Asymptotic Learning with Ambiguous Information*

Pëllumb Reshidi †　　João Thereze ‡　　Mu Zhang §

November 2021

Abstract

We study asymptotic learning when the decision maker is ambiguous about 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, ambiguity regarding information sources induces ambiguity about the state. We show that this induced ambiguity does not vanish even as the number of information sources grows indefinitely, and characterize the limit set of posteriors. The decision maker's asymptotic estimate of the state is generically incorrect. We consider several applications. Among them, (i) we provide a foundation for disagreement among agents with access to the same large dataset; (ii) we show that a small amount of ambiguity can exacerbate the effect of model misspecification on learning; and (iii) we analyze a setting in which the decision maker learns from observing others' actions.

*We thank Roland Benabou, 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 and BEAT Conference (Tsinghua University) for their helpful comments and discussions. An earlier version of this paper was circulated under the title "Information Aggregation under Ambiguity".

†Department of Economics, Princeton University preshidi@princeton.edu

‡Department of Economics, Princeton University joaothereze@princeton.edu

§Department of Economics, Princeton University muz@princeton.edu

1 Introduction

Consider an agent who relies on multiple information sources to learn a payoff relevant state. A voter may depend on poll results and advertising to find out a politician's competence and agenda; an investor uses reports of different analysts to forecast the future returns of a stock. A common assumption in the literature is that the decision maker has beliefs about the qualities of their information sources, and that these beliefs are correctly specified. In such cases, asymptotic learning is successful. While these assumptions are reasonable, in many settings forming beliefs might not be straightforward. For example, consider a prospective customer consulting online reviews before making a purchasing decision. She may not have particular beliefs about the quality of each reviewer, as they are being consulted for the first time. Such settings are widespread, yet little is known about learning in these environments. This paper addresses this gap.

We analyze asymptotic learning when the decision maker lacks particular beliefs about her information sources. We study a decision maker who estimates a state by minimizing a loss function. She observes monotonic transformations of multiple unbiased signals. The state and the signals are jointly normally distributed, but the decision maker does not know the signals' precisions — i.e. the inverse of their variances. The decision maker is not probabilistically sophisticated; instead, she is ambiguous regarding 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 decision maker, a joint distribution over signals and the state. Thus, an interval of perceived precisions induces a set of beliefs. We assume the decision maker updates her beliefs prior-by-prior. That is, upon observing information, she updates each belief according to Bayes' rule, obtaining a posterior about the state. In doing so, ambiguity about precisions induces ambiguity about the state, in that the agent considers multiple posteriors. Finally, she takes a robust approach and evaluates expected loss according to the worst case across all posteriors.

This setup encompasses a broad range of environments. By modeling observables as monotonic transformations of signals we cover cases in which the decision maker may observe unbiased signals directly, actions of other agents, and many other possibilities. Furthermore, our assumptions on both distributions and ambiguity attitudes are for the sake of tractability. Our main insights go through in much more general settings.[1]

Our first result shows that the induced ambiguity over the state does not vanish asymptotically. That is, the posterior beliefs of the decision maker do not converge to the same distribution as the number of information sources grows. We characterize this asymptotic set of posteriors. Just as in standard Bayesian learning, the variance of each posterior converges to zero. However, different beliefs lead to different weighting of signals, and consequently to different posterior means. For example, for any realization of signals, there is a belief in the agent's belief set 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 is a belief that leads to a relatively low posterior mean. Considering the set of all agent's beliefs generates an interval of posterior means. The set of asymptotic posteriors is the set of Dirac measures over that interval. Importantly, this set is independent of the objective of the decision maker.

Our second result characterizes the decision maker's asymptotic estimate of the state. Her decision problem can be interpreted as a zero-sum game against nature. First, the decision maker receives information and chooses the estimate that minimizes her expected loss. Afterwards, nature chooses the precision of each source with the aim of maximizing the agent's loss. In doing so, nature affects the agent'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 agent chooses a relatively low value within the interval, nature will maximize her loss by choosing the highest value possible, and vice versa. Given this, to minimize the maximal loss, the agent chooses an estimate that makes nature indifferent between choosing the highest or the lowest value in the interval of posterior means.

We show that, in our setting, asymptotic learning typically fails. That is, the agent's estimate is not consistent. Thus, we complement the vast literature on model misspecification, which obtains similar results by assuming that the true parameter values are not in the support of the decision maker's prior distribution. In contrast, we maintain the assumption that the true precisions are in the set of beliefs the agent deems possible. In fact we show that the agent's estimate is typically inconsistent *even* in cases where a misspecified Bayesian decision maker would learn the truth.

These results have several implications. First, our model provides a foundation for disagree-

[1] See Section 6 for a detailed discussion of these extensions.

ment among agents even when information is abundant. In particular, we show that two agents with the same prior beliefs, who evaluate estimation errors differently, could disagree significantly asymptotically. Note that a Bayesian decision maker's posterior belief converges to a Dirac measure. Hence, regardless of the loss function, her estimate will be equal to this value. In contrast, our agent's asymptotic estimate depends on the particular form of the loss function. Thus, ambiguity about the precision of information sources might rationalize disagreement even between informed experts who aim to find out the truth — e.g. scientists with access to the same large dataset.

Second, we argue that a small amount of ambiguity can significantly amplify the effect of model misspecification. Concretely, assume that the agent observes the unbiased signals directly. In that case, a Bayesian decision maker estimates the state correctly even when she holds misspecified beliefs about the precision of information sources. In contrast, we show that for any amount of prior ambiguity, however small, there exist misspecified agents that, when faced with this ambiguity, experience arbitrarily large losses and estimation errors. Thus, the interaction between ambiguity and model misspecification might prove to be a fruitful direction for future research.

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

Last, we consider an application where the decision maker learns from others' actions, instead of observing signals directly. An ambiguity averse econometrician observes choices by Bayesian decision makers who attempt to estimate a payoff relevant state given their private information. She aims to estimate this state but does not know the precision of their private signals. For a concrete 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. We show that the econometrician generically 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 prior beliefs and the true level of precisions.

Related Literature Our paper follows the literature on learning under ambiguity. Epstein and Schneider (2007) introduce a framework where an agent seeking to learn the state of the world, lacks confidence in their information about the environment. They consider the MaxMin Expected Utility model (MEU) following Gilboa and Schmeidler (1989) and a general updating rule for ambiguity that encompasses both prior-by-prior (full Bayesian) updating (Pires, 2002) and maximum likelihood updating (Gilboa and Schmeidler, 1993). Epstein and Schneider (2008) study an application to a financial market where the representative agent observes one signal with ambiguous precision, and updates her beliefs prior by prior. They show how this ambiguity affects reactions to information and the asset price. Follow-up papers extend these results by incorporating ambiguity on the mean of the signals, and by considering equilibrium portfolio choices as well as general utility functions (Illeditsch, 2011; Gollier, 2011; Condie and Ganguli, 2017). In this paper, we consider a similar setup as Epstein and Schneider (2008) but focus on whether ambiguity vanishes, and whether the agent can estimate the state correctly, when the number of signals she observes goes to infinity.

Of relevance is also the literature on single-agent misspecified learning, which is another possible driving force for the failure of asymptotic learning. In this literature, a misspecified agent typically has a prior 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 agent's beliefs converge, although not to the true state. Other works have focused on settings where the signals can be affected by the actions of the agent and are hence endogenous. Nyarko (1991) and Fudenberg et al. (2017) provide examples where the convergence of beliefs fails. Similar to our setup Heidhues et al. (2019) consider the convergence of beliefs and actions with a Gaussian prior and signals. Frick et al. (2020), Esponda et al. (2019), and Fudenberg et al. (2020) focus on the convergence results in general models with finite actions. Our paper differs from the existing work in three ways. First, the agent in the misspecified learning literature is a Bayesian learner, while in our setup, the decision maker holds multiple beliefs and adopts prior-by-prior updating. Second, the decision maker in our model is not misspecified in the sense that the true model is contained in her set of priors. 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 which produce good performance even with deviations from assumptions on the data generation 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 one 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 in order to reconcile the asymptotic disagreement between Bayesian and frequentist inferences. By comparison, our focus is on whether information aggregation is successful as the number of sources grows without bound. Instead of hinging on set-identified parameters, ambiguity does not vanish in our model since the precisions of different information sources are allowed to be different and those parameters are not payoff-relevant. This is in contrast with Marinacci (2002) where ambiguity vanishes since all observations are drawn with replacement from the same ambiguous urn.

