Distrust in Experts and the Origins of Disagreement

Alice Hsiaw, International Business School, Brandeis University

Ing-Haw Cheng, Tuck School of Business, Dartmouth University

Working Paper Series

2016 | 110R

Distrust in Experts and the Origins of Disagreement[*]

Ing-Haw Cheng[†] Alice Hsiaw[‡]

First draft: October 2016
This draft: November 2016

Abstract

Disagreements about substance and expert credibility often go hand-in-hand and are hard to resolve, even when people share common information, on a wide range of issues ranging from economics, climate change, to medicine. We argue that a learning bias helps explain disagreement in environments such as these where both the state of the world and the credibility of information sources (experts) are uncertain. Individuals with our learning bias overinterpret how much they can learn about two sources of uncertainty from one signal, leading them to over-infer expert quality. People who encounter information or experts in different order disagree about substance because they endogenously disagree about the credibility of each others' experts. Disagreement persists because first impressions about experts have long-lived influences on beliefs about the state. These effects arise even though agents share common priors, information, and biases, providing a theory for the origins of disagreement.

[*] The authors thank Botond Kőszegi, Dustin Tingley, Davide Pettenuzzo, Uday Rajan, Wei Xiong, and seminar participants at Brandeis University and Harvard University for comments.

[†] Dartmouth College, Tuck School of Business. Email: ing-haw.cheng@tuck.dartmouth.edu.

[‡] Brandeis University, International Business School. Email: ahsiaw@brandeis.edu.

Disagreement appears everywhere. In topics ranging from the effects of climate change to the consequences of immigration, people square off against each other with very strong opinions, even though people increasingly share the same information. Instead, the disagreement is often not just about substance ("Do humans affect climate change?"), but also about the credibility of different sources of information ("How much faith should we put in the scientists and their data?"). One side typically expresses supreme confidence in its preferred experts while dismissing the other side's sources.

A core feature of many of these situations is that individuals are uncertain about both about the state of the world and the credibility of the information sources who provide signals about the state. For example, a policymaker is called on to make a decision ("should I support a climate agreement?") when he must evaluate source quality ("how reliable is the scientist and her information?") as well as the given claim ("humans cause climate change").

There are several reasons why information about expert quality is often limited. First, individuals usually do not have the skills to collect and evaluate primary-source evidence themselves, making experts necessary. Second, independent signals of expert credibility may be unavailable, or even if they were available, are difficult to evaluate themselves. An individual might know that an expert has a Ph.D., but might know little about the granting school's quality or how well a Ph.D. qualifies someone to make certain statements. Finally, it is often difficult to independently evaluate expert quality through repeated tests - climate change is slow-moving and climate patterns are highly variable.

Our thesis is that disagreement about substance fundamentally reflects disagreement about expert quality, due to a learning bias that arises when there is uncertainty about expert quality. In our framework, agents overinterpret how much they can learn about two unknowns from one source, leading to distorted beliefs about quality which feed through to distorted beliefs about substance. This bias leads to disagreement even when agents share common priors, information, and biases. While the literature has pointed out that heterogeneity in any of the above may generate disagreement (e.g., Scheinkman and Xiong, 2003; Acemoglu, Chernozhukov and Yildiz, 2016), the source of heterogeneity is less clear. We argue that biased learning about expert quality provides a theory for the *origins* of disagreement.

Section 1 introduces the information environment and learning. Our baseline model features an agent who learns about a binary state (A or B) using signals from an expert who is of either high or low quality. A high quality expert is more likely to report the true state than a low quality expert. To focus on the effect of learning on disagreement, experts are simply signal sources and are not strategic.

Because quality is unknown, the agent faces a problem when trying to learn about the unknown state: the true weight the agent should apply to the expert's signal is unknown. A Bayesian recognizes this uncertainty and applies her prior belief about the expert's quality to the value of the information. As we discuss, individuals in reality may use what the expert says to first fine-tune their beliefs about expert quality and the weight in the signal. Indeed, the use of "empirical Bayes" methods in certain applications in statistics closely parallels this process (Carlin and Louis, 2000). We model this in the following quasi-Bayesian manner. First, after seeing a signal, the agent evaluates the expert's quality using Bayes' rule. For example, she forms a belief about whether a scientist is competent based on whether the claim is plausible (e.g., how it compares with his prior on climate change). Second, having formed this belief about the expert's quality, the agent then applies this weight to the signal(s) from the expert to form his posterior belief. That is, he uses his updated belief about the scientist's quality to weigh all of her claims.

This process is intuitive in that it generates the same posterior belief as Bayes' rule in the canonical case where agents observe one signal and have common neutral priors with independent states and expert quality. Although this process, which we call *pre-screening*, is innocuous in the canonical case, Section 2 shows how it generally violates Bayes' rule because a pre-screener over-infers quality.

With more signals, a pre-screener and Bayesian share the same belief about the state in expectation, but disagree about both states and quality along nearly every ex-post realized signal path. We use the following terminology: we say that a pre-screener overtrusts (undertrusts) an expert if his posterior belief that the expert is high quality is higher (lower) than the Bayesian's. A pre-screener is optimistic (pessimistic) if his posterior belief in the objectively more likely state given all observed information is higher (lower) than the Bayesian's.

Along these paths, pre-screeners are optimistic if and only if they overtrust the expert, and are pessimistic if and only if they under-trust. Intuitively, believing too strongly in the expert's high (low) quality means that the pre-screener overweights (underweights) his information content, and vice versa. In other words, disagreement between a Bayesian and pre-screener about states and expert quality go hand-in-hand. The same is true when comparing two pre-screeners who have received the same information in different order. The only exception is if evidence about the state is perfectly mixed, in which case the pre-screener can correctly discern that the expert is conveying no new information about the state.

Section 3 develops further predictions about how disagreement about expert quality ("trust") and states dynamically evolve along these paths. We focus on disagreement between a Bayesian and pre-screener as a benchmark. A Bayesian's beliefs are completely summarized by the total number of signals received and the fraction of those signaling state A. Holding this fixed, a Bayesian's beliefs are invariant to signal order. In contrast, a pre-screener's beliefs are highly path-dependent in that first impressions about expert credibility have an outsized influence on beliefs about the state. An early sequence with few signal reversals generates a positive first impression in that it leads to significant overtrust and optimism; conversely, an early sequence of many signal reversals generates a negative first impression, and therefore under-trust and pessimism.

These first impressions create persistent disagreement, particularly if the first impression is negative. The overtrust and optimism from a positive first impression, while persistent, can be undone given enough subsequent mixed signals, which suggest low quality. In contrast, the under-trust and pessimism from a negative first impression can cast a long shadow, persisting even when subsequent signals are all identical and indicate high quality. This fundamental asymmetry between the effects of positive and negative first impressions arises because mixed signals are relatively worse news for expert quality than identical signals are good news. In general, the model predicts that negative first impressions are very difficult to overcome.

Pre-screening also implies that the order in which experts report matters for beliefs, not just the order of signals themselves. There is an asymmetry between "inside" experts (those who have already reported) and new unknown "outsiders" (those just reporting). If a pre-screener has a strong positive first impression of an inside expert, information from the same expert which contradicts current beliefs will help resolve disagreement. However, a pre-screener will actively discredit the same information if delivered by an outsider; the contrary outsider information will actually bolster her optimism and overtrust in the insider. Crucially, the model suggests a way in which outsiders can mitigate this effect: deliver those contrary signals together in a "data dump" rather than sequentially.

Our exercise highlights the role of uncertainty about expert quality in generating endogenously biased trust in experts. Section 4 discusses how this generates testable predictions that set our proposed bias apart from those in the literature. The closest learning bias to ours is confirmation bias (Griffin and Tversky, 1992; Rabin and Schrag, 1999), which predicts that individuals interpret information in a way that confirms preconceived beliefs, so that first impressions about the state matter. With pre-screening, the first impression *about the expert* matters, a distinction which has significant economic bite and helps explain when behavior which looks like confirmation bias (as well as its opposite) endogenously arises. This distinction also implies that the order in which experts arrive is important for beliefs, not just the signal order itself as in Rabin and Schrag (1999).

More broadly, the literature has generally recognized that individuals misperceive the informativeness of their signals, often due to overconfidence (e.g., Scheinkman and Xiong, 2003; Ortoleva and Snowberg, 2015). Phenomena which look like over- and under-confidence in signals endogenously arise in our framework due to learning about credibility. Our bias is distinct from inattention (Schwartzstein, 2014; Kominers, Mu and Peysakhovich, 2016), heterogeneous priors about signal quality (Acemoglu et al., 2016), model uncertainty (Gilboa and Schmeidler, 1989, 1993), and models of expert or media slant (Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006). The latter focus on the response of strategic news "supply" to heterogeneous beliefs. We complement this approach by considering the "demand" perspective, showing how disagreement can arise even without biased information.

Overall, the key distinguishing feature of our framework is that agents disagree about substance because of the more fundamental disagreement about which experts are believable, even when agents are paying attention to all experts and information is widely available. Section 5 argues that this describes the essence of real-world disagreements over several economic questions ("What is the value of stimulus spending?") as well as other debates

(“Are vaccinations safe for children?,” and “Why is it hard to debunk fake news?”).

Section 6 concludes with more speculative implications of our model. For concreteness, the bulk of the paper focuses on the case where signals come from external sources. However, our theory does not require this. Alternatively, the source can be the individual’s own experiences, with uncertainty arising because the individual does not know how informative her experiences are about the true state. This interpretation is broadly related to the idea that people’s life experiences may be particularly important for how they form beliefs (Malmendier and Nagel, 2011, 2016). Overall, learning when source quality is uncertain and its implications for trust in experts can shed light on the foundations of disagreement.

1 Model

1.1 Environment

An agent learns about an unknown state $\theta \in \{A, B\}$ by observing binary signals $s_t \in \{a, b\}$ in each period t from an expert. The expert has quality $q \in \{L(ow), H(igh)\}$, which the agent also does not know. The high quality expert has a higher probability of correctly reporting the state than the low quality expert, making his signals more informative: $P(s_t = a|q, A) = P(s_t = b|q, B) = p_q$, where $1 > p_H > p_L > 1 - p_H$ and $p_H > 1/2$. Experts are not strategic, and nature draws true expert quality independently from the true state. Conditional on state and expert quality, signals are independent and identically distributed. For clarity, we assume the agent observes one signal per period, but the model easily generalizes to multiple signals per period.

In this environment, the agent does not know both the state and the quality of the expert, yet only observes the signal(s) about the state from the expert. This type of situation departs from the canonical setup by assuming both that 1) the quality of the expert is uncertain, and 2) there are fewer signals than sources of uncertainty.

Before proceeding with the model, we discuss why this type of situation is particularly relevant for several real-world decision problems. Consider, as one example, economics. Few individuals have the expertise or training in theory or data analysis to evaluate primary evidence on issues such as trade policy, suggesting a need for economists. Yet, from the individual's perspective, the economist's ability is uncertain, and the individual must form beliefs about it.

