Online Appendix for

"Asymptotic Learning with Ambiguous Information"

Pëllumb Reshidi [*] João Thereze [†] Mu Zhang [‡]

January 2022

Contents

[*]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 Learning with Ambiguous Variance and Bias

In this section we consider an extension of our model that incorporates ambiguity about the signals' bias. We show that this model delivers implications that are similar to the model with unbiased signals. Then, we argue that the model in the main text is the most natural and parsimonious departure from the Bayesian learning model.

For simplicity, we focus on the case where signals are directly observable: i.e. $\boldsymbol{a}(s_i, \rho_i) = s_i$. Concretely, we assume that signals satisfy $s_i = \theta + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}\left(\delta_i, \frac{1}{\rho_i}\right)$. The decision maker does not know the bias or the precision of the shock ε_i, but rather considers a perceived bias interval $[-\delta, \delta]$, with $\delta > 0$, in addition to the perceived precision interval $[\underline{\rho}, \overline{\rho}]$. The rest of the problem is as described in the main text, except that now nature also chooses the bias of each signal so as to maximize the decision maker's loss. By following the same steps of Theorem 1, we can characterize the limit set of posteriors in this model.

Remark OA 1. *With ambiguity about bias and variance, the limit set of posteriors satisfies, for almost all realizatons s:*

$$\mathbb{P}^{\infty}(s) = \{\delta_b : \underline{m} - \delta \leq b \leq \overline{m} + \delta\}$$

where $\underline{m}$, $\overline{m}$ are as in Theorem 1.

In words, the addition of bias simply increases the interval of posterior means that the decision maker asymptotically considers. Again, for each allocation of bias and precisions to signals, the variance of the decision maker's beliefs goes to zero as the number of information sources grows. We know that, by using unbiased signals, nature can obtain any posterior mean between the bounds $\underline{m}$ and $\overline{m}$, so the interval $[\underline{m}, \overline{m}]$ remains achievable. Because nature's objective is to make the interval of posterior means as large as possible, the optimal strategy is simple. For example, for obtaining the highest possible posterior means, she chooses the assignment of precisions that generates $\overline{m}$ and, on top of that, makes all signals biased downwards. By the symmetric argument, we obtain the lower bound.

Note that, in the absence of ambiguity, a Bayesian agent facing only biased signals may fail to learn correctly. Consider, for example, the case in which all signals have the same bias. The Bayesian agent knows the variance of each signal, and that this bias is drawn from some nondegenerate distribution. In this case, the model is unidentified, as the agent does not know the realized value of the bias and even infinite signals do not let her learn this value. Thus, adding ambiguity in this case convolutes it's effect with that of unidentifiability. This is in contrast with the case of a Bayesian who knows the signals are unbiased but is unsure about the precision of the signals. In this case, even a misspecified Bayesian learns correctly. Thus, after adding ambiguity, any departure from perfect learning is due to this addition. This motivates us to focus our main analysis on the case with ambiguity only on signal precisions.

2 Learning from Finite Signals

2.1 One Signal Source

In this section we study the problem of an observer who has access to exactly one information source. The assumption is that signals are directly observable, that is, $\boldsymbol{a}(s_i, \rho_i) = s_i$, and the loss function is mean-squared errors. As in the main text, $s = \theta + \varepsilon$, 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:

$$\min_{g} \max_{\hat{\rho} \in [\underline{\rho}, \overline{\rho}]} \left\{ (g - \mathbb{E}[\theta | s, \hat{\rho}])^2 + \mathbb{V}[\theta | s, \hat{\rho}] \right\} \tag{1}$$

For a fixed precision $\hat{\rho}$, $\mathbb{E}[\theta|s,\hat{\rho}] = \frac{\hat{\rho}s + \rho_\mu \mu}{\hat{\rho} + \rho_\mu}$ 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).

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

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

Proposition OA 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 $\overline{z}$, 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.

