Information Aggregation under Ambiguity*

Pëllumb Reshidi † João Thereze ‡ Mu Zhang §

October 2020

Abstract

We study information aggregation when an observer is ambiguous about the precisions of her information sources. The observer estimates a payoff-relevant state by minimizing quadratic loss according to MaxMin Expected Utility, and updates her beliefs prior by prior, which induces ambiguity regarding the state. We show that this ambiguity does not vanish even if the number of information sources grows indefinitely, and characterize the asymptotic set of posteriors the observer entertains. When information sources are unbiased signals, the observer learns the state correctly. In contrast, when the observer has access only to other agent's guesses, her estimate converges away from the truth with probability one.

*We thank Pietro Ortoleva, Roland Benabou, Leeat Yariv, and Wolfgang Pesendorfer for very helpful suggestions and feedback. We also thank audiences at Princeton University for their helpful comments and discussions.

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

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

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

1 Introduction

1.1 Overview

In a wide range of economic problems, an observer relies on multiple information sources to learn a payoff relevant state. For example, a voter may count on news outlets and advertising to find out a politician's agenda and competence; a student may consult her friends and refer to course reviews to learn the quality of an optional course; an investor may use reports of different analysts and historical performances to forecast the future returns of a stock. The vast literature on information aggregation assumes that observers have beliefs about the qualities of all their information sources, and that these beliefs are correctly specified. In fact, by and large, individuals are assumed to have complete knowledge of these qualities. Under such assumptions, the observer learns the exact state given enough information.

However, these assumptions are often unrealistic. Violations of correct specification have been extensively studied in the literature of misspecified learning following Berk (1966). In contrast, to the best of our knowledge no papers analyze information aggregation when observers might fail to form a belief about the quality of their information sources. Consider a financial investor deciding whether to invest in a new company. To make an informed decision, she consults several of her colleagues, besides many sources in the target firm, and might resort to news outlets or even her social network for advice. While she might have an idea that some information sources might be more reliable than others, it is unlikely that she knows the quality of any of them with complete certainty. Similarly, forming beliefs about the precision of each source is likely impossible: some of them may be consulted for the very first time. Given these observations, how is information aggregated when the quality of information sources is unknown?

In this paper, we address precisely this question. We study how an agent aggregates information when her beliefs regarding the precision of the information sources are not represented by a single belief. Rather, the observer deems possible a set of precisions for each source and has no probabilistic assessment over them. Faced with this situation, she takes a robust approach and evaluates her outcome according to the worst-case scenario in this set. In other words, she minimizes the largest estimation error across all possible signal precisions. We implement this robustness approach by modeling the observer as an ambiguity-averse agent. Concretely, the observer is assumed to be a MaxMin Expected Utility maximizer (MEU, see Gilboa and Schmeidler (1989)), who minimizes the mean squared error of her estimate for the state according to the worst belief in her belief set.

The observer's belief set is given by the product space of precisions she considers for each information source. Explicitly, each allocation of precisions to information sources corresponds to a particular belief, that is, a joint probability over the state and signals given those precisions. For a given precision allocation, the state and signals are normally distributed. We assume that the observer updates her beliefs prior-by-prior. In other words, upon observing information from the information sources, she updates each distribution in her belief set according to Bayes rule. Because the observer lacks a probabilistic assessment over her priors, to her, the set of possible precisions amounts to unquantifiable uncertainty, that is, ambiguity. This contrasts with the quantifiable uncertainty represented by the belief about the state she aims to estimate, i.e. risk, which is also present under the standard Bayesian learning problem.

The observer's estimation problem can alternatively be thought of as a zero-sum game against Nature. The timing of the game is as follows. First, the observer receives information from her sources; afterward, given this information, she chooses the estimate that minimizes the mean squared error. Finally, Nature acts, choosing the precision of each source with the aim of maximizing the estimation errors. In practice, by manipulating precisions, Nature has two instruments: it can tamper with the variance of the agent's posterior, and it can affect the mean of the agent's posterior so as to increase the estimation bias.

In order to construct the building blocks for the information aggregation results, we first partially solve the estimation problem for a finite number of information sources. The solution is always characterized by the trade-off between the two instruments above. When signals realize near the prior mean, Nature cannot generate a large bias, so it resorts to maximizing the variance. In that case, the optimal guess for the observer coincides with the posterior mean obtained when all the signals are the least precise. When some signals realize far from the prior, however, Nature may want to maximize estimation error by increasing the bias. We show how the observer's guess changes accordingly.

We then let the number of information sources go to infinity to obtain the information aggregation results. Our first main result is that the observer cannot get rid of ambiguity even asymptotically. In other words, the agent still considers multiple possible posteriors on the state even after observing a very large number of signals. We characterize this set of posteriors as a function of the true distribution of precisions. Essentially, as the number of information sources grow, the observer becomes increasingly certain about a value for the state of the world under each of these posteriors - this value is the posterior mean. That is, just as in the standard Bayesian learning problem, the variance of each belief converges to zero. However, each belief concentrates in a different point, because information has different interpretations under different beliefs. It turns out that the set of posterior means is an interval that can be obtained by imposing correlations between the possible precisions and signal realizations. For example, by making precisions positively correlated with signal realizations, one can obtain a posterior that concentrates on a high expected mean.

The second set of aggregation results concerns the asymptotic quality of the observer's estimate. We prove that, despite the fact that ambiguity persists, the observer can consistently estimate the state when she observes unbiased signals directly. Because signals are normal, the distribution of signals is symmetric around the truth. This fact, on top of the symmetry of mean squared error, plays a major role in the consistency result. In fact, in the context of the zero-sum game, as the number of information sources grows, Nature cannot punish the observer with risk, as variance vanishes. The only available action for Nature is in ambiguity space: essentially, to choose within the interval of posterior means. Then, the best estimator of the agent must be the center of that interval that, under the normality assumptions, is the truth. We then consider a setting where signals are not directly observed by the observer. Rather, each signal is seen by a Bayesian agent, and the observer has access to the best estimate of the state for each of these agents. The agents and the observer share the same prior belief about the state. We prove that in this setting the observer generically fails to aggregate information correctly . Here, the information the observer receives is not unbiased anymore, as each Bayesian best estimate incorporates the prior mean. Because that breaks symmetry around the true state, consistency fails.

1.2 Related Literature

Our paper follows the literature on learning under ambiguity. Epstein and Schneider (2007) introduce a framework where the agent, who wants 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 and maximal likelihood updating. Epstein and Schneider (2008) study the 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 the ambiguity affects reactions to information and the asset price. Follow-up papers extend their results to incorporate ambiguity in mean of signals, equilibrium portfolio choices and 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 the problem whether ambiguity vanishes and the agent can guess the state correctly when the number of signals she observes goes to infinity.

Our paper is also related to the literature on single-agent misspecified learning, which is another possible driving force for failure of successful asymptotic learning. In those papers, a misspecified agent has a prior that assigns probability 0 to (a neighborhood of) the true model. Berk (1966) and Shalizi (2009) show that under mild conditions and exogenous information, the agent's beliefs converge, although not to the true model. Economic applications have been focusing on the extension where the signals can be affected by the actions of the agent and are hence endogenous. For example, Nyarko (1991) and Fudenberg et al. (2017) provide examples where the convergence of beliefs fails. Heidhues et al. (2019) consider the convergence of beliefs and actions with Gaussian prior and signals, which is similar to our setup. 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 those papers in three ways. First, the agent in the misspecified learning literature is a Bayesian learner, while in our setup, the observer holds multiple beliefs and adopts prior-by-prior updating rule. Second, the observer in our model is not misspecified in the sense that the true model is contained in her set of priors. Third, as a starting point, we assume that the observer is passively learning from exogenous information like Berk (1966) and Shalizi (2009). However, we show that her belief set diverges almost surely.

2 Setup

An observer aims to learn the state of the world $\theta \in \Theta := \mathbb{R}$, and has at her disposal N information sources. Denote the set of information sources as $I =: \{1, ..., N\}$. The prior distribution P_0 of the state θ is a normal distribution $\mathcal{N}\left(\mu, \frac{1}{\rho_\mu}\right)$, where the precision $\rho_\mu > 0$ is the inverse of the variance. The prior is common knowledge. Each information source $i \in I$ features a signal $s_i = \theta + \epsilon_i$, where the noise ϵ_i is normally distributed with mean 0 and precision $\rho_i > 0$, that is, $\epsilon_i \sim \mathcal{N}\left(0, \frac{1}{\rho_i}\right)$. For now, we assume that the prior and all signal errors are mutually independent. Denote the profile of precisions as $\rho^N := (\rho_1, ..., \rho_N)$, the profile of signals as $s^N := (s_1, ..., s_N)$, and for each $n \geq 1$ the set of distributions over $\mathbb{R}^n$ as $\Delta(\mathbb{R}^n)$.

We consider two cases of information sources. In the first case, the observer has direct access to the signals s_i from each information source. This corresponds to the situation where the observer has access to multiple signals. In the second case, there are agents in each information source, who utilize the signals to make certain decisions, and the observer can only observe the actions of these agents. This is similar to the problem faced by an econometrician, who can only observe the choices of different agents, but not their private information. We assume that there exists a bijective mapping from the set of signals to the set of actions. For example, in Section 4, we assume that agent i aims to match the state θ using the prior and the private signal s_i. For a Bayesian agent the action would be $a_i = \frac{\rho_\mu \mu + \rho_i s_i}{\rho_\mu + \rho_i}$. Denote the profile of actions as $a^N := (a_1, ..., a_N)$.

When the observer knows the precision ρ_i of each information source, the two cases are equivalent as she can figure out the signal s_i from the action a_i.[1] In this paper, we assume that the observer faces ambiguity with regards to the precision of the information sources, and does not have a specific belief about the distribution of these precisions. She only knows that the precision of each information source must lie in $[\underline{\rho}, \overline{\rho}]$ with $0 < \underline{\rho} < \overline{\rho}$ and $\underline{\rho} \leq \rho_i \leq \overline{\rho}$ for each i. One important feature of our setup is that the observer is not misspecified as she does not deem the actual precisions as impossible ex ante. This follows from the assumption that the true precision ρ_i lies in the ambiguous set $\underline{\rho} \leq \rho_i \leq \overline{\rho}$ for each source i. Denote $\hat{\rho}_i$ as the observer's perceived precision for information source i and $\hat{\rho}^N := (\hat{\rho}_1, ..., \hat{\rho}_N)$.

Following Epstein and Schneider (2007) and Epstein and Schneider (2008), we define $L^o(\hat{\rho}^N, \theta) \in$

[1] When the observer does not know the precisions and believes that they are i.i.d. according to a prior distribution, then the two cases differ when N is finite but they converge to the same limit as N goes to infinity.

$\mathbb{R}^n$ as the likelihood function for the observed variables, which is the conditional distribution for those variables given perceived precisions $\hat{\rho}^N$ and the realized state θ. In the case where signals are observable, i.e., $o = s$, $L^s(\hat{\rho}^N, \theta)$ is the likelihood function for the profile of signals s^N. In the case where actions are observable, i.e., $o = a$, $L^a(\hat{\rho}^N, \theta)$ is the likelihood function for the profile of actions a^N. Then the set of likelihood functions of the observer can be represented by $\mathcal{L}_N^o$, $o = s$ or a, where

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

We assume that the observer adopts prior-by-prior (full Bayesian) updating to derive posteriors using the prior P_0 and the set of likelihood functions $\mathcal{L}_N^o$. Given the realized profile of observed variables o^N and a likelihood function $L^o(\hat{\rho}^N, \theta)$, the posterior over the states $P_N(o^N, \hat{\rho}^N) \in \Delta(\mathbb{R})$ is obtained by applying the Bayes rule. Then the posteriors of the observer can be represented by the following set:

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

After observing either the profile of signals or the profile of actions, the observer makes a guess with regards to the state, under a quadratic loss function. Given multiple beliefs, she is assumed to be a maxmin expected utility (MEU) maximizer following Gilboa and Schmeidler (1989), and evaluates her guess based on the worst possible belief. This might be a result of the observer being ambiguity averse, or the observers intention to derive a robust upper bound for the expected loss. That is, the observer's objective is to minimize the maximal mean squared errors across all beliefs in the set of posteriors. She picks a guess g to solve the following problem:

$$\min_{g} \max_{p \in \mathbb{P}^o(o^N)} \left\{\mathbb{E}_p\left[(g - \theta)^2\right]\right\}, \quad o = s, a. \tag{1}$$

Due to the properties of a quadratic loss function, the above optimization problem can be simplified to one which only depends on the conditional mean and variance of the state. Denote them as $\mathbb{E}[\theta | o^N, \rho^N]$ and $\mathbb{V}[\theta | o^N, \rho^N]$ respectively. Then the objective of the observer becomes

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

3 Learning from Signals

3.1 One Signal Source

In this section we study the problem of an observer who has access to exactly one information source. The information source consists of a signal $s = \theta + \varepsilon$, with $\varepsilon \sim \mathcal{N}\left(0, \frac{1}{\rho}\right)$. Recall that the true ρ is unknown to the observer, who entertains a set of possible precisions $[\underline{\rho}, \overline{\rho}]$. After observing the single signal s, she chooses a guess g to solve:

$$\min_{g} \max_{\hat{\rho} \in [\underline{\rho}, \overline{\rho}]} \left\{ (g - \mathbb{E}[\theta|s, \hat{\rho}])^2 + \mathbb{V}[\theta|s, \rho] \right\} \tag{3}$$

where $\mathbb{E}[\theta|s, \hat{\rho}] = \frac{\hat{\rho}s + \rho_\mu \mu}{\hat{\rho} + \rho_\mu}$ and $\mathbb{V}[\theta|s, \hat{\rho}] = \left(1 - \frac{\hat{\rho}}{\hat{\rho} + \rho_\mu}\right)\frac{1}{\rho_\mu}$, can be derived by the joint normality of (θ, s), for a fixed precision $\hat{\rho}$.

This optimization problem can be interpreted as a zero-sum game between the observer and Nature. First, after the signal realization, the observer chooses an estimate for the state so as to minimize the mean-square error. Subsequently, with knowledge of the guess, Nature is free to choose any precision for the signal distribution within the uncertainty set. The objective of the observer is, then, to guarantee the smallest mean quadratic error conditional on the fact that Nature acts after her and to her detriment.

By controlling the precision of the signal, Nature affects both terms of the observer's utility: the squared bias and the variance. It affects the bias since a very precise signal implies an expected state that is close to the signal realization: that is, by making the signal more precise, Nature drives the expected value of the posterior closer to s and away from μ. The exact way by which Nature prefers to change the expected posterior depends on the guess of the agent, g. If g is close to the signal, then Nature benefits from choosing a low precision, as an expected state closer to μ increases bias.

Nature affects the variance because an imprecise information source makes for an imprecise estimator. In contrast to the bias, Nature's preference with respect to variance are independent on the choice of the observer. In particular, in the absence of the effect on bias, it would be always optimal for Nature to assign the smallest precision to the signal, so as to maximize the posterior variance. The actual choice of Nature, $\hat{\rho}^*(g)$ balances the incentives for increasing posterior bias with the incentives for increasing posterior variance. In its turn, the observer chooses her guess, taking into account how it affects Nature incentives.

A Bayesian observer who believes the precision of the signal to be $\hat{\rho}$, minimizes the mean square error by choosing the posterior mean as her optimal guess. The posterior mean consists of a convex combination between the signal realization and the prior mean, weighing the signal realization s by the relative precision $\frac{\hat{\rho}}{\hat{\rho}+\rho_\mu}$, and the prior mean with the complementary weight. Denoting the relative precision by z, the posterior mean can be written as $zs+(1-z)\mu$. The following proposition characterizes the optimal guess of the non-Bayesian observer in our model. Define the relative precisions $\bar{z} = \frac{\overline{\rho}}{\overline{\rho}+\rho_\mu}$ and $\underline{z} = \frac{\underline{\rho}}{\underline{\rho}+\rho_\mu}$.

Proposition 1. *When the observer has access to one information source, the optimal guess $g^* : \mathbb{R} \to \mathbb{R}$ satisfies:*

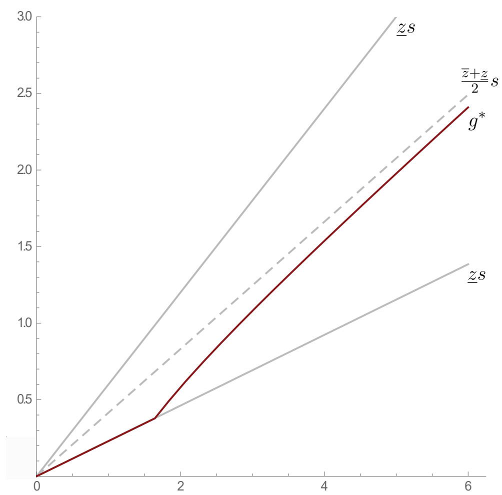
$$g^*(s) = \begin{cases} \underline{z}s + (1-\underline{z})\mu, & \text{if } (s-\mu)^2 \leq \frac{1}{\rho_\mu}\frac{1}{\bar{z}-\underline{z}} \\ \frac{\bar{z}+\underline{z}}{2}s + \left(1 - \frac{\bar{z}+\underline{z}}{2}\right)\mu - \frac{1}{2}\frac{1}{\rho_\mu}\frac{1}{(s-\mu)}, & o.w. \end{cases}$$

Proposition 1 shows that the optimal guess of the observer may fall into two categories. When the signal realization s is close to the prior mean, μ, the observes guesses as if she was a Bayesian believing the signal to have precision $\underline{\rho}$ - that is, relative precision $\underline{z}$. In contrast, when the signal is far from the prior mean, the guess can be divided in two parts - the guess of a Bayesian agent that believes the relative precision of the signal is the mean between $\underline{z}$ and $\bar{z}$, and an adjustment term. The optimality of this guess can be understood by again resorting to the interpretation of the game as a game against Nature. Recall that, by choosing the precision of the signal, Nature can tamper with the bias and the variance of the observer's posterior.