2 Setup

An decision maker aims to learn the state of the world $\theta \in \Theta := \mathbb{R}$, and has access to N information sources. Denote the set of information sources as $I =: \{1, \ldots, N\}$. The prior distribution P_0 of the state θ is a normal distribution $\mathcal{N}\left(\mu, \frac{1}{\rho_\mu}\right)$, where $\rho_\mu > 0$ is the inverse of the variance. We call ρ_μ the *precision* of the prior. The prior is common knowledge among the decision maker and all information sources. Each information source $i \in I$ features a signal $s_i = \theta + \varepsilon_i$, where the noise ε_i is normally distributed with mean 0 and precision $\rho_i > 0$, that is, $\varepsilon_i \sim \mathcal{N}\left(0, \frac{1}{\rho_i}\right)$.[2] We assume that the state and all noises are independent from each other given precisions.

We further assume that the actual precisions of information sources are drawn i.i.d. from

[2]In the Online Appendix, we consider the case of biased signals. There, we argue that the unbiased signals case not only delivers all the intuition, but also allows us to isolate the effect of ambiguity. Thus, we focus our analysis on the case with no signal bias.

some distribution function G on $[\underline{\rho},\overline{\rho}]$ with $\overline{\rho} > \underline{\rho} > 0$. However, the decision maker in our model is ambiguous about the precisions of her information sources. 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. The decision maker can form conjectures about the precision of any information source i. We denote the decision maker's conjectured precision as $\hat{\rho}_i \in [\underline{\rho},\overline{\rho}]$. Finally, notice that the decision maker is not misspecified since she does not deem the actual precisions as impossible ex ante. This follows from the assumption that the actual precision ρ_i of information source i lies in the perceived precision set $[\underline{\rho},\overline{\rho}]$.

As mentioned in the introduction, the decision maker might be unable to observe the realized signals of the information sources. Instead, the observable for each information source i is $\boldsymbol{a}(s_i,\rho_i)$, which depends on the the realized signal and the precision of information source i. We assume that for each precision ρ_i, the observable $\boldsymbol{a}(s_i,\rho_i)$ is invertible as a function of s_i. Denote the realized observable as a_i and the inverted function for signals as $\boldsymbol{s}^{\boldsymbol{a}}(\rho_i,a_i)$. Given the observable a_i and the conjectured precision $\hat{\rho}_i$, the decision maker's conjectured signal is $\hat{s}_i = \boldsymbol{s}^{\boldsymbol{a}}(\hat{\rho}_i,a_i)$, which might be different from the actual realized signal s_i. Moreover, conditional on the realized state θ, the actual observables are i.i.d. according to the distribution function F on $\mathbb{R}$ where

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

with $F_\rho \sim \mathcal{N}\left(\theta,\frac{1}{\rho}\right)$ for each $\rho \in [\underline{\rho},\overline{\rho}]$. Later in this paper, we will discuss several different observables. For instance, the unbiased signals sources might be directly observable — $\boldsymbol{a}(s_i,\rho_i) = s_i$. We also study the case where decision maker can observe estimates of Bayesian agents based on their prior and private signals — $\boldsymbol{a}(s_i,\rho_i) = \frac{\rho_i s_i + \rho_\mu \mu}{\rho_i + \rho_\mu}$, assuming a common prior.

Belief Updating Denote the profile of precisions as $\rho^N := (\rho_1,...,\rho_N)$, the profile of conjectured precisions as $\hat{\rho}^N := (\hat{\rho}_1,...,\hat{\rho}_N)$ and the profile of observables as $a^N := (a_1,...,a_N)$, and for each $n \geq 1$ the set of distributions over $\mathbb{R}^n$ as $\Delta(\mathbb{R}^n)$. Following Epstein and Schneider (2007) and Epstein and Schneider (2008), we define $L^{\boldsymbol{a}}(\hat{\rho}^N,\theta) \in \Delta(\mathbb{R}^n)$ as the likelihood function for the profile of observables, which is the conditional distribution for observables given conjectured precisions $\hat{\rho}^N$ and the realized state θ. Then the set of likelihood functions of the decision maker can be represented by $\mathcal{L}_N^{\boldsymbol{a}}$, where

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

It is worthwhile to mention that 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 given the profile of conjectured precisions.

We assume that the decision maker adopts full Bayesian updating (Pires, 2002) to derive posteriors using the prior P_0 and the set of likelihood functions $\mathcal{L}_N^{\boldsymbol{a}}$. In other words, given the realized profile of observables a^N, and a vector of conjectured precisions $\hat{\rho}^N$, the posterior over the states $P_N^{\boldsymbol{a}}(a^N, \hat{\rho}^N) \in \Delta(\mathbb{R})$ is obtained by applying Bayes' rule.[3] Then the posteriors of the decision maker can be represented by the following set:

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

After observing the profile of observables, the decision maker makes an estimation g with regards to the state θ in order to minimize some loss function $u(g-\theta)$. We assume that $u : \mathbb{R} \to \mathbb{R}$ is strictly convex and minimized at 0. Given multiple beliefs, the decision maker is a maxmin expected utility (MEU) maximizer following Gilboa and Schmeidler (1989), and she evaluates her estimate based on the worst possible belief. This might be a result of the decision maker being ambiguity averse, or the decision maker's intention to derive a robust upper bound for the expected loss. That is, the decision maker's objective is to minimize the maximal expected loss across all distributions in the set of posteriors. She picks an estimate g to solve the following min-max problem:

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

To maintain tractability, we made several assumptions. In Section 6, we discuss how our re-

[3]Under this assumption, for each conjectured precision, the decision maker updates as if she was certain that the conjecture is correct. Alternatively, we could allow the decision maker to update her beliefs given a conjectured non-degenerate distribution about the precision of each signal. This does not change our analysis: all our results can be obtained using degenerate distributions alone.

sults depend on these assumptions. We find that our main insights continue to hold for general signal distributions, as well as for ambiguity attitudes beyond MEU. In contrast, we highlight that our belief updating rule plays a crucial role in our results.

In the rest of the paper, we will focus on the limiting case where the number of information sources N goes to infinity.[4]

3 Asymptotic Beliefs

In this section, we characterize how the agent's posterior set behaves as the number of observables grows large. In particular, recall that for any N, the agent observes a^N. Given a profile of conjectured precisions $\hat{\rho}^N$, his posterior belief is $P_N^{\mathbf{a}}(a^N, \hat{\rho}^N)$. We are interested in the asymptotic behavior of the agent's posterior set: $\mathbb{P}^{\mathbf{a}}(a^N)$. For that, we define the limit set of posteriors:

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

Notice that $\mathbb{P}_{\infty}^{\mathbf{a}}(a)$ is defined as the set of limits of posteriors that can be generated by some profile of precisions. That definition is silent about which posterior beliefs converge. In fact, there are many non-converging sequences of posterior beliefs, but, as will become clear in Section 4, these sequences are immaterial for our discussion.

In order to characterize this set, we start by interpreting the optimization problem described in Section 2 as a zero-sum game between the decision maker and nature. Under this interpretation, after signals are realized, the decision maker chooses an estimate for the state so as to minimize her loss function. Subsequently, with knowledge of the estimate, nature is free to choose, for each signal, any precision within the uncertainty set of the decision maker. The decision maker's objective is, then, to guarantee the lowest loss conditional on the fact that nature acts after her and to her detriment.