Recent evidence suggests that American households view economists skeptically. Sapienza and Zingales (2013) show that average American households have sharply different views than economists on questions ranging from whether it is hard to predict stock prices to whether the North American Free Trade Agreement (NAFTA) increased welfare. They find this difference tends to be large even when there is strong consensus among economists. Of course, economists may be wrong, as there is substantial uncertainty about theoretical models and evidence. Assuming that economists are type H, we capture this as $p_H < 1$. If there were no uncertainty about quality, individuals would also view economists as type H, and form beliefs about economic issues accordingly. But when told that economists agree that the stock market is unpredictable, average beliefs among households hardly moved – if anything, an even larger percentage of households thought that the market *was* predictable. This suggests the more troubling possibility that households view economists as type L.

Why don't independent signals about expert quality, or "credentials," resolve this gap, which persists despite doctorates, chaired professorships, and Nobel prizes? We conjecture at least three reasons. First, in some settings, credentials may not be objectively very informative about the quality of specific signals or forecasts. DellaVigna and Pope (2016) run a large experiment estimating how different incentive schemes affect effort, and ask economists to forecast the effectiveness of each treatment ex-ante. They find that, despite clear differences in treatment effectiveness, the forecast error of experts is disperse and that objective measures of expertise are not correlated with forecast accuracy.

Second, the informativeness of a credential may itself be uncertain to households, even if they are very informative to other experts. Many people have economics Ph.D.'s, from many different schools whose Ph.D. programs have different strengths, in many different areas in economics, with very different publication records and impact, and with varying practical experience in, say, trade. Even if an individual with no expertise knew all of these facts about a particular expert, they would have little idea of how to evaluate them together. A trade economist might, but that is of no help to the lay person, because the economist's quality is precisely what is unknown.

Third, perhaps the most useful tool for evaluating expert reliability - the ability to compare predictions to outcomes through repeated controlled experiments - is unavailable in many economic settings. Economists may predict that X (free trade) will cause Y (growth, improved welfare) ex-ante, but evaluating whether any given instance of X (NAFTA) did actually cause Y ex-post is difficult, *even for economists*, because of the difficulty of empirical identification from observational data.

In summary, households face substantial uncertainty about the informativeness of experts, and credentials which should help resolve this uncertainty are often of uncertain value themselves. In this case, such uncertainty may prevail even if one expands the set of experts an agent observes, because each source carries with it additional uncertainty. To isolate how learning might occur, we focus on a particularly stark environment with no signals that explicitly convey expert quality.

1.2 Learning

Suppose the agent has the prior that the state and quality are independent with marginal probabilities $(\omega_0^\theta, \omega_0^q) \in (0,1) \times (0,1)$, respectively, so that his joint prior over both, ω_0, is given by Table 1.

	$\theta = A$	$\theta = B$
$q = H$	$\omega_0^H \omega_0^A$	$\omega_0^H (1 - \omega_0^A)$
$q = L$	$(1 - \omega_0^H)\omega_0^A$	$(1 - \omega_0^H)(1 - \omega_0^A)$

Table 1: Joint prior beliefs ω_0

Our thesis is that the following learning bias, which we call *pre-screening*, tends to arise when agents must infer both source quality and the state. Suppose the agent observes a sequence of signals of length n, denoted $\mathbf{s}^n = (s_1, s_2, \dots s_n)$, where one signal is observed each period. When forming his joint posterior beliefs about the quality and state, the biased decision-maker follows two steps to make inferences. First, he applies Bayes Rule to update on his belief about the expert's quality by combining the signal's content with his joint prior on expert quality and state, denoted $\kappa_q(\mathbf{s}^n)$. Second, he subsequently weights all observed information by using his first-stage *updated* belief about the expert's quality $\kappa_q(\mathbf{s}^n)$ to form posterior beliefs on the joint distribution of state and quality.

To illustrate the pre-screener's updating algorithm, suppose he observes two signals, one in each period. After observing the first signal (s_1), the biased agent's first-stage updated belief about the expert's quality, $\kappa_q(s_1)$, is:

$$\kappa_q(s_1) = \frac{\omega_0^q \sum_\theta P(s_1|q,\theta)\omega_0^\theta}{\sum_q \sum_\theta P(s_1|q,\theta)\omega_0^\theta \omega_0^q}.$$

Using $\kappa_q(s_1)$ to form his joint posterior belief on the state and quality, $P^b(q,\theta|s_1)$, yields his posterior beliefs after the first signal:

$$P^b(q,\theta|s_1) = \frac{P(s_1|q,\theta)\kappa_q(s_1)\omega_0^\theta}{\sum_q \sum_\theta P(s_1|q,\theta)\kappa_q(s_1)\omega_0^\theta}.$$

After observing the second signal (s_2), the biased agent's first-stage updated belief about the expert's quality, $\kappa_q(s_1, s_2)$ is

$$\kappa_q(s_1, s_2) = \frac{\sum_\theta P(s_2|q,\theta)P^b(q,\theta|s_1)}{\sum_q \sum_\theta P(s_2|q,\theta)P^b(q,\theta|s_1)}.$$

The agent then uses $\kappa(s_1, s_2)$ to form his joint posterior belief on the state and quality by re-weighting all the information from the expert. The updated posterior, $P^b(q,\theta|s_1,s_2)$, equals:

$$P^b(q,\theta|s_1,s_2) = \frac{P(s_2|q,\theta)P(s_1|q,\theta)\kappa_q(s_1,s_2)w_0^\theta}{\sum_q \sum_\theta P(s_2|q,\theta)P(s_1|q,\theta)\kappa_q(s_1,s_2)\omega_0^\theta}.$$

Iterating on the biased agent's updating process allows us to characterize his posterior beliefs.

Definition 1 (Pre-screener's beliefs) *After observing a sequence of n signals $\mathbf{s}^n$ from an expert, the **pre-screener's first-stage updated belief** about expert quality, $\kappa_q(\mathbf{s}^n)$, is given by:*

$$\kappa_q(\mathbf{s}^n) = \frac{\kappa_q(\mathbf{s}^{n-1}) \sum_\theta \left(\prod_{t=1}^n P(s_t|q,\theta)\omega_0^\theta\right)}{\sum_q \kappa_q(\mathbf{s}^{n-1}) \sum_\theta \left(\prod_{t=1}^n P(s_t|q,\theta)\omega_0^\theta\right)}, \tag{1}$$

where $\kappa_q(\emptyset) = \omega_0^q$. The ***pre-screener's final joint posterior*** *on expert quality and the state, $P^b(q, \theta|\mathbf{s}^n)$, is given by:*

$$P^b(q, \theta|\mathbf{s}^n) = \frac{\left(\prod_{t=1}^{n} P(s_t|q, \theta)\right) \kappa_q(\mathbf{s}^n)\omega_0^\theta}{\sum_q \sum_\theta \left(\prod_{t=1}^{n} P(s_t|q, \theta)\right) \kappa_q(\mathbf{s}^n)\omega_0^\theta}. \tag{2}$$

This definition assumes ex-ante independence of states and quality. We maintain this assumption for the bulk of our analysis both for simplicity and because it isolates the how pre-screening affects the evolution of correlated beliefs about states and quality without assuming any correlation ex-ante. We provide a generalized definition in Appendix A.1.

The biased individual is quasi-Bayesian in that he otherwise applies Bayes Rule correctly within each step of his updating process. However, the process is erroneous in that, by updating on quality first to produce $\kappa_q(\mathbf{s}^n)$ and re-weighting all information according to updated beliefs, it uses the same signal content multiple times. Thus, pre-screening causes the biased agent to overinfer expert quality. In contrast, Bayesian forms a joint posterior on the state and quality together in one step, using the ex-ante prior quality of the expert.

The information processing mechanism of (erroneously) using updated beliefs to form posterior beliefs was initially conjectured by Lord, Ross and Lepper (1979) [p.2107] to explain subjects' differential interpretation of disconfirming versus confirming evidence: "[Our subjects'] sin lay in their readiness to use evidence already processed in a biased manner to bolster the very theory or belief that initially 'justified' the processing bias." Indeed, the closest related work to ours is confirmation bias (Rabin and Schrag, 1999), although our learning bias puts the role of the expert front and center.

This learning process is also analogous to the use of "empirical Bayes" methods in statistics (see Carlin and Louis, 2000, for a review), where a researcher first uses the data to estimate a prior before using the data to estimate a larger model around this prior. Several critics have noted that "double-dipping" the data this way can lead to erroneous inference (Gelman et al., 2003); Lindley (1969) famously noted that "there is no one less Bayesian than an empirical Bayesian."

More recently, Enke and Zimmermann (2016) demonstrate that environmental complexity is relevant to the degree to which individuals make inference errors. In an experimental setting where subjects receive signals from multiple sources with common underlying information, subjects fail to recognize the double-counting issue when it is not obvious, but are able to compute beliefs correctly when this is pointed out. Though their setting is quite different, such behavior is suggestive that individuals tend to make errors with applying the appropriate weight to information when the inference problem is complex.

2 Disagreement

A Bayesian's posterior belief $P^u(q,\theta|\mathbf{s}^n)$ equals:

$$P^u(q,\theta|\mathbf{s}^n)=\frac{\left(\prod_{t=1}^{n}P(s_t|q,\theta)\right)\omega_0^{\theta}\omega_0^{q}}{\sum_q\sum_\theta\left(\prod_{t=1}^{n}P(s_t|q,\theta)\right)\omega_0^{\theta}\omega_0^{q}}. \tag{3}$$

An immediate implication of Equations 1, 2, and 3 is that the biased and Bayesian's posterior beliefs coincide, $P^b(q,\theta|\mathbf{s}^n)=P^u(q,\theta|\mathbf{s}^n)$, in the canonical case when the prior on the state is neutral ($\omega_0^{\theta}=1/2$) and the agent observes only one signal ($n=1$), as then $\kappa_q(\mathbf{s}^n)=w_0^q$. Intuitively, the first step is innocuous here because the ex-ante belief is that both states are equally likely and independent of quality, so that the pre-screener finds a single signal about the state uninformative about the expert's quality. That these two cases coincide in such a simple case indicates the subtlety of the bias.