When the signal is near the prior mean Nature's ability to increase the bias is quite limited. Intuitively, Nature manipulates the weights that compose the posterior mean, which, regardless of the chosen weights, must lie between μ and s. As a consequence, the guess of the agent must also be located between these two values, implying that the square bias that Nature can generate cannot be larger than $(s-\mu)^2$. In contrast, Nature's ability to tamper with the posterior variance is unencumbered by the signal realization. Thus, when the potential to build squared bias is small, Nature prefers to default to maximal variance, regardless of the observer's guess - which is obtained by choosing the lowest possible precision for the signal, $\underline{\rho}$. In turn, the best the observer can do is to act as if the signal she observes has the lowest precision $\underline{\rho}$.

On the other hand, if the signal is far from the prior mean, Nature can greatly affect both the posterior mean and the posterior variance. In such cases, Nature's decision depends on the observer's guess. As an example, assume that the observer's guess is equal to the signal. In this case, Nature has a high incentive to choose the lowest signal precision: doing so increases the variance and the bias, as it moves the posterior mean close to the prior mean. On the other hand, if the observer guess is equal to the prior mean, Nature can choose a high precision, increasing the bias at the expense of a smaller posterior variance. It turns out that it is always optimal for Nature to attribute to the signal either the highest or the lowest precision, and when Nature's strategy is not dominant - when the signal is not close to the prior mean - it is optimal for the agent to make Nature indifferent between these two options. This is achieved by the guess in Proposition 1.

Figure 1: Optimal Guess with One Information Source



This figure compares the optimal guess for the observer with the posterior means for Bayesian agents with relative precisions $\underline{z}$, $\frac{\underline{z}+\overline{z}}{2}$, and $\overline{z}$, as the signal realization s varies. In this example the prior has mean $\mu = 0$ and precision $\rho_\mu = 1$. Finally, the uncertainty set is: $[\underline{\rho}, \overline{\rho}] = [.3, 1.5]$.

Figure 1 illustrates the optimal guess when the signal realization varies and compares it with the optimal guesses of a Bayesian agent with three different signal precisions. As noted before, when s is not too far from μ, the behavior of the observer is the same as that of a Bayesian agent that believes the signal to be minimally precise, which in the figure is represented by the line with the lowest slope. The dashed, intermediate line corresponds to the guess of a Bayesian agent who believes that the signal's relative precision is $\frac{\underline{z}+\overline{z}}{2}$.[2] For all realizations of the signal the observer's guess is always closer to the mean of the prior than the guess of the latter Bayesian agent; but converge to the latter as the signal realization converges to infinity. In this sense, $\frac{\bar{z}+\underline{z}}{2}$ is the tighter upper bound for the level of confidence the agent places on the signal, regardless of the signal realization. Finally, the line with the steepest slope represents the actions of a Bayesian agent that believes the precision of the signal to be $\overline{\rho}$.

Last, from Proposition 1, notice that the set of signal realizations in which the observer acts as an imprecise Bayesian shrinks as the precision of the prior ρ_μ grows. Intuitively, a more precise prior decreases the impact Nature can have in the posterior variance. Therefore, for a given signal realization, Nature now has a higher relative benefit from tampering with the bias.

The interpretation of a zero-sum game against Nature provides an intuitive lens to interpret the results in this section. Depending on the signal realization, either Nature has maximizing the posterior variance as a dominant strategy - and therefore chooses the lowest precision regardless of the observer's guess; or the observer guesses so as to make Nature indifferent between two precision assignments. As we show in the next subsections, this intuition continues to hold when the observer has access to $N > 1$ information sources.

3.2 N Signals

In this section we generalize the previous results by assuming that the observer has access to N signals. Each signal $s_i = \theta + \varepsilon_i$ is such that $\varepsilon_i \sim \mathcal{N}\left(0, \frac{1}{\rho_i}\right)$, where the observer believes precisions to be in the uncertainty set $[\underline{\rho}, \overline{\rho}]$. We denote by s^N the vector of the N observed signals. As in the last section, the problem of the observer is:

$$\min_{g} \max_{\hat{\rho}^N \in [\underline{\rho}, \overline{\rho}]^N} \{ \left(g - \mathbb{E}[\theta | s^N, \hat{\rho}^N] \right)^2 + \mathbb{V}[\theta | s^N, \rho^N] \} \tag{4}$$

Where, once more, by joint normality of (θ, s^N), $\mathbb{E}[\theta | s^N, \hat{\rho}^N] = \frac{\rho^N \cdot s^N + \rho_\mu \mu}{\rho^N \cdot \mathbb{1}^N + \rho_\mu}$ and $\mathbb{V}[\theta | s^N, \rho^N] = \left(1 - \frac{\rho^N \cdot \mathbb{1}^N}{\rho^N \cdot \mathbb{1}^N + \rho_\mu}\right) \frac{1}{\rho_\mu}$.

Under the game against Nature interpretation, just as in the previous section, Nature assigns

[2]This relative precision corresponds to a signal with precision $\left(\frac{1}{2}\left(\frac{1}{\overline{\rho}} + \frac{1}{\underline{\rho}}\right)\right)^{-1}$.

precisions to signals so as to tamper with the posterior mean and variance of the observer. Also as in the one signal case, it turns out that Nature never benefits from assigning an interior precision to any signal: in practice, $\hat{\rho}^N \in \{\underline{\rho}, \overline{\rho}\}^N$. This implies that for any s^N, Nature has a complex problem with 2^N assignments of precision to choose from. Definition 1 and Lemma 1 below show that attention can be restricted to a subset of precision assignments.

Definition 1. *An assignment $\rho^N : s^N \to [\underline{\rho}, \overline{\rho}]$ 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 monotone if it is order-preserving or order-reversing.*

Monotone precision assignments are simply monotone functions of signal realizations. Because Nature finds it optimal to always choose extreme values in the precision set for each signal, these assignments can be understood as threshold strategies: which amount to choosing a threshold such that signal values above (below) this threshold receive minimal (maximal) precision. From a practical perspective, they consist of a subset of Nature's strategies that is much smaller than its whole strategy set. In particular, for any number of observed signals, N, there are only $2N$ monotone assignments. The following Lemma implies that it is without loss of generality to exclusively consider threshold strategies.

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

In order to understand the previous result, recall $\hat{\rho}^N \in \{\underline{\rho}, \overline{\rho}\}$ and notice that, due to the normality assumption, the posterior variance depends on the precision assignment only through the number of signals with each precision. In other words, the value of each signal has no effect on the posterior variance, which is completely determined by the proportion of high precision versus low precision signals. As a consequence, if we restrict attention to strategies with a fixed proportion of precise signals, Nature's choice of which signal to attribute each precision affects exclusively the posterior mean.

As a consequence, for each fixed proportion of precisions, Nature assigns 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: as an example, if the observer's guess is rather low, Nature finds it optimal to maximize the posterior mean; to do so, she assigns high valued signals high precisions and low valued signals low precisions, thereby moving the expectation towards higher signal realizations. In other words, for a fixed proportion of precise signals, Nature's allocation of precisions to signals is always monotonic. Because this argument is true for all fixed proportions of precisions, it must be true for the optimal precisions assignment. Because they attain extreme posterior means, threshold strategies have an important role for this and all subsequent sections.

Proposition 2 is the main result of this section. It partially characterizes the observer's optimal guess and, consequently, Nature's strategy. Just as in Proposition 1, there are two possible scenarios: either Nature has a dominant strategy or she is made indifferent between two strategies. Nature has an optimal strategy when the signal realizations turn out to be close to the prior mean. In such cases, just as with one signal, Nature cannot benefit enough by creating bias, so she finds it optimal to maximize variance by assigning the lowest possible precision to all signals. With N signals the notion of proximity to the prior mean is defined by s^N being in an appropriate polytope around μ that is fully characterized below.

In contrast, when s^N is not close to the prior mean, Nature assigns precisions to adversely affect the posterior mean to the observer. As noted above, Nature does so by using threshold strategies that monotonically attribute precisions to signals. In parallel with the result for one signal, the observer reacts optimally by making Nature indifferent between two such strategies - Nature threatens the observer with a large bias, and the observer reacts by minimizing the bias given the threat.

In addition, Proposition 2 also shows that these two strategies, towards which Nature is made indifferent, have opposing monotonicity - one is order-preserving, the other order-reversing. Intuitively, Nature wants one of the strategies to generate a large expected mean and the other to generate a low expected mean so as to threaten the agent with a large bias. The best way to obtain these means is by assigning high precision to high signal realizations and low precision to low signal realizations, for the former, and vice versa for the later.

Again, the results are expressed in terms of relative precisions. For an assignment $\hat{\rho}^N$, z^N is the associated assignment of relative precisions - i.e. $z_i = \frac{\hat{\rho}_i}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}$. Notice that if $\hat{\rho}^N$ is order-preserving or order reversing, then the associated z^N keeps this property. Let Z^N be the set of relative precision vectors compatible with some $\rho^N \in [\underline{\rho}, \overline{\rho}]^N$. Finally, let $\underline{z}^N$ be the relative precision associated with $\underline{\rho}^N$, that is, the precision vector with $\underline{\rho}$ at each entry.

Proposition 2. *When the observer has access to N information sources, the optimal guess* $g^* : \mathbb{R}^N \to \mathbb{R}$ *satisfies:*

$$g^*(s^N) = \begin{cases} \underline{z}^N \cdot s^N + (1 - \underline{z}^N \cdot \mathbb{1}^N)\mu, & \text{if } s - \mu \in K^N \\ \frac{z^N + z^N{}'}{2} \cdot s + (1 - \frac{z'^N + z''^N}{2} \cdot \mathbb{1}^N) - \frac{1}{2}\frac{1}{\rho_\mu}\frac{(z'^N - z''^N)\cdot\mathbb{1}}{(z'^N - z''^N)\cdot(s-\mu)}, & o.w. \end{cases}$$

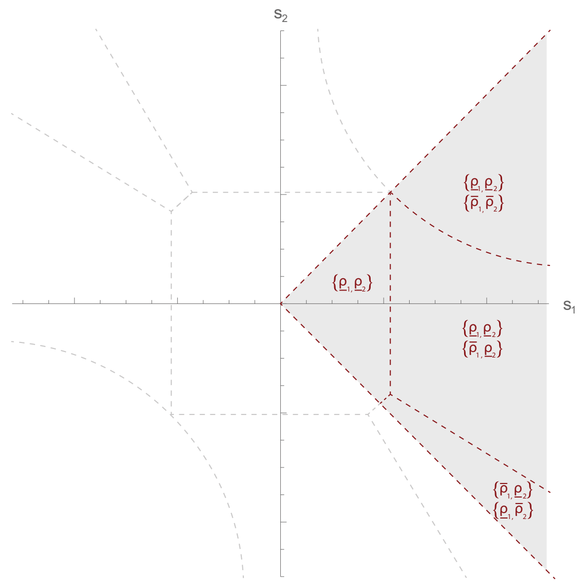
where $K^N = \{x \in \mathbb{R}^N : |\frac{z^N - \underline{z}^N}{\sqrt{(z^N - \underline{z}^N)\cdot\mathbb{1}}} \cdot x|\frac{1}{\rho_\mu} \leq 1 \ \forall z^N \in Z^N\}$ *and* z'^N, z''^N *are functions of* s^N *and are order-preserving and order-reversing, respectively.*

Proposition 2 is short of a full characterization in the sense that $z^N{}'$ and $z^N{}''$ are not explicitly derived. In general, these two assignments vary with the signal realization s^N in a way that is intuitive but difficult to formalize, due to the discreteness of the problem. Figure 2 illustrates Nature's strategies and regions in which they are optimal when $N = 2$. In two dimensions, signal realizations are points in the plane (s_1, s_2). Nature's equilibrium strategy divides the plane into regions, delimited by the dashed curves. We describe next the rationale for the equilibrium strategies within the shaded triangle, which can be extended to the rest of the plane by symmetry. Within each region, the values within the brackets show Nature's equilibrium precision assignments.

The prior mean, μ is centered on the origin. When both signal realizations are close to the origin, as argued above, Nature has maximizing posterior variance as a dominant strategy, so that $\hat{\rho}^* = (\underline{\rho}, \underline{\rho})$; she chooses minimal precision for both signals.

The lower-right region of the figure depicts Nature's strategy in cases when s_1 and s_2 both have high realizations with different signs, in particular $s_2 < 0 < s_1$. The fact that the signal realizations are far apart gives Nature an excellent opportunity to create a large bias. If the observer chooses a high guess, greatly following the positive valued signal, nature places a high precision on the negative valued signal and a low precision on the positive valued signal. If the observer chooses a low guess, greatly following the negative valued signal, nature does the opposite. In equilibrium the observer submits a guess that makes nature indifferent between the two extremes. Because she signals are far apart, the reduction in variance that follows by giving up the highest variance assignment is compensated by the increase in bias.

Figure 2: Nature's Equilibrium Strategy



This figure illustrates the equilibrium strategies of Nature when $N = 2$. The dashed curves divide the signal realization plane into regions in which Nature's strategy is the same. The equilibrium strategies for each region are in parentheses within the dark triangle and, by symmetry, can be extrapolated to the whole plane. In this figure the origin corresponds to the prior mean, while the prior precision is $\rho_\mu = 1$. Finally, the uncertainty set is: $[\underline{\rho}, \overline{\rho}] = [.3, 1.5]$.

The upper-right region represents the case in which both signals have relatively high and similar realizations. In this case, if the observer greatly follows the signals, nature chooses minimal precision for both signals. If, however, the observer guesses values close to the prior, to maximize the bias, nature chooses high precisions for both signals.

Finally, the middle-right section represents cases in which the value of the first signal is relatively far away from the prior mean, while the value of the second signal is relatively close to the mean. Similar to the previous cases, if the observer greatly follows the first signal, nature makes that signal minimally precise, and the signal with the realization close to the mean maximally precise. And vice versa if the observer's guess is close to the prior mean. In equilibrium, once more, the observer chooses a guess that makes nature indifferent.

3.3 $N \to \infty$ Signals

In this section we study the asymptotic properties of information aggregation by the observer. Formally, the agent observes a sequence of signals s^N. As N grows unbounded we study the behavior of the belief set $\mathbb{P}^s(s^N)$, and the guess $g^*(s^N)$. We assume that, in reality, all signals have the same precision, drawn from a distribution $G \in \Delta([\underline{\rho}, \overline{\rho}])$ (precision iid w.r.t. G). Recall that the likelihood function of signals is represented by $L^s(\rho, \theta)$. The distribution of a signal given the true distribution of precisions can be written as:

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

When the signal precisions are known, it is well understood that all quantifiable uncertainty (i.e. risk) vanishes asymptotically, and that the true realization of θ is revealed almost surely. That is, if the observer is a Bayesian who knows the precision of signals, her guess approaches the truth as the number of signals increases. Notice, however, that the robust observer is concerned with unquantifiable uncertainty, reflecting the fact that she considers the outcome under multiple priors. In order to assess risk and ambiguity, define the following notation. Given a conjecture $\hat{\rho}^N$ of precisions, the observer's posterior upon observing signals s^N is $P_N(s^N, \hat{\rho}^N)$. Consider s as the infinite sequence such that $s|_N = s^N$. We define the set of limiting beliefs:

$$\mathbb{P}_\infty(s) = \{P \in \Delta(\mathbb{R}) : P_N(s^N, \hat{\rho}^N) \xrightarrow{w} P,\ \hat{\rho}^N \in [\underline{\rho}, \overline{\rho}]^N \text{ and } N \in \mathbb{N}\}$$

That is, given the realization s, $\mathbb{P}_\infty(s)$ is the set of probability measures that are weak limits of posteriors. As N grows, the beliefs of the agent either do not converge or they converge to the ones in the limiting set of beliefs. Proposition 3 shows that the convergence of posterior beliefs cannot be taken for granted. In particular, for a sequence of signal realizations, there are typically many sequences of precisions that generate non-converging beliefs.

Proposition 3 (Non-Convergence)**.** *For almost all sequences of realized signals, s, there exists some infinite sequences $\hat{\rho} \in [\underline{\rho}, \overline{\rho}]^\infty$ such that $P_N(s^N, \hat{\rho}^N)$ does not converge.*

Specifically, for a generic sequence of signal realizations, we can construct a sequence of precisions that cause the sequence of posterior means to oscillate indefinitely. This is done by dividing the sequence of signals in batches, with the cardinality of each batch greatly larger than the previous ones. Because signals are i.i.d., by choosing large batches of signals, the distribution of signals within each selected group should resemble the real distribution. Then, by overweighting (underweighting) high signals in odd (even) batches, one can make the sequence of posterior means climb above the real mean within a batch, and decrease below it at the end of the next batch. Because their means oscillate, it must be that these posterior distribution sequences do not converge.

Despite the ubiquity of non-converging posterior sequences, we show, later in this section, that the posterior limits, $\mathbb{P}_\infty(s)$ are sufficient for characterizing the observer's asymptotic beliefs. In particular, we show that the following holds:

$$\lim_{N\to\infty} \arg\min_g \max_{p\in\mathbb{P}^s(s^N)} \mathbb{E}_p\left[(g-\theta)^2\right] = \arg\min_g \max_{p\in\mathbb{P}_\infty(s)} \mathbb{E}_p\left[(g-\theta)^2\right]$$

In words, as N grows, the observer's problem converges to the behavior of an agent that maximizes the worst-case scenario mean-squared error under all distributions in the limiting belief set $\mathbb{P}_\infty(s)$. In this sense, the structure of the latter set captures whether risk and ambiguity are still asymptotically relevant for the observer. Below, we specifically formalize what it means for risk and ambiguity to disappear asymptotically.

Definition 2. *For a sequence of signals s, we say that* ***Risk Vanishes*** *if $\mathbb{P}_\infty(s) \subset \{\delta_a : a \in \mathbb{R}\}$. We say that* ***Ambiguity Vanishes*** *if $\mathbb{P}_\infty(s)$ is a singleton.*

The first definition above states that no quantifiable risk is left in the belief set of the observer when she observes enough signals - i.e., under each posterior, she is certain of the state of the world, much like a Bayesian. The second definition deals with unquantifiable uncertainty. Ambiguity vanishes only if the set of limiting posteriors is a singleton. The intuition is that as long as there are more than one posteriors in the observer's set, she cannot get rid of the uncertainty concerning which of them is correct. Proposition 4 shows that, as a Bayesian, the observer faces no risk in the limit, but that, importantly, the same is not true for ambiguity. It also says that, although they may be non-convergent, all sequences of beliefs satisfy some limiting properties.