Example 1: Quadratic Loss with Observable Signals To help build intuition, and as a rough sketch of the proof of our more general results, we describe and partially analyze the special case in which the loss function is quadratic, while the observables are simply equal to the realized

[4] In the Online Appendix we solve the estimation problem for finite N when the loss function is quadratic.

signal values: i.e. $u(g-\theta) = (g-\theta)^2$ and $\boldsymbol{a}(\rho_i, s_i) = s_i$. That is, the decision maker has access to N unbiased and normally distributed signals $s_i = \theta + \varepsilon_i$, with $\varepsilon_i \sim \mathcal{N}\left(0, \frac{1}{\rho_i}\right)$. Recall that the true ρ_i are unknown to the decision maker, who entertains an interval of perceived precision $[\underline{\rho}, \overline{\rho}]$. The decision maker's objective is to minimize the maximal mean squared errors across all distributions in the set of posteriors. We denote by s^N the vector of the N observed signals. She picks an estimate g to solve the following problem:

$$\min_{g} \max_{p \in \mathbb{P}^s(s^N)} \left\{ \mathbb{E}_p\left[(g-\theta)^2\right] \right\}$$

Due to the properties of the quadratic loss function, the above optimization problem can be simplified to one which only depends on the conditional mean and variance of the state, which can be calculated in closed from due to the assumption of joint normality. Denote them as $\mathbb{E}[\theta | s^N, \hat{\rho}^N] = \frac{\hat{\rho}^N \cdot s^N + \rho_\mu \mu}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}$ and $\mathbb{V}[\theta | s^N, \hat{\rho}^N] = \left(1 - \frac{\hat{\rho}^N \cdot \mathbb{1}^N}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}\right)\frac{1}{\rho_\mu}$ respectively. Then the objective of the decision maker becomes

$$\min_{g} \max_{\hat{\rho} \in [\underline{\rho}, \overline{\rho}]^N} \left\{ \left(g - \mathbb{E}[\theta | s^N, \hat{\rho}^N]\right)^2 + \mathbb{V}[\theta | s^N, \hat{\rho}^N] \right\}$$

By changing the precision of each signal, nature affects both the squared bias and the 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. From the definition of the posterior variance, it becomes clear that as long as each signal is somewhat informative ($\underline{\rho} > 0$), as the number of available signals N increases, the posterior variance converges to 0. Hence, the more signals the decision maker receives, the more nature focuses on affecting the decision maker's loss function via the squared bias. In the extreme case, in which $N \to \infty$, for any choice of precisions that nature may consider, the posterior variance is equal to 0 and nature utilizes the square bias as its only lever. We next 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.

Definition 1. *An assignment $\hat{\rho} : \mathbb{R}^\infty \to [\underline{\rho}, \overline{\rho}]^\infty$ is order-preserving if $s_i \leq s_j \implies \rho_i \leq \rho_j$ and it is order-reversing if $s_i \leq s_j \implies \rho_i \geq \rho_j$. An assignment is a threshold assignment if it is order-preserving* *or order-reversing and* $\mathrm{Im}(\hat{\rho}) \in \left\{\underline{\rho}, \overline{\rho}\right\}^{\infty}$

Lemma 1. *Let* $\hat{\rho}^*$ *solve* $\max_{\hat{\rho}\in[\underline{\rho},\overline{\rho}]^\infty} \left(g - \mathbb{E}[\theta|s,\hat{\rho}]\right)^2$ *for some* $g \in \mathbb{R}$. *Then* $\hat{\rho}^*$ *is a threshold assignment.*

Nature finds it optimal to assign precisions to signals so as to maximize the squared bias. Intuitively, the way to do so is to either maximize or minimize the posterior mean: if the decision maker's estimate g is relatively low, nature finds it optimal to maximize the posterior mean, and vice versa. The intuition for Lemma 1 can be derived by analyzing the expression of the posterior mean when N goes to infinity. In that case, given an observed empirical distribution of signals, H, the expression for the posterior mean can be written as: $\mathbb{E}[\theta|s,\hat{\rho}] = \frac{\int s\hat{\rho}(s)dF(s)}{\int \hat{\rho}(s)dF(s)}$. Consider nature's choice to maximize this expression, while keeping the same sum of precisions $\int \hat{\rho}(s)dF(s) = c \in [\underline{\rho}, \overline{\rho}]$. Since c pins down the denominator of the expression for the posterior mean, nature chooses an assignment to maximize $\int s\hat{\rho}(s)dF(s)$. It is clear that, 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 do that, therefore justifying the optimality of threshold strategies. Naturally, an analogous strategy is optimal to minimize the posterior mean.

To summarize this example, as the number of signals goes to infinity nature focuses on affecting the agent's bias by strategically assigning precisions to signal realizations. This is true because asymptotically this is the only way nature can affect the agent's loss, as variance goes to zero regardless of the decision maker's conjecture. Finally, nature can implement this bias-maximizing behavior applying threshold strategies: monotonic precision assignments that use only extreme precisions.

3.1 Ambiguity does not Vanish

When signals are observable and the loss function is quadratic, we have argued in the last section that (i) nature can restrict attention to threshold strategies; and (ii) as the number of signals goes to infinity, the set of posteriors converges to a set of degenerate distributions. We now show that these two insights generalize.

First, we provide sufficient conditions on observables that guarantee that nature can still restrict attention to threshold strategies. Recall that under general observables the agent cannot observe the realized signal, but rather has to backtrack those signals based on their conjectured precision. In particular, given conjectures precision $\hat{\rho}_i$ and observable realization a_i, the agent believes the signal realization is $\hat{s}_i = \mathbf{s}^{\mathbf{a}}(\hat{\rho}_i, a_i)$. In contrast to the observable signals case, when associating a particular precision to an observable realization, the agent changes his interpretation about the signal realization. The following assumption makes sure that the effect of this association does not break the monotonicity between observables and inverted signals that is required for simple threshold strategies to be optimal.

Assumption 1. *Define the weighted inverted signal function* $g(\rho, x) \equiv \rho \mathbf{s}^{\mathbf{a}}(\rho, x)$*. g is affine in ρ and strictly supermodular.*

This assumption allows for a broad range of observables relevant in several economics applications. The two examples described in the setup — directly observable unbiased signals and observable estimates from Bayesians with private information — satisfy this assumption. In the context of financial markets, the demand of CARA investors for an asset with value θ also satisfies Assumption 1. In particular, when the unbiased signals are the investor's private information, their demand for the risky asset is $\boldsymbol{a}(s_i, \rho_i) = \frac{1}{\alpha}(\rho_i s_i + \rho_\mu \mu)$, where α is their absolute risk aversion.

Lemma 2. *Let $\hat{\rho}^*$ solve* $\max_{\hat{\rho} \in [\underline{\rho}, \overline{\rho}]^\infty} \left(g - \mathbb{E}[\theta | a, \hat{\rho}]\right)^2$ *for some $g \in \mathbb{R}$. Under Assumption 1, $\hat{\rho}^*$ is a threshold assignment.*

We now address the asymptotic behavior of posteriors. As previously discussed, each strategy of nature corresponds to a plausible belief in the agent's belief set. In last section, as the number of observables went to infinity, we argued that nature loses the ability to influence the posterior variance, as aggregate information becomes infinitely precise. However, by assigning precisions to signals, nature could still affect the agent's bias. With a general loss functions, higher moments of the posterior distribution are payoff-relevant for the agent. Nevertheless, the above rationale is preserved: all moments but the posterior mean become irrelevant asymptotically, and, in the limit, nature can only command the interval of posterior means. As a consequence, the set of posterior beliefs converges to an interval of degenerate distributions regardless of the loss function.

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

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

Then, for almost all sequences, a, of realized observables:

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 s.*

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

Theorem 1 formalizes the observation above. It 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 $\overline{m}$. Finally, Theorem 1 characterizes the values of these boundaries. For example, $\overline{m}$ is generated by the following strategy of nature: 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 highest such posterior mean is expressed by the fixed point $\overline{m}$.

The fundamental consequence of Theorem 1 is that induced ambiguity on the state does not vanish asymptotically. Rather, the agent still entertains a wide range of values for the state θ even when he has access to an arbitrarily large number of informative observables. This 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 around their means. In the next section we show how the presence of ambiguity in the limit set of posteriors affects the optimal estimate of the agent.

4 Asymptotic Estimate

In this section, we characterize the asymptotic behavior of the decision maker. In particular, we are interested in analyzing how ambiguity with regards to the decision maker's information sources affects her ability to correctly estimate the state as the number of observables grows. Recall that, for each realization of observables a^N, her estimate $g^*(a^N)$ minimizes her loss function, considering the worst case posterior in $\mathbb{P}^{\boldsymbol{a}}(a^N)$. Because observables and loss functions are arbitrary, obtaining an explicit solution to $g^*(a^N)$ for finite N is not an easy task, which makes a direct attempt at characterizing the solution intractable. Rather, we leverage on Theorem 1 to solve this problem. The main result of this section characterizes the asymptotic estimate by showing that the following limit exchange holds.

$$\lim_{N\to\infty} \arg\min_{g} \max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p\Big[u(g-\theta)\Big] = \arg\min_{g} \max_{p\in\mathbb{P}_\infty(a)} \mathbb{E}_p\Big[u(g-\theta)\Big]$$

Theorem 1 states that, for almost all realizations of observables $\mathbb{P}_\infty(a) = \{\delta_m : m \in [\underline{m}, \overline{m}]\}$. The limit swap above suggests that, as N grows, the optimal estimate converges to the estimate of an agent who does not know the mean of θ, but wants to guarantee the minimal loss in the interval $[\underline{m}, \overline{m}]$. This greatly simplifies the characterization: the asymptotic behavior of the estimate is pinned down by an extremely simple optimization problem. In this problem, the agent only cares about how biased her estimate is in the worst case scenario. Recall that her loss is larger the farther from the true state her estimate is. If her estimate is too far from $\underline{m}$, then 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 to these two extreme possible values of the state. This intuition is formalized in the next result.