However, the pre-screener makes two conceptual errors relative to the Bayesian. First, she uses the latest signal s_n to form a belief about expert quality $\kappa_q(\mathbf{s}^n)$ in Equation 1 *before* forming an updated belief about the state. A Bayesian does this in one step. Having formed an erroneous opinion of expert quality, the pre-screener makes a second error by re-evaluating the informativeness of all prior signals, as one can see by comparing Equations 2 and 3. A Bayesian naturally lets her posterior about expert quality evolve without explicitly re-visiting the informativeness of previous signals given her opinion about expert quality today. A pre-screener's posterior beliefs are also identical to the Bayesian's for any sequence of signals if there is no uncertainty about expert quality (i.e., $p_L=p_H$), since the pre-screening step has no bite in this case. The overall result is that whenever there is uncertainty about expert quality, the pre-screener over-infers expert quality, leading to biased beliefs and disagreement with a Bayesian outside of the canonical case.[1]

If agents begin with the prior that both states are equally likely, the ex-ante difference between the pre-screener and Bayesian's posterior marginal beliefs about the state equals zero. This is because beliefs about states are ex-ante symmetric around A and B and are ex-ante independent of quality, as we show in:

Proposition 1 (No average disagreement about θ) *Let a Bayesian and pre-screener share a common prior of $(\omega_0^\theta, \omega_0^H) = (1/2, \hat{q})$ for any $\hat{q} \in (0,1)$, and suppose this represents the true distribution from which nature draws (θ, q). Then $E_0[P^b(\theta = A|\mathbf{s}^n) - P^u(\theta = A|\mathbf{s}^n)] = 0$, where the expectation E_0 is taken over this distribution and all signal paths $\mathbf{s}^n$. However, $E_0\left[\left(P^b(\theta = A|\mathbf{s}^n) - P^u(\theta = A|\mathbf{s}^n)\right)^2\right] > 0$.*

However, disagreement arises along several paths. As a corollary, the expected squared (or absolute) difference in marginal posteriors about θ is strictly positive. To characterize disagreement, we define the following terms to simplify exposition.

Definition 2 (Information content) *The **information content** of any sequence of signals $\mathbf{s}^n$ is given by the number of "a" signals n_a and the number of "b" signals, n_b.*

Definition 3 (Optimism and trust) *Fix the information content with $n_a > n_b$ without loss of generality. Given a signal sequence $\mathbf{s}^n$,*

1. *A pre-screener is **optimistic** if $Pr^b(\theta = A|\mathbf{s}^n) > Pr^u(\theta = A|\mathbf{s}^n)$, and **pessimistic** if strictly less than (<).*

2. *A pre-screener **overtrusts** if $Pr^b(q = H|\mathbf{s}^n) > Pr^u(q = H|\mathbf{s}^n)$, and **under-trusts** if strictly less than (<).*

[1] One can ask what happens with either error without the other. For example, one can make the first error only apply the updated belief to the latest signal by re-scaling each period's prior belief appropriately, so that beliefs are Markov. One can also potentially apply the second error without the first by re-weighting all past signals by last period's posterior belief. Generally, the first error relates to how today's information is mistakenly processed, while the second error relates how that mistake feeds back into beliefs. Both relate to the general concept of over-inferring expert quality: having formed an erroneous opinion using the most recent signal (the first error), it is intuitively natural to wish to explicitly re-evaluate all previous signals in light of this belief (the second). Most of our results stem from the combination of the two errors.

An important feature of disagreement is that, for any sequence of signals, whether or not a pre-screener is optimistic or pessimistic is determined by whether or not he over- or under-trusts the expert.

Proposition 2 (Correlated disagreement) *For any $\mathbf{s}^n$ with $n_a > n_b$, and for all $\omega_0^\theta \in (0,1)$ and $\omega_0^q \in (0,1)$, the agent under-trusts the expert if and only if he is pessimistic about the more likely state: $P^b(q = H|\mathbf{s}^n) < P^u(q = H|\mathbf{s}^n)$ if and only if $P^b(\theta = A|\mathbf{s}^n) < P^u(\theta = A|\mathbf{s}^n)$. Likewise, the agent overtrusts the expert if and only if he is optimistic in beliefs about the more likely state: $P^b(q = H|\mathbf{s}^n) > P^u(q = H|\mathbf{s}^n)$ if and only if $P^b(\theta = A|\mathbf{s}^n) > P^u(\theta = A|\mathbf{s}^n)$.*

Intuitively, if the pre-screener under-trusts the expert, he is too skeptical about the information content of the expert's signals. If the signals imply that A is objectively likely, then the pre-screener will believe A is less likely than it actually is. Conversely, if the pre-screener thinks A is less likely than the Bayesian, it must be because he under-trusts the expert and therefore underweights his information content. Analogously, overtrust is positively correlated with optimism. Note that Proposition 2 does not depend on assuming a neutral prior about the state, as it holds for any $\omega_0^\theta \in (0,1)$.

Proposition 2 shows that disagreement about states and expert quality between a pre-screener and Bayesian go hand in hand. However, two pre-screeners who experience the same set of signals *in different order* also disagree. In Proposition 3, we compare the beliefs of two pre-screeners who have identical priors and observe sequences with identical information content, but may observe different orderings. The analog of Proposition 2 applies - a pre-screener who trusts the expert more (less) must also believe the reported state is more (less) likely, and vice versa. Disagreement arises on many paths, so that the expected squared difference in beliefs is positive. Thus, our framework generates disagreement even when agents share identical information content, learning biases, and priors, providing a foundations for the origins of disagreement.

Proposition 3 (Origins of disagreement) *Suppose two pre-screeners, J and M, have identical common priors $(\omega_0^\theta, \omega_0^H) = (\hat{\theta}, \hat{q})$ for any $\hat{\theta} \in (0,1)$ and $\hat{q} \in (0,1)$, and observe* *signal sequences $\mathbf{s}_J^n$ and $\mathbf{s}_M^n$ that have identical information content, where $n_a + n_b = n$ and $n_a > n_b$.*

1. *(Correlated disagreement) Agent J trusts the expert more than agent M does if and only if agent J believes state A is more likely than agent M does: $P^b(q = H|\mathbf{s}_J^n) > P^b(q = H|\mathbf{s}_M^n)$ if and only if $P^b(\theta = A|\mathbf{s}_J^n) > P^b(\theta = A|\mathbf{s}_M^n)$. Likewise, Agent J trusts the expert less than agent M if and only if agent J believes state A is less likely than agent M does: $P^b(q = H|\mathbf{s}_J^n) < P^b(q = H|\mathbf{s}_M^n)$ if and only if $P^b(\theta = A|\mathbf{s}_J^n) < P^b(\theta = A|\mathbf{s}_M^n)$.*
2. *(Squared disagreement about θ) $E_0\left[\left(P^b(\theta = A|\mathbf{s}_J^n) - P^b(\theta = A|\mathbf{s}_M^n)\right)^2\right] > 0$, where the expectation E_0 is taken over the distribution of all signal paths $\mathbf{s}_i^n$ where each path i has identical fixed information content.*

3 How Do Trust and Disagreement Evolve?

For simplicity, we focus on disagreement between a Bayesian and a pre-screener. Proposition 1 shows that, even though there is no ex-ante disagreement about θ, ex-post there is substantial disagreement among realized paths. One important reason this occurs is because the pre-screener's beliefs are path-dependent, while the Bayesian's are not.

To see this, re-arrange Equations 1 and 2 to obtain:

$$P^b(q, \theta|\mathbf{s}^n) = \frac{b_q(\mathbf{s}^n)\left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^\theta \omega_0^q}{\sum_q b_q(\mathbf{s}^n) \sum_\theta \left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^\theta \omega_0^q}, \tag{4}$$

where $b_q(\mathbf{s}^n)$ is defined as:

$$\begin{aligned} b_q(\mathbf{s}^n) &\equiv \left(\textstyle\sum_\theta P(s_1|q,\theta)\omega_0^\theta\right) \times \left(\textstyle\sum_\theta P(s_1|q,\theta)P(s_2|q,\theta)\omega_0^\theta\right) \times \ldots \times \left(\textstyle\sum_\theta P(s_1|q,\theta)P(s_2|q,\theta)\ldots P(s_n|q,\theta)\omega_0^\theta\right) \\ &= \prod_{m=1}^n \left(\sum_\theta \left(\prod_{t=1}^m P(s_t|q,\theta)\right)\omega_0^\theta\right), \end{aligned} \tag{5}$$

and where $b_q(\emptyset) \equiv 1$. In these equations, $b_q(\mathbf{s}^n)$ reflects the cumulative effect of pre-screening and re-weighting information by updated beliefs about quality after every signal. Equation 5 makes clear that early signals tend to have a larger influence on b_q than later signals, because each subsequent signal is evaluated relative to the preceding ones.

As an example, consider the signal sequence $\{a, a, b\}$. Swapping the order in which the agent observes s_1 and s_3 generates different posteriors: $P^b(q, \theta|\{a, a, b\}) \neq P^b(q, \theta|\{b, a, a\})$, because $b_q(\{a, a, b\}) \neq b_q(\{b, a, a\})$. Under the sequence $\{a, a, b\}$, the pre-screener over-infers that the expert is type H after the second signal, while under $\{b, a, a\}$, he over-infers that the expert is type L. The difference in this early part of the signal sequence colors how the pre-screener interprets the final signal and generates path-dependence.

In contrast, a Bayesian's posterior beliefs are not path dependent, and depend solely on the information content of the signals. In the previous example, $P^u(q, \theta|\{a, a, b\}) = P^u(q, \theta|\{b, a, a\})$. The signal sequence does not matter, because while the Bayesian also infers that the expert is more likely to be H after the first two signals in $\{a, a, b\}$, he does not over-infer this likelihood.

This simple example highlights the important nature of *first impressions about the expert.* Holding the information content fixed, changing the order of signals changes the level of overtrust and under-trust with the expert. To characterize this, we show that there is a unique sequence of signals which generates the maximal over- and under-trust for any fixed information content:

Lemma 1 (First impressions about experts) *Consider a given combination of n_a a signals and n_b b signals, where $n_a > n_b \geq 1$. The sequence in which n_a consecutive a signals is followed by n_b consecutive b signals generates the maximal degree of trust in the expert. The sequence in which n_b a signals alternating with n_b b signals is followed by $n_a - n_b$ a signals generates the minimal degree of trust in the expert.*

Lemma 1 shows that pre-screeners erroneously value early consistency. While fewer reversals objectively suggests that the expert is high quality, the pre-screener over-infers this, leading to overtrust. Likewise, more initial reversals suggest the expert is low quality, which the pre-screener over-infers, leading to under-trust. Holding information content fixed, re-ordering the signals so that all of those that favor the objectively more likely state into a consistent string first generates the most trust in the expert, while alternating the signals early generates the least trust.

3.1 The origins of disagreement: first impressions

Proposition 4 shows that first impressions about the expert have a large influence on later inferences about the state and quality. Start with a positive first impression. Even if subsequent signals are uninformative, and would have by themselves generated a *negative* first impression, the pre-screener may still overtrust the expert. Suppose the pre-screener observes $n_a > 1$ consecutive a signals, creating overtrust. Any negative information about quality conveyed by ensuing mixed signals must be sufficiently strong to unravel this overtrust, which is stronger for larger n_a. How much negative information is required depends on how informatively the mixed signals indicate low quality, which itself depends on p_L and p_H. When p_L and p_H are sufficiently low, mixed signals are not as informative about lower quality because neither expert type is reliable. But if p_H is high relative to p_L, mixed signals strongly indicate low quality. Likewise, if both p_L and p_H are sufficiently high, mixed signals strongly indicate lower quality precisely because both types are highly reliable. Thus, mixed signals can be sufficiently strong evidence to unravel positive first impressions quickly in the latter two cases.