Proposition 4 (Ambiguity Does Not Vanish)**.** *Let $\underline{m} < \overline{m}$ be defined as:*

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

Then, for almost all sequences of realized signals s:

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

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

$$\lim_{N \to \infty} \mathbb{V}_{P_N(s^N, \hat{\rho}^N)}[\theta | s^N, \hat{\rho}^N] = 0$$

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

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

Proposition 4 has two parts. First, it bounds the oscillation of any sequence of posterior beliefs. Despite the fact that many of the sequences do not converge, it shows that their variances are asymptotically null, and that their expected values are bounded in an interval that is independent of s. These established bounds on posterior means are important, as they will allow us to eventually focus solely on the beliefs that do converge. However, the central result in Proposition 4 is the second part.

The second part of the proposition states that, given a conjecture over precisions, and provided that the sequence of posterior mean converges, the observer is certain about the true state of the world after observing a large number of signals — i.e. each posterior is degenerate on a state. However, the observer considers multiple conjectures. Each of these conjectures determines a state and the observer takes into account all of them when making a guess and, therefore, the observer remains ambiguous.

The second important remark of the proposition is that the boundaries of the posterior belief interval can be computed directly using the true distribution of signals, and are therefore independent of the specific signal realizations. This is a consequence of the independence of signals conditional on the precision. Under independence, through the Glivenko-Cantelli Theorem, we know that the empirical distribution of signals converges uniformly to the true distribution for almost all signal realizations. It turns out that, even after weighting the signals with arbitrary weights, the (empirical) posterior averages converges almost surely to the counterpart average in the true distribution — which follows from the strong law of large numbers.

As a consequence, the bounds of $\mathbb{P}_{\infty}(s)$ are the maximal and minimal posterior means that can be attained by using the true distribution of signals as the observables, and attributing precisions with values in $[\underline{\rho}, \overline{\rho}]$. As an example, to generate the largest posterior expectation, we start by ordering the signals by their realized values, on the real line. Then, given some threshold, we assign the highest permitted precision to the signals with realizations above the threshold, and the lowest permitted precision to the signals with realized values below the threshold. Doing so increases the posterior mean beyond the posterior mean in which each signal is assigned identical precision. By choosing the threshold optimally $\overline{m}$ is attained. Conversely, for $\underline{m}$, signals with realizations below some threshold are given maximal precision, while signals with realizations above the threshold are given minimal precisions. Then, by choosing this threshold optimally $\underline{m}$ is attained.

Proposition 5 below formalizes the previous argument that the asymptotic behavior of the observer is equivalent to the behavior she would have by considering the worst-case distribution in set $\mathbb{P}_\infty(s)$. It also shows that, perhaps surprisingly, the optimal guess converges, asymptotically, to the truth.

Proposition 5 (The Observer Guesses Correctly)**.** $g^*(s^N) \xrightarrow{a.s.} \frac{\overline{m}+\underline{m}}{2} = \theta$.

The intuition for Proposition 5 is, once again, better understood as a consequence of the game against Nature. As the number of signals increases, since each signal contains at least some information, the posterior variance gets arbitrarily small for every possible conjectured precision. Accordingly, Nature loses its ability to reduce the observer's utility through the variance channel. This happens since Nature's ability to threaten the observer with posterior variance is related to quantifiable risk, which was shown to vanish asymptotically. Therefore, Nature relies on the only tool she has left, and thus, maximizes squared bias. For all possible decisions of the observer, Nature's optimal action must be either the lower or the upper bound of the limiting belief interval. Given that, and as was the case for finite signals, the observer's optimal action is to make Nature indifferent, thereby guessing the median of the interval.

Notice that the intuitive argument above is made at the limit, as if the agent had observed infinite signals, whereas the result concerns convergence as N grows large. The proof of Proposition 5 consists of showing that as N goes to infinity, the observer's optimal guess becomes arbitrarily close to her optimal guess under the limiting belief set, $\mathbb{P}_\infty(s)$. Formally:

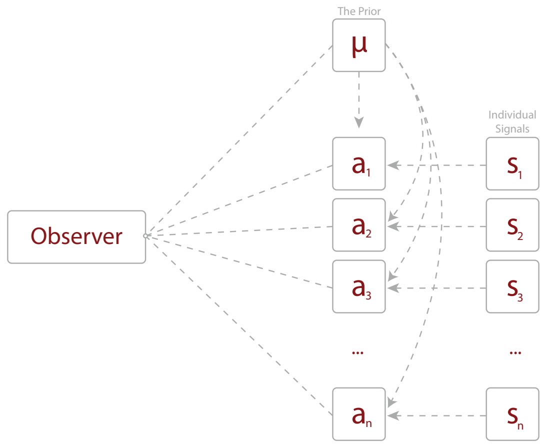
Technically, the effective change of limits above relies on an extension of the Glivenko-Cantelli theorem. A key implication of this result is that the non-converging beliefs are innocuous: we can characterize the limiting behavior of the optimal guess without addressing them. The boundedness of non-converging beliefs is essential for this to go through. 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.

It is worth noting that the consistency of the observer's guess is a knife-edge case, in the sense that it relies on the double symmetry provided by the Gaussian Family and the loss function. In particular, if the observer's loss function were not symmetric, consistency would fail almost everywhere. Asymmetric loss functions are often relevant in practice, for example, when economic agents have a preference for overestimating rather than underestimating a state.

4 Learning from Actions

We now consider a case in which the information sources of the observer consist of other agents' actions. That is, instead of directly receiving signals, the observer has access to actions of agents who in turn have observed private information. We assume that all agents and the observer share the same prior beliefs about the state θ. As in the previous section, $\theta \sim \mathcal{N}\left(\mu, \frac{1}{\rho_\mu}\right)$, according to the prior. 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. The agents know the precision of their private signals. We consider the case in which each agent attempts to estimate the realized value of θ in order to minimize the mean squared error. Given the prior and the private signal, the optimal Bayesian guess for agent i would then be $a_i = \mathbb{E}[\theta|\rho_i, s_i] = \frac{\rho_\mu \mu + \rho_i s_i}{\rho_\mu + \rho_i}$. It is exactly these actions that the observer has access to. The setup studied in this section is graphically depicted in Figure 3.

Figure 3: Learning From Actions Setup



While each agent exactly knows the precision of their private signal, this is not the case for the observer. We once more assume that the observer entertains a set of possible precisions $\left[\underline{\rho},\overline{\rho}\right]$. Each action is a convex combination of the private signal s_i and the mean of the prior μ. An observer who intends to guess the value of θ correctly, will first have to transform the actions back to signals. While the distribution of signals is centered around θ, this will not be the case for the distribution of actions. For a conjectured precision $\hat{\rho}_i$, the recovered signal will be $\hat{s_i}\left(a,\hat{\rho}_i\right) = a_i + \frac{\rho_\mu}{\hat{\rho}_i}(a_i - \mu)$. We assume that the observer solves

$$\min_g \max_{\hat{\rho}\in[\underline{\rho},\overline{\rho}]} \{\left(g - \mathbb{E}(\theta|s(a,\hat{\rho}),\hat{\rho})\right)^2 + \mathbb{V}(\theta|s,\hat{\rho})\} \tag{5}$$

From now on we assume that G is degenerate, that is, all signals have the same precision in reality: $\rho_i = \rho^*$. Recall that $P_N^a(a^N,\hat{\rho}^N)$ is the set of observer's posteriors given action sequence a^N and precision sequence $\hat{\rho}^N$. In parallel with last section, the asymptotic behavior of the agent's guess hinges on his set of limiting beliefs, defined by:

$$\mathbb{P}_\infty(a) = \{P \in \Delta(\mathbb{R}) : P_N^a(a^N,\hat{\rho}^N) \xrightarrow{w} P,\ \hat{\rho}^N \in [\underline{\rho},\overline{\rho}]^N \text{ and } N \in \mathbb{N}\}$$

Just as in last section, not all sequences of posterior beliefs converge. Rather, given a realization of actions, one can allocate precisions so as to make the sequence of posterior means oscillate between two fixed values. As the logic is exactly the same as in the previous section, we just state the result for completeness.

Proposition 6 (Non-Convergence)**.** *For almost all sequences of realized actions, a, there exists* $\hat{\rho} \in [\underline{\rho}, \overline{\rho}]^{\infty}$ *such that* $P_N^a(a^N, \hat{\rho}^N)$ *does not converge.*

We now study the properties of the set of limit posteriors the observer entertains. We denote the real distribution of actions by H. The next result parallels Proposition 4.

Proposition 7 (Ambiguity Does Not Vanish)**.** *Let* $\underline{m}_a < \overline{m}_a$ *be defined as:*

$$\overline{m}_a = \frac{\underline{\rho}\int_{-\infty}^{\overline{m}_a} x dH(x) + \bar{\rho}\int_{\overline{m}_a}^{\infty} x dH(x) + c}{\underline{\rho}H(\overline{m}_a) + \bar{\rho}(1 - H(\overline{m}_a))}, \quad \underline{m}_a = \frac{\overline{\rho}\int_{-\infty}^{\underline{m}_a} x dH(x) + \underline{\rho}\int_{\underline{m}_a}^{\infty} x dH(x) + c}{\overline{\rho}H(\underline{m}_a) + \underline{\rho}(1 - H(\underline{m}_a))}$$

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

Then, for almost all sequences of realized actions a:

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

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

$$\lim_{N\to\infty} \mathbb{V}_{P_N^a(a^N, \hat{\rho}^N)}[\theta | a^N, \hat{\rho}^N] = 0$$

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

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

Proposition 7 shows that as was the case in Section 3, as the number of actions available to the observer increases to infinity, her set of posterior beliefs does not shrink to a singleton. Rather, that limiting set of degenerate distributions, the support of which form an interval. The intuition of why this happens is similar to the discussion in Section 3: different allocations of precisions to a sequence of realizations generates different posterior means. The upper and lower bounds differ from the bounds found when the observer learns from signals.

The difference between the bounds in Proposition 7 and the ones in Proposition 4 stems from c. Under the true distribution of signals, given by precision ρ^*, one can rewrite an observed action exactly as:

$$s_i = a_i + \frac{\rho_\mu}{\rho^* + \rho_\mu}(s_i - \mu)$$

The term $\frac{\rho_\mu}{\rho^*+\rho_\mu}(s_i - \mu)$ is, then, an adjustment in the transformation of a_i to s_i given the real precision, that takes into account the difference between the signal realization and the prior mean. Notice that c is proportional to the expectation of this adjustment. It dictates whether actions have to be adjusted upwards (downwards) when trying to estimate θ, depending exactly on whether the common prior is driving all the actions down (up). This term does not depend on the consideration set of the observer, but only on the real distribution. The relationship between the prior mean and the state realization plays a key role in our subsequent analysis, starting from the next result.

Proposition 8 (The Observer Guesses Incorrectly). *The agent's optimal guess* $g(a^N) \xrightarrow{a.s.} (\overline{m}_a + \underline{m}_a)/2 := g^*$.

Proposition 8 reveals that the observer once more finds it optimal to choose the median of the interval as her optimal guess. However now, for almost all realizations of signals, this guess will no longer be equal to the state θ. This is the main difference in the optimal behavior of the observer learning from actions compared to the case in which she directly has access to signals. When faced with the task of recovering signals from actions, although we continue to assume that the distribution of signals is symmetric, the final guess is no longer correct. We proceed by characterizing exactly how the guess of the observer differs from the optimal decision. As we soon make formal, if $\theta = \mu$ then $g^* = \theta = \mu$. Since the correction term c moves in the same direction as θ, both the upper and lower bounds move in the direction of θ. Hence, if $\theta > \mu$, then $g^* > \mu$, and vice versa. We say that g^* *overreacts* if $|g^* - \mu| > |\theta - \mu|$ and *underreacts* it if the inequality is reversed.

Proposition 9 (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 9 reveals that whether the observer over or underreacts depends on the true precision of the signals and possibly on the realization of the state θ. We first discuss the intuition behind the cases in which the observer over or underreacts regardless of the state. For any distribution of actions, the observer considers all possible combinations of signal precisions and identifies the ones that would lead to the highest and lowest possible guesses. Having done so, given her optimization problem, or alternatively, given the fact that nature will choose the worse outcome for the observer, she finds it optimal to choose the median value. Since the observer considers all possible combinations of precisions when forming the upper and lower bound, the optimal choice given the true precision must also lie within these bounds.

Roughly speaking, the optimal robust guess corresponds to the observer trying to backtrack the mean of the unobservable signals from the mean of observed actions. Because signals are unbiased, their mean is effectively θ, the state the observer desires to learn. When ρ^* is high, θ is relatively close to the mean of actions, as the agents put a high weight on their unbiased signals when choosing their actions. However, the robust agent does not know the real precision, so she backtracks signals from actions using, roughly, the same method regardless of what ρ^* is. If ρ^* is sufficiently high, it is as if she was assuming the signals were less precise than ρ^*. In order to justify this precision - that is lower than the true one - she has to conjecture that the state is farther away from the mean of actions than it really is. In effect, that implies that she overestimates the mean of signals and, hence, overreacts to the actions. The flipside of this argument explains underreaction when ρ^* is small enough.

When the true precision ρ^* takes relatively intermediate values, the comparison is no longer as clear cut. Rather, whether the observer under or overreacts depends on the realization of the state θ. When the realized state is relatively close to the prior, the observer undereacts, whereas when the realized state is relatively far away from the prior, the observer ends up overreacting once more. Thus, in this case the error of the observer is not monotonic in the realized state θ. As θ is further away from the mean of the prior μ, the mistake of the observer increases, afterwards it starts decreasing, and once more, having entered the overeaction interval, the observer's error increases for any further increase in the difference between θ and the mean of the prior μ. In the previous two cases, in which the observer always under or overreacts, her error always increases as the difference between θ and μ increases.

5 Extensions

5.1 Tying Nature's Hands

In the previous sections, we assume that the observer only knows the range of the precision of each information source. In practice, however, the observer might have more information ex-ante. For instance, although she might not observe the identity of the sources, she might know that a fraction of information sources are better informed, and hence their signals are more precise than the others'. Recall that the optimization problem of the observer can be interpreted as a game between her and Nature, who chooses the perceived precision of each information source after the observer's guess, in an attempt to maximize the observers expected losses. Hence, the additional information that the observer has will tie Nature's hands by serving as a restriction on its choices of precisions. In this section, we focus on the limiting case as the number of information sources goes to infinity, with the additional assumptions on what the observer knows about the precisions of information sources. Specifically, we assume that there are two groups of information sources. Group 1 consists of $\alpha \in [0,1]$ fraction of information sources with shared precision $\hat{\rho}_1$ and Group 2 consists of $1-\alpha$ fraction with shared precision $\hat{\rho}_2$. We will consider the following cases with observable signals and observable actions.

5.1.1 Case 1: $\hat{\rho}_1 = \overline{\rho}$ and $\hat{\rho}_2 = \underline{\rho}$

Suppose that a fraction α of the information sources have precision $\overline{\rho}$ and the rest have precision $\underline{\rho}$. Consider the limiting case as N goes to infinity, as in Section 3, the Nature's choice can be represented by a mapping $\hat{\rho}$ from the signal space $\mathbb{R}$ to the set of possible precisions $[\underline{\rho}, \overline{\rho}]$. In

Section 3, any such mapping is feasible for Nature. Now, however, the observer's prior knowledge places certain restrictions on the distribution of perceived precision that Nature can choose. Recall that F is the limiting empirical distribution of signals with density function f. Denote Φ_1 as the set of feasible mappings of perceived precisions.

$$\Phi_1 = \{\hat{\rho} : \hat{\rho}(x) \in \{\underline{\rho}, \overline{\rho}\}, \int \hat{\rho}(x) f(x) dx = \alpha \overline{\rho} + (1-\alpha)\underline{\rho}\}$$

We start with observable signals. For any feasible precision mapping $\hat{\rho} \in \Phi_1$, the posterior mean of the state is

$$v(\hat{\rho}) = \int \frac{\hat{\rho}(x)}{\int \hat{\rho}(y) f(y) dy} x f(x) dx. \tag{6}$$

Based on similar arguments as in Section 3, Nature's problem is to maximize or minimize the posterior mean over the set of precision mappings Φ_1.

Recall that for fixed value of $\int \hat{\rho}(x) f(x) dx$, the optimal solution for Nature's problem takes a threshold form. Denote $\overline{s}_\alpha$ and $\underline{s}_\alpha$ such that $F(\overline{s}_\alpha) = 1-\alpha, F(\underline{s}_\alpha) = \alpha$. Then the maximal posterior mean is achieved when the $\hat{\rho}(x) = \overline{\rho}$ for $x \geq \overline{s}$ and $\hat{\rho}(x) = \underline{\rho}$ for $x < \overline{s}$, that is, when signals with high realizations are assigned higher precisions, while signals with relatively low realizations are assigned low precisions; and vice versa for the minimal posterior mean. Notice that in the limiting case, every possible posterior belief of the observer regarding the state is a degenerate distribution at some feasible posterior mean. Then the upper and lower bounds of the set of degenerate limit posteriors are

$$\overline{m}^s_{\alpha,1} = \frac{\underline{\rho} \int_{-\infty}^{\overline{s}_\alpha} x dF(x) + \bar{\rho} \int_{\overline{s}_\alpha}^{\infty} x dF(x)}{\alpha \overline{\rho} + (1-\alpha)\underline{\rho}}, \quad \underline{m}^s_{\alpha,1} = \frac{\overline{\rho} \int_{-\infty}^{\underline{s}_\alpha} x dF(x) + \underline{\rho} \int_{\underline{s}_\alpha}^{\infty} x dF(x)}{\alpha \overline{\rho} + (1-\alpha)\underline{\rho}}$$

As a result, the optimal guess of the observer is the median of the two bounds $g^s_{\alpha,1} = (\underline{m}^s_{\alpha,1} + \underline{m}^s_{\alpha,1})/2$.

As the limiting empirical distribution F is symmetric around the true state θ, $g^s_{\alpha,1} \equiv \theta$ and the observer will still guess correctly just as the benchmark case in Section 3 where the observer has no information about the precision other than the range. However, the set of degenerate limit posteriors, i.e., the amount of ambiguity, does shrink with additional prior knowledge of the observer.