Theorem 2. $g^*(s^N) \xrightarrow{a.s.} g^*$*, 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 mentioned above. A priori, it is not clear that the limit swap holds. First, 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 are always sequences of precisions such that posterior beliefs diverge. Still, our result confirms that the limit exchange is valid and that the heuristic argument we gave above goes through formally. We make this argument in two steps, addressing each of the concerns highlighted above.

The first step is to show that the decision maker's optimization can be approximated by an optimization that considers only the mean of posterior distributions, as N grows large. For any finite N, it is clear that the decision maker's loss is 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 $g^*(a^N)$ are bounded. Recall, from part 1 of Theorem 1, that non-converging posteriors are bounded. Thus, intuitively, N by N, the payoff obtained by a non-converging sequence can be bounded by the payoff of two converging sequences, so that restricting attention to the converging ones turns out to be without loss of generality. As a consequence, non-converging beliefs are innocuous: we can characterize the asymptotic behavior of the agent's estimate without addressing them. We prove that these two steps are sufficient to guarantee the convergence of g^*.

It is clear by Theorem 2 that the asymptotic estimate is typically incorrect. To see this, recall that $[\underline{m}, \overline{m}]$ in Theorem 1 are independent of the particular choice of the loss function. Rather, they are determined by the initial ambiguity and the observable function $\boldsymbol{a}$. In contrast, the asymptotic estimate is a consequence of the behavior of the loss function, u on the interval of posterior means $[\underline{m}, \overline{m}]$. This suggests that the decision maker estimates the state correctly asymptotically only in the knife-edge case in which their loss function coincides with the observable function in a very particular way. Moreover, when that is the case, perturbing either of these functions would once again lead to an incorrect limit estimate. This is particularly striking when compared to the behavior of a Bayesian agent who knows the precision of each source.[5] Because observables map one-to-one to signals conditional on those precisions, a Bayesian's asymptotic estimate would be equal to the state, regardless of loss function. The next two examples illustrate this result.

Example 1: Quadratic Loss with Observable Signals We revisit the example from last section to show a case in which the agent correctly estimates the state asymptotically. By Theorem 2, we

[5] We could also consider the comparison with a Bayesian agent who does not know the precision of each information source, but rather entertains a distribution over those precisions. The comparison remains the same as long as their statistical model is identified.

have that, under quadratic losses, $g^* = \frac{\overline{m}+\underline{m}}{2}$. Because signals are observable, Theorem 1 states that $\underline{m}$ and $\overline{m}$ are defined by:

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

Normality implies that the real distribution of signals F is symmetric around the true state θ. Thus, in this case, the decision maker estimates the state correctly. This follows as a result of the symmetric loss function, as well as the symmetry of the normal distribution, and the assumption of observable signals.

Corollary 1. $g^*(s^N) \xrightarrow{a.s.} \theta$.

The next example breaks this coincidence to show that this result is knife-edge.

Example 2: Asymmetric Loss with Observable Signals In this example we maintain the assumption of observable signals, but 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 decision maker's loss is different for over and under estimating the state θ. If $\lambda > 1$, for example, the agent is less concerned with losses when she overestimates the true state, compared to when she underestimates it. There are many reasons why this could be the case. For instance, a health official who wants to learn about the prevalence of a disease in a population would likely be more affected if he believes the transmission rate is lower than it really is than the other way around. Conversely, a product developer may face a much higher personal loss if they believe demand to be higher than it actually is and end up developing a costly product that fails to be marketed.

Following Theorem 2, we have that the optimal estimate satisfies $g^* = \frac{\underline{m}+\sqrt{\lambda}\overline{m}}{1+\sqrt{\lambda}}$. However, because the results of Theorem 1 do not depend on the loss function, we still have that $\frac{\underline{m}+\overline{m}}{2} = \theta$. Thus, the agent estimates incorrectly for any $\lambda \neq 1$. In particular, if $\lambda < 1$, her optimal estimate is below the real value of the state: $g^* < \theta$. The example above shows how an environment in which an agent estimates correctly can be easily perturbed so that the agent no longer estimates the state correctly. This is done by maintaining the assumption that signals are observable, but departing from the assumption of symmetric losses. In the next section we explore applications of our model. In one of these applications, we depart from Example 1 by changing the assumption of signal observability instead of the symmetry assumption. We characterize the optimal estimate for that case and show how, again, the decision maker generically fails to correctly estimate the state.

A Foundation for Disagreement The above example also highlights that the loss function directly affects the agent's asymptotic estimate. Note that a Bayesian decision maker's posterior belief converges to a Dirac measure, and consequently, her estimate of the state will be equal to this value. Thus, with multiple Bayesian agents, as the available information grows, regardless of their loss functions, Bayesian agents will agree on the optimal estimation of the state.[6] In contrast, our agent's asymptotic estimate continues to depend on the particular form of the loss function. Thus, ambiguity about the precision of information sources might rationalize disagreement even between informed experts who aim to find out the truth — e.g. scientists with access to the same large dataset. This offers a foundation for disagreement among agents even when information is abundant.

5 Applications

In this section, we will explore different applications of our main results by changing the set of observables available to the decision maker or her loss function.

5.1 Comparative Statics of Ambiguity

First, we revisit Example 1 on quadratic loss with observable signals. We claim that, contrary to intuition, making all signals more precise is not necessarily beneficial to the decision maker.

[6] With observable signals, this holds true even for misspecified Bayesian agents who wrongly perceive the precision of the signals. As the amount of information grows without bounds, their estimates converge to the same value.

Recall that by Theorem 1, the limit set of posteriors is a set of degenerate distributions δ_b with $\underline{m} \leq b \leq \overline{m}$ where

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

To see how the set of precisions deemed possible by the decision maker affects the limit set of posterior beliefs, first note that $\overline{m}$ and $\underline{m}$ only depend on the fraction of the highest and the lowest possible precisions, instead of their absolute values, since we can rewrite $\overline{m}$ and $\underline{m}$ as

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

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

Proposition 1. *Let $\eta = \frac{\overline{\rho}}{\underline{\rho}} \in (1, +\infty)$. Then $\overline{m}$ 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 1, $\eta = \frac{\overline{\rho}}{\underline{\rho}}$ can be interpreted as the size of ambiguity in the set of possible precisions $[\underline{\rho}, \overline{\rho}]$. When there is more ambiguity regarding precisions of signals ex ante, the limit set of posteriors also expands and hence there is more ambiguity regarding states ex post.

Now we explore the welfare implication of such comparative statics. By Corollary 1 of Theorem 2, the decision maker always estimates correctly at the limit when signals are observable. Thus, the optimal utility solely depends on the size of the limit set of posterior means, that is, $\overline{m} - \underline{m}$. Corollary 2 directly follows from Proposition 1.

Corollary 2. *As η increases, asymptotically the decision maker is strictly worse off.*

Corollary 2 has two possibly counter-intuitive implications. First, it implies that if we fix $\underline{\rho}$ and increase $\overline{\rho}$ the decision maker is strictly worse off. That is, she prefers all of her signals to be imprecise rather than being ambiguous that some signals might be more precise. Second, consider two decision problems with the set of possible precisions given by $[\underline{\rho}_1, \overline{\rho}_1]$ and $[\underline{\rho}_2, \overline{\rho}_2]$ respectively. If $\underline{\rho}_2 > \overline{\rho}_1$ and $\eta_1 = \frac{\overline{\rho}_1}{\underline{\rho}_1} < \eta_2 = \frac{\overline{\rho}_2}{\underline{\rho}_2}$, then the decision maker believes that any signal in the second decision problem is more precise than any signal in the first one, but she is strictly worse off in the second decision problem. This suggests that making all signals more precise is not necessarily beneficial to the decision maker.