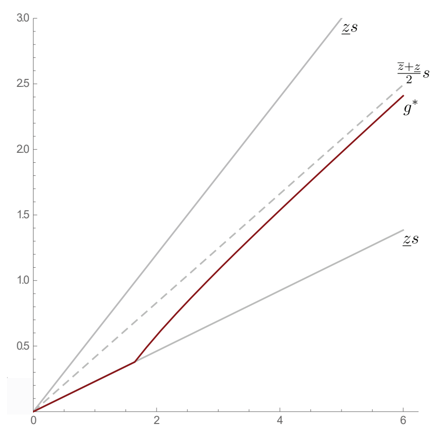
When the signal is near the prior mean Nature's ability to increase the bias is quite limited. Intuitively, Nature manipulates the weights that compose the posterior mean, which, regardless of the chosen weights, must lie between μ and s. As a consequence, the guess of the agent must also be located between these two values, implying that the square bias that Nature can generate

[1]The proof of Proposition OA 1 is subsumed in the proof of Proposition OA 2.

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 OA 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 $\frac{\underline{z}+\bar{z}}{2}$.[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 $\overline{\rho}$.

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 subsection, this intuition continues to hold when the observer has access to $N > 1$ information sources.

2.2 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:

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

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

Under the game against Nature interpretation, as in the previous section, Nature assigns precisions to signals so as to tamper with the posterior mean and variance of the observer. Also as in

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

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. Lemma OA 1 below shows that attention can be restricted to the subset of precision assignments that is either order-reversing or order-preserving, in the spirit of Lemma 1 in the main text.

Lemma OA 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 either order-preserving or order-reversing.*[3]

Proof of Lemma OA 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 5, $\mathbb{V}[\theta | s^N, \hat{\rho}^N]$ depends only on $\#\hat{\rho}^N$. Then, for any $d \in \{1, ..., N\}$:

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

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 1 in the main text. The problem of Nature is then:

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

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

□

The next result, Proposition OA 2, generalizes the optimal guess for 1 signal to a multisignal case. Just as in Proposition OA 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,

[3] For the definition of order-preserving and order-reversing look at Definition 1 in the main text.

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 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.

Again, the results are expressed in terms of relative precisions. The proof of Proposition OA 2 is at the end of this section. 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 OA 2. *When the observer has access to N information sources, the optimal guess $g^* : \mathbb{R}^N \to \mathbb{R}$ satisfies:*

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

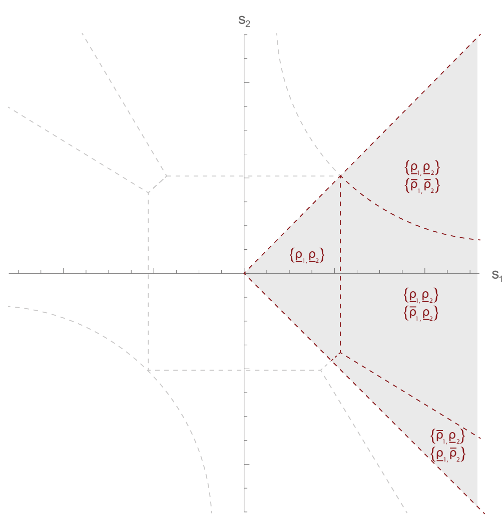
where $K^N = \{x \in \mathbb{R}^N : |\frac{z^N - \underline{z}^N}{\sqrt{(z^N - \underline{z}^N)\cdot \mathbb{1}}} \cdot x| \frac{1}{\rho_\mu} \leq 1 \ \forall z^N \in Z^N\}$ and z'^N, z''^N are functions of s^N and are order-preserving and order-reversing, respectively.

Proposition OA 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.