Now we consider learning from actions. Notice that in Section 4, we assume that the actual precision is ρ for all information sources. However, in the current setup, this is not consistent with the observer's prior about precisions generically and hence the observer will be misspecified. To avoid misspecification, we assume that, in this case, the precisions of information sources are independently and identically distributed with probability α being $\overline{\rho}$ and probability $1-\alpha$ being $\underline{\rho}$. Then the limiting empirical distribution of actions H can be represented by a mixture of two normal distributions:

$$H_1(x) \sim \mathcal{N}\left(\frac{\rho_\mu \mu + \overline{\rho}\theta}{\rho_\mu + \overline{\rho}}, \frac{\overline{\rho}}{(\rho_\mu + \overline{\rho})^2}\right), \quad H_2(x) \sim \mathcal{N}\left(\frac{\rho_\mu \mu + \underline{\rho}\theta}{\rho_\mu + \underline{\rho}}, \frac{\underline{\rho}}{(\rho_\mu + \underline{\rho})^2}\right)$$

Denote h as the density function for H. The mapping from signals to actions remains unchanged given the perceived precisions and the posterior mean is given by

$$v'(\hat{\rho}) = \int \frac{\hat{\rho}(x) + c'}{\int \hat{\rho}(y) f(y) dx} f(x) x dx. \tag{7}$$

where $c' = \alpha \frac{\overline{\rho}\rho_\mu}{\rho_\mu + \overline{\rho}}(\theta - \mu) + (1-\alpha)\frac{\underline{\rho}\rho_\mu}{\rho_\mu + \underline{\rho}}(\theta - \mu)$.

Again, for fixed value of $\int \hat{\rho}(x) h(x) dx$, the optimal solution for Nature's problem takes a threshold form. Denote $\overline{a}_\alpha$ and $\underline{a}_\alpha$ such that $H(\overline{a}_\alpha) = 1-\alpha, H(\underline{a}_\alpha) = \alpha$. Then the bounds of the set of posterior means are

$$\overline{m}^a_{\alpha,1} = \frac{\underline{\rho}\int_{-\infty}^{\overline{a}_\alpha} x dH(x) + \bar{\rho}\int_{\overline{a}_\alpha}^{\infty} x dH(x) + c'}{\alpha\overline{\rho} + (1-\alpha)\underline{\rho}}, \quad \underline{m}^a_{\alpha,1} = \frac{\overline{\rho}\int_{-\infty}^{\underline{a}_\alpha} x dH(x) + \underline{\rho}\int_{\underline{a}_\alpha}^{\infty} x dH(x) + c'}{\alpha\overline{\rho} + (1-\alpha)\underline{\rho}}$$

and the optimal guess is

$$\begin{aligned}
g^a_{\alpha,1} := \frac{1}{2}(\overline{m}^a_{\alpha,1} + \underline{m}^a_{\alpha,1}) &= \frac{c'}{\alpha\overline{\rho} + (1-\alpha)\underline{\rho}} + \alpha\frac{\rho_\mu\mu + \overline{\rho}\theta}{\rho_\mu + \overline{\rho}} + (1-\alpha)\frac{\rho_\mu\mu + \underline{\rho}\theta}{\rho_\mu + \underline{\rho}} \\
&= \theta + \frac{\alpha(1-\alpha)(\overline{\rho} - \underline{\rho})\rho_\mu}{\alpha\overline{\rho} + (1-\alpha)\underline{\rho}}\Big(\frac{1}{\rho_\mu + \overline{\rho}} - \frac{1}{\rho_\mu + \underline{\rho}}\Big)(\theta - \mu) \\
&= \theta - \frac{\alpha(1-\alpha)(\overline{\rho} - \underline{\rho})^2\rho_\mu}{[\alpha\overline{\rho} + (1-\alpha)\underline{\rho}](\rho_\mu + \overline{\rho})(\rho_\mu + \underline{\rho})}(\theta - \mu)
\end{aligned}$$

Hence, when there is ex-ante ambiguity ($\alpha \neq 0, 1$ and $\overline{\rho} > \underline{\rho}$), the observer still guess incorrectly whenever $\theta \neq \mu$. Moreover, unlike the benchmark case, the observer will always underreact to the observed actions, that is, the optimal guess always lies between the realized state and prior mean of the state.

An immediate observation is that a larger fraction of information sources with the highest precision does not necessarily make the observer better off. Note that when $\theta = 0$ or 1, the observer will always guess correctly and obtains the maximal payoff. When α is close to 0, the guessing error is increasing as α increases and hence the observer is worse off even though she is faced with a more precise pool of information sources.

Intuitively, when $\alpha = 1$ or 0, there is no ambiguity and information will always be aggregated efficiently as the number of information sources grows to infinity, regardless of the precision level. For finitely many information sources, it is true that higher precision helps as it can reduce the ex-post variance of the observer's belief. However, the ex-post variance vanishes as the number of information sources grows, and the observer's faces only ambiguity at the limit.

5.1.2 Case 2: $\hat{\rho}_1, \hat{\rho}_2 \in [\underline{\rho}, \overline{\rho}]$ and $\hat{\rho}_1 \geq \hat{\rho}_2$

We assume that the observer believes there are two groups of information sources, with one group endowed with weakly higher precision. Again, we are focusing on the limiting case with infinitely many information sources. Denote Φ_2 as the set of feasible mappings of perceived precisions.

$$\Phi_2 = \bigcup_{\substack{\hat{\rho}_1 \geq \hat{\rho}_2; \\ \hat{\rho}_1, \hat{\rho}_2 \in [\underline{\rho}, \overline{\rho}]}} \{\hat{\rho} : \hat{\rho}(x) \in \{\hat{\rho}_1, \hat{\rho}_2\}, \int \hat{\rho}(x) f(x) dx = \alpha \hat{\rho}_1 + (1-\alpha)\hat{\rho}_2\}$$

We start with observable signals and find extreme values for the posterior mean of the state. For any fixed value of $\hat{\rho}_1 \geq \hat{\rho}_2$, the cutoff structure is preserved, with the same cutoffs as in Case 1. Then the bounds for the set of degenerate posteriors by fixing $\hat{\rho}_1$ and $\hat{\rho}_2$ are

$$\overline{m}^s_{\alpha,2}(\hat{\rho}_1, \hat{\rho}_2) = \theta + \frac{\hat{\rho}_2 \int_{-\infty}^{\overline{s}_\alpha} (x-\theta) dF(x) + \hat{\rho}_1 \int_{\overline{s}_\alpha}^{\infty} (x-\theta) dF(x)}{\alpha \hat{\rho}_1 + (1-\alpha)\hat{\rho}_2} \geq \theta,$$

$$\underline{m}^s_{\alpha,2}(\hat{\rho}_1, \hat{\rho}_2) = \theta + \frac{\hat{\rho}_1 \int_{-\infty}^{\underline{s}_\alpha} (x-\theta) dF(x) + \hat{\rho}_2 \int_{\underline{s}_\alpha}^{\infty} (x-\theta) dF(x)}{\alpha \hat{\rho}_1 + (1-\alpha)\hat{\rho}_2} \leq \theta.$$

We need to range over different values of $\hat{\rho}_1 \geq \hat{\rho}_2$ to find the global upper bound and lower bound. It is easy to see that $\overline{m}^s_{\alpha,2}(\hat{\rho}_1, \hat{\rho}_2)$ is increasing in $\hat{\rho}_1$ and decreasing in $\hat{\rho}_2$, while $\underline{m}^s_{\alpha,2}(\hat{\rho}_1, \hat{\rho}_2)$ is increasing in $\hat{\rho}_2$ and decreasing in $\hat{\rho}_1$. Hence, the universal bounds of the set of degenerate posteriors are still given by

$$\overline{m}^s_{\alpha,2}(\overline{\rho},\underline{\rho}) = \theta + \frac{\underline{\rho}\int_{-\infty}^{\overline{s}_\alpha}(x-\theta)dF(x) + \overline{\rho}\int_{\overline{s}_\alpha}^{\infty}(x-\theta)dF(x)}{\alpha\overline{\rho}+(1-\alpha)\underline{\rho}},$$

$$\underline{m}^s_{\alpha,2}(\overline{\rho},\underline{\rho}) = \theta + \frac{\overline{\rho}\int_{-\infty}^{\underline{s}_\alpha}(x-\theta)dF(x) + \underline{\rho}\int_{\underline{s}_\alpha}^{\infty}(x-\theta)dF(x)}{\alpha\overline{\rho}+(1-\alpha)\underline{\rho}}.$$

and the observer still guesses correctly.

Then we consider the case of observable actions. We assume that all information sources share the same precision ρ, which is consistent with the prior knowledge of the observer. Recall that H is the limiting empirical distribution of actions, $H(\overline{a}_\alpha) = 1-\alpha, H(\underline{a}_\alpha) = \alpha$ and $c = \frac{\rho\rho_\mu}{\rho_\mu+\rho}(\theta-\mu)$, the bounds for the set of degenerate posteriors by fixing $\hat{\rho}_1$ and $\hat{\rho}_2$ are

$$\overline{m}^a_{\alpha,2}(\hat{\rho}_1,\hat{\rho}_2) = \theta + \frac{\hat{\rho}_2\int_{-\infty}^{\overline{a}_\alpha}(x-\theta)dH(x) + \hat{\rho}_1\int_{\overline{a}_\alpha}^{\infty}(x-\theta)dH(x)+c}{\alpha\hat{\rho}_1+(1-\alpha)\hat{\rho}_2}$$

$$\underline{m}^a_{\alpha,2}(\hat{\rho}_1,\hat{\rho}_2) = \theta + \frac{\hat{\rho}_1\int_{-\infty}^{\underline{a}_\alpha}(x-\theta)dH(x) + \hat{\rho}_2\int_{\underline{a}_\alpha}^{\infty}(x-\theta)dH(x)+c}{\alpha\hat{\rho}_1+(1-\alpha)\hat{\rho}_2}$$

The monotonicity of the two bounds depends on the sign of c. For instance, if $c \geq 0$, that is, the realized state θ is higher than the prior mean μ, then $\overline{m}^a_{\alpha,2}(\hat{\rho}_1,\hat{\rho}_2)$ is decreasing in $\hat{\rho}_2$ and can be either increasing or decreasing in $\hat{\rho}_1$ depending on the value of c. Similar properties hold for the lower bound and the case with $c<0$. As a summary, both bounds are increasingly or decreasingly monotonic in $\hat{\rho}_1$ and $\hat{\rho}_2$, and the global bounds across all feasible pairs of $\hat{\rho}_1$ and $\hat{\rho}_2$ are given by

$$\begin{aligned}\max_{\substack{\hat{\rho}_1\geq\hat{\rho}_2;\\ \hat{\rho}_1,\hat{\rho}_2\in[\underline{\rho},\overline{\rho}]}} \overline{m}^a_{\alpha,2} = \max\Big\{&\theta + (\frac{1}{\overline{\rho}}-\frac{1}{\rho})c,\\ &\theta + (\frac{1}{\underline{\rho}}-\frac{1}{\rho})c,\\ &\theta + \frac{\underline{\rho}\int_{-\infty}^{\overline{a}_\alpha}(x-\theta)dH(x) + \overline{\rho}\int_{\overline{a}_\alpha}^{\infty}(x-\theta)dH(x)+c}{\alpha\overline{\rho}+(1-\alpha)\underline{\rho}}\Big\},\end{aligned}$$

$$\min_{\substack{\hat{\rho}_1 \geq \hat{\rho}_2; \\ \hat{\rho}_1, \hat{\rho}_2 \in [\underline{\rho}, \overline{\rho}]}} \underline{m}^a_{\alpha,2} = \min\Big\{\theta + (\frac{1}{\overline{\rho}} - \frac{1}{\rho})c,$$

$$\theta + (\frac{1}{\underline{\rho}} - \frac{1}{\rho})c,$$

$$\theta + \frac{\overline{\rho}\int_{-\infty}^{\underline{a}_\alpha}(x-\theta)dH(x) + \underline{\rho}\int_{\underline{a}_\alpha}^{\infty}(x-\theta)dH(x) + c}{\alpha\overline{\rho} + (1-\alpha)\underline{\rho}}\Big\}.$$

After a little algebra, it is possible to prove that all the combinations can generate $\frac{\underline{m}^a_{\alpha,2} + \overline{m}^a_{\alpha,2}}{2} = \theta$ only in knife-edge cases. Thus, the agents guesses incorrectly generically, just as in the main model.

6 Conclusions

In this paper, we study the information aggregation problem of an observer who wants to learn a state of interest, while being ambiguous about the precision of her information sources. We assume that the prior and all signals are normally distributed, and that the observer adopts max-min expected utility. The optimization problem can be interpreted as a zero-sum game between the observer and Nature. In the case where the signals are directly observable, we partially characterize the observer's optimal guess with N information sources. We highlight Nature's trade-offs between the posterior bias and the posterior variance, channels utilized to minimize the observer's payoff. As the number of signals goes to infinity, we show that there exist posterior beliefs that diverge. However, with regards to the observer's optimal guess these divergent sequences are innocuous, allowing us to restrict our attention to sequences that converge. In turn, the guesses of the observer converge to the true state.

We also consider a situation in which signals cannot be observed directly. Instead, the observer has access to the optimal actions of agents who rely on their prior and private independent signals. Unlike in the Bayesian case, an ambiguity averse observer fails to perfectly invert signals from actions. We show that once more there exist posterior beliefs that diverge, while once more the guess of the observer converges. However, under this specification, the point to which the guess of the observer converges is generically different from the true state. Specifically, given the assumption that all information sources share the same true precision, we show that the observer will overreact to the observed actions if the true precision level is relatively high and underreact if the true precision level is relatively low. For intermediate values of the true precision, whether the observer overreacts or underreacts depends on how much the realized value of the true state differs from the prior mean. Our results suggest that under ambiguity, learning can fail even though the number of independent information sources goes to infinity.

For future work, one can think about extending our assumption of max-min expected utility to general ambiguity-averse preferences akin to the ones analyzed in Cerreia-Vioglio et al. (2011). We expect similar qualitative results to hold; that is, although there will be non-convergent posterior beliefs, the observer's optimal guess at the limit will depend on the set of convergent beliefs. One might also divert from prior-by-prior updating. There is an ongoing debate with regards to the most reasonable updating rules under ambiguity.[3] Our conjecture is that, as long as the updating rule itself does not exclude ambiguity automatically after updating (say, like the maximal likelihood updating rule), then similar results should be found. Other extensions can involve studying a multi-agent learning problem, such as social learning, which is now one of the current central topics in misspecified learning.

7 Appendix

7.1 Proofs

Proof of Proposition 3. Since the prior as well as all signals are normally distributed, for any $\hat{\rho}^N$, $P_n(s^N, \hat{\rho}^N)$ is also normally distributed with mean

$$\mathbb{E}[\theta|s^N, \hat{\rho}^N] = \frac{\sum_i \hat{\rho}_i s_i + \rho_\mu \mu}{\sum_i \hat{\rho}_i + \rho_\mu}$$

For the first $\tilde{N}_1$ signals, consider the following sequence

$$\hat{\rho}_i = \begin{cases} \overline{\rho} & if \quad s_i \in S_1 \\ \underline{\rho} & if \quad s_i \in S_1' \end{cases} \qquad S_1 = \{s_i \leq \theta : i \leq \tilde{N}_1\} \qquad S_1' = \{s_i >: i \leq \tilde{N}_1\}$$

[3] See Shishkin and Ortoleva (2019) for a detailed discussion.