Negative first impressions are persistent in that the pre-screener may still under-trust the expert even if subsequent signals all identically favor one state. Positive information about quality conveyed by ensuing identical signals must be sufficiently strong to unravel the initial under-trust. In contrast to the case of positive first impressions, the persistence of negative impressions is not conditional on the distribution of expert types. This is because there is a fundamental asymmetry in the informativeness of mixed versus identical evidence: mixed signals are relatively worse news for quality than identical signals are good news. For example, consider the extreme case of $p_L = 1/2$ and $p_H \approx 1$. Mixed signals almost immediately rule out the possibility of a high type. In contrast, identical signals do not imply high quality so obviously, because it is always possible for a low quality type to draw identical signals by chance. Thus since the mixed evidence is both more informative and is overweighted, more ensuing consistency is needed to unravel the negative first impression.

Proposition 4 (Persistent effects of first impressions) *First impressions of expert quality persist in the face of contrary information about quality:*

1. *Positive first impressions: Suppose the agent observes $n_a \geq 1$ consecutive a signals, followed by m pairs of (b, a) signals: $\mathbf{s}^n = (a, a, a, \ldots, b, a, b, a)$.*

   (a) *If $n_a \leq 2$, the pre-screener under-trusts and is pessimistic about the most likely state for all $m \geq 1$.*

   (b) *If $n_a \geq 3$, then there exists some $m' > 3$ and $p' \in (\frac{1}{2}, 1)$ such that when $m < m'$ and $p_L < p_H \leq p'$, the pre-screener overtrusts and is optimistic about the most likely state, where m' increases with n_a.*

2. *Negative first impressions: Suppose the agent observes $n_b \geq 1$ pairs of (a, b) signals, followed by $m \geq 1$ consecutive a signals, where $m \geq 1$: $\mathbf{s}^n = (a, b, a, b, \ldots, a, a, a)$. Then there exists some $m^* > 3$ such that when $m < m^*$, the pre-screener under-trusts and is pessimistic about the most likely state, where m^* increases with n_b.*

Proposition 5 shows that this asymmetry between mixed and identical evidence affects the degree to which first impressions persist in the limit. Enough mixed signals can always unravel a positive first impression, because mixed signals are relatively bad news for quality. In contrast, arbitrarily high levels of persistence can arise for negative first impressions, depending on the distribution of expert types. If p_L is sufficiently low and p_H is sufficiently high, the under-trust created by even the simple sequence (a, b) is very persistent: given any m, we can always find such a combination of (p_L, p_H) such that the under-trust survives m identical signals of a afterwards. The same is true if p_L and p_H are both very high. In both of these cases, mixed signals are strong evidence of lower quality, which is disproportionately overweighted by the pre-screener when initially observed, while later consistent signals are inherently weaker evidence.

Proposition 5 (Persistent effects after short sequences) *Positive first impressions can eventually be undone, but negative first impressions may be arbitrarily persistent:*

1. *Positive first impressions: Suppose the agent observes $n_a \geq 1$ consecutive a signals, followed by $m \geq 1$ pairs of (b, a) signals: $\mathbf{s}^n = (a, a, a, \ldots, b, a, b, a)$. For a given n_a, there exists $\hat{m}$ such that when $m > \hat{m}$, the pre-screener under-trusts and is pessimistic about the most likely state for any (p_L, p_H).*

2. *Negative first impressions: Suppose the agent observes $n_b \geq 1$ pairs of (a, b) signals, followed by $m \geq 1$ consecutive a signals: $\mathbf{s}^n = (a, b, a, b, \ldots, a, a, a)$. For a given $n_b \geq 1$ and $m \geq 1$, there exists some $\check{p} > \frac{1}{2}$ and $\hat{p} < 1$ such that the pre-screener under-trusts and is pessimistic about the most likely state if (p_L, p_H) satisfies one of the following sufficient conditions:*

   (a) *$\hat{p} \leq p_L < p_H$, or*

   (b) *$p_L \leq \check{p}$ and $p_H < \hat{p}$.*

Because over- and under-trust are intricately linked with optimism and pessimism by Proposition 2, Propositions 4 and 5 imply that persistent disagreement about quality filters through to persistent disagreement about states.

3.2 Resolving disagreement: insiders vs outsiders

First impressions create persistent disagreement. What resolves disagreement? Suppose the pre-screener begins with a flat prior on the state, and therefore has a positive first impression after k signals, all identically a. After an additional k identical signals of all b, he will have the correct posterior on the state, though not necessarily on the expert's quality. Intuitively, even the pre-screener understands that all signals originate from the same source, so he will realize that $n_a = n_b = k$ is equivalent to having no new information about the state, even if he is incorrect about the expert's quality due to overinference. This intuition holds more generally: given any prior on the state, $\omega_0^\theta \in (0, 1)$, observing $n_a = n_b = k$ signals in any order will lead the pre-screener back to her prior on the state, which is the objectively correct marginal posterior.

Proposition 6 (Resolving disagreement from an overtrusted expert) *After observing $n_a = k > 1$ consecutive a's, a successive sequence of $n_b = k$ consecutive b's returns the disagreement about the state to zero.*

A more realistic situation is one in which an agent receives additional signals from another expert, or a "second opinion." Intuitively, a second opinion should help resolve disagreement between two agents. However, this is not necessarily the case with a pre-screener.

Suppose the pre-screener now receives signals from two independently drawn experts, $j = 1, 2$, with qualities q_j. Let s_{tj} be the t^{th} signal sent by expert $j \in \{1, 2\}$. Each expert j sends a sequence of n_j signals, $\mathbf{s}^{n_j}$. Denote expert 1 as the "inside" expert who is first expert to report, and expert 2 as the "outside" expert. Let $\mathbf{s}^{n_1,n_2}$ be the sequence of observed signals from both experts, where $\mathbf{s}^{n_1,n_2} = (\mathbf{s}^{n_1}, \mathbf{s}^{n_2})$. Let $\mathbf{s}^{n_1,0}$ denote the sequence of signals from expert 1 when expert 2 has not said anything yet.

In this model, the agent now has three sources of uncertainty - the quality of each expert and the state of the world. Since the expert quality is independent and identically distributed, $\omega_0^{q_j} = \omega_0^q$ for all j. Since the reliability of a signal t from expert j is independent of the other expert k's quality, note that $P(s_{tj}|q_j, q_k, \theta) = P(s_{tj}|q_j, \theta)$ for all t where $j \neq k$.

The biased agent's pre-screening procedure extends naturally from one source to multiple sources. First, he updates on the joint belief about the experts' qualities, denoted $\kappa_{q_1q_2}(\mathbf{s}^{n_1,n_2})$, by combining the signal's content with his joint prior on expert qualities and state. Second, he subsequently uses her updated belief about the experts' quality to form a joint posterior beliefs on the state and qualities. Iterating on the biased agent's updating process allows us to characterize his posterior beliefs when he receives any set of signals from both experts, $\mathbf{s}^{n_1,n_2}$. The pre-screener's beliefs after observing expert 1 (but not expert 2) are:

$$\kappa_{q_1q_2}(\mathbf{s}^{n_1,0}) = \frac{\kappa_{q_1q_2}(\mathbf{s}^{n_1-1,0}) \left(\sum_\theta \left(\prod_{t=1}^{n_1} P(s_{t1}|q_1,\theta)\right) \omega_0^\theta\right)}{\sum_{q_1} \sum_{q_2} \kappa_{q_1q_2}(\mathbf{s}^{n_1-1,0}) \left(\sum_\theta \left(\prod_{t=1}^{n_1} P(s_{t1}|q_1,\theta)\right) \omega_0^\theta\right)}, \tag{6}$$

where $\kappa_{q_1q_2}(\emptyset) = \omega_0^{q_1}\omega_0^{q_2}$, and:

$$P^b(q_1, \theta|\mathbf{s}^{n_1,0}) = \frac{\left(\prod_{t=1}^{n_1} P(s_{t1}|q_1,\theta)\right) \kappa_{q_1q_2}(\mathbf{s}^{n_1,0})\omega_0^\theta}{\sum_q \sum_\theta \left(\prod_{t=1}^{n_1} P(s_{t1}|q_1,\theta)\right) \kappa_{q_1q_2}(\mathbf{s}^{n_1,0})\omega_0^\theta} \tag{7}$$

After observing expert 2, beliefs are:

$$\kappa_{q_1q_2}(\mathbf{s}^{n_1,n_2}) = \frac{\left(\sum_\theta \left(\prod_{t=n_1+1}^{n_1+n_2} P(s_{t2}|q_2,\theta)\right) \left(\prod_{t=1}^{n_1} P(s_{t1}|q_1,\theta)\right) \omega_0^\theta\right) \kappa_{q_1q_2}(\mathbf{s}^{n_1,n_2-1})}{\sum_{q_2} \sum_{q_1} \left(\sum_\theta \left(\prod_{t=n_1+1}^{n_1+n_2} P(s_{t2}|q_2,\theta)\right) \left(\prod_{t=1}^{n_1} P(s_{t1}|q_1,\theta)\right) \omega_0^\theta\right) \kappa_{q_1q_2}(\mathbf{s}^{n_1,n_2-1})}, \tag{8}$$

and:

$$P^b(q_1, q_2, \theta|\mathbf{s}^{n_1,n_2}) = \frac{\left(\prod_{t=n_1+1}^{n_1+n_2} P(s_{t2}|q_2, \theta)\right)\left(\prod_{t=1}^{n_1} P(s_{t1}|q_1, \theta)\right)\kappa_{q_1 q_2}(\mathbf{s}^{n_1,n_2})\omega_0^\theta}{\sum_{q_2}\sum_{q_1}\sum_\theta \left(\prod_{t=n_1+1}^{n_1+n_2} P(s_{t2}|q_2, \theta)\right)\left(\prod_{t=1}^{n_1} P(s_{t1}|q_1, \theta)\right)\kappa_{q_1 q_2}(\mathbf{s}^{n_1,n_2})\omega_0^\theta}. \tag{9}$$

As before, a Bayesian's posterior beliefs $P^u(q_1, q_2, \theta|\mathbf{s}^{n_1,n_2})$ depend purely on the information content delivered by each expert, not on the order in which signals are received from a given expert.

We now examine the role of the second outside expert in resolving disagreement. Consider the case where the pre-screener has a positive first impression about expert 1, due to $k > 1$ identical signals of a from her. Suppose expert 2 delivers k identical signals of b. Again, we assume that the pre-screener has a neutral prior on the state.

A Bayesian who begins with a neutral prior on the state infers that the experts cannot both be high quality: despite both delivering consistent messages, the signals contradict, and one expert must be wrong. However, he also understands that there is insufficient evidence to deduce which expert is wrong. As a result, given two groups of k opposing signals, he concludes that neither state is more likely than the other.