5.2 Misspecification and Ambiguity

In this section, we show that a small amount of ambiguity might substantially amplify the effect of model misspecification. We say that an agent is misspecified if the real precisions of information sources are not in her consideration set — i.e. $\operatorname{supp} G \cap [\underline{\rho}, \overline{\rho}] = \emptyset$, where G is the true distribution of precisions. To focus on a simple environment, we proceed under the assumption that the agent observes unbiased signals directly, but has an asymmetric loss function u as in Example 2. Note that the impact of prior ambiguity can still be captured by the ratio η defined in the last section. Note that for the special case of no ambiguity $\eta = 1$. We say that an agent is a misspecified Bayesian, if she is misspecified and faces no ambiguity, $\eta = 1$.

Proposition 2. *Assume u is as in Example 2. For any $\eta > 1$, and any constant $C > 0$, there exist true distributions of precisions G such that $|g^* - \theta| > C$.*

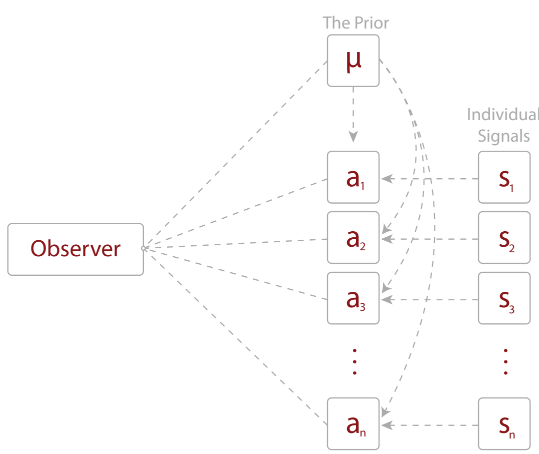
To see the connection of this proposition with misspecified Bayesians, note that the misspecified Bayesian agent believes the precision of information source i to be $\hat{\rho}_i \notin \operatorname{supp} G$. However, since signals are unbiased, this misspecification has no effect on asymptotic learning, regardless of the distribution G. In particular, for any loss function, the agent estimates the state correctly and, therefore, experiences zero loss. Now, consider this same agent when facing a small amount of ambiguity, η very close to 1. By Proposition 2, there exist true distributions over precisions, G, such that this agent makes arbitrarily large estimation errors. Thus, any amount of ambiguity, no matter how small, might be sufficient to cause large deviations from the correct estimate. These large deviations occur in settings in which the true precision of the signals is rather low. This is inconsequential for the misspecified Bayesian agent, who still estimates the state correctly. However, in the presence of ambiguity, the less informative the signals are in reality, the bigger this gap will be.

5.3 Aggregating Estimates

We study next the problem of an ambiguity averse econometrician who aims to estimate the state by aggregating observable estimations by numerous Bayesian agents. The agents share the same prior but have access to different information sources. While the econometrician knows the prior distribution of the state, she does not know the precision of the individual sources. This environment is reasonable in many applications. For instance, 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.

Formally, we assume that all agents and the decision maker share the same prior beliefs about the state θ. As in the previous section according to the prior $\theta \sim \mathcal{N}\left(\mu, \frac{1}{\rho_\mu}\right)$. Conditional on the realization of θ, agent i receives a private signal $s_i = \theta + \varepsilon_i$ where $\varepsilon_i \sim \mathcal{N}\left(0, \frac{1}{\rho_i}\right)$. That is, each agent receives an unbiased signal about the state. We consider the case in which each agent attempts to estimate the realized value of θ in order to minimize mean squared error. Given the prior and the private signal, the optimal Bayesian estimate for agent i would then be $\mathbf{a}(s_i, \rho_i) = \mathbb{E}[\theta|\rho_i, s_i] = \frac{\rho_\mu \mu + \rho_i s_i}{\rho_\mu + \rho_i}$. It is exactly these actions that the econometrician has access to. The setup studied in this section is graphically depicted in Figure 1.

Figure 1: Learning From Actions Setup



While each agent knows the precision of their private signal, this is not the case for the econometrician. We once more assume that for each signal the econometrician considers a set of possible precisions $\left[\underline{\rho}, \overline{\rho}\right]$. Since each action is a convex combination of the private signal s_i and the mean of the prior μ, an econometrician who intends to estimate the value of θ, will first have to transform the actions back to signals. For a conjectured precision $\hat{\rho}_i$, the recovered signal will be $\mathbf{s^a}(a_i, \hat{\rho}_i) = a_i + \frac{\rho_\mu}{\hat{\rho}_i}(a_i - \mu)$. We again assume that the loss function of the econometrician is quadratic. We start by utilizing Theorem 1 to characterize the limit set of posteriors of the econometrician in this example.

Proposition 3. *The boundaries of the limit set of posteriors for the econometrician are:*

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

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

The boundaries of the limit set of posteriors are defined by a fixed point similar to the one from the example with observable signals. However, here the econometrician has to backtrack realized signals from observed estimates, which leads to an adjustment term c. The next result is a corollary of Theorem 2.

Corollary 3. *For almost all sequences a and values of the state* θ, $\lim_{N\to\infty} |g^*(a^N) - \theta| > 0$.

That is, the econometrician's estimation converges away from the truth almost surely. This is exactly because inverting from observables to signals depends on the conjectured precisions and of the prior mean. The lack of knowledge about the former makes it impossible for the econometrician to distinguish signal realization from the prior mean, thus generating a bias in the recovered signals. The following assumption allows us to clearly characterize the optimal estimate and to analyze comparative statics.

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. This assumption allows us to characterize how the econometrician estimate differs from the true parameter value. We say that the econometrician *overreacts* if $|g^* - \mu| > |\theta - \mu|$ and *underreacts* it if the inequality is reversed. In other words, an estimation overreacts to information if it is farther from the prior mean than the real state is.

Proposition 4 (Guess Characterization). *$g(A^n) \xrightarrow{a.s.} g^*$ where*

1. *If $\mu = \theta$, $g^* = \theta$*

2. *If $\mu \neq \theta$, then $\exists \tilde{\rho} < \tilde{\tilde{\rho}}$ such that*

   - *If $\rho^* \leq \tilde{\rho}$, then $|g^* - \mu| > |\theta - \mu|$ and the agent underreacts to observed actions*

   - *If $\rho^* \geq \tilde{\tilde{\rho}}$, then $|g^* - \mu| < |\theta - \mu|$ and the agent overreacts to observed actions*

   - *If $\tilde{\rho} < \rho^* < \tilde{\tilde{\rho}}$, underreacting if $|\theta - \mu|$ is small and overreacting if $|\theta - \mu|$ is large*

   where:

$$\tilde{\rho} = \frac{2\underline{\rho}\overline{\rho}}{\underline{\rho} + \overline{\rho}} \qquad\qquad \tilde{\tilde{\rho}} = \underline{\rho} H\left(\overline{m}(\tilde{\tilde{\rho}}, \mu)\right) + \overline{\rho}\left(1 - H\left(\overline{m}(\tilde{\tilde{\rho}}, \mu)\right)\right)$$

Proposition 4 reveals that whether the decision maker over or underreacts depends on the true precision of the signals and possibly the realization of the state θ. Roughly speaking, the optimal robust estimate corresponds to the decision maker trying to backtrack the mean of the unobservable signals from the mean of observed actions. Because signals are unbiased, their unobservable mean is effectively θ, the state the econometrician aims to estimate. When ρ^* is high, θ is relatively close to the mean of actions, as the agents place a high weight on their unbiased signals when choosing their actions. However, the econometrician does not know the real precision, so she backtracks signals from actions using, roughly, the same method regardless of what ρ^* is. Because of that, the direction of her estimation error depends on the true precision.

Finally, as the prior precision ρ_μ changes, there is non-monotonicity in the accuracy of the optimal estimate. Note that the estimation error is related to how actions are contaminated by the prior, making it impossible for the agent to disentangle the effect of the prior from the effect of individual information. When the prior is extremely imprecise, $\rho_\mu \approx 0$, this contamination is extremely small, and thus, the optimal estimate is approximately equal to the one characterized in the observable signals example: the econometrician estimates correctly. On the other hand, when the precision of the prior grows to infinity, $\rho_\mu \to \infty$, the agent also estimates correctly by essentially disregarding the information in the observed actions. For intermediate values, however, Corollary 3 implies that the estimate is wrong almost surely. In other words, the accuracy of the estimate is not monotonic with the precision of the prior: 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 decision maker only knows the highest possible and lowest possible precisions of each information source but nothing else; (ii) both the state and signals follow normal distributions; (iii) the decision maker is a MEU maximizer regarding ambiguity; (iv) the decision maker adopts full Bayesian updating. In this section, we will briefly argue that our main result — that ambiguity does not vanish asymptotically — remains valid when we relax the first three assumptions. Hence, the essential assumption is the updating rule under ambiguity.