Where θ is the mean, as well as the median of the distribution of signals. After the first $\tilde{N}_1$ signals, the mean of the posterior will then be

$$\mathbb{E}[\theta|s^{\tilde{N}_1}, \hat{\rho}^{\tilde{N}_1}] = \frac{\overline{\rho}\sum_{i\in S_1} s_i + \underline{\rho}\sum_{i\in S_1'} s_i + \rho_\mu \mu}{|S_1|\overline{\rho} + |S_1'|\underline{\rho} + \rho_\mu}$$

Where $|S_1|$ and $|S_1'|$ represent the cardinality of the set S_1 and S_1' respectively. We can rewrite the above as

$$\mathbb{E}[\theta|s^{\tilde{N}_1}, \hat{\rho}^{\tilde{N}_1}] = \frac{\overline{\rho}\frac{|S_1|}{\tilde{N}_1}\frac{\sum_{i\in S_1} s_i}{|S_1|} + \underline{\rho}\frac{|S_1'|}{\tilde{N}_1}\frac{\sum_{i\in S_1'} s_i}{|S_1'|} + \frac{\rho_\mu \mu}{\tilde{N}_1}}{\frac{|S_1|}{\tilde{N}_1}\overline{\rho} + \frac{|S_1'|}{\tilde{N}_1}\underline{\rho} + \frac{\rho_\mu}{\tilde{N}_1}}$$

As $\tilde{N}_1$ increases, the strong law of large numbers implies that the above expression converges almost surely to

$$e_1 := \frac{\overline{\rho}F(\theta)\mathbb{E}[s|s \leq \theta] + \underline{\rho}(1-F(\theta))\mathbb{E}[s|s > \theta]}{F(\theta)\overline{\rho} + (1-F(\theta))\underline{\rho}} = \frac{\overline{\rho}\mathbb{E}[s|s \leq \theta] + \underline{\rho}\mathbb{E}[s|s > \theta]}{\overline{\rho} + \underline{\rho}}$$

Where the last equality follows from the fact that θ is the median of the distribution. Further define

$$e_2 := \frac{\underline{\rho}\mathbb{E}[s|s \leq \theta] + \overline{\rho}\mathbb{E}[s|s > \theta]}{\overline{\rho} + \underline{\rho}}$$

And note that $e_1 < e_2$. Denote M_1 as the set of realized signal sequences such that $\mathbb{E}[\theta|s^{\tilde{N}_1}, \hat{\rho}^{\tilde{N}_1}]$ converges to e_1 and we know $\mathbb{P}(M_1) = 1$. Fix $0 < \epsilon < (e_2 - e_1)/2$, for any signal sequence $s \in M_1$, then we can find $\tilde{N}_1^*$ such that

$$\mathbb{E}[\theta|s^{\tilde{N}_1^*}, \hat{\rho}^{\tilde{N}_1^*}] < e_1 + \epsilon.$$

Now, for the next $\tilde{N}_2 - \tilde{N}_1^*$ signals after the first $\tilde{N}_1^*$ signals, consider the following sequence

$$\hat{\rho}_i = \begin{cases} \underline{\rho} & if \quad s_i \in S_2 \\ \overline{\rho} & if \quad s_i \in S_2' \end{cases} \qquad S_2 = \{s_i \leq \theta : \tilde{N}_1^* < i \leq \tilde{N}_2\} \qquad S_2' = \{s_i > \theta : \tilde{N}_1^* < i \leq \tilde{N}_2\}$$

That is, we simply flip the weights assigned to signals lower than, or greater than θ. The expected value conditioning on signals up to $\tilde{N}_2$ can be expressed as

$$\mathbb{E}[\theta|s^{\tilde{N}_2}, \hat{\rho}^{\tilde{N}_2}] = \mathbb{E}[\theta|s^{\tilde{N}_1^*}, \hat{\rho}^{\tilde{N}_1^*}]\frac{d_1}{d_1 + d_2} + \frac{\underline{\rho}\sum_{i\in S_2} s_i + \overline{\rho}\sum_{i\in S_2'} s_i}{d_2}\frac{d_2}{d_1 + d_2}$$

Where $d_1 = |S_1|\overline{\rho} + |S_1'|\underline{\rho} + \rho_\mu$ and $d_2 = |S_2|\underline{\rho} + |S_2'|\overline{\rho}$. Once more, by the strong law of large numbers, as $\tilde{N}_2$ increases, $\frac{\underline{\rho}\sum_{i\in S_2} s_i + \overline{\rho}\sum_{i\in S_2'} s_i}{d_2}$ converges almost surely to e_2. Furthermore, keeping $\tilde{N}_1^*$ fixed, as we increase $\tilde{N}_2$

$$\frac{d_1}{d_1+d_2} \to 0 \qquad\qquad \frac{d_2}{d_1+d_2} \to 1$$

Thus, for a fixed $\tilde{N}_1^*$, as $\tilde{N}_2$ increases $\mathbb{E}[\theta|s^{\tilde{N}_2}, \hat{\rho}^{\tilde{N}_2}] \xrightarrow{a.s.} e_2$. Denote M_2 as the set of realized signal sequences such that $\mathbb{E}[\theta|s^{\tilde{N}_2}, \hat{\rho}^{\tilde{N}_2}]$ converges to e_2. Again, $\mathbb{P}(M_2) = 1$. Given previously defined ϵ, for any signal sequence $s \in M_1 \cap M_2$, then we can find $\tilde{N}_2^*$ such that

$$\mathbb{E}[\theta|s^{\tilde{N}_2^*}, \hat{\rho}^{\tilde{N}_2^*}] > e_2 - \epsilon.$$

Continuing in this fashion, for any $n > 2$, we can define M_n similarly and, for any signal sequence $s \in \bigcap_{j=1}^n M_j$, then we can find $\tilde{N}_n^*$ such that $\mathbb{E}[\theta|s^{\tilde{N}_n^*}, \hat{\rho}^{\tilde{N}_n^*}] > e_2 - \epsilon$ if n is even, and $\mathbb{E}[\theta|s^{\tilde{N}_n^*}, \hat{\rho}^{\tilde{N}_n^*}] < e_1 + \epsilon$ if n is odd. As $e_1 + \epsilon < e_2 - \epsilon$, we know that, for all $s \in \bigcap_{j\geq 1} M_j$, the sequence of $\mathbb{E}[\theta|s^K, \hat{\rho}^K]$ has two subsequences which cannot converge to the same limit point, which implies the original sequence will not converge. Notice that $\mathbb{P}(M_n) = 1$ for each n. By countable additivity, we know $\mathbb{P}(\bigcap_{j\geq 1} M_j) = 1$. Hence for almost surely any sequence of realized signals s, given the constructed precision sequence $\hat{\rho}$, $P_N(s^K, \hat{\rho}^K)$ does not converge.

Lemma 2. *Let $\hat{\rho}^*$ solve $\max_{\hat{\rho}^N \in [\underline{\rho}, \overline{\rho}]^N} \mathbb{E}[\theta|s^N, \hat{\rho}^N]$. Then, $\hat{\rho}^*$ is order-preserving and $\hat{\rho}^* \in \{\underline{\rho}, \overline{\rho}\}^N$.*

Proof of Lemma 2. First note that

$$\mathbb{E}[\theta|s^N, \hat{\rho}^N] = \frac{\sum_i \hat{\rho}_i s_i + \rho_\mu \mu}{\sum_i \hat{\rho}_i + \rho_\mu}$$

Without loss of generality assume that $s_i \leq s_j$ if $i < j$. Assume by contradiction that $\hat{\rho}_j^* > \underline{\rho}$ and $\hat{\rho}_k^* < \overline{\rho}$ for some $k > j$. Let $\triangle = \min\{\hat{\rho}_i^* - \underline{\rho}, \overline{\rho} - \hat{\rho}_k^*\}$. Define a new vector $\tilde{\rho}$ to be equal to $\hat{\rho}^*$ except in entry j and k. Let $\tilde{\rho}_j = \hat{\rho}_j^* - \triangle$, and $\tilde{\rho}_k = \hat{\rho}_k^* + \triangle$. We will then have

$$\mathbb{E}[\theta|s^N, \tilde{\rho}] = \frac{\sum_i \tilde{\rho}_i s_i + \rho_\mu \mu}{\sum_i \tilde{\rho}_i + \rho_\mu} = \frac{\sum_{i\neq j,k} \hat{\rho}_i^* s_i + (\hat{\rho}_j^* - \triangle)s_j + (\hat{\rho}_k^* + \triangle)s_k + \rho_\mu \mu}{\sum_i \hat{\rho}_i^* + \rho_\mu}$$

The difference in the expected θ is

$$\mathbb{E}[\theta|s^N, \tilde{\rho}] - \mathbb{E}[\theta|s^N, \hat{\rho}^*] = \frac{\triangle(s_k - s_j)}{\sum_i \hat{\rho}_i^* + \rho_\mu} > 0$$

Contradicting the claim that the initial vector $\hat{\rho}^*$ was optimal. Then, for the optimal $\hat{\rho}^*$ it can not be the case that some signal receives weight strictly above $\underline{\rho}$ if some other higher signal receives weight strictly below $\overline{\rho}$. This implies that $\hat{\rho}^*$ is order preserving. The only values $\hat{\rho}^*$ can take, as to not trigger the above contradiction, are $\hat{\rho}_i^* = \underline{\rho}\ \forall\, i < j$ and $\hat{\rho}_i^* = \overline{\rho}\ \forall\, i > j$, for some j.

Once more by contradiction assume that $\hat{\rho}_j^* \neq \{\underline{\rho}, \overline{\rho}\}$. We calculate the following derivative

$$\frac{\partial \mathbb{E}[\theta|s^N, \hat{\rho}^*]}{\partial \hat{\rho}_j^*} = \frac{\left((j-1)\underline{\rho} + (N-1)\overline{\rho} + \rho_\mu\right)s_j - \left(\underline{\rho}\sum_{i<j} s_i + \overline{\rho}\sum_{i>j} s_i + \rho_\mu \mu\right)}{\left((j-1)\underline{\rho} + \rho_j + (N-1)\overline{\rho} + \rho_\mu\right)^2}$$

The sign of the derivative is positive when s_j is higher than the weighed average of all other signals and the prior, $s_j > \frac{\left(\underline{\rho}\sum_{i<j} s_i + \overline{\rho}\sum_{i>j} s_i + \rho_\mu \mu\right)}{\left((j-1)\underline{\rho} + (N-1)\overline{\rho} + \rho_\mu\right)^2}$, negative if s_j is lower than this value and equal to 0 only in the knife-edge case in which s_j is exactly equal to this value. When the derivative is positive or negative, $\mathbb{E}[\theta|s^N, \hat{\rho}^*]$ increases as $\hat{\rho}_j^*$ shifts towards $\overline{\rho}$ or $\underline{\rho}$ respectively, contradicting the claim that $\hat{\rho}^*$ was optimal. When the derivative is equal to 0, shifting $\hat{\rho}_j^*$ all the way to $\underline{\rho}$ or $\overline{\rho}$ has no impact, thus, for this knife-edge case we can simply define $\hat{\rho}_j^*$ to be equal to $\overline{\rho}$.

By symmetry we have the following corollary.

Corollary 1. *Let $\tilde{\rho}^*$ solve $\min_{\hat{\rho}^N \in [\underline{\rho}, \overline{\rho}]^N} \mathbb{E}[\theta|s^N, \hat{\rho}^N]$. Then, $\tilde{\rho}^*$ is order-reversing and $\tilde{\rho}^* \in \{\underline{\rho}, \overline{\rho}\}^N$.*

Proof of Proposition 2. Recall we want to solve the problem:

$$\min_{g} \max_{\hat{\rho}^N \in [\underline{\rho}, \overline{\rho}]^N} \{\left(g - \mathbb{E}[\theta|s^N, \hat{\rho}^N]\right)^2 + \mathbb{V}[\theta|s^N, \rho^N]\} \tag{8}$$

Start by associating, to each vector of precisions $\hat{\rho}^N$, a vector of relative precisions: $z^N = \frac{1}{\hat{\rho}^N \cdot \mathbb{1}^N + \rho_\mu}\hat{\rho}^N$.

Under the assumption of normality, Bayesian updating generates the following expressions for the posterior mean and variances:

$$\mathbb{E}[\theta|s^N, z^N] = \mu + z^N \cdot (s^N - \mu) \tag{9}$$

and

$$\mathbb{V}[\theta|s^N, z^N] = (1 - z^N \cdot \mathbb{1}^N)\frac{1}{\rho_\mu} \tag{10}$$

We now transform the constraint set in the optimization from the space of precisions to relative precisions. For that, notice that for $j \in \{2, .., N\}$:

$$z_j = z_1 \frac{\hat{\rho}_j}{\hat{\rho}_1}$$

For that j, then, the bounds can be rewritten as:

$$\underline{\rho} \leq \hat{\rho}_j \leq \overline{\rho} \iff \frac{\underline{\rho}}{\rho_1} z_1 \leq z_j \leq \frac{\overline{\rho}}{\rho_1} z_1$$

and $\underline{\rho} \leq \rho_1 \leq \overline{\rho}$. Applying the latter on the conditions for z grants, for $j \in \{2, ..., N\}$

$$\frac{\underline{\rho}}{\overline{\rho}} z_1 \leq z_j \leq \frac{\overline{\rho}}{\underline{\rho}} z_1$$

or, rewriting in matrix form, there is a matrix A such that: $A \cdot z^N \leq 0$. Therefore, the constraint set consists of a bounded polytope in $\mathbb{R}^N$. Let $P^N = \{z : A \cdot z^N \leq 0\}$.

The inner maximization problem, in terms of relative precisions is then to:

$$\max_{z^N \in P^N} \{\left(g - \mathbb{E}[\theta|s^N, z^N]\right)^2 + \mathbb{V}[\theta|s^N, z^N]\}$$

By using the expressions in 9 and 10, it is easy to see that the objective function is (strictly) convex, and is therefore maximized in the extreme points of the convex polyhedron P^N, $ex(P^N)$. The following lemma proves that all extreme points of P are extreme points of the hypercube $[\underline{\rho}, \overline{\rho}]^N$, guaranteeing that the solution of the maximization problem is an assignment $\hat{\rho}^N(g) \in \{\underline{\rho}, \overline{\rho}\}^N$.

Lemma 3. *Let T be the transformation defined by* $T(\rho^N) = \frac{1}{\rho^N \cdot \mathbb{1}^N + \rho_\mu} \rho^N$. *Then* $ex(P^N) \subset T\left(ex([\underline{\rho}, \overline{\rho}]^N)\right)$.

Proof of Lemma. Take $z^N \in ex(P)$. Because z^N is an extreme point of the compact polytope P^N, it is the unique solution to some linear programming problem, say:

$$z^N = \arg\max\{c \cdot z : z \in P\}$$

Because T is surjective, we can rewrite this problem in terms of ρ^N:

$$\max\{c \cdot \frac{1}{\rho \cdot \mathbb{1}^N + \rho_\mu} \rho : \rho \in [\underline{\rho}, \overline{\rho}]^N\}$$

Lemma 2 shows that at least one solution to this problem is obtained at $\rho^N \in \{\underline{\rho}, \overline{\rho}\}^N$. This means $\rho^N \in ex([\underline{\rho}, \overline{\rho}]^N)$. Since $z^N = T(\rho^N)$ we are done.

Lemma 3 makes sure that the optimization can be done in the space of relative precisions and then transferred back to the space of precisions. Importantly, this implies that the optimal precision assignments satisfy $\hat{\rho}^{N*}(g) \in \{\underline{\rho}, \overline{\rho}\}^N$.

We now characterize the two types of solutions. Let $\underline{z}^N$ be the relative precision of the precision vector $\underline{\rho}^N = \underline{\rho}\mathbb{1}^N$. First, by the minmax inequality:

$$\begin{aligned}
&\min_g \max_{z^N \in P} \{\left(g - \mathbb{E}[\theta|s^N, z^N]\right)^2 + \mathbb{V}[\theta|s^N, z^N]\} \geq \\
&\max_{z^N \in P} \min_g \{\left(g - \mathbb{E}[\theta|s^N, z^N]\right)^2 + \mathbb{V}[\theta|s^N, z^N]\} = \\
&\qquad \max_{z^N \in P} \mathbb{V}[\theta|s^N, z^N] = \mathbb{V}[\theta|s^N, \underline{z}^N]
\end{aligned} \tag{11}$$

This is a lower bound for the observer's loss, so the best they can expect to achieve. This is actually achieved when, at g, the optimal strategy for nature is set $\underline{z}^N$. That is:

$$\mathbb{V}[\theta|s^N,\underline{z}^N] = \max_{z^N}\{\left(g - \mathbb{E}[\theta|s^N,z^N]\right)^2 + \mathbb{V}[\theta|s^N,z^N]\}$$

Because for each z^N the objective function is a parabola for g, this condition can be rewritten using 9 and 10 as:

$$-1 \leq \frac{z^N - \underline{z}^N}{\sqrt{(z^N - \underline{z}^N)\cdot \mathbb{1}}} \cdot \frac{1}{\rho_\mu} \leq 1$$

for all $z^N \in P$, justifying the definition of the set K^N in the statement of the proposition. Under these circumstances, as pointed out before, the optimal guess is $g^* = \mathbb{E}[\theta|s^N,\underline{z}^N]$, which is the statement in the proposition, per equation 9.

Finally, assume that $s^N \notin K^N$, so the optimal guess is not $g^* = \mathbb{E}[\theta|s^N,\underline{z}^N]$. Notice that, in that case, $g^* \neq \mathbb{E}[\theta|s^N,z^{N*}(g^*)]$. Indeed, if that was the case, Nature could increase it's payoff by deviating to $\underline{z}^N$. Then, assume, without loss of generality, that $\mathbb{E}[\theta|s^N,z^{N*}(g^*)] - g^* = \epsilon > 0$. If, for any $0 < \delta < \epsilon$, $z^{N*}(g^* + \delta) = z^{N*}(g^*)$, then this $g^* + \delta$ is a deviation for the observer, as it reduces her squared bias without affecting the posterior variance.

The argument above shows that $z^{N*}(g^* + \delta) \neq z^{N*}(g^*)$ for any $\epsilon > \delta > 0$. By the fact that Nature has only a finite number of strategies and continuity of Nature's value function, $z^{N*}(g^*)$ is non-unique. That is, Nature is indiferent between two strategies at g^*. Noting them z'^N and z''^N and imposing indifference for Nature obtains the second part of the guess in the statement.

What is left to prove are the properties of z'^N, z''^N. For that, we first prove Lemma 1.

Proof of Lemma 1. Fix g. First, recall that the solution to Nature's problem is $\hat{\rho}^{N*} \in \{\underline{\rho},\overline{\rho}\}^N$. For a vector in this space, define $\#x = \#\{i : x_i = \overline{\rho}\}$. Finally, from equation 10, $\mathbb{V}[\theta|s^N,\hat{\rho}^N]$ depends only on $\#\hat{\rho}^N$. Then, for any $d \in \{1,\ldots,N\}$:

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

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

The solution to the above maximization problem is obtained by either maximizing or minimizing $\mathbb{E}[\theta|s^N,\rho^N]$. Either way, the precision assignment that solves it is monotonic, from Lemma 2. The problem of Nature is then:

$$\max_{d \in \{1,\ldots,N\}}\{\left(g - \mathbb{E}[\theta|s^N,\hat{\rho}^N(d)]\right)^2 + \mathbb{V}[\theta|s^N,\hat{\rho}^N(d)]\}$$

which is a choice over monotonic assignments and so is, again, a monotonic assignment.

Lemma 1 guarantees that the optimal strategy for Nature is monotonic for any g. We conclude by showing that, at the optimal g^*, the two optimal strategies: $z^{'N}$ and $z^{''N}$ are one order preserving, the other order-reversing. Let $\hat{\rho}'^N$ and $\hat{\rho}''^N$ be the precisions that generate them, respectively. Assume $\mathbb{E}[\theta|s^N, \hat{\rho}'^N] > \mathbb{E}[\theta|s^N, \hat{\rho}''^N]$.

First notice that $g^* \in \left[\mathbb{E}[\theta|s^N, \hat{\rho}''^N)], \mathbb{E}[\theta|s^N, z'^N]\right]$. Indeed, that is implied by the argument that the two strategies guarantee the agent the same utility. Formally, by strict convexity of Nature's objective, it is indifferent to at most two strategies. Because there are only finite strategies, that means that there is a neighborhood around g^* where at least one of the strategies is still optimal for Nature. If both posterior means are, say, lower than g^*, the observer could reduce their guess and increase her payoff by decreasing the squared bias, while keeping the variance constant.

Assume, to obtain a contradiction, that the two functions are monotonically increasing. We construct a profitable deviation for Nature. Consider the order-reversing assignment $\tilde{\rho}^N$ such that $\#\tilde{\rho}^N = \#\hat{\rho}''^N$ - that is, this assignment takes the same number of high (and low) precisions as $\hat{\rho}''^N$, but inverts their order, attributing low precisions to high-valued signals.