However, the following proposition shows that the pre-screener incorrectly trusts the first expert more than the outsider, and therefore incorrectly believes A is more likely:

Proposition 7 (Outsider rejection) *Let the agent observe k a signals from expert 1, followed by k b signals from expert 2: $\mathbf{s}^{n_1} = (a, \ldots, a)$ and $\mathbf{s}^{n_2} = (b, \ldots, b)$ where $n_1 = n_2 = k$. Given $\omega_0^A = 1/2$ and $k > 1$, the biased agent believes that state A is more likely than B, and that the first expert is more likely to be high quality than the second expert: $P^b(\theta = A|\mathbf{s}^{n_1,n_2}) > 1/2$ and $P^b(H_1|\mathbf{s}^{n_1,n_2}) > P^b(H_2|\mathbf{s}^{n_1,n_2})$.*

Intuitively, the first impression creates overtrust in the first expert. So the pre-screener's interpretation of the second expert's consistent messages are biased by the fact that they contradict the overtrusted first expert, leading the pre-screener to believe too strongly in the possibility that the second expert is low quality and the first expert is high quality. Thus there is an initial, overly strong drop in trust in the second expert, in contrast to the initial overtrust in the first expert. This asymmetry means that the second expert cannot completely unravel the first expert's messages. The positive impression from the first expert's consistency inflates the pre-screener's trust in the first expert and deflates trust in the second expert, relative to the Bayesian inference. Thus, the pre-screener also realizes that the experts are most likely to be two different qualities, but incorrectly concludes that the first expert is more credible than the second and therefore differentially weights information in favor of the first expert.

Proposition 8 shows that this persists in the limit: information that should lead to more uncertainty about qualities and no change in beliefs about the state instead leads the biased agent to be *more sure of and more wrong in* his beliefs along both dimensions when he observes information from different experts sequentially.

Proposition 8 (Outsider rejection in the limit) *Let the agent observe k a signals from expert 1, followed by k b signals from expert 2: $\mathbf{s}^{n_1} = (a, \ldots, a)$ and $\mathbf{s}^{n_2} = (b, \ldots, b)$ where $n_1 = n_2 = k$. Given $\omega_0^A = 1/2$ and $k > 1$, $\lim_{k\to\infty} P^b(\theta = A|\mathbf{s}^{n_1,n_2}) = 1$ and $\lim_{k\to\infty} P^b(H_1, L_2|\mathbf{s}^{n_1,n_2}) = 1$.*

As the two experts send an increasing equal number of opposing signals, the Bayesian infers that either state is equally likely, and each expert's quality is increasingly uncertain. In contrast, the pre-screener becomes *more* certain that the first expert is high quality and becomes increasingly incorrect about the true state. Intuitively, the pre-screener's first impression from the first expert's increasing consistency leads to even greater overtrust in the first expert and under-trust in the second expert.

How then can signals from outsiders resolve disagreement? The fundamental problem is that the pre-screener over-infers expert quality at each step. This suggests that the outsider can better resolve disagreement were she to deliver all k opposing signals in one "blast". We show this in Proposition 9 by extending the model to allow multiple signals to be observed in a given period.

Proposition 9 (Overcoming outsider rejection) *Consider a sequence of $2k$ observed signals such that expert 1 sends the first k a signals, then expert 2 sends k b signals from expert 2, where $k > 1$.*

1. *The pre-screener overtrusts expert 1 even more when expert 1's signals are sent sequentially rather than simultaneously.*

2. *Expert 2's credibility is higher when sending his signals simultaneously rather than sequentially, but the pre-screener still believes that state A is more likely than B.*

Whether or not the first expert's identical a signals are sent simultaneously or sequentially, the pre-screener overinfers the good news about his quality and therefore overtrusts him. But sequential signals imply overinference that compounds upon each signal, leading to more overtrust than observing simultaneous signals, where overinference occurs one time.

If expert 2 delivers all countervailing signals simultaneously, the pre-screener believes it is relatively less likely that expert 2 is low quality because he compares all k b signals against his beliefs based on expert 1's k signals. Even though the pre-screener overtrusts expert 1's quality and is optimistic about A, this is better for expert 2's credibility than sending each signal sequentially, in which at each step the pre-screener would overweight the inference that the first expert is likely to be high quality and the second to be low quality because he has observed a longer sequence of a's from expert 1 than b's from expert 2.

Propositions 7 through 9 suggest that the order in which experts present themselves, not just information, is highly relevant for persuasion. There is an asymmetry between "inside" and "outside" experts, with a clear first mover advantage for the inside expert when sources consistently disagree. However, while overtrust in an expert is difficult to undo by an outsider, it is more fragile to internal contradictions by the insider.

Moreover, expert order is highly relevant for the timing of information release to bolster credibility. In our framework, experts are exogenous sources of signals and are not strategic. We deliberately take this modeling approach to isolate the effect of the bias. But our analysis suggests that an expert who wants to convince the agent of his quality should release information slowly in order to "build up trust" if he is the first mover and has consistent evidence. In contrast, if he is the second mover and knows that *that same information* is contrary to some other first mover, he should instead release all of it simultaneously because he has to "disprove incompetence." Alternatively, if the second mover has preliminary evidence against the prevailing theory, he should wait to amass more countervailing evidence before disclosing all of it.

4 The Central Role of Experts

4.1 Confirmation bias and pre-screening

Our framework assigns a central role to expert quality for biased learning. This distinguishes it from several other well-known biases, the closest of which is confirmation bias (Lord, Ross and Lepper, 1979; Griffin and Tversky, 1992; Rabin and Schrag, 1999), or the tendency for individuals to misinterpret new information as confirming existing beliefs. In Rabin and Schrag (1999), individuals probabilistically flip signals which oppose current beliefs. This makes "first impressions matter" because early signals over-influence how individuals interpret subsequent signals.

In our framework, the first impression *about the expert* matters, rather than the state. This distinction helps clarify when confirmation bias arises, in two ways. First, it helps explain why confirmation bias may be more likely to arise in more ambiguous settings (Lord, Ross and Lepper, 1979; Griffin and Tversky, 1992), such as when source quality is uncertain. Second, it makes concrete, testable predictions about when confirmation bias arises within such an environment, and when the opposite behavior occurs.

Proposition 10 considers when confirmation bias arises in response to the marginal signal. It conducts the following thought experiment: suppose an agent has observed signals $\mathbf{s}^n$ and has posterior ω_n^b. For the sake of contrast with confirmation bias (although not required by the proposition), assume that $n_A > n_B$, so that the weight of the existing evidence suggests A. Does the agent over- or under-update in response to the marginal signal s_{n+1}, compared to a Bayesian endowed with ω_n^b? Broadly speaking, relative to the Bayesian, an agent in Rabin and Schrag (1999) under-updates on the marginal signal if it contradicts current beliefs, and correctly updates on the marginal signal if it confirms current beliefs.

For a pre-screener, what matters is not whether the signal confirms or contradicts current beliefs about the state, but rather how it affects the pre-screener's beliefs about the expert quality in conjunction with the existing evidence. Suppose the marginal signal confirms current beliefs ($s_{n+1} = a$). If trust is high (part 1a), she will over-update towards A, consistent with confirmation bias. However, if trust is low, she will under-update even though the signal confirms beliefs (part 1b).

Now suppose the marginal signal contradicts current beliefs ($s_{n+1} = b$) but that the combined evidence ($\{\mathbf{s}^n, s_{n+1}\}$) objectively still suggests A. The pre-screener will under-update towards B if trust is high (part 1a), consistent with confirmation bias, because the pre-screener over-values the weight of the combined evidence from the expert suggesting A. In contrast, if trust is low, she will over-update towards B, because she assigns too low of a weight to the combined evidence. Part 1c of the proposition provides a knife-edge case when the combined evidence suggests neither A nor B is objectively more likely.

Thus, in response to new information, the pre-screener exhibits confirmation bias when his trust in the expert is high, and the opposite of confirmation bias when his trust in the expert is low. Importantly, this trust is generated endogenously by his evaluation of the observed sequence of signals.

Intuitively, the pre-screener understands that all information comes from the same expert, and over-infers quality in two ways: she updates her belief about quality based on s_{n+1} before re-evaluating the combined evidence based on this belief. A Bayesian endowed with the same distorted belief also understands this, but she does not over-infer expert quality, as the effect of how expert quality changes the value of signals works completely through Bayes' rule.

Part 2 of the proposition clarifies this. A Bayesian updates identically irrespective of whether she is endowed with a belief or observe a history of signals consistent with that belief: $P^u[\theta = A|\{\mathbf{s}^n, s_{n+1}\}] = P^u[\theta = A|prior = \omega_n^u, \{s_{n+1}\}]$, where ω_n^u equals the Bayesian posterior generated by $\mathbf{s}^n$. However, the effect of a new signal on a pre-screener's beliefs cannot be summarized simply by its effect on the prior. It requires knowledge of the entire history of signals to compute $\kappa_H(\mathbf{s}^n)$. This is the sense in which first impressions about experts matter, rather than the state.

Proposition 10 (Reaction to subsequent signals) *Let $\mathbf{s}^n$ be a sequence of n observed signals (with an arbitrary number of a's and b's), let s_{n+1} be the $(n+1)$th observed signal, and let ω_n^b equal the pre-screener's joint posterior after the sequence $\mathbf{s}^n$. WLOG, let the number of a's be greater than or equal to the number of b's in $\{\mathbf{s}^n, s_{n+1}\}$.*

1. *Relative to Bayesian:*

   (a) $P^b[\theta = A|\{\mathbf{s}^n, s_{n+1}\}] > P^u[\theta = A|prior = \omega_n^b, \{s_{n+1}\}]$ *if* $\{\mathbf{s}^n, s_{n+1}\}$ *has strictly more a's than b's and* $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) > \kappa_H(\mathbf{s}^n)$,

   (b) $P^b[\theta = A|\{\mathbf{s}^n, s_{n+1}\}] < P^u[\theta = A|prior = \omega_n^b, \{s_{n+1}\}]$ *if* $\{\mathbf{s}^n, s_{n+1}\}$ *has strictly more a's than b's and* $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) < \kappa_H(\mathbf{s}^n)$,

   (c) $P^b[\theta = A|\{\mathbf{s}^n, s_{n+1}\}] = P^u[\theta = A|prior = \omega_n^b, \{s_{n+1}\}]$ *if* $\{\mathbf{s}^n, s_{n+1}\}$ *has an equal number of a's and b's or* $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) = \kappa_H(\mathbf{s}^n)$,

   where

$$\begin{aligned}
&\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) > \kappa_H(\mathbf{s}^n) \text{ if and only if } P^u(q = H|\{\mathbf{s}^n, s_{n+1}\}) > \omega_0^H \\
&\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) = \kappa_H(\mathbf{s}^n) \text{ if and only if } P^u(q = H|\{\mathbf{s}^n, s_{n+1}\}) = \omega_0^H \\
&\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) < \kappa_H(\mathbf{s}^n) \text{ if and only if } P^u(q = H|\{\mathbf{s}^n, s_{n+1}\}) < \omega_0^H.
\end{aligned}$$

2. *History-dependence:* $P^b[q, \theta|\{\mathbf{s}^n, s_{n+1}\}] = P^b[q, \theta|prior = \omega_n^b, \{s_{n+1}\}]$ *if and only if* $P^b[q|\mathbf{s}^n] = \omega_0^q$.