Information about Precisions Consider the case where the decision maker has more information about the precisions of her information sources ex-ante. Specifically, the decision maker knows 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 low precision $\underline{\rho}$. The decision maker does not know which group a particular information source belongs to. Recall that the optimization problem of the decision maker can be interpreted as a zero-sum game between her and nature. The decision maker's additional information heavily restricts nature's choices on precisions. However, when $\alpha \in (0,1)$, nature can still induce the decision maker 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, even with the additional restrictions, ambiguity will not asymptotically vanish. The asymptotic estimate of the decision maker will be correct only with observable signals and a symmetric loss function, and incorrect otherwise.[7] Regardless of whether her estimate is correct or not, for any $\alpha \in (0,1)$ the decision maker faces ambiguity, and thus, suffers from a loss. Consequently, an decision maker who believes all her information sources to have minimal precision $\underline{\rho}$, is better off than an decision maker who believes

[7] It is worth mentioning that if $\alpha = 1$ or 0, that is, the decision maker knows that all of her information sources are precise or imprecise, asymptotic learning is successful, and the decision maker has asymptotic loss 0 since there is no ambiguity.

a fraction of her information sources have precision $\overline{\rho} > \underline{\rho}$. [8]

Distributions Typically, ambiguity does not vanish even when the state and signals are not normally distributed. In particular, as long as the decision maker's belief set is rich enough, ambiguity about the state will persist asymptotically. We expand on that next. For general distributions, the precision of each signal is no longer fully captured by the reciprocal of its variance. In order to extend our model to other distributions, assume that the decision maker considers a set of likelihood functions for each information source. Just as in the main model, each allocation of likelihoods to information sources defines a belief for the agent. Under Full Bayesian Updating for each belief the agent forms a posterior on the state. As long as two different beliefs result in two different posterior means, our results hold. This condition is generic: if a non singleton set of likelihood functions does not satisfy this property, one of the functions can be perturbed so that the property holds and ambiguity does not vanish. Finally, departing from normal distributions, it is no longer necessarily the case that higher moments of the posterior vanish. Thus, in addition the induced ambiguity, the agent might face risk even asymptotically.

Ambiguity Preferences We can extend the decision maker's preference under ambiguity. So long as the ambiguity faced by the decision maker takes the form of a set of beliefs over the state and signals and she adopts the full Bayesian updating rule upon receiving signals, then Theorem 1 still holds. To see that, note that our analysis on asymptotic beliefs does not rely on the specific utility function of the decision maker. For instance, ambiguity does not vanish if the decision maker has an uncertainty averse preference introduced by Cerreia-Vioglio et al. (2011), which incorporates most ambiguity-averse preferences used in the literature, including variational preferences (Klibanoff et al., 2005), smooth ambiguity preferences (Maccheroni et al., 2006), and, of course, MEU.

Updating Rule. Finally, we note that our result does rely on the updating rule under ambiguity. An alternative to the full Bayesian updating rule is the maximum likelihood rule. Unlike full Bayesian updating, where the decision maker applies Bayes' rule to the entire set of priors, the decision maker with the maximum likelihood rule would discard priors that do not ascribe the

[8] For a detailed analysis, refer to the Online Appendix.

maximal probability to the observed signals and update the remaining priors according to Bayes' rule. Hence, the maximum likelihood rule suggests that ambiguity might vanish even with one single signal.

Appendix: Proofs

Proof of Lemma 2

By the form of the objective function, it is easy to see that $\hat{\rho}^*$ solves either $\max_{\hat{\rho}[\underline{\rho},\overline{\rho}]^\infty} \mathbb{E}[\theta|a,\hat{\rho}]$ or $\min_{\hat{\rho}[\underline{\rho},\overline{\rho}]^\infty} \mathbb{E}[\theta|a,\hat{\rho}]$.

Given a distribution of observables, H with density h, recall that

$$v(\hat{\rho}) \equiv \mathbb{E}[\theta|a,\hat{\rho}] = \int \frac{\hat{\rho}(x)\hat{s}(x,\hat{\rho}(x))}{\int \hat{\rho}(x)h(x)dx} h(x)dx$$

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

$$\begin{aligned} &\max_{\hat{\rho}[\underline{\rho},\overline{\rho}]^\infty} \{v(\hat{\rho}) : \int \hat{\rho}(x)h(x)dx = M\} \\ = \frac{1}{M} &\max_{\hat{\rho}[\underline{\rho},\overline{\rho}]^\infty} \{\int \hat{\rho}(x)\hat{s}(x,\hat{\rho}(x))h(x)dx : \int \hat{\rho}(x)h(x)dx = M\} \end{aligned}$$

where the last equality is justified because we are equating the denominator of v to M. By Lagrange multiplier Theorem in Banach spaces, we obtain that there is λ such that, for each x:

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

By the supermodularity in Assumption 1, we know the objective function of each of these optimizations is supermodular, so $\hat{\rho}(x)$ is increasing with x, according to Topkis' lemma. By affinity, 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. Thus, maximizing over M's the solution must also be a threshold. Clearly, the same result holds for minimization and the proof is concluded.

□

Proof of Theorem 1

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:

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

and

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

$\underline{m}^N$ and $\overline{m}^N$ are (random) bounds on posterior means. Assume that $\underline{\rho}^N$ and $\overline{\rho}^N$ are the respective maximizers.

Let $\hat{\rho} : \mathbb{R} \to [\underline{\rho}, \overline{\rho}]$ be a precision assignment. Let H be the real distribution of observables. Again, given a precision assignment, signals are jointly normally distributed with the state, so we can write the posterior mean 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:

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

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

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

By the proof of Lemma 2, $\underline{m}$ is solved by a threshold strategy. We can then write the optimization that determines it by:

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

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

The first order condition leads to:

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

Which implicitly defines the value $\underline{m}$ that solves that maximization. We show that the objective function is single-peaked, so that the first order condition is necessary and sufficient. The first derivative of v can be written as:

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

First, notice that because the second term is positive for all $a \in \mathbb{R}$, the sign of v' is determined by $v(a)-a$. This immediately implies v is quasiconcave: if there is $\underline{a}$ such that $v'(\underline{a}) > 0$, then $v'(a) > 0$ for all $a \leq \underline{a}$; similarly, if there is $\overline{a}$ such that $v'(\overline{a}) < 0$, then $v'(a) < 0$ for all $a \geq \overline{a}$. We prove the second, the first follows by symmetry. Assume there is $\overline{a}$ such that $v'(\overline{a}) < 0$ and, to obtain a contradiction, let there be $a > \overline{a}$ with $v'(a) > 0$. Since v' is continuous, there must be $\overline{a} < b < a$ with $v'(b) = 0$, which implies $v(b) = b$. Choose the smallest such $b > \overline{a}$, so for $\overline{a} \leq x < b$, $v'(x) < 0$. We then have:

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

since $v'(\overline{a}) < 0$ implies $v(\overline{a}) < \overline{a}$. We have thus obtained a contradiction.

Because v is quasiconcave, the first order condition is necessary and sufficient. We now prove that the solution exists and is unique.

As $a \to -\infty$, $v(a) \to \int_{-\infty}^{\infty} x f(x) dx$, as all signals are assigned precision $\overline{\rho}$, leading to uniform weighting. Because we know F has a finite mean, that implies that we can find a sufficiently small number $\underline{a}$ such that $v(\underline{a}) - \underline{a} > 0$, implying $v'(\underline{a}) > 0$. Notice that the same should be true for all $a \leq \underline{a}$, so that v is an increasing function in $(-\infty, \underline{a}]$.