It easy to see that $\mathbb{E}[\theta|s^N, \tilde{\rho}^N] < \mathbb{E}[\theta|s^N, \hat{\rho}''^N]$, as the former weights more the low realizations of the signals. On top of that, their ex-post variances coincide. Because, as previously argued, $g^* > \mathbb{E}[\theta|s^N, \hat{\rho}''^N]$, this precision assignment increases the squared bias without affecting the variance, proving that it would be suboptimal for Nature to choose both strategies to be monotonically increasing. A symmetric argument proves that they cannot be both order-reversing either, so the proof is finished.

□

Proof of Proposition 4. Start defining, for any realization of signals s^N:

$$\underline{m}^N \equiv \min_{\hat{\rho}\in[\underline{m},\overline{m}]^N} \mathbb{E}[\theta|s^N, \hat{\rho}^N] = \frac{\overline{\rho}\int_{-\infty}^{\underline{m}^N} x dF^N(x) + \underline{\rho}\int_{\underline{m}^N}^{\infty} x dF^N(x)}{\overline{\rho}F^N(\underline{m}^N) + \underline{\rho}(1 - F^N(\underline{m}^N))}$$

and

$$\overline{m}^N \equiv \max_{\hat{\rho}\in[\underline{m},\overline{m}]^N} \mathbb{E}[\theta|s^N, \hat{\rho}^N] = \frac{\underline{\rho}\int_{-\infty}^{\overline{m}^N} x dF^N(x) + \overline{\rho}\int_{\overline{m}^N}^{\infty} x dF^N(x)}{\underline{\rho}F^N(\overline{m}^N) + \overline{\rho}\left(1 - F^N(\overline{m}^N)\right)}$$

where F^N is the empirical distribution of s^N and the equality is justified by the proof of Lemma 2.[4] $\underline{m}^N$ and $\overline{m}^N$ are (random) bounds on posterior means. Let $\hat{\rho} : \mathbb{R} \to [\underline{\rho}, \overline{\rho}]$ be a precision assign-

[4]Explicitly, in that proof, it is shown that if $u = \max_{\hat{\rho}^N\in[\underline{\rho},\overline{\rho}]^N}\{\mathbb{E}[\theta|s^N, \hat{\rho}^N]\}$, then by setting $g(x) = \underline{\rho}$, for $x < u$ and $g(x) = \overline{\rho}$ otherwise, and setting: $\tilde{\rho}_i^N = g(s_i)$, we obtain:

$$u = \mathbb{E}[\theta|s^N, \tilde{\rho}^N]$$

ment. Recalling that F is the real distribution of signals, define:

$$\mathbb{E}[\theta|\hat{\rho}] = \int \frac{\hat{\rho}(x) x dF(x)}{\int \hat{\rho}(x) dF(x)}$$

In **Step 1** and **Step 2** below, we prove two auxiliary results. We start the proof by showing, in **Step 1** that $\underline{m} = \min_{\hat{\rho}:\mathbb{R}\to[\underline{\rho},\overline{\rho}]} \mathbb{E}[\theta|\hat{\rho}]$ and $\overline{m} = \max_{\hat{\rho}:\mathbb{R}\to[\underline{\rho},\overline{\rho}]} \mathbb{E}[\theta|\hat{\rho}]$, where $\underline{m}$ and $\overline{m}$ are as in the statement of Proposition 4. In that step, we also show that for any value $\underline{m} \le m \le \overline{m}$, there is $\hat{\rho}_m$ such that $\mathbb{E}[\theta|\hat{\rho}_m] = m$. We then proceed to show, in **Step 2**, that $\underline{m}^N$ and $\overline{m}^N$ converge almost surely to $\underline{m}$ and $\overline{m}$, respectively. After establishing these auxiliary properties on the bounds, the proof of Proposition 4 is direct. We demonstrate it in **Step 3** and **Step 4**. In the former we prove the asymptotic bounds on all posterior sequences, whereas in the latter we characterize the limiting belief set.

Step 1. Properties of $\underline{m}$, $\overline{m}$

We begin this step of the proof by showing that the $\hat{\rho}$ that maximizes or minimizes $\mathbb{E}[\theta|s,\hat{\rho}]$ takes threshold forms, assigning all signals with low values minimal precision, and signals with high values maximal precision, or vice versa. This is shown in **Step 1.1**. We continue in **Step 1.2** by identifying the unique point where the optimal precision flips from minimal to maximal, or vice versa. There we also identify the value of the highest and lowest attainable $\mathbb{E}[\theta|\hat{\rho}]$. These values are $\overline{m}$, $\underline{m}$, respectively. Finally, in **Step 1.3** we show that the interval has full support, and that by using only extreme precisions $\hat{\rho}(x) \in \{\underline{\rho}, \overline{\rho}\}$ any value in the interval is attainable.

From now on, we focus on $\overline{m}$. A symmetric version of all arguments holds for $\underline{m}$. Then, define $\hat{\rho}^* \in \arg\max_{\hat{\rho}:\mathbb{R}\to[\underline{\rho},\overline{\rho}]} \mathbb{E}[\theta|\hat{\rho}]$.

Step 1.1. $\hat{\rho}^*$ is order-preserving and $\hat{\rho}^*(x) \in \{\underline{\rho}, \overline{\rho}\}$ for all $x \in \mathbb{R}$.

Let $\tilde{\rho} : \mathbb{R} \to \left[\underline{\rho}, \overline{\rho}\right]$. Then, recall that:

$$\mathbb{E}[\theta|\tilde{\rho}] = \int \frac{\tilde{\rho}(x) x}{\int \tilde{\rho}(x) f(x) dx} f(x) dx$$

For an arbitrary value δ let $L(\delta) = \{x < \delta : \tilde{\rho}(x) > \underline{\rho}\}$ and $R(\delta) = \{x > \delta : \tilde{\rho}(x) < \overline{\rho}\}$. Furthermore define

$$\mu_L(\delta) = \int_{L(\delta)} f(x) dx \qquad \mu_R(\delta) = \int_{R(\delta)} f(x) dx$$

Assume that δ is a value for which $\mu_L(\delta) > 0$ and $\mu_R(\delta) > 0$. Further define

$$\hat{\mu}_L = \int_{L(\delta)} \left(\tilde{\rho}(x) - \underline{\rho}\right) f(x) dx \qquad \hat{\mu}_R = \int_{R(\delta)} (\bar{\rho} - \tilde{\rho}(x)) f(x) dx$$

Without loss of generality, in what follows we assume that $\hat{\mu}_L(\delta) < \hat{\mu}_R(\delta)$. Then, define $\tilde{\tilde{\rho}}(x)$ as follows

$$\tilde{\tilde{\rho}}(x) = \begin{cases} \tilde{\rho}(x) & if \quad x \notin \{L(\delta) \cup R(\delta)\} \\ \underline{\rho} & if \quad x \in L(\delta) \\ \tilde{\rho}(x) + g(x) & if \quad x \in R(\delta) \end{cases}$$

Where $g(x)$ is positive for any x and is chosen such that:

$$\int_{R(\delta)} g(x) f(x) dx = \hat{\mu}_L$$

Note that

$$\begin{aligned} \int \tilde{\rho}(x) f(x) dx &= \int_{L(\delta)} \tilde{\rho}(x) f(x) dx + \int_{R(\delta)} \tilde{\rho}(x) f(x) dx + \int_{x \notin (L(\delta) \cup R(\delta))} \tilde{\rho}(x) f(x) dx \\ &= \int_{L(\delta)} \underline{\rho} f(x) dx + \int_{L(\delta)} \left(\tilde{\rho}(x) - \underline{\rho}\right) f(x) dx + \int_{R(\delta)} \tilde{\rho}(x) f(x) dx + \int_{x \notin (L(\delta) \cup R(\delta))} \tilde{\rho}(x) f(x) \\ &= \int_{x \in L} \tilde{\tilde{\rho}}(x) f(x) dx + \hat{\mu}_L + \int_{x \in R} \tilde{\rho}(x) f(x) dx + \int_{x \notin \{L \cup R\}} \tilde{\tilde{\rho}}(x) f(x) \\ &= \int \tilde{\tilde{\rho}}(x) f(x) dx \end{aligned}$$

Hence, the value of the denominator of $\mathbb{E}[\theta|\tilde{\rho}]$ is the same under $\tilde{\rho}$ and $\tilde{\tilde{\rho}}$. Under $\tilde{\tilde{\rho}}$ the mean of the posterior will be

$$\mathbb{E}[\theta|\tilde{\tilde{\rho}}] = \int_{L(\delta)} \frac{\underline{\rho} x}{\int \tilde{\rho}(x) f(x) dx} f(x) dx + \int_{R(\delta)} \frac{(\tilde{\rho}(x) + g(x)) x}{\int \tilde{\rho}(x) f(x) dx} f(x) dx + \int_{x \notin (L(\delta) \cup R(\delta))} \frac{\rho(x) x}{\int \tilde{\rho}(x) f(x) dx} f(x) dx$$

We then have

$$\mathbb{E}[\theta|\tilde{\tilde{\rho}}] - \mathbb{E}[\theta|\tilde{\rho}] = \int_{R(\delta)} \frac{g(x)}{\int \tilde{\rho}(x) f(x) dx} x f(x) dx - \int_{L(\delta)} \frac{\tilde{\rho}(x) - \underline{\rho}}{\int \tilde{\rho}(x) f(x) dx} x f(x) dx \geq 0$$

To see why the inequality above holds, note that

$$\int_{R(\delta)} \frac{g(x)}{\int \tilde{\rho}(x) f(x) dx} x f(x) dx \geq \hat{\mu}_L \underline{x}_R \geq \hat{\mu}_L \bar{x}_L \geq \int_{L(\delta)} \frac{\tilde{\rho}(x) - \underline{\rho}}{\int \tilde{\rho}(x) f(x) dx} x f(x) dx$$

Where $\underline{x}_R = \{x \in R(\delta) : x \leq x' \; \forall x' \in R(\delta)\}$ and $\bar{L}_R = \{x \in L(\delta) : x \geq x' \; \forall x' \in L(\delta)\}$. Hence, for any $\tilde{\rho}(x)$, if we can find some point δ, such that for a positive mass of $x < \delta$ we have $\tilde{\rho}(x) > \underline{\rho}$, and for a positive mass of $x > \delta$ we have $\tilde{\rho}(x) < \bar{\rho}$, we can ensure a higher mean of the posterior by shifting some of the mass from the x values that are lower than δ to the x values that are larger than δ. As a consequence it must be that $\rho^*(x)$ has the following threshold form

$$\rho^*(x) = \begin{cases} \underline{\rho} & if \quad x \leq a^* \\ \bar{\rho} & if \quad x > a^* \end{cases}$$

For some a^*. This concludes **Step 1.1**

Step 1.2. $\overline{m} = \arg\max_{\hat{\rho}:\mathbb{R}\to[\underline{\rho},\overline{\rho}]} \mathbb{E}[\theta|\hat{\rho}]$.

From the previous step, we know that to maximize the expected value of θ, nature finds it optimal to assign precision $\underline{\rho}$ to signals with values lower than some value a and $\overline{\rho}$ to signals with values higher than a. Let $[\rho_a]_{a\in\mathbb{R}}$ be the set of all such presision assignments. In this part we determine the value of a, which results from the maximization problem below

$$\max_{a\in\mathbb{R}} \mathbb{E}[\theta|\rho_a]$$

The first order condition leads to:

$$a^* = \frac{\underline{\rho}\int_{-\infty}^{a^*} xf(x)dx + \bar{\rho}\int_{a^*}^{\infty} xf(x)dx}{\underline{\rho}F(a^*) + \bar{\rho}(1 - F(a^*))}$$

Which implicitly defines the value a^* that solves that maximization. We show that the objective function is single-peaked, so that the first order condition is necessary and sufficient. Denote $v(a) = \mathbb{E}[\theta|\rho_a]$. 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:

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

which is a contradiction with $v'(b) = 0$ - i.e. with $v(b) = b$.

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

As $a \to -\infty$, $v(a) \to \int_{-\infty}^{\infty} xf(x)dx$, 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 $v(a) \to \int_{-\infty}^{\infty} xf(x)dx$, 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 $a \geq \bar{(}a)$.

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:

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

again, yielding a contradiction. Therefore a^* is unique. Noticing that $\mathbb{E}[\theta|\rho_{a^*}] = v(a^*) = a^*$, and redefining $a^* = \overline{m}$ we conclude **Step 1.2**.

By symmetry we have the following corollary

Corollary 2. *Let $\tilde{\rho}^*$ solve $\min_{\hat{\rho}} \mathbb{E}[\theta|s, \hat{\rho}]$. Then, $\tilde{\rho}^*$ is order-reversing and $\tilde{\rho}^*(x) \in \{\underline{\rho}, \overline{\rho}\}$. In particular*

$$\tilde{\rho}^*(x) = \begin{cases} \bar{\rho} & if \quad x \leq \tilde{a}^* \\ \underline{\rho} & if \quad x > \tilde{a}^* \end{cases}$$

Where $\tilde{a}^*$ is implicitly defined as follows:

$$\tilde{a}^* = \frac{\bar{\rho}\int_{-\infty}^{\tilde{a}^*} xf(x)dx + \underline{\rho}\int_{\tilde{a}^*}^{\infty} xf(x)dx}{\bar{\rho}F(\tilde{a}^*) + \underline{\rho}(1 - F(\tilde{a}^*))}$$

Step 1.3. The mapping $\hat{\rho} \to \mathbb{E}[\theta|\hat{\rho}]$ is onto in $[\underline{m}, \overline{m}]$ Having identified the highest and lowest achievable $\mathbb{E}[\theta|\hat{\rho}]$ we now show that we have full support within this interval. That is, we can find $\hat{\rho}$ such $\mathbb{E}[\theta|\hat{\rho}]$ is equal to any value within this interval.

Once more let

$$v(a) = \int_{-\infty}^{a} \frac{\underline{\rho}}{\underline{\rho}F(a) + \bar{\rho}(1 - F(a))} xf(x)dx + \int_{a}^{\infty} \frac{\bar{\rho}}{\underline{\rho}F(a) + \bar{\rho}(1 - F(a))} xf(x)dx$$

That is, $v(a)$ represent the mean of the posterior when all signals with values lower than a are assigned precision $\underline{\rho}$ while all signals with values higher than a are assigned precision $\overline{\rho}$. It is straightforward to see that this function is continuous in a. From the analysis in **Step 1.2** we know that function is maximized when $a = a^*$, in which case $v(a^*) = a^*$, the upper bound of the interval. If instead we set $a = -\infty$

$$v(a = -\infty) = \int_{-\infty}^{\infty} xf(x)dx = \mu$$

Which is necessarily lower than a^*. By continuity of $v(a)$ with a, we know that any point between μ and the upper bound a^* is attainable. Now let

$$v(a) = \int_{-\infty}^{a} \frac{\overline{\rho}}{\overline{\rho}F(a) + \underline{\rho}(1 - F(a))} x f(x) dx + \int_{a}^{\infty} \frac{\underline{\rho}}{\overline{\rho}F(a) + \underline{\rho}(1 - F(a))} x f(x) dx$$

That is, $\tilde{v}(a)$ represent the mean of the posterior when all signals with values lower than a are assigned precision $\overline{\rho}$ while all signals with values higher than a are assigned precision $\underline{\rho}$. Once more, it is straightforward to see that this function is continuous in a. From symmetry and the analysis in **Step 1.2** we know that function is minimized when $a = \tilde{a}^*$, in which case $\tilde{v}(\tilde{a}^*) = \tilde{a}^*$, the lower bound of the interval. If instead we set $a = -\infty$

$$v(a = \infty) = \int_{-\infty}^{\infty} x f(x) dx = \mu$$

Which is necessarily higher than $\tilde{a}^*$. By continuity of $\tilde{v}(a)$ with a, we know that any point between μ and the upper bound a^* is attainable. Thus, with values of $\hat{\rho} \in \{\underline{\rho}, \overline{\rho}\}$ alone we can swipe the entire interval. This concludes **Step 1**.

Step 2. $\underline{m}^N \xrightarrow{a.s.} \underline{m}$ **and** $\overline{m}^N \xrightarrow{a.s.} \overline{m}$ From now on, we will focus on proving $\underline{m}^N \xrightarrow{a.s.} \underline{m}$. Symmetric arguments can be used to prove $\overline{m}^N \xrightarrow{a.s.} \overline{m}$. Denote

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

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

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

Step 2.1. $\sup_{\mathbf{a} \in \mathbb{R}} |\boldsymbol{\Psi}^{\mathbf{N}}(\mathbf{a}) - \boldsymbol{\Psi}(\mathbf{a})| \xrightarrow{\mathbf{a.s.}} \mathbf{0}.$ Given the Glivenko–Cantelli theorem, we know the empirical distribution function converges to the true cumulative distribution function pointwise and uniformly over x, that is,

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

For each real-valued function v, denote

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

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

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

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

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

Given two real-valued functions l and u and $\epsilon > 0$, a *ϵ-bracket* $[l,u]$ is the set of all functions f such that $l \le f \le u$ and $\|u - l\|_{L_1(F)} \le \epsilon$. The *bracketing number* $N(\epsilon, \mathcal{V}, \|\cdot\|_{L_1(F)})$ 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 1. *(Blum-DeHardt (Blum, 1955; DeHardt, 1971)) If $N(\epsilon, \mathcal{V}, \|\cdot\|_{L_1(F)}) < \infty$ for any $\epsilon > 0$, then $\mathcal{V}$ is a P-Glivenko-Cantelli class.*

Denote

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

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

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 $\epsilon > 0$. 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

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

This is feasible as P 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 $N(\epsilon, \mathcal{V}_1, \|\cdot\|_{L_1(F)}) \le M - 1 < \infty$. 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

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

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 $N(\epsilon, \mathcal{V}_2, \|\cdot\|_{L_1(F)}) \le M' - 1 < \infty$. Hence $\mathcal{V}_2$ is a P-Glivenko-Cantelli class.