Proposition 11 provides conditions under which pre-screeners exhibit confirmation bias based on the relative proportion of a's and b's, regardless of the observed order of the signals. That is, even if we did not know the order in which pre-screeners experienced signals, we can still characterize cases in which they exhibit confirmation bias or its opposite.

Proposition 11 (Over- and under-trust without knowing signal order) *Whether pre-screeners exhibit confirmation bias or the opposite depends on the relative proportion of a's and b's and the distribution of beliefs about quality:*

1. *There exists some* n_a^* *and* $\check{p} > \frac{1}{2}$ *such that the agent overtrusts the expert and is optimistic that the state is A for* any *sequence with fixed* n_a, n_b *when* $n_b < n_a^* < n_a$ *and* $p_L < p_H \leq \check{p}$.

2. *There exists some* $\hat{n}_b$, $\underline{p} > \frac{1}{2}$ *and* $\overline{p} < 1$ *such that the agent under-trusts the expert and is pessimistic that the state is A for* any *sequence with fixed* n_a, n_b *when* $0 \leq \hat{n}_b < n_b < n_a$ *and one of the following sufficient conditions is met:*

(a) $\overline{p} \leq p_L < p_H$*, or*

(b) $p_L \leq \underline{p}$ *and* $p_H \geq \overline{p}$*.*

Part 1 says that, when the proportion of a's is much greater than the proportion of b's, the pre-screener will exhibit confirmation bias in that she may always be optimistic about A relative to the Bayesian. This is true even for the sequence that generates the lowest possible trust in the expert (when holding fixed the information content), so long as mixed signals do not sufficiently distinguish between high and low quality (p_L and p_H sufficiently low). In this case, the initial negative impression from mixed signals is relatively weak, so the ensuing consistency of many a's countervails the initial under-trust, generating optimism and overtrust.

Part 2 says that, when the proportion of a's is sufficiently similar to the proportion of b's, the biased agent may be under-trusting and pessimistic given *any* observed order, including the sequence that generates the highest degree of trust in the expert. This is true when p_L is sufficiently low and p_H is sufficiently high, or when both p_L and p_H are high, as this information content then strongly suggests that the expert is low quality. In this case, fixing the information content, ensuing contrary signals therefore overcome even the most positive first impression, resulting in under-trust and pessimism.

Thus, our framework sheds light on when and whether or not behavior that appears to be confirmation bias can arise when expert quality remains uncertain. While existing work has focused on signal content alone, we show that instead the distribution of content (i.e., the ratio of a's to b's) and the distribution of expert quality (i.e., (p_L, p_H)) are relevant. In particular, the effects that we predict are more likely to arise when source quality is difficult to learn or verify.

Our model also distinguishes how confirmation bias is affected by the source of information, a distinction not addressed by Rabin and Schrag (1999), and addressed in Section 3.2. Our framework predicts that confirmation bias arises when multiple sources contradict one another, but not when a single source is self-contradictory. Finally, Rabin and Schrag (1999) assume that the severity of confirmation bias does not depend on the strength of existing beliefs, while our framework fully endogenizes expert trust. Together, these observations emphasize the central role of expert trust in determining whether confirmation bias arises.

Similarly to Rabin and Schrag (1999), Fryer, Harms and Jackson (2016) consider a setting where signals sometimes deliver an ambiguous signal of ab, which agents interpret in favor of their prior beliefs. As a result, disagreement can arise when there are many ambiguous ab signals observed in different order. But because this bias is also based on beliefs about the state alone, it does not predict qualitatively different effects that vary with source and source quality. It is also less suited to describe situations in which signals are unambiguous but disagreement nevertheless is often strong. Several of the examples in Section 5 fit this description.

4.2 Comparison with other frameworks

We compare how our framework distinguishes our predictions from theories of biased learning and disagreement other than confirmation bias. The unifying theme is that agents disagree about substance because of the more fundamental disagreement about which experts are believable, even when agents are paying attention to all experts.

This focus on experts sets us apart from several other frameworks. For example, although one can understand our setting as one where there is model uncertainty (Gilboa and Schmeidler, 1989, 1993) about the informativeness of signals, understanding it in the context of experts delivers several new insights. This focus also distinguishes us from environments which require heterogeneous priors (e.g., Acemoglu, Chernozhukov and Yildiz, 2016), private signals, or other distortions such as anticipatory utility (Brunnermeier and Parker, 2005; Brunnermeier, Gollier and Parker, 2007) to generate disagreement. Our framework generates disagreement even when agents share the same bias, information content, and priors.

4.2.1 Overconfidence

A large strand of literature studies how agents may misinterpret signals because they misperceive their accuracy, or equivalently how correlated signals are with the underlying state, often due to overconfidence. Scheinkman and Xiong (2003) provide a review of a portion of this literature and argue that disagreement arises because agents “agree to disagree” about how correlated signals are with the state. Models of overconfidence in which the agent has mistaken beliefs about the environment but is otherwise Bayesian predict biased beliefs that are nonetheless invariant to signal and source order. One such example is an agent who mistakenly believes that experts are more reliable than they actually are. Another is an agent who mistakenly interprets a sequence of signals from a single expert as though it were from multiple independent experts. Furthermore, there he would exhibit pessimism (optimism) in the face of consistent (mixed) information content, the opposite of our prediction.

Ortoleva and Snowberg (2015) and Enke and Zimmermann (2016) consider correlation neglect, where agents under-estimate correlation among signals. They predict no inference errors in our environment, where each expert and her signals are independent draws. Disagreement also requires heterogeneity in the information agents observe or in the degree of correlation neglect, whereas agents in our environment can share the same information content and bias.

Despite the importance of overconfidence, its source is often less clear. In our framework, the correlation of signals with the underlying state is precisely what agents are trying to learn. The behavior following positive and negative first impressions resembles over- and under-confidence endogenously. The most related paper that endogenizes overconfidence is Gervais and Odean (2001), where successful traders are overconfident in their financial trading skills due to a form of self-attribution bias. In contrast, our pre-screener is a passive observer who learns about an exogenous state from experts.

4.2.2 Inattention

A growing literature considers individuals who are boundedly rational and have limited ability to process information (Sims, 2003, 2006; Gabaix, Laibson, Moloche and Weinberg, 2006). Schwartzstein (2014) considers a setting where agents must learn to selectively pay attention to variables based on their predictive ability. Since they may persistently fail to incorporate information from signals they mistakenly perceive as inaccurate, they may also over-infer from the signals to which they pay attention, a form of omitted variables bias. Wilson (2014) shows that confirmation bias can arise when agents have bounded memory and can forget the realized history of signals, so might ignore mildly informative signals.

Kominers, Mu and Peysakhovich (2016) assume that agents trade off attention costs and belief accuracy. Because they decide whether or not to pay a cost and internalize observed signals, they screen out uninformative signals with low decision value.[2] Since contrary signals have particularly high value in this framework, agents do not exhibit confirmation bias. While some form of inattention can potentially explain the persistence of a first impression or why overtrust in an insider is difficult to undo by a contrary outsider, it is difficult to reconcile simultaneously with our prediction that prevailing beliefs unravel with an insider's internal contradictions.

Fundamentally, inattention biases revolve around agents not paying enough attention to certain signals. The central feature of our framework is that agents disagree about the credibility of signals *to which they all pay attention.* This is an important distinction which we argue captures the essence of several important real-world disagreements, as we discuss below in Section 5.

4.2.3 Media and Persuasion

The literature has also focused on the role of information supply in disagreement and polarization, by showing that the media will slant news to build reputation (Gentzkow and Shapiro, 2006) or to cater to consumers' preferences for beliefs (Mullainathan and Shleifer, 2005). Because such models assume Bayesian consumers, they imply that signal and source order are irrelevant to final beliefs, given fixed information content.

In these frameworks, exogenous heterogeneous priors drive why the media endogenously respond with biased information even when covering the same consumers. Instead, we ask how heterogeneous priors arise endogenously even if sources are unbiased. Our model suggests that people may perceive media to be more or less credible than they actually are, due to the difficulty of simultaneously learning about source quality and substance. Our focus on biased learning about quality also differentiates us from the large literature on strategic experts (e.g., Hong, Scheinkman and Xiong, 2008) and persuasion more generally (Gentzkow and Kamenica, 2011, 2016).

[2] Kominers et al. (2016) use the term "pre-screen" at one point to describe this. The overlap in terminology is purely accidental. In our framework, pre-screening refers to evaluation of expert quality based on current beliefs.

5 Discussion

5.1 Real-world examples of disagreement

We argue that our theme of disagreement about experts describes essence of several real-world disagreements over several economic questions as well as other important debates. The literature has in general recognized that disagreement is important for welfare (Brunnermeier, Simsek and Xiong, 2013).

Consider the example of disagreement about economics from Section 1. The topic of whether government stimulus promotes growth generates disagreement among both economists (e.g., Krugman, 2009; Cochrane, 2009; The Economist Magazine, 2013) and the public (Sapienza and Zingales, 2013). Regardless of who is right, disagreement amongst the public is likely correlated with which economists they believe are credible. For example, opponents and proponents of stimulus on the editorial pages of the Wall Street Journal and New York Times routinely disagree about the credibility of the opposing side's economists, despite obviously paying attention to what each others' experts say (e.g., Moore, 2011; New York Times, 2014).

Indeed, inattention is unlikely to be a good description of several hot-button economic issues where two sides disagree strongly. The opposition to free trade is quite plausibly due to the belief that economists don't know what they are talking about, rather than simply not noticing what experts have said. Sapienza and Zingales (2013) report that providing economists' opinion that NAFTA increased welfare to the surveyed households changed their opinions very little about its merits, even though surveyors actively told households what experts thought, mitigating inattention. The widespread support among economists for free trade suggests these disagreements are also not well-described by the selective interpretation of ambiguous signals as in Fryer, Harms and Jackson (2016), as there is little ambiguity about the signal. Similarly, one-hundred percent of experts surveyed in Sapienza and Zingales (2013) thought stock prices were hard to predict; yet when told experts' opinions on the topic, survey respondents if anything thought prices were easier to predict.

Our interpretation is simply that trust in economists is low. In areas where economists strongly disagree amongst themselves (stimulus spending), what is legitimate open scientific debate may nevertheless lead to polarized opinions among different factions of the public due to differences in the order in which they were exposed to these experts. Put more simply, learning about a topic early from one expert (say, a teacher or mentor) tends to overly color opinions about the credibility of other experts encountered down the line. In other areas such as free trade where economists agree more, households may "be their own expert" and trust their own experiences over that of outside experts, an interpretation of the model we discuss in the conclusion. Moreover, a lay person may be more open to the anti-trade views of non-economist figures exactly because they distrust how much valuable expertise economists actually provide.

Notably, economics also naturally lends itself to exactly the type of environment where learning is difficult because of uncertainty regarding source quality. Macroeconomics in particular is a technical subject where repeated experimentation to determine expert credibility is difficult as history does not repeat in a controlled environment. In finance, investors often rely on advice from financial analysts whose expertise can be difficult to ascertain. Jia, Wang and Xiong (2016) find that local investors react more to recommendations of local analysts and foreign investors react more to those of foreign analysts, consistent with Propositions 2 and 3. These types of environments are more likely to generate our bias.