On the other hand, as $a \to \infty$, again we have $v(a) \to \int_{-\infty}^{\infty} x f(x) dx$, this time because all signals are receiving precision $\underline{\rho}$. Then, there is a sufficiently high number $\overline{a}$ with $v(a) - a < 0$, 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:

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

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 the expectation $\overline{m}^N$ by the expectation generated by a threshold strategy as N grows large. For any realization of actions, a^N, let F^N be the associated empirical distribution of actions. We then define:

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

Call the objective function of the problem above $\Psi^N(a)$. At the same time, using the proof of Lemma 2 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:

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

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 that, notice that for almost all sequences a, it must be that $\sup_a\left\{F^N(a)-F^N(a-)\right\}\leq\frac{1}{N}$. Applying that, the uniform convergence result is direct.

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 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

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

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 3. *((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

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

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

Easy to see

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

Then we want to show that $\mathcal{V}_1$ and $\mathcal{V}_2$ are both 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 $\{a_1, ..., a_M\}$ on the extended real line such that $a_1 = -\infty$, $a_M = \infty$ and

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

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

$$\|F^N - F\|_{\mathcal{V}_1} = \sup_{v \in \mathcal{V}_1} |F^N(v) - F(v)| = \sup_{a \in \mathbb{R}} |F^N(v_1^a) - F(v_1^a)| \xrightarrow{a.s.} 0. \tag{1}$$

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

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 1 and 2, 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 3. *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 3. 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

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

Hence by condition (1), we know $|\tilde{\Psi}^N(\overline{m}_N) - \Psi(\overline{m})| \xrightarrow{a.s.} 0$. Finally, suppose by contradiction that $\overline{m}_N$ does not converge to $\overline{m}$ almost surely. Then there exists an event M with positive probability such that for all $\omega \in M$, $\overline{m}_N(\omega) \nrightarrow \overline{m}(\omega)$. As $\overline{m}$ is the unique minimum of Ψ by condition (3), $\tilde{\Psi}(\overline{m}_N(\omega)) \nrightarrow \Psi(\overline{m}(\omega))$. Again condition (1) implies that $|\tilde{\Psi}^N(\overline{m}_N) - \Psi(\overline{m}_N)| \xrightarrow{a.s.} 0$. Hence we know that there exists $M' \subseteq M$ with positive probability such that for all $\omega \in M'$, $\tilde{\Psi}^N(\overline{m}_N(\omega)) \nrightarrow \Psi(\overline{m}(\omega))$, which contradicts with $|\tilde{\Psi}^N(\overline{m}_N) - \Psi(\overline{m})| \xrightarrow{a.s.} 0$. Thus, we have $\overline{m}_N \xrightarrow{a.s.} \overline{m}$.

□

Now it suffices to show that the conditions in Lemma 3 holds 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 belief means. For any N, with observables a^N and conjectured precisions $\hat{\rho}^N$, recall we have:

$$\theta | s^N, \hat{\rho}^N \sim \left(\frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s^a}(a_i, \hat{\rho}_i) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}, \left(1 - \frac{\sum_{i=1}^N \hat{\rho}_i}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}\right) \frac{1}{\rho_\mu} \right) \tag{3}$$

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 $\underline{m}^N$, $\overline{m}^N$:

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

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 decision maker converge to, $\mathbb{P}_\infty(a)$. By 3, it is clear that a necessary condition for weak convergence is that the posterior mean $\frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s^a}(a_i, \hat{\rho}_i) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}$ converges. We can then focus on sequences with converging means. Define $b = \lim_{N\to\infty} \frac{\sum_{i=1}^N \hat{\rho}_i \boldsymbol{s^a}(a_i, \hat{\rho}_i) + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}$.

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

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

By Step 3, the variance converges to zero. We then have, for all t:

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

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 $\{\rho(a_i)\boldsymbol{s}^{\boldsymbol{a}}(a_i, \rho(a_i))\}_{i=1,\ldots}$ is a sequence of independent signals with uniformly bounded variance. Then, by the strong law of large numbers:

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

We finish this step by showing that by appropriately choosing the function ρ, $\frac{\int \rho(x) s dF(x)}{\int \rho(x) dF(x)}$ can achieve any point between $\underline{m}$ and $\overline{m}$. To see that, recall that $\overline{m} = \max_a \Psi(a)$. It should be clear that $\mu = \min_a \Psi(a)$. Since Ψ is continuous, by choosing different a's any number in $[\mu, \overline{m}]$ can be achieved. Because any a corresponds to a particular threshold assignment ρ, this means that $\frac{\int \rho(x)\boldsymbol{s}^{\boldsymbol{a}}(x, \rho(x)) dF(x)}{\int \rho(x) dF(x)}$ can achieve any value in $[\mu, \overline{m}]$. With the symmetric argument for $\underline{m}$ we obtain the result and complete Step 3.

$\square$

Proof of Theorem 2

Define

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

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

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

Also denote

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

where $\underline{m}$ and $\overline{m}$ are defined in Theorem 1.

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:

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

where $\overline{m}$ and $\underline{m}$ are defined in Theorem 1 and the equality follows from the fact that u is convex.

The result of the proposition is a consequence of the following lemma.

Lemma 4. *Let f^N be a sequence of 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 uniformly bounded almost everywhere.*

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

Proof of Lemma 4. 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 3, 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 Γ^N, Γ, $g^N \equiv g^*(s^N)$ and g^* solving $u(g^* - \overline{m}) = u(g^* - \underline{m})$ satisfy the conditions of Lemma 4. 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. $\mathbf{\sup_{g \in C} |\Gamma(g) - \Gamma^N(g)| \xrightarrow{a.s.} 0}$**, as $n \to \infty$, for all compact sets $C \subset \mathbb{R}$**

Step 1.1. $\sup_{\mathbf{g}\in\mathbb{R}} |\tilde{\boldsymbol{\Gamma}}^{\mathbf{N}}(\mathbf{g}) - \boldsymbol{\Gamma}^{\mathbf{N}}(\mathbf{g})| \xrightarrow{\text{a.s.}} \mathbf{0}$

We start by using the auxiliary function $\tilde{\Gamma}^N$. As N grows to infinity, the gap between Γ^N and $\tilde{\Gamma}^N$ shrinks uniformly. We prove this statement next. Start by noticing that, for fixed P^N, for any $p \in \mathbb{P}^a(a^N)$:

$$\Gamma^N(g) \geq \mathbb{E}_p\left[u(g-\theta)\right] \geq u\left(g - \mathbb{E}_p[\theta]\right)$$

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

Now, for each g, $\tilde{\theta}$ and $q \in [\underline{m}, \overline{m}]$, Taylor's rule implies existence of $\omega(g, \tilde{theta}, q)$:

$$u(g-\tilde{\theta}) = u\left(g-q\right) + u'(g-\omega(g,\tilde{\theta},q))\left(\tilde{\theta}-q\right)$$

By the implicit function theorem, $\omega(\cdot,\theta,\cdot)$ is a differentiable, and thus continuous function.

Now, if there exists $p \in \mathbb{P}^{\boldsymbol{a}}(a^N)$ with $\mathbb{E}_p[\theta] = q$, we can take expectations with respect to p in the above equation to obtain:

$$\mathbb{E}_p[u(g-\tilde{\theta})] = u(g-\mathbb{E}_p[\theta]) + \mathbb{E}_p\left[u'(g,\omega(g,\tilde{\theta},\mathbb{E}_p[\theta]))\left(\tilde{\theta}-\mathbb{E}_p[\theta]\right)\right]$$

We can then use subadditivity of the max operator to obtain:

$$\begin{aligned}\Gamma^N(g) &\leq \max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} u(g-\mathbb{E}_p[\theta]) + \max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}\left[u'\left(g-\omega(g,\tilde{\theta},\mathbb{E}_p[\theta])\right)\left(\tilde{\theta}-\mathbb{E}_p[\theta]\right)\right] \\ &= \tilde{\Gamma}^N(g) + \max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}\left[u'\left(g-\omega(g,\tilde{\theta},\mathbb{E}_p[\theta])\right)\left(\tilde{\theta}-\mathbb{E}_p[\theta]\right)\right]\end{aligned}$$

Now, fix a compact set C. Define $v(\tilde{\theta}) = \max_{g\in C, q\in[\underline{m},\overline{m}]} u'\left(g-\omega(g,\tilde{\theta},q)\right)$. Which is guaranteed to be well-defined by continuity of u' and ω. We then have:

$$0 \leq \Gamma^N(g) - \tilde{\Gamma}^N(g) \leq \max_{p\in\mathbb{P}^{\boldsymbol{a}}(a^N)} \mathbb{E}_p\left[v(\tilde{\theta})\left(\tilde{\theta}-\mathbb{E}_p[\theta]\right)\right]$$

Notice that neither bound depends on g within this compact set. On top of that, the upper bound converges to zero. To see that, recall that all signals are informative - $\underline{\rho} > 0$. That implies every $p^N \in \mathbb{P}^{\boldsymbol{a}}(a^N)$ have an almost-sure convergent subsequence to a degenerate distribution. Therefore, $\tilde{\theta} - \mathbb{E}_p[\theta] \to 0$ almost surely. That implies:

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

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

Recall that we can write $\tilde{\Gamma}^N(g) = \max\{u(g - \overline{m}^N), u(g - \underline{m}^N)\}$. Also by Theorem 1, $\overline{m}^N \xrightarrow{a.s.} \overline{m}$ and $\underline{m}^N \xrightarrow{a.s.} \underline{m}$.