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

$$\|F^N - F\|_{\mathcal{V}_1} = \sup_{v \in \mathcal{V}_1} |F^N(v) - F(v)| = \sup_{a \in \mathbb{R}} |F^N(v_1^a) - F(v_1^a)| \xrightarrow{a.s.} 0. \tag{12}$$

$$\|F^N - F\|_{\mathcal{V}_1} = \sup_{v \in \mathcal{V}_1} |F^N(v) - F(v)| = \sup_{a \in \mathbb{R}} |F^N(v_1^a) - F(v_1^a)| \xrightarrow{a.s.} 0. \tag{13}$$

Now we can show the convergence of Ψ^N.

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

Notice that $0 < \underline{\rho} \le F(v_1^a) \le \overline{\rho} < \infty$ and $0 < \underline{\rho} \le F^N(v_1^a) \le \overline{\rho} < \infty$ 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,

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

By equations 12 and 13, we know

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

Step 2.2. $\underline{\mathbf{m}}^{\mathbf{N}} \xrightarrow{\mathbf{a.s.}} \underline{\mathbf{m}}$ This result follows directly from the following standard results about consistency of M- estimators. We include the proof for completeness.

Lemma 4. *Suppose that*

1. $\sup_{a\in\mathbb{R}} |\Psi^N(a) - \Psi(a)| \xrightarrow{a.s.} 0$,
2. $\underline{m}_N \in \arg\min_{a\in\mathbb{R}} \Psi^N(a)$ *for each* N,
3. $\underline{m} \in \arg\min_{a\in\mathbb{R}} \Psi(a)$ *is the unique minimum of* Ψ,

Then $\underline{m}_N \xrightarrow{a.s.} \underline{m}$.

Proof of Lemma 4. By conditions (2) and (3), we know $\Psi^N(\underline{m}_N) \leq \Psi^N(\underline{m})$ and $\Psi(\underline{m}) \leq \Psi(\underline{m}_N)$ for each N. Using these inequalities we have

$$\Psi^N(\underline{m}_N) - \Psi(\underline{m}_N) \leq \Psi^N(\underline{m}_N) - \Psi(\underline{m}) \leq \Psi^N(\underline{m}) - \Psi(\underline{m})$$

Therefore from the above we have

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

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

Now it suffices to show that the conditions in Lemma 4 holds in our case. Condition (1) is shown in Step 1.1. Condition (2) holds by the definition of $\underline{m}_N$. Condition (3) is shown in the proof of Proposition 4. This completes the proof for $\underline{m}_N \xrightarrow{a.s.} \underline{m}$. The same arguments apply for showing $\overline{m}_N \xrightarrow{a.s.} \overline{m}$.

Step 3. Constraints on beliefs

For any N, we consider the sequence of signals s^N and an associated sequence of precisions $\hat{\rho}^N$. Recall that, for $N \in \mathbb{N}$:

$$\theta|s^N, \hat{\rho}^N \sim \left(\frac{\sum_{i=1}^N \hat{\rho}_i s_i + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}, \left(1 - \frac{\sum_{i=1}^N \hat{\rho}_i}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu}\right)\frac{1}{\rho_\mu}\right) \tag{14}$$

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

$$\underline{m}^N \leq \frac{\sum_{i=1}^N \hat{\rho}_i s_i + \rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i + \rho_\mu} \leq \overline{m}^N$$

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 4. Characterization of the set of limit posteriors.

Fix a sequence of realizations s. We want to characterize the set of distributions the posterior beliefs of the observer converge to, $\mathbb{P}_\infty(s)$. By 14, it is clear that a necessary condition for weak convergence is that the expected mean $\frac{\sum_{i=1}^N \hat{\rho}_i s_i+\rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i+\rho_\mu}$ converges. We can then focus on sequences with converging means. Define $b = \lim_{N\to\infty} \frac{\sum_{i=1}^N \hat{\rho}_i s_i+\rho_\mu \mu}{\sum_{i=1}^N \hat{\rho}_i+\rho_\mu}$.

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

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

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 monotonic function $\rho : \mathbb{R} \to \{\underline{\rho}, \overline{\rho}\}$. Then $\{\rho(s_i)s_i\}_{i=1,...}$ is a sequence of independent signals with uniformly bounded variance. Then, by the strong law of large numbers:

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

Then, Step 1 shows 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 $\overline{m}$. This finishes the proof.

Proof of Proposition 5 . Denote

$$\Gamma^N(g) = \max_{z^N \in P^N} \left\{(g-\mu-z^N\cdot(s^N-\mu\mathbb{1}))^2 + (1-z^N\cdot\mathbb{1})\sigma^2\right\}$$

where $P^N = \{z^N : A \cdot z^N \leq 0\}$.

By previous arguments, as the number of signals goes to infinity, the limiting optimal guess is

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

We also denote

$$\Gamma(g) = \max\{(g-\overline{m})^2, (g-\underline{m})^2\}$$

where $\underline{m}$ and $\overline{m}$ are defined in Proposition 4.

We start with introducing an auxiliary problem with finitely many signals by ignoring the effect of the posterior variances. Denote

$$\tilde{\Gamma}^N(g) = \max_{z^N \in P^N} \left\{(g-\mu-z^N \cdot (s^N - \mu\mathbb{1}))^2\right\}$$

Following the same arguments as Proposition 4,

$$\tilde{\Gamma}^N(g) = \max\{(g-\overline{m}^N)^2, (g-\underline{m}^N)^2\}$$

where

$$\begin{aligned}
\underline{m}^N = \min_{z^N \in P^N} \mu + z^N \cdot (s-\mu\mathbb{1}) &= \frac{\overline{\rho}\int_{-\infty}^{\underline{m}^N} x dF^N(x) + \underline{\rho}\int_{\underline{m}^N}^{\infty} x dF^N(x)}{\overline{\rho}F^N(\underline{m}^N) + \underline{\rho}(1-F^N(\underline{m}^N))} \\
&= \arg\min_{a\in\mathbb{R}} \frac{\overline{\rho}\int_{-\infty}^{a} x dF^N(x) + \underline{\rho}\int_{a}^{\infty} x dF^N(x)}{\overline{\rho}F^N(a) + \underline{\rho}(1-F^N(a))}.
\end{aligned}$$

$$\begin{aligned}
\overline{m}^N = \max_{z^N \in P^N} \mu + z^N \cdot (s-\mu\mathbb{1}) &= \frac{\underline{\rho}\int_{-\infty}^{\overline{m}^N} x dF^N(x) + \overline{\rho}\int_{\overline{m}^N}^{\infty} x dF^N(x)}{\underline{\rho}F^N(\overline{m}^N) + \overline{\rho}\left(1-F^N(\overline{m}^N)\right)} \\
&= \arg\max_{a\in\mathbb{R}} \frac{\underline{\rho}\int_{-\infty}^{a} x dF^N(x) + \overline{\rho}\int_{a}^{\infty} x dF^N(x)}{\underline{\rho}F^N(a) + \overline{\rho}(1-F^N(a))}.
\end{aligned}$$

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

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

In the remainder of this proof, we aim to show that Γ^N, Γ, $g^N \equiv g^*(s^N)$ and $g^* = \frac{\overline{m}+\underline{m}}{2}$ satisfy the conditions of Lemma 5. We do so in three steps, one for each condition in the lemma. This allows us to obtain that $g^*(s^N) \xrightarrow{a.s} \frac{\underline{m}+\overline{m}}{2}$. We then use **Step 4** to prove that $\frac{\underline{m}+\overline{m}}{2} = \theta$.

Step 1. $\sup_{\mathbf{g}\in\mathbf{C}} |\boldsymbol{\Gamma}(\mathbf{g}) - \boldsymbol{\Gamma}^{\mathbf{N}}(\mathbf{g})| \xrightarrow{\text{a.s.}} \mathbf{0}$**, 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 $\tilde{\Gamma}^N$, obtained by ignoring the variance in Nature's problem. As N grows to infinity, the gap between Γ^N and $\tilde{\Gamma}^N$ shrinks uniformly. To see that formally, start by noticing that $\Gamma^N(g) \geq \tilde{\Gamma}^N(g)$ for all g, since the variance term is always positive. Additionally, by subadditivity of the max:

$$\begin{aligned}\Gamma^N(g) &\leq \max_{z^N\in P^N} (g - \mu - z^N \cdot (s^N - \mu))^2 + \max_{z^N\in P^N} (1 - z^N \cdot \mathbb{1})\frac{1}{\rho_\mu} \\ &= \tilde{\Gamma}^N(g) + \max_{z^N\in P^N} (1 - z^N \cdot \mathbb{1})\frac{1}{\rho_\mu}\end{aligned}$$

Because $\underline{\rho} > 0$, the last term on the right hand side converges to zero for any signal distribution. Clearly this convergence is independent of g. That means, for every realization of signals, $\limsup \sup_g \{\Gamma^N(g) - \tilde{\Gamma}^N(g))\} \leq 0$. But we proved this difference is non-negative for all N, then the uniform convergence result follows.

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

Before we proceed, recall from Part 1 that we can write $\tilde{\Gamma}^N(g) = \max\{(g - \overline{m}^N)^2, (g - \underline{m}^N)^2\}$. Also by part one, $\overline{m}^N \xrightarrow{a.s.} \overline{m}$ and $\underline{m}^N \xrightarrow{a.s.} \underline{m}$.

For any $g \in \mathbb{R}$, let $m^N(g) \in \{\underline{m}^N, \overline{m}^N\}$ be such that $\tilde{\Gamma}^N(g) = (g - m^N(g))^2$ and $m(g) \in \{\underline{m}, \overline{m}\}$ such that $\Gamma(g) = (g - m(g))^2$. Finally, let $\mathcal{A}^N = \{g : m^N(g) = \overline{m}^N \wedge m(g) = \overline{m} \text{ or } m^N(g) = \underline{m}^N \wedge m(g) = \underline{m}\}$. Fix a compact set $C \subseteq \mathbb{R}$. Then, for any $g \in C$:

$$
\begin{aligned}
|\Gamma(g) - \tilde{\Gamma}^N(g)| = \Big|2g\Big(m^N(g) - m(g)\Big) + (m^2(g) - m^{N2}(g))\Big| = \\
\Big|2\Big(g - \frac{m(g) + m^N(g)}{2}\Big)\Big(m^N(g) - m(g)\Big)\Big| \leq \\
2\Big|g - \frac{m(g) + m^N(g)}{2}\Big|\Big|m^N(g) - m(g)\Big|
\end{aligned}
\tag{15}
$$

We proceed by finding uniform bounds for the expression on the right hand side of (15), and showing these bounds converge to zero almost surely.

Start with a bound for $g \in \mathcal{A}^N$. If $g \in \mathcal{A}^N$:

$$
\begin{aligned}
2\Big|g - \frac{m(g) + m^N(g)}{2}\Big|\Big|m^N(g) - m(g)\Big| \leq 2\max_{g\in C}\Big\{\Big|g - \frac{m(g) + m^N(g)}{2}\Big|\Big\}\max\{|\overline{m}^N - \overline{m}|, |\underline{m}^N - \underline{m}|\} \\
\xrightarrow{a.s.} 0
\end{aligned}
$$

The bound in the right hand side is clearly uniform in $g \in C$. The convergence result holds because the first term in the r.h.s. is bounded and, since, $\overline{m}^N \xrightarrow{a.s.} \overline{m}$ and $\underline{m}^N \xrightarrow{a.s.} \underline{m}$, the second converges almost surely.

In contrast, notice that $g \notin \mathcal{A}^N$ implies g between $\frac{\underline{m}^N + \overline{m}^N}{2}$ and $\frac{\underline{m} + \overline{m}}{2}$. Then:

$$
\begin{aligned}
2\Big|g - \frac{m(g) + m^N(g)}{2}\Big|\Big|m^N(g) - m(g)\Big| &\leq 2\max_{g\notin \mathcal{A}^N}\Big\{\Big|g - \frac{m(g) + m^N(g)}{2}\Big|\Big\}\max\{|\overline{m}^N - \underline{m}|, |\underline{m}^N - \overline{m}|\} \\
&\leq 2\max\Big\{\Big|\frac{\overline{m} - \overline{m}^N}{2}\Big|, \Big|\frac{\underline{m} - \underline{m}^N}{2}\Big|\Big\}\max\{|\overline{m}^N - \underline{m}|, |\underline{m}^N - \overline{m}|\} \\
&\xrightarrow{a.s.} 0
\end{aligned}
$$

Where the last inequality comes from the observation above and the fact that, for $g \notin \mathcal{A}^N$, $m(g) + m^N(g) \in \{\underline{m} + \overline{m}^N, \overline{m} + \underline{m}^N\}$. Now, the second term in the right hand side above is bounded and the first converges to zero almost everywhere, justifying the convergence.

We then have found bounds on (15) that are uniform in $g \in C$ and converge almost surely to zero, finishing the proof of **Step 1.2.**

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. $\mathbf{g}^* \in \arg\min_{\mathbf{g}\in\mathbb{R}} \boldsymbol{\Gamma}(\mathbf{g})$ **is the unique minimum of** Γ.

Recall that $\Gamma(g) = \max\{(g - \underline{m})^2, (g - \overline{m})^2\}$, it is clear that $g^* = \frac{\overline{m} + \underline{m}}{2}$ is the unique global minimizer of Γ.

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

For a signal realization s^N, recall that $\underline{m}^N = \min_{z^N \in P^N} \mathbb{E}[\theta|s^N, z^N]$ and, symmetrically, $\overline{m}^N = \max_{z^N \in P^N} \mathbb{E}[\theta|s^N, z^N]$. That implies $g^N \in [\underline{m}^N, \overline{m}^N]$. To see that, assume first that $g^N < \underline{m}^N$. Since, for any choice of nature $z^N \in P^N$, $\underline{m}^N \leq \mathbb{E}[\theta|s^N, z^N] \leq \overline{m}^N$, we have:

$$(g^N - \mathbb{E}[\theta|s^N, z^N])^2 + \mathbb{V}[\theta|s^N, z^N] > (\underline{m}^N - \mathbb{E}[\theta|s^N, z^N])^2 + \mathbb{V}[\theta|s^N, z^N]$$

Then, g^N is strictly dominated by $\underline{m}^N$, violating the assumption of optimality. The symmetric argument can be made for $g^N > \overline{m}^N$. Therefore, $\underline{m}^N \leq g^N \leq \overline{m}^N$.

For almost all signal realizations, we know $\underline{m}_N$ and $\overline{m}_N$ are convergent sequences, and therefore bounded. Then, for almost all signal realizations, there exists some $\overline{M}$, $\underline{M}$ such that:

$$\underline{M} \leq g^N \leq \overline{M}$$

Given steps 1, 2 and 3, Lemma 5 shows that $g^N \xrightarrow{a.s.} g^* = \frac{\overline{m}+\underline{m}}{2}$. All that is left is to prove that $\frac{\overline{m}+\underline{m}}{2} = \theta$.

Step 4. $\frac{\overline{m}+\underline{m}}{2} = \theta$. 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 $\overline{\zeta}$:[5] $\overline{m} = 2\theta - \underline{m}$, and we are done.

Proof of Proposition 6. The construction is completely the same as the proof for Proposition 3 except for that the conditional mean given sequence of actions a^N and sequence of precisions $\hat{\rho}^N$ is

$$\mathbb{E}[\theta|a^N, \hat{\rho}^N] = \frac{\sum_i \hat{\rho}_i s_i + \rho_\mu \mu + c}{\sum_i \hat{\rho}_i + \rho_\mu}$$

[5] See the Proof of Proposition 4.

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

Proof of Proposition 7. Given that H is the empirical distribution of actions, the posterior mean given the sequence of actions a and sequence of precisions $\hat{\rho}$ is

$$\mathbb{E}[\theta|a,\hat{\rho}] = \int \frac{\hat{\rho}(x)x + c}{\int \hat{\rho}(x)dH(x)} dH(x)$$

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

First, we want to show that $\underline{m}_a = \min_{\hat{\rho}:\mathbb{R}\to[\underline{\rho},\overline{\rho}]} \mathbb{E}[\theta|a,\hat{\rho}]$ and $\overline{m} = \max_{\hat{\rho}:\mathbb{R}\to[\underline{\rho},\overline{\rho}]} \mathbb{E}[\theta|a,\hat{\rho}]$. We focus on the maximization problem for the illustration. In the proof for Proposition 4, we have shown that there exists a threshold solution for maximizing $\int \frac{\hat{\rho}(x)x}{\int \hat{\rho}(x)dH(x)} dH(x)$, where $\hat{\rho}(x) = \underline{\rho}$ under the threshold and $\hat{\rho}(x) = \overline{\rho}$ equal to or above the threshold. The additional term $\int \frac{c}{\int \rho(x)dH(x)} dH(x)$ is affected only by the total mass $\int \rho(x)dH(x)$. Hence, the maximizer still adopts a threshold structure. The rest of the proof for Proposition 4 follows and we are done.

Proof of Proposition 8. The proof will be exactly the same as the proof for Proposition 5 after we adjust the definitions of $\overline{m}_a^N$ and $\underline{m}_a^N$.

Proof of Proposition 9 .

Recall that the bounds of the limiting posterior set are given by

$$\overline{m}_a = \frac{\underline{\rho}\int_{-\infty}^{\overline{m}_a} xdH(x) + \bar{\rho}\int_{\overline{m}_a}^{\infty} xdH(x) + c}{\underline{\rho}H(\overline{m}_a) + \bar{\rho}(1 - H(\overline{m}_a))}, \quad \underline{m}_a = \frac{\overline{\rho}\int_{-\infty}^{\underline{m}_a} xdH(x) + \underline{\rho}\int_{\underline{m}_a}^{\infty} xdH(x) + c}{\overline{\rho}H(\underline{m}_a) + \underline{\rho}(1 - H(\underline{m}_a))} \tag{16}$$

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

The optimal guess is $m_a = \frac{\overline{m}_a + \underline{m}_a}{2}$. When $\theta = \mu$, $c = 0$ and by Proposition 5, $m_a = \theta = \mu$ and the observer guesses correctly. From now on, we first focus on the case where $\theta > \mu$.