Proof of Proposition OA 2

Recall we want to solve the problem:

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

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:

$$\mathbb{E}[\theta | s^N, z^N] = \mu + z^N \cdot (s^N - \mu) \tag{4}$$

and

$$\mathbb{V}[\theta | s^N, z^N] = (1 - z^N \cdot \mathbb{1}^N) \frac{1}{\rho_\mu} \tag{5}$$

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 4 and 5, 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 OA 2. *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 OA 1 shows 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 OA 2 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:

$$\begin{aligned} &\min_g \max_{z^N \in P} \{\left(g - \mathbb{E}[\theta|s^N, z^N]\right)^2 + \mathbb{V}[\theta|s^N, z^N]\} \geq \\ &\max_{z^N \in P} \min_g \{\left(g - \mathbb{E}[\theta|s^N, z^N]\right)^2 + \mathbb{V}[\theta|s^N, z^N]\} = \\ &\max_{z^N \in P} \mathbb{V}[\theta|s^N, z^N] = \mathbb{V}[\theta|s^N, \underline{z}^N] \end{aligned} \tag{6}$$

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 4 and 5 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 4.

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 nonunique. 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. Lemma OA 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 and 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 this assignment takes the same number of high (and low) precisions as $\hat{\rho}''^N$ — that is, $\#\tilde{\rho}^N = \#\hat{\rho}''^N$ —, 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. □

3 Extensions: 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.

3.1 $\hat{\rho}_1, \hat{\rho}_2 \in [\underline{\rho}, \overline{\rho}]$ 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, $F(\overline{a}_\alpha) = 1-\alpha, F(\underline{a}_\alpha) = \alpha$ 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

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

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

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

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

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

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

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

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

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

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.

4 Remaining Proofs

Proof of Proposition 3

Recall from **Proposition 2** that the bounds of the limiting posterior set are given by

$$\overline{m}_a = \frac{\underline{\rho}\int_{-\infty}^{\overline{m}_a} x dF(x) + \bar{\rho}\int_{\overline{m}_a}^{\infty} x dF(x) + c}{\underline{\rho}F(\overline{m}_a) + \bar{\rho}(1 - F(\overline{m}_a))}, \qquad \underline{m}_a = \frac{\overline{\rho}\int_{-\infty}^{\underline{m}_a} x dF(x) + \underline{\rho}\int_{\underline{m}_a}^{\infty} x dF(x) + c}{\overline{\rho}F(\underline{m}_a) + \underline{\rho}(1 - F(\underline{m}_a))} \tag{7}$$

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 **Corollary 1** $m_a = \theta = \mu$ and the observer guesses correctly. From now on, we first focus on the case where $\theta > \mu$.

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

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

This implies

$$\underline{\rho}\int_{-\infty}^{\overline{m}_a} F(x)dx - \overline{\rho}\int_{\overline{m}_a}^{\infty}(1-F(x))\,dx = c. \tag{8}$$

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

$$\overline{\rho}\int_{-\infty}^{\underline{m}_a} F(x)dx - \underline{\rho}\int_{\underline{m}_a}^{\infty}(1-F(x))\,dx = c. \tag{9}$$

Taking the derivative with respect to the state θ on both sides of equation 8 and equation 9, 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 derivative of the optimal guess $m_a = \frac{\overline{m}_a+\underline{m}_a}{2}$ with respect to θ is then:

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

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

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

It then follows that

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

Lemma OA 3. $\left(\frac{1}{\overline{G}^2(\overline{m}_a)} - \frac{1}{\underline{G}^2(\underline{m}_a)}\right) > 0$ *whenever* $\theta > \mu$

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

$$\overline{\zeta}(z,u) = \frac{\underline{\rho}\int_{-\infty}^{z} xdF(x) + \bar{\rho}\int_{z}^{\infty} xdF(x) + u}{\underline{\rho}F(z) + \bar{\rho}(1-F(z))}, \; \underline{\zeta}(z,u) = \frac{\overline{\rho}\int_{-\infty}^{z} xdF(x) + \underline{\rho}\int_{z}^{\infty} xdF(x) + u}{\overline{\rho}F(z) + \underline{\rho}(1-F(z))} \tag{11}$$

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

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

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

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

This conclusion the proof of the lemma. □

Therefore,

$$\left(\frac{f(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) \geq 0 \implies \frac{d^2 m_a}{d\theta^2} > 0. \tag{12}$$

