rho,\theta) > \theta$ for $\theta > \theta^*$.*

Proof. Suppose there exists $\theta^* > \mu$ such that $m_a(\rho,\theta^*) = \theta^*$ and $m_a(\rho,\theta) < \theta$ for $\mu < \theta < \theta^*$. This implies $\frac{dm_a}{d\theta}(\rho,\theta^*) \geq 1$. Again by equation (10), we know $\left(\frac{f(\overline{m}_a)}{\overline{G}(\overline{m})} - \frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) > 0$, which leads to $\frac{d^2 m_a}{d\theta^2}(\rho,\theta^*) > 0$ by (9). Then for any θ in a small neighborhood to the right of θ^*, $\frac{dm_a}{d\theta}(\rho,\theta) > 1$ and $m_a(\rho,\theta) > \theta$. By Lemma 8. This concludes the proof of the lemma and the proposition. □

References

Al-Najjar, N. I. (2009). Decision makers as statisticians: Diversity, ambiguity, and learning. *Econometrica*, 77(5):1371–1401.

Auster, S., Che, Y.-K., and Mierendorff, K. (2024). Prolonged learning and hasty stopping: the wald problem with ambiguity. *American Economic Review*, 114(2):426–461.

Auster, S. and Kellner, C. (2022). Robust bidding and revenue in descending price auctions. *Journal of Economic Theory*, 199:105072.

Auster, S. and Kellner, C. (2023). Timing decisions under model uncertainty. Technical report, ECONtribute Discussion Paper.

Battigalli, P., Francetich, A., Lanzani, G., and Marinacci, M. (2019). Learning and self-confirming long-run biases. *Journal of Economic Theory*, 183:740–785.

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

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, pages 51–58.

Blum, J. R. (1955). On the convergence of empiric distribution functions. *The Annals of Mathematical Statistics*.

Bose, S. and Daripa, A. (2009). A dynamic mechanism and surplus extraction under ambiguity. *Journal of Economic theory*, 144(5):2084–2114.

Bose, S. and Renou, L. (2014). Mechanism design with ambiguous communication devices. *Econometrica*, 82(5):1853–1872.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., and Montrucchio, L. (2013). Ambiguity and robust statistics. *Journal of Economic Theory*, 148(3):974–1049.

Chen, J. Y. (2023). Sequential learning under informational ambiguity. Technical report, Working paper.

Cheng, X. (2022). Relative maximum likelihood updating of ambiguous beliefs. *Journal of Mathematical Economics*, 99:102587.

Cheng, X. (2024). Improving robust decisions with data. Technical report, Working paper.

Condie, S. and Ganguli, J. (2017). The pricing effects of ambiguous private information. *Journal of Economic Theory*, 172:512–557.

DeHardt, J. (1971). Generalizations of the glivenko-cantelli theorem. *The Annals of Mathematical Statistics*, 42(6):2050–2055.

Epstein, L. G. and Schneider, M. (2007). Learning under ambiguity. *The Review of Economic Studies*, 74(4):1275–1303.

Epstein, L. G. and Schneider, M. (2008). Ambiguity, information quality, and asset pricing. *The Journal of Finance*, 63(1):197–228.

Esponda, I., Pouzo, D., and Yamamoto, Y. (2021). Asymptotic behavior of bayesian learners with misspecified models. *Journal of Economic Theory*, 195:105260.

Frick, M., Iijima, R., and Ishii, Y. (2023). Belief convergence under misspecified learning: A martingale approach. *The Review of Economic Studies*, 90(2):781–814.

Fudenberg, D., Lanzani, G., and Strack, P. (2021). Limit points of endogenous misspecified learning. *Econometrica*, 89(3):1065–1098.

Fudenberg, D., Romanyuk, G., and Strack, P. (2017). Active learning with a misspecified prior. *Theoretical Economics*, 12(3):1155–1189.

Gershkov, A., Moldovanu, B., Strack, P., and Zhang, M. (2023). Optimal insurance: Dual utility, random losses, and adverse selection. *American Economic Review*, 113(10):2581–2614.

Ghosh, G. and Liu, H. (2021). Sequential auctions with ambiguity. *Journal of Economic Theory*, 197:105324.

Giacomini, R. and Kitagawa, T. (2020). Robust bayesian inference for set-identified models. *Econometrica*, Forthcoming.

Giacomini, R., Kitagawa, T., and Uhlig, H. (2019). Estimation under ambiguity. Technical report, Cemmap working paper.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153.

Gilboa, I. and Schmeidler, D. (1993). Updating ambiguous beliefs. *Journal of Economic Theory*, 59(1):33–49.

Gollier, C. (2011). Portfolio choices and asset prices: The comparative statics of ambiguity aversion. *The Review of Economic Studies*, 78(4):1329–1344.

Gul, F. and Pesendorfer, W. (2021). Evaluating ambiguous random variables from choquet to maxmin expected utility. *Journal of Economic Theory*, 192:105129.

Heidhues, P., Kőszegi, B., and Strack, P. (2021). Convergence in models of misspecified learning. *Theoretical Economics*, 16(1):73–99.

Huber, P. J. (2004). *Robust statistics*, volume 523. John Wiley & Sons.

Hurwicz, L. (1951). Some specification problems and applications to econometric models. *Econometrica*, 19(3):343–344.

Illeditsch, P. K. (2011). Ambiguous information, portfolio inertia, and excess volatility. *The Journal of Finance*, 66(6):2213–2247.

Kellner, C. and Le Quement, M. T. (2018). Endogenous ambiguity in cheap talk. *Journal of Economic Theory*, 173:1–17.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335.

Lambert, N. S., Ostrovsky, M., and Panov, M. (2018). Strategic trading in informationally complex environments. *Econometrica*, 86(4):1119–1157.

Marinacci, M. (2002). Learning from ambiguous urns. *Statistical Papers*, 43(1):143.

Nyarko, Y. (1991). Learning in mis-specified models and the possibility of cycles. *Journal of Economic Theory*, 55(2):416–427.

Pires, C. P. (2002). A rule for updating ambiguous beliefs. *Theory and Decision*, 53(2):137–152.

Seidenfeld, T. and Wasserman, L. (1993). Dilation for sets of probabilities. *The Annals of Statistics*, 21(3):1139–1154.

Shalizi, C. R. (2009). Dynamics of bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics*, 3:1039–1074.

Shishkin, D. and Ortoleva, P. (2023). Ambiguous information and dilation: An experiment. *Journal of Economic Theory*, 208:105610.

Tang, R. (2024). A theory of contraction updating. Technical report, Working paper.

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

Tillio, A. d., Kos, N., and Messner, M. (2017). The design of ambiguous mechanisms. *The Review of Economic Studies*, 84(1):237–276.