The question of whether humans contribute significantly to climate change is another example where uncertainty over source quality may generate disagreement. Disagreement between so-called climate "deniers" and supporters of the proposition is largely about the credibility of the consensus among the scientific community, which supports the proposition, versus a small minority of scientists who do not support the consensus. In this setting, repeated experimentation to verify which experts are more credible is difficult, almost by definition. Inattention would suggest that climate deniers have simply not paid attention to the consensus. More plausibly, climate deniers have paid attention to what mainstream scientists say, yet actively deny their credibility, because they believe an alternative set of "experts," or their own intuition, as our model suggests. The fact that the consensus is so strong may itself contribute to the disagreement (Proposition 7).

A final example of an area prone to bias due to the difficulty of learning about expert quality is medical advice, such as the "debate" over childhood vaccination. The unknown true state is whether that vaccination is safe for a child, but a parent does not know how well her doctor's advice correlates with the true state. There are limited opportunities for repeated experimentation to learn about the doctor's quality, and the credentials are difficult to evaluate. Unlike a medical professional, a lay person is unlikely to understand the difference between medical training and alternative medicine, let alone medical degrees from various schools in various specialties.

In 1998, Andrew Wakefield and co-authors held a press conference describing the results of a study they would publish later that year linking the measles vaccination with autism (Wakefield et al., 1998). Although subsequent research (by others) discredited this research, leading the journal to retract the article, the idea has lingered. The role of expert quality uncertainty is laid bare by the observation that Jennifer McCarthy Wahlberg ("Jenny McCarthy") - an entertainer with no discernible objective expertise in medical science - became an influential figure in the ensuing anti-vaccination movement. Indeed, our model is consistent with the evidence that rumors and misinformation are stubbornly resistant to fact-checking or debunking by outsiders (Berinsky, 2012; Nyhan and Reifler, 2010). Instead, consistent with our model, notable exceptions in which corrections to initial misinformation are successful include retractions from the original source (Simonsohn, 2011) or from sources whose credibility is likely highly correlated with the original source (Berinsky, 2012).

The proliferation of "fake news" over social media platforms such as Facebook and Twitter during the recent U.S. presidential election highlights the importance of uncertain expert quality to learning (The Economist Magazine, 2016). Recent evidence suggests that young individuals in particular may attach credibility to content even when there are obvious indicators that the source is not trustworthy (Stanford History Education Group, 2016; Shellenbarger, 2016, for a summary). Our theory suggests that impressions of credibility - either under-trust in traditional news sources, or overtrust in alternative sources - may be particularly difficult to unwind with this group.

5.2 Testable Implications

Several of our propositions have empirically testable analogues. Propositions 2 and 3 suggest that, in a cross-section of individuals, beliefs about source quality should predict beliefs about a given claim. This effect should also be pronounced when source quality is more uncertain and difficult to objectively ascertain. These ideas are testable given appropriately rich data on individuals' opinions about a topic, where they get their information, and who they get their information from.

Propositions 4 and 5 suggest that individuals who receive an initial sequence of signals from an expert which are consistent should tend to overtrust those same experts and be optimistic about a given topic later in life. Conversely, those who received inconsistent signals should tend to under-trust experts and be pessimistic later in life. These ideas are testable if the data described above were available in a long panel, as these predictions speak to variation within individuals rather than across individuals.

Propositions 6-9 suggest that individuals who overtrust an existing expert tend to discount new sources of information. This suggests the following natural experiment. Suppose individuals had a positive first impression of an expert and are thus also optimistic about a given state. Divide them into two groups: one group which receives subsequent information from an outsider which contradicts the initial expert (the "treatment"), while another group receives the same contradictory information from the initial expert (the "control"). The treatment group should under-trust the outsider and increase their overtrust in the initial expert, and becoming more optimistic about the state, relative to the control group, who should eventually lose trust in the initial expert and reverse their opinions about the state.

Proposition 10 suggests a way to measure our bias experimentally. An experimenter could measure how the beliefs of an individual change in response to a new signal and directly compare them with how a Bayesian's belief would have changed. Proposition 11 suggests a less direct way of testing this in a non-experimental setting. Situations with sufficiently mixed signals should generate under-trust and pessimism, and situations with sufficiently consistent signals should generate overtrust and optimism.

6 Conclusion

We argue that learning in environments where source quality is uncertain is important for understanding the origins of disagreement. If individuals tend to over-infer expert quality, they will disagree with each other about substance because they endogenously disagree about which signals are credible.

This helps us understand why people strongly disagree across several fields ranging across economics, climate science, and medicine, often despite sharing common information. We can also broaden the interpretation of the model to consider how experiences affect beliefs. Although we interpret experts as external sources of signals, an alternative interpretation is that individuals have noisy experiences that inform them about the unknown true state. They learn both about about how accurately their experiences reflect the true state as well as the state itself. Here, the "unknown expert quality" is the reliability of the experience-generating process, which varies by individual.

Malmendier and Nagel (2011) find that experiences affect whether individuals trust the stock market, potentially through a beliefs channel. For example, individuals born in the Depression tend to have lower stock market participation than younger generations born who more recently experienced a boom. Koudijs and Voth (2016) also find that personal experiences affect risk-taking. Malmendier and Nagel (2016) find that differences in experienced inflation explain disagreement about inflation. Our model suggests that people may endogenously trust their own experiences more or less based on the consistency of those experiences, and that this may affect their trust in external sources of information. We leave a deeper exploration of this speculative link for future research.

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz, "Fragility of Asymptotic Agreement under Bayesian Learning," *Theoretical Economics*, 2016, *11*, 187–227.

Berinsky, Adam, "Rumors, Truths, and Reality: A Study of Political Misinformation," 2012. Massachusetts Institute of Technology.

Brunnermeier, Markus K., Alp Simsek, and Wei Xiong, "A welfare criterion for models with distorted beliefs," *Quarterly Journal of Economics*, 2013, *129* (4), 1753–1797.

_ **and Jonathan A. Parker**, "Optimal expectations," *American Economic Review*, 2005, *95* (4), 1092–1118.

_ **, Christian Gollier, and Jonathan A. Parker**, "Optimal beliefs, asset prices, and the preference for skewed returns," *American Economic Review Papers and Proceedings*, 2007, *97*, 159–165.

Carlin, Bradley P. and Thomas A. Louis, "Empirical Bayes: Past, Present and Future," *Journal of the American Statistical Association*, 2000, *95* (452), 1286–1289.

Cochrane, John H., "How did Paul Krugman get it so wrong?," September 16 2009. Available online: https://faculty.chicagobooth.edu/john.cochrane/research/papers/krugman_response.htm [Last accessed: September 2016].

DellaVigna, Stefano and Devin Pope, "Predicting Experimental Results: Who Knows What?," 2016.

Enke, Benjamin and Florian Zimmermann, "Correlation Neglect in Belief Formation," 2016.

Fryer, Roland G., Philipp Harms, and Matthew O. Jackson, "Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization," 2016.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg, "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 2006, *96* (4), 1043–1068.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, *Bayesian Data Analysis*, Boca Raton, Florida: Chapman and Hall/CRC, 2003.

Gentzkow, Matthew and Emir Kamenica, "Bayesian Persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

_ **and** _ , "Bayesian Persuasion with Multiple Senders and Rich Signal Spaces," 2016.

_ **and Jesse M. Shapiro**, "Media Bias and Reputation," *Journal of Political Economy*, 2006, *114* (2), 280–316.

Gervais, Simon and Terrance Odean, "Learning to be Overconfident," *Review of Financial Studies*, 2001, *14*, 1–27.

Gilboa, Itzhak and David Schmeidler, "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, 1989, *18*, 141–153.

_ **and** _ , "Updating Ambiguous Beliefs," *Journal of Economic Theory*, 1993, *59*, 33–49.

Griffin, Dale and Amos Tversky, "The weighing of evidence and the determinants of confidence," *Cognitive Psychology*, July 1992, *24* (3), 411–435.

Hong, Harrison, José A. Scheinkman, and Wei Xiong, "Advisors and asset prices: a model of the origins of bubbles," *Journal of Financial Economics*, 2008, *89* (2), 268–287.

Jia, Chunxin, Yaping Wang, and Wei Xiong, "Market segmentation and differential reactions of local and foreign investors to analyst recommendatons," *Review of Financial Studies*, 2016, *Forthcoming.*

Kominers, Scott Duke, Xiaosheng Mu, and Alexander Peysakhovich, "Paying (for) Attention: The Impact of Information Processing Costs on Bayesian Inference," 2016.

Koudijs, Peter and Hans-Joachim Voth, "Leverage and beliefs: personal experience and risk-taking in margin lending," *American Economic Review*, 2016, *106* (11), 3367–3400.

Krugman, Paul, "How Did Economists Get It So Wrong?," *New York Times Magazine*, September 2 2009. Available online: http://www.nytimes.com/2009/09/06/magazine/06Economic-t.html [Last accessed: September 2016].

Lindley, Dennis Victor, "Compound Decisions and Empirical Bayes: Discussion," *Journal of the Royal Statistical Society*, 1969, *31* (3), 397–425.

Lord, Charles G., Lee Ross, and Mark R. Lepper, "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence," *Journal of Personality and Social Psychology*, 1979, *37* (11), 2098–2109.

Malmendier, Ulrike and Stefan Nagel, "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?," *Quarterly Journal of Economics*, 2011, *126* (1), 373–416.

_ **and** _ , "Learning from Inflation Experiences," *Quarterly Journal of Economics*, 2016, *131* (1), 53–87.

Moore, Stephen, "Why Americans Hate Economists," August 19 2011. http://www.wsj.com/articles/SB10001424053111903596904576514552877388610 [Last accessed: September 2016].

Mullainathan, Sendhil and Andrei Shleifer, "The Market for News," *The American Economic Review*, 2005, *95* (4), 1031–1053.

New York Times, "What the stimulus accomplished," February 22 2014. http://www.nytimes.com/2014/02/23/opinion/sunday/what-the-stimulus-accomplished.html [Last accessed: September 2016].

Nyhan, Brendan and Jason Reifler, "When Corrections Fail: The Persistence of Political Misperceptions," *Political Behavior*, 2010, *32* (2), 303–330.

Ortoleva, Pietro and Erik Snowberg, "Overconfidence in Political Behavior," *American Economic Review*, 2015, *105* (2), 504–535.

Rabin, Matthew and Joel L. Schrag, "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics*, 1999, *114* (1), 37–82.

Sapienza, Paola and Luigi Zingales, "Economic Experts versus Average Americans," *American Economic Review Papers and Proceedings*, 2013, *103* (3), 636–642.

Scheinkman, José A. and Wei Xiong, "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 2003, *111* (6), 1183–1220.

Schwartzstein, Joshua, "Selective Attention and Learning," *Journal of the European Economic Association*, 2014, *12* (6), 1423–1452.