We next consider the partial derivative of the optimal guess with respect to ρ. We start with an alternative implicit function of $\overline{m}_a$ and $\underline{m}_a$. Notice that if f as the density function of a normal distribution with mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$, then $\frac{\partial f(x)}{\partial x} = -\frac{x-\tilde{\mu}}{\tilde{\sigma}^2} f(x)$. This implies $xf(x) = \tilde{\mu}f(x) - \tilde{\sigma}^2 \frac{\partial f(x)}{\partial x}$. Plugging this into the initial implicit functions 7, we get

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

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

By definition of m_a, we have

$$m_a = \theta + (\theta - \mu)\left(\frac{dm_a}{d\theta} - 1\right) + \frac{(\overline{\rho} - \underline{\rho})\rho}{2(\rho_\mu + \rho)^2}\left(\frac{f(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right). \tag{13}$$

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.

$$\theta > \mu \quad \text{and} \quad m_a \leq \theta \quad \Longrightarrow \quad \frac{dm_a}{d\rho} > 0. \tag{14}$$

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:

$$\frac{d^2m_a}{d\rho\, d\theta} = \rho_\mu\kappa_1\frac{dm_a}{d\theta} - \rho\kappa_1\left(\frac{dm_a}{d\theta} - 1\right) + \frac{\rho\rho_\mu}{\rho_\mu + \rho}\kappa_2\left(\frac{dm_a}{d\theta} - \frac{\rho}{\rho_\mu + \rho}\right) + c\kappa_2\frac{d^2m_a}{d\theta^2} \tag{15}$$

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:

$$\theta = \mu \quad \text{and} \quad \frac{dm_a}{d\theta} \leq 1 \quad \Longrightarrow \quad \frac{d^2m_a}{d\rho\, d\theta} > 0. \tag{16}$$

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

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

Lemma OA 4. *For any given ρ, if $m_a(\rho,\hat{\theta}) > \hat{\theta}$ and $\frac{dm_a}{d\theta}(\rho,\hat{\theta}) > 1$, then $m_a(\rho,\theta) > \theta$ for all $\theta > \hat{\theta}$.*

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

At θ', equation (13) implies $\left(\frac{f(\overline{m}_a)}{\overline{G}(\overline{m}_a)} - \frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) > 0$, which guarantees $\frac{d^2 m_a}{d\theta^2}(\rho,\theta') > 0$. This implies that for a neighborhood to the right of θ', $\frac{dm_a}{d\theta} > 1$. Notice that this holds for any $\theta \in [\hat{\theta},\overline{\theta}]$ with $\frac{dm_a}{d\theta}(\rho,\theta) = 1$. Thus $\frac{dm_a}{d\theta}(\rho,\theta) \geq 1$ for all $\theta \in [\hat{\theta},\overline{\theta}]$, which contradicts the assumption that $m_a(\rho,\overline{\theta}) = \overline{\theta}$. As a result, we know $m_a(\rho,\theta) > \theta$ for $\theta > \hat{\theta}$. This concludes the proof of the lemma.

□

Lemma OA 5. *For any given ρ, if there exists $\theta^* > \mu$ such that $m_a(\rho,\theta^*) = \theta^*$ and $m_a(\rho,\theta) < \theta$ for all $\mu < \theta < \theta^*$, then $m_a(\rho,\theta) > \theta$ for $\theta > \theta^*$.*

Proof. Suppose there exists $\theta^* > \mu$ such that $m_a(\rho,\theta^*) = \theta^*$ and $m_a(\rho,\theta) < \theta$ for $\mu < \theta < \theta^*$. This implies $\frac{dm_a}{d\theta}(\rho,\theta^*) \geq 1$. Again by equation (13), we know $\left(\frac{f(\overline{m}_a)}{\overline{G}(\overline{m})} - \frac{f(\underline{m}_a)}{\underline{G}(\underline{m}_a)}\right) > 0$, which leads to $\frac{d^2 m_a}{d\theta^2}(\rho,\theta^*) > 0$ by (12). 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 OA 4. This concludes the proof of the lemma and the proposition.

□