We use the following lemma:

Lemma 5. *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:

$$\begin{aligned} h(x) &\leq (f^N(x) + \varepsilon)\mathbb{1}_{f(x) \geq g(x)} + (g^N(x) + \varepsilon)\mathbb{1}_{g(x) \geq f(x)} \\ &\leq h^N(x) + \varepsilon \end{aligned}$$

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

$$\begin{aligned} h^N(x) &\leq (f(x) + \varepsilon)\mathbb{1}_{f^N(x) \geq g^N(x)} + (g(x) + \varepsilon)\mathbb{1}_{g^N(x) \geq f^N(x)} \\ &\leq h(x) + \varepsilon \end{aligned}$$

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, 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. $\sup_{\mathbf{g}\in\mathbf{C}}|\boldsymbol{\Gamma}(\mathbf{g}) - \boldsymbol{\Gamma}^{\mathbf{N}}(\mathbf{g})| \xrightarrow{\mathbf{a.s.}} \mathbf{0}$**, 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** Γ.

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

Step 3. The sequence g^N **is uniformly bounded almost everywhere** .

For an observable realization a^N, recall that $\underline{m}^N = \min_{p\in\mathbb{P}^a(a^N)}\mathbb{E}[\theta]$ and, symmetrically, $\overline{m}^N = \max_{p\in\mathbb{P}^a(a^N)}\mathbb{E}[\theta]$. Assume, for a contradiction, that there is an event M with probability 1, 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:

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

Now, because $\overline{m}^N \xrightarrow{a.s.} \overline{m}$, we can choose an event $M' \subset M$, also with probability one, in which m^N. That implies, with the unboundedness of g^N and strict convexity of u, that the lower bound above diverges, so $\Gamma(g^N)$ is unbounded. To show that g^N cannot be optimal, it suffices to show that there is a sequence x^N such that $\Gamma(x^N)$ is bounded in this event. For any real a, take the sequence $x^N = a$ for all N. Because $\Gamma^N \xrightarrow{a.s.} \Gamma$ uniformly in any compact set, we have that, for a further event $M'' \subset M'$, with probability 1, that for any ε, for sufficiently large N,

$$\Gamma^N(a) < \Gamma(a) + \varepsilon$$

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

$$\Gamma^N(a) < \Gamma^N(g^N)$$

which is the contradiction that we were seeking.

□

Proof of Corollary 1

Define

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

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

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

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 $\overline{\zeta}$:[9] $\overline{m} = 2\theta - \underline{m}$, and we are done.

□

[9]See the Proof of Theorem 1

Proof of Proposition 1

$\overline{m}(\underline{m})$ **monotonically increases(decreasing) 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

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

Then, via the envelope theorem we have

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

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 last 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.

□

Proof of Proposition 2

We use k_η as define in the Proof of Proposition 1. Fix any $\eta > 1$ and $C > 0$ and consider the family of distributions over precisions $G_\rho = \delta_\rho$, $\rho > 0$. Given the state, the real signals are distributes according to $s \sim \mathcal{N}(\theta, \frac{1}{\rho})$. Let F_ρ be the distribution of s.

We use k_η as defined in the Proof of Proposition 1, but we call it k_ρ here, as we are interested in precisions instead. We can then use the formula for conditional means of normals to write:

$$k_\rho(a) = \theta + \frac{\frac{1}{\rho}(\eta - 1) f_\rho(a)}{F_\rho(a) + \eta(1 - F_\rho(a))}$$

For each a, as ρ converges to 0, the denominator converges to $\frac{1+\eta}{2}$, while the numerator converges to infinity. Because $\overline{m}_\rho = \max_a k_\rho(a)$, there is a sufficiently low ρ so that $\overline{m}_\rho$ is arbitrarily large. A similar argument can be made for $\underline{m}_\rho$. Now, by Example 2, we have that:

$$g^* = \frac{\underline{m}_\rho + \lambda \overline{m}_\rho}{1 + \sqrt{\lambda}}$$

Thus, $|g^* - \theta| = \frac{|1-\sqrt{\lambda}|(\overline{m}_\rho - \underline{m}_\rho)}{1+\sqrt{\lambda}}$ that can be made arbitrarily large for ρ sufficiently close to zero, which concludes our proof. Note that, although this result was proved using a degenerate G, it can easily be extended to continuous G's.

$\square$

Proof of Proposition 3

We go through the proof for $\overline{m}$, a symmetric argument holds for $\underline{m}$. From Theorem 2 we know that

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

Where $\boldsymbol{s^a}(x, \rho)$ is the inverted signal given action x and conjectured precision ρ, and $F(x)$ is the distribution of the observables. In the observable actions case the inverted signal is simply $\mathbf{s^a}(a_i, \hat{\rho}_i) =$

$a_i + \frac{\rho_\mu}{\hat{\rho}_i}(a_i - \mu)$, thus the above equation becomes

$$\overline{m}_a = \frac{\underline{\rho} \int_{-\infty}^{\overline{m}_a} \left(x + \frac{\rho_\mu}{\underline{\rho}}(x-\mu)\right) dF(x) + \overline{\rho} \int_{\overline{m}_a}^{\infty} \left(x + \frac{\rho_\mu}{\overline{\rho}}(x-\mu)\right) dF(x)}{\underline{\rho} F(\overline{m}_a) + \bar{\rho}(1 - F(\overline{m}_a))}$$
$$= \frac{\underline{\rho} \int_{-\infty}^{\overline{m}_a} x dF(x) + \overline{\rho} \int_{\overline{m}_a}^{\infty} x dF(x) + \int_{-\infty}^{\infty} \rho_\mu (x-\mu) dF(x)}{\underline{\rho} F(\overline{m}_a) + \bar{\rho}(1 - F(\overline{m}_a))}$$

Recall that actions are $\mathbf{a}(s_i, \rho_i) = \frac{\rho_\mu \mu + \rho_i s_i}{\rho_\mu + \rho_i}$, where ρ_i is the true not conjectured precision of the agent. Thus, given ρ_i the expected value of the observable is $\frac{\rho_\mu \mu + \rho_i \theta}{\rho_m u + \rho_i}$, since the signals normally distributed around θ. Recall from the setup that

$$F(x) = \int_{[\underline{\rho}, \overline{\rho}]} F_\rho\left(\boldsymbol{s}^{\boldsymbol{a}}(x, \rho)\right) dG(\rho)$$

Leading to

$$\int_{-\infty}^{\infty} \rho_\mu (x - \mu) dF(x) = (\theta - \mu) \int \frac{\rho \rho_\mu}{\rho_\mu + \rho} dG(\rho) = c$$

□

Proof of Proposition 4

The proof of this proposition relies on numerous algebraic calculations which we delegate to the online appendix.

References

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*.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., and Montrucchio, L. (2011). Uncertainty averse preferences. *Journal of Economic Theory*, 146(4):1275–1330.

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

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. (2019). Asymptotic behavior of bayesian learners with misspecified models. *Working Paper*.

Frick, M., Iijima, R., and Ishii, Y. (2020). Stability and robustness in misspecified learning models. *Working Paper*.

Fudenberg, D., Lanzani, G., and Strack, P. (2020). Limits points of endogenous misspecified learning. *Working Paper*.

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

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.

Heidhues, P., Koszegi, B., and Strack, P. (2019). Convergence in misspecified learning models with endogenous actions. *Working Paper*.

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

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

Klibanoff, P., Marinacci, M., and Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892.

Maccheroni, F., Marinacci, M., and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498.

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.

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