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

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

This implies

$$\underline{\rho}\int_{-\infty}^{\overline{m}_a} H(x)dx - \overline{\rho}\int_{\overline{m}_a}^{\infty}(1-H(x))\,dx = c. \tag{17}$$

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

$$\overline{\rho}\int_{-\infty}^{\underline{m}_a} H(x)dx - \underline{\rho}\int_{\underline{m}_a}^{\infty}(1-H(x))\,dx = c. \tag{18}$$

Recall that H is normally distributed and denote its density function as h. (This two equations are derived after the derivative of $\overline{m}_a$ with respect to θ. Also, there is a negative sign before $\frac{\rho_\mu}{\rho_\mu+\rho}$.)

$$\frac{dH(\overline{m}_a)}{d\theta} = \frac{\partial H(\overline{m}_a)}{\partial \overline{m}_a}\frac{d\overline{m}_a}{d\theta} + \frac{\partial H(\overline{m}_a)}{\partial\theta} = \frac{\rho_\mu}{\rho_\mu+\rho}\frac{\rho}{\overline{G}(\overline{m}_a)}h(\overline{m}_a)$$
$$\frac{dH(\underline{m}_a)}{d\theta} = \frac{\partial H(\underline{m}_a)}{\partial \underline{m}_a}\frac{d\underline{m}_a}{d\theta} + \frac{\partial H(\underline{m}_a)}{\partial\theta} = \frac{\rho_\mu}{\rho_\mu+\rho}\frac{\rho}{\underline{G}(\underline{m}_a)}h(\underline{m}_a)$$

Taking the derivative withe respect to the state θ on both sides of equation 17 and equation 18, we get

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

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

$$\frac{dm_a}{d\theta} = \frac{\rho}{\rho_\mu+\rho} + \frac{\rho_\mu}{\rho_\mu+\rho}\frac{\rho}{2}\left(\frac{1}{\overline{G}(\overline{m}_a)} + \frac{1}{\underline{G}(\underline{m}_a)}\right) \tag{19}$$

It then follows

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

Lemma 6. $\left(\frac{1}{\overline{G}^2(\overline{m}_a)} - \frac{1}{\underline{G}^2(\underline{m}_a)}\right) > 0$ *whenever* $\theta > \mu$

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

Define

$$\overline{\zeta}(z,u) = \frac{\underline{\rho}\int_{-\infty}^{z} x dH(x) + \bar{\rho}\int_{z}^{\infty} x dH(x) + u}{\underline{\rho}H(z) + \bar{\rho}(1 - H(z))}, \ \underline{\zeta}(z,u) = \frac{\overline{\rho}\int_{-\infty}^{z} x dH(x) + \underline{\rho}\int_{z}^{\infty} x dH(x) + u}{\overline{\rho}H(z) + \underline{\rho}(1 - H(z))} \tag{20}$$

We know $\overline{m}_a = \overline{\zeta}(\overline{m}_a, c)$, and it was previously proved that $\overline{m}_a$ maximizes $\overline{\zeta}(\overline{m}_a, c)$.

By the envelope theorem we have:

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

A similar argument implies that $\frac{\underline{\zeta}(\underline{m}_a, u)}{du} > 0$, for all $u \in \mathbb{R}$. Finally, by an equivalent argument to Step 4 in the proof of prop: signals guess characterization, we have $\frac{\overline{\zeta}(\overline{m}_a,0)+\underline{\zeta}(\underline{m}_a,0)}{2} = \int x dH = \frac{\rho\theta+\rho_\mu\mu}{\rho+\rho_m u}$.

We then have, if $\theta > \mu$ - which implies $c > 0$:

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

.

This finishes the proof.

Therefore,

$$\left(\frac{h(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{h(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) \geq 0 \implies \frac{d^2 m_a}{d\theta^2} > 0. \tag{21}$$

Then we want to consider the partial derivative of the optimal guess with respect to ρ. We start with an alternative implicit function of $\overline{m}_a$ and $\underline{m}_a$. Notice that if f as the density function of a normal distribution with mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$, then $\frac{\partial f(x)}{\partial x} = -\frac{x-\tilde{\mu}}{\tilde{\sigma}^2} f(x)$. This implies $xf(x) = \tilde{\mu}f(x) - \tilde{\sigma}^2 \frac{\partial f(x)}{\partial x}$. Plugging this into the initial implicit functions 16, we get

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

By definition of m_a, we have

$$m_a = \theta + (\theta - \mu)\left(\frac{dm_a}{d\theta} - 1\right) + \frac{(\overline{\rho} - \underline{\rho})\rho}{2(\rho_\mu + \rho)^2}\left(\frac{h(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{h(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right). \tag{22}$$

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

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

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

$$\theta > \mu \quad \text{and} \quad m_a \le \theta \quad \Longrightarrow \quad \frac{dm_a}{d\rho} > 0. \tag{23}$$

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

$$\frac{d^2m_a}{d\rho\, d\theta} = \rho_\mu\kappa_1\frac{dm_a}{d\theta} - \rho\kappa_1\left(\frac{dm_a}{d\theta}-1\right) + \frac{\rho\rho_\mu}{\rho_\mu+\rho}\kappa_2\left(\frac{dm_a}{d\theta}-\frac{\rho}{\rho_\mu+\rho}\right) + c\kappa_2\frac{d^2m_a}{d\theta^2} \tag{24}$$

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

$$\theta = \mu \quad \text{and} \quad \frac{dm_a}{d\theta} \le 1 \quad \Longrightarrow \quad \frac{d^2m_a}{d\rho\, d\theta} > 0. \tag{25}$$

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

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

Lemma 7. *For any given ρ, if $m_a(\rho,\hat{\theta}) > \hat{\theta}$ and $\frac{dm_a}{d\theta}(\rho,\hat{\theta}) > 1$, then $m_a(\rho,\theta) > \theta$ for all $\theta > \hat{\theta}$.*

Proof of Lemma 7. Fix ρ. Assume that there exists $\hat{\theta}$, $m_a(\rho,\hat{\theta}) > \hat{\theta}$ and $\frac{dm_a}{d\theta}(\rho,\hat{\theta}) > 1$. Suppose by contradiction that there exists some $\overline{\theta} > \hat{\theta}$ such that $m_a(\rho,\overline{\theta}) = \overline{\theta}$. By continuity of $\frac{dm_a}{d\theta}$, there exists $\theta' < \theta'' \in (\hat{\theta},\overline{\theta}]$ where $\frac{dm_a}{d\theta}(\rho,\theta') = 1$ and $\frac{dm_a}{d\theta}(\rho,\theta'') < 1$. By continuity of m_a, $m_a(\rho,\theta') > \theta'$.

At θ', equation (22) implies $\left(\frac{h(\overline{m}_a)}{\overline{G}(\overline{m}_a)}-\frac{h(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) > 0$, which guarantees $\frac{d^2m_a}{d\theta^2}(\rho,\theta') > 0$. This implies that for a neighborhood to the right of θ', $\frac{dm_a}{d\theta} > 1$. Notice that this holds for any $\theta \in [\hat{\theta},\overline{\theta}]$ with $\frac{dm_a}{d\theta}(\rho,\theta) = 1$. Thus $\frac{dm_a}{d\theta}(\rho,\theta) \ge 1$ for all $\theta \in [\hat{\theta},\overline{\theta}]$, which contradicts the assumption that $m_a(\rho,\overline{\theta}) = \overline{\theta}$. As a result, we know $m_a(\rho,\theta) > \theta$ for $\theta > \hat{\theta}$.

Lemma 8. *For any given ρ, if there exists $\theta^* > \mu$ such that $m_a(\rho,\theta^*) = \theta^*$ and $m_a(\rho,\theta) < \theta$ for all $\mu < \theta < \theta^*$, then $m_a(\rho,\theta) > \theta$ for $\theta > \theta^*$.*

Proof of Lemma 8. Suppose there exists $\theta^* > \mu$ such that $m_a(\rho,\theta^*) = \theta^*$ and $m_a(\rho,\theta) < \theta$ for $\mu < \theta < \theta^*$. This implies $\frac{dm_a}{d\theta}(\rho,\theta^*) \ge 1$. Again by equation (22), we know $\left(\frac{h(\overline{m}_a)}{\overline{G}(\overline{m})}-\frac{h(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) > 0$, which leads to

$\frac{d^2 m_a}{d\theta^2}(\rho, \theta^*) > 0$ by (21). Then for any θ in a small neighborhood to the right of θ^*, $\frac{dm_a}{d\theta}(\rho, \theta) > 1$ and $m_a(\rho, \theta) > \theta$. By Lemma 7, we are done.

Lemma 9. *$\tilde{\tilde{\rho}}$ is well-defined.*

Proof of Lemma 9. Recall that $\tilde{\tilde{\rho}} = \underline{\rho} H(\overline{m}(\tilde{\tilde{\rho}}, \mu)) + \overline{\rho}(1 - H(\overline{m}(\tilde{\tilde{\rho}}, \mu))$ is equivalent to the statement that at $\rho = \tilde{\tilde{\rho}}$ and $\theta = \mu$, $\frac{dm_a}{d\theta}(\tilde{\tilde{\rho}}, \mu) = 1$. Equation (19) implies that $\frac{dm_a}{d\theta}(\underline{\rho}, \mu) < 1$ and $\frac{dm_a}{d\theta}(\overline{\rho}, \mu) > 1$. Also, by equation (24), whenever $\frac{dm_a}{d\theta}(\rho, \mu) < 1$, we have $\frac{d^2 m_a}{d\rho d\theta}(\rho, \mu) > 0$ and thus $\frac{dm_a}{d\theta}(\rho, \mu)$ is strictly increasing in ρ. Then we can find some $\tilde{\tilde{\rho}}$ such that $\frac{dm_a}{d\theta}(\tilde{\tilde{\rho}}, \mu) = 1$. We then show this is the unique ρ satisfying $\frac{dm_a}{d\theta}(\rho, \mu) = 1$. Suppose not, without loss of generality, assume $\tilde{\tilde{\rho}}$ is the smallest one among them. By continuity of $\frac{d^2 m_a}{d\rho d\theta}$, $\frac{dm_a}{d\theta}$ and the fact that ρ is bounded, there exists $k > 1$ such that for any ρ with $\frac{dm_a}{d\theta}(\rho, \mu) \leq k$, we have $\frac{d^2 m_a}{d\rho d\theta}(\rho, \mu) > 0$. Thus, for all $\rho > \tilde{\tilde{\rho}}$, $\frac{dm_a}{d\theta}(\rho, \mu) > 1$, which leads to a contradiction.

Now we show the results for ρ different from $\tilde{\rho}$ and $\tilde{\tilde{\rho}}$.

Lemma 10. *(i) If $\rho < \tilde{\rho}$, $m_a(\rho, \theta) < \theta$ for all $\theta > \mu$;*

(ii) If $\rho > \tilde{\tilde{\rho}}$, $m_a(\rho, \theta) > \theta$ for all $\theta > \mu$;

(iii) If $\tilde{\rho} < \rho < \tilde{\tilde{\rho}}$, there exists $\theta^* > \mu$ such that, $m_a(\rho, \theta^*) = \theta^*$, $m_a(\rho, \theta) > \theta$ for $\theta > \theta^*$ and $m(\rho, \theta) < \theta$ for $\mu < \theta < \theta^*$.

Proof of Lemma 10. For $\rho > \tilde{\tilde{\rho}}$, by Lemma 9, we know $\frac{dm_a}{d\theta}(\rho, \mu) > 1$. This implies that for any $\theta > \mu$ that is sufficiently close to μ, $m_a(\rho, \theta) > \theta$ and $\frac{dm_a}{d\theta}(\rho, \theta) > 1$. By Lemma 7, we know $m_a(\rho, \theta) > \theta$ for all $\theta > \mu$. This finishes the proof for (ii).

We claim that $\tilde{\tilde{\rho}} \geq \tilde{\rho}$. Otherwise, for $\tilde{\tilde{\rho}} < \rho < \tilde{\rho}$, we know $m_a(\rho, \theta) > \theta$ for all $\theta > \mu$. However, as $\frac{dm_a}{d\theta}(\rho, \theta)$ converges to some number strictly smaller than 1 when θ goes to $+\infty$, we know when θ is large enough, $m_a(\rho, \theta) < \theta$. Contradiction!

Recall in Lemma 9, we showed that $\frac{dm_a}{d\theta}(\rho, \mu)$ is strictly increasing in ρ for $\rho \geq \underline{\rho}$ and ρ close to $\underline{\rho}$ until it reaches 1 at $\rho = \tilde{\tilde{\rho}}$. That is, for all $\rho < \tilde{\tilde{\rho}}$, $\frac{dm_a}{d\theta}(\rho, \mu) < 1$. As a result, for $\rho < \tilde{\tilde{\rho}}$, m_a will be strictly smaller than θ when θ is sufficiently close to μ.

For $\tilde{\rho} < \rho < \tilde{\tilde{\rho}}$, we know as θ goes to $+\infty$, $\frac{dm_a}{d\theta}(\rho, \theta)$ converges to some number strictly larger than 1. This implies that when θ is large enough, $m_a(\rho, \theta) > \theta$. Thus, there must be a $\theta^* > \mu$ such that $m_a(\rho, \theta^*) = \theta^*$ and $m_a(\rho, \theta) < \theta$ for $\mu < \theta < \theta^*$. By Lemma 8, we know $m_a(\rho, \theta) > \theta$ for $\theta > \theta^*$. Thus (iii) holds.

Finally, for $\rho < \tilde{\rho}$, suppose that there exists $\theta^* > \mu$ such that $m_a(\rho, \theta^*) = \theta^*$ and $m_a(\rho, \theta) < \theta$ for $\mu < \theta < \theta^*$. By Lemma 8, we know $m_a(\rho, \theta) > \theta$ for $\theta > \theta^*$. However, as $\frac{dm_a}{d\theta}(\rho, \theta)$ converges to some number strictly smaller than 1 when θ goes to $+\infty$, we know when θ is large enough, $m_a(\rho, \theta) < \theta$. This leads to a contradiction and hence (i) holds.

We complete the proof of the proposition by considering the thresholds $\rho = \tilde{\rho}$ and $\rho = \tilde{\tilde{\rho}}$.

Lemma 11. *$m(\tilde{\rho},\theta) < \theta$ and $m(\tilde{\tilde{\rho}},\theta) > \theta$ for all $\theta > \mu$.*

Proof of Lemma 11. By the proof of Lemma 10, we know $\tilde{\tilde{\rho}} \geq \tilde{\rho}$. We first show that $\tilde{\rho} \neq \tilde{\tilde{\rho}}$. Suppose by contradiction that $\tilde{\rho} = \tilde{\tilde{\rho}}$. By definition, at $\rho = \tilde{\rho}$, $\frac{dm_a}{d\theta}(\tilde{\rho},\mu) = 1$, $m_a(\tilde{\rho},\mu) = \mu$ and $\frac{d^2 m_a}{d\theta^2}(\tilde{\rho},\mu) = 0$. Consider a neighborhood to the right of μ as $(\mu, \mu+\epsilon)$ where $\epsilon > 0$ is small enough. There are three possible cases:

- Case (a): Suppose $m_a(\tilde{\rho},\theta) = \theta$ for all $\theta \in (\mu, \mu+\epsilon)$. Then $\frac{dm_a}{d\theta}(\tilde{\rho},\theta) = 1$ and $\frac{d^2 m_a}{d\theta^2}(\tilde{\rho},\theta) = 0$ for all $\theta \in (\mu, \mu+\epsilon)$. By equation (22) and $m_a(\tilde{\rho},\theta) = \theta$ for all $\theta \in (\mu, \mu+\epsilon)$, $\left(\frac{h(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{h(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) = 0$, which guarantees $\frac{d^2 m_a}{d\theta^2}(\tilde{\rho},\theta) > 0$ by (21), which is a contradiction.

- Case (b): Suppose $m_a(\tilde{\tilde{\rho}},\theta^*) > \theta^*$ for some $\theta^* \in (\mu, \mu+\epsilon)$. By continuity of m with respect to ρ, we know that there exists $\rho' < \tilde{\tilde{\rho}} = \tilde{\rho}$ such that $m_a(\rho',\theta^*) > \theta^*$, which contradicts with part (i) of Lemma 10.

- Case (c): Suppose $m_a(\tilde{\rho},\theta^*) < \theta^*$ for some $\theta^* \in (\mu, \mu+\epsilon)$. By continuity of m with respect to ρ, we know that there exists $\rho' > \tilde{\tilde{\rho}} = \tilde{\rho}$ such that $m_a(\rho',\theta^*) < \theta^*$, which contradicts with part (ii) of Lemma 10.

Now we have shown that $\tilde{\tilde{\rho}} > \tilde{\rho}$. If $\rho = \tilde{\tilde{\rho}}$, by the above case (c), we can find some $\theta^* > \mu$ with $m_a(\tilde{\tilde{\rho}},\theta^*) > \theta^*$, $\frac{dm_a}{d\theta}(\tilde{\tilde{\rho}},\theta^*) > 1$ and $m_a(\tilde{\tilde{\rho}},\theta) > \theta$ for $\theta \in (\mu,\theta^*)$. By Lemma 10, we know $m_a(\tilde{\tilde{\rho}},\theta) > \theta$ for all $\theta > \mu$.

If instead $\rho = \tilde{\rho}$, then $\frac{dm_a}{d\theta}(\tilde{\rho},\mu) < 1$ and for θ in a small neighborhood to the right of μ, $m_a(\tilde{\rho},\theta) < \theta$. Suppose by contradiction that $m_a(\tilde{\rho},\theta) \geq \theta$ for some $\theta > \mu$. By continuity, denote θ^1 as the smallest one such that $m_a(\tilde{\rho},\theta^1) = \theta^1$. By Lemma 8, we know $m_a(\tilde{\rho},\theta) > \theta$ for all $\theta > \theta^1$. Fix some $\theta^2 > \theta^1$. As m_a is continuous with respect to ρ, we can find $\rho' < \tilde{\rho}$ such that $m_a(\rho',\theta^2) > \theta^2$, which contradicts with part (i) of Lemma 10. Thus $m_a(\tilde{\rho},\theta) < \theta$ for all $\theta > \mu$. This completes the proof.

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.

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.

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.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with a non-unique prior.

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*.

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

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

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

Shishkin, D. and Ortoleva, P. (2019). Ambiguous information and dilation: An experiment. Technical report, Tech. rep., Mimeo, Princeton University.