Shellenbarger, Sue, "Most Students Don't Know When News is Fake, Stanford Study Finds," *The Wall Street Journal*, November 21 2016. Available online: http://www.wsj.com/articles/most-students-dont-know-when-news-is-fake-stanford-study-finds-1479752576 [Last accessed: November 2016].

Simonsohn, Uri, "Lessons from an "Oops" at *Consumer Reports*: Consumers Follow Experts and Ignore Invalid Information," *Journal of Marketing Research*, February 2011, *48*, 1–12.

Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50*, 665–690.

_ , "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review*, 2006, *96* (2), 158–163.

Stanford History Education Group, "Evaluating Information: The Cornerstone of Civic Online Reasoning," 2016. Available online: https://sheg.stanford.edu/upload/V3LessonPlans/Executive%20Summary%2011.21.16.pdf [Last accessed: November 2016].

The Economist Magazine, "Sovereign doubts," September 28 2013. Available online: http://www.economist.com/news/schools-brief/21586802-fourth-our-series-articles-financial-crisis-looks-surge-public [Last accessed: September 2016].

_ , "The role of technology in the presidential election," November 20 2016. Available online: http://www.economist.com/news/united-states/21710614-fake-news-big-data-post-mortem-under-way-role-technology [Last accessed: November 2016].

Wakefield, AJ, SH Murch, A Anthony, J Linnell, DM Casson, M Malik, M Berelowitz, AP Dhillon, MA Thomson, P Harvey, A Valentine, SE Davies, and JA Walker-Smith, "Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children," *The Lancet*, 1998, *351*, 637–641.

Wilson, Andrea, "Bounded Memory and Biases in Information Processing," *Econometrica*, 2014, *82* (6), 2257–2294.

A Appendix

A.1 Generalized Pre-Screening

Let $\omega_0^{q\theta}$ be the prior belief on quality q and state θ, where $\sum_q \sum_\theta \omega_0^{q\theta} = 1$.

When the prior beliefs about the quality and state can potentially be correlated, we cannot apply the first-stage updated belief $\kappa_q(\mathbf{s}^n)$, which is a marginal belief on quality, directly to the second stage in place of a prior belief on quality because the joint priors on quality and state are not independent. Therefore, the generalized pre-screening algorithm requires the second stage to apply Bayes' Rule to a belief whose marginal prior about quality sums to $\kappa(\mathbf{s}^n)$. Thus, in the second stage, we assume that the agent applies the *weighted* first-stage updated belief $\kappa_q(\mathbf{s}^n)\left(\frac{\omega_0^{q\theta}}{\sum_\theta \omega_0^{q\theta}}\right)$ in place of the existing joint prior. If the prior beliefs about quality and state are independent ($\omega_0^{q\theta} = \omega_0^q \omega_0^\theta$ for all q and θ), then Equations (11), (12), and (13) reduce to Equations (1), (4), and (5), respectively.

To illustrate the pre-screener's updating algorithm, suppose he observes two signals, one in each period. After observing the first signal (s_1), the biased agent's updated belief about the expert's quality, $\kappa_q(s_1)$, is:

$$\kappa_q(s_1) = \frac{\sum_\theta P(s_1|q,\theta)\omega_0^{q\theta}}{\sum_q \sum_\theta \sum_\theta P(s_1|q,\theta)\omega_0^{q\theta}}.$$

Using the weighted first-stage updated belief $\kappa_q(s_1)\left(\frac{\omega_0^{q\theta}}{\sum_\theta \omega_0^{q\theta}}\right)$ to form his joint posterior belief on the state and quality, $P^b(q,\theta|s_1)$, yields his posterior beliefs after the first signal:

$$P^b(q,\theta|s_1) = \frac{P(s_1|q,\theta)\kappa_q(s_1)\left(\frac{\omega_0^{q\theta}}{\sum_\theta \omega_0^{q\theta}}\right)}{\sum_q \sum_\theta P(s_1|q,\theta)\kappa_q(s_1)\left(\frac{\omega_0^{q\theta}}{\sum_\theta \omega_0^{q\theta}}\right)}.$$

After observing the second signal (s_2), the biased agent's updated belief about the expert's quality, $\kappa_q(s_1,s_2)$ is

$$\kappa_q(s_1,s_2) = \frac{\sum_\theta P(s_2|q,\theta)P^b(q,\theta|s_1)}{\sum_q \sum_\theta P(s_2|q,\theta)P^b(q,\theta|s_1)}.$$

Using the weighted first-stage updated belief $\kappa(s_1,s_2)\left(\frac{\omega_0^{q\theta}}{\sum_\theta \omega_0^{q\theta}}\right)$ to form his joint posterior belief on the state and quality, $P^b(q,\theta|s_1,s_2)$, yields:

$$P^b(q,\theta|s_1,s_2) = \frac{P(s_2|q,\theta)P(s_1|q,\theta)\kappa_q(s_1,s_2)\left(\frac{\omega_0^{q\theta}}{\sum_\theta \omega_0^{q\theta}}\right)}{\sum_q \sum_\theta P(s_2|q,\theta)P(s_1|q,\theta)\kappa_q(s_1,s_2)\left(\frac{\omega_0^{q\theta}}{\sum_\theta \omega_0^{q\theta}}\right)}.$$

Iterating on the biased agent's updating process allows us to characterize his posterior beliefs:

Applying the generalized pre-screening procedure described above to prior beliefs $\omega_0^{q\theta}$ yields:

$$\kappa_q(\mathbf{s}^n) = \frac{\left(\frac{\kappa_q(\mathbf{s}^{n-1})}{\sum_\theta \omega_0^{q\theta}}\right)\sum_\theta\left(\prod_{t=1}^n P(s_t|q,\theta)\omega_0^{q\theta}\right)}{\sum_q \left(\frac{\kappa_q(\mathbf{s}^{n-1})}{\sum_\theta \omega_0^{q\theta}}\right)\sum_\theta\left(\prod_{t=1}^n P(s_t|q,\theta)\omega_0^{q\theta}\right)}, \tag{10}$$

where $\kappa_q(\emptyset) = \sum_\theta \omega_0^{q\theta}$.

$$\begin{aligned} P^b(q,\theta|\mathbf{s}^n) &= \frac{\left(\prod_{t=1}^n P(s_t|q,\theta)\right)\left(\frac{\kappa_q(\mathbf{s}^n)}{\sum_\theta \omega_0^{q\theta}}\right)\omega_0^{q\theta}}{\sum_q\sum_\theta\left(\prod_{t=1}^n P(s_t|q,\theta)\right)\left(\frac{\kappa_q(\mathbf{s}^n)}{\sum_\theta \omega_0^{q\theta}}\right)\omega_0^{q\theta}} & (11) \\ &= \frac{b_{q\theta}(\mathbf{s}^n)\left(\frac{1}{\sum_\theta \omega_0^{q\theta}}\right)^n\left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^{q\theta}}{\sum_q\left(\frac{1}{\sum_\theta \omega_0^{q\theta}}\right)^n\sum_\theta b_{q\theta}(\mathbf{s}^n)\left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^{q\theta}}. & (12) \end{aligned}$$

where $b_{q\theta}(\mathbf{s}^n)$ is given by:

$$\begin{aligned} b_{q\theta}(\mathbf{s}^n) &= \left(\sum_\theta P(s_1|q,\theta)\omega_0^{q\theta}\right)\times\left(\sum_\theta P(s_1|q,\theta)P(s_2|q,\theta)\omega_0^{q\theta}\right)\times\ldots\times\left(\sum_\theta P(s_1|q,\theta)P(s_2|q,\theta)\ldots P(s_n|q,\theta)\omega_0^{q\theta}\right) \\ &= \prod_{m=1}^n\left(\sum_\theta\left(\prod_{t=1}^m P(s_t|q,\theta)\right)\omega_0^{q\theta}\right), & (13) \end{aligned}$$

A.2 Proof of Proposition 1

Define $D(\mathbf{s}^n) = P^b(\theta = A|\mathbf{s}^n) - P(\theta = A|\mathbf{s}^n)$ as the ex-post realized disagreement after any signal path. The proposition is that $E_0[D(\mathbf{s}^n)] = 0$, where the expectation E_0 is taken by the econometrician over the common prior of states and quality, which we assume reflects the true ex ante distribution of (θ, q). Note that the common prior on states and quality generate a common distribution on the probability of any given signal path.

Divide the set of all possible signal paths $\{\mathbf{s}^n\}$ into two groups: one group $\{\mathbf{g}^n\}$ where the first signal is a and another group $\{\mathbf{h}^n\}$ where the first signal is b. Because there are two states, there are the same number of signal paths in each group, and the union of these two groups equals $\{\mathbf{s}^n\}$.

It is clear that taking any signal path $\mathbf{g}^n$ and flipping all the a's to b and b's to a defines a one-to-one and onto mapping F of $\{\mathbf{g}^n\}$ into $\{\mathbf{h}^n\}$. This mapping has two properties:

1. $P(\mathbf{g}^n|q,\theta) = P(F(\mathbf{g}^n)|q,-\theta)\ \forall(q,\theta)$, and

2. $P^b(\theta = A|F(\mathbf{g}^n)) - P(\theta = A|F(\mathbf{g}^n)) = -\left(P^b(\theta = A|\mathbf{g}^n) - P(\theta = A|\mathbf{g}^n)\right)\ \forall\mathbf{g}^\mathbf{n}$,

where $-\theta$ is the opposite state as θ. The first property says that the probability of the flipped signal sequence is the same as the original signal sequence, once the true state is flipped. The second property can be re-written as $D(F(\mathbf{g}^n)) = -D(\mathbf{g}^n)$ and says that disagreement under the flipped signal path equals the opposite disagreement under the original signal path. Intuitively, these properties follow because, starting from a neutral prior about the state which is independent from quality, the model is symmetric in A and B irrespective of the true expert type.

More precisely, the first property follows because:

$$\begin{aligned} P(\mathbf{g}^n|q, A) &= p_q^{n_a^g}(1-p_q)^{n_b^g} \\ &= p_q^{n_b^h}(1-p_q)^{n_a^h} \\ &= P(F(\mathbf{g}^n))|q, B), \end{aligned}$$

where n_θ^g, n_θ^h represent the number of times a signal indicating state θ appears in signal sequence $\mathbf{g}^n$ and $F(\mathbf{g}^n)$, respectively, and $n_a^g = n_b^h, n_b^g = n_a^h$ by construction. Similarly, $P(\mathbf{g}^n|q, B) = P(F(\mathbf{g}^n)|q, A)$.

To prove the second property, note that, for the Bayesian, $P(\theta = A|\mathbf{g}^n) = P(\theta = B|F(\mathbf{g}^n)) = 1 - P(\theta = A|F(\mathbf{g}^n))$. The first equality follows from applying the first property and $\omega_0^A = \omega_0^B = 0.5$ to Equation 4, noting that a Bayesian has constant $b_q(F(\mathbf{g}^n))$.

Now consider the biased individual. Given any sequence $\mathbf{g}^n$, let g_i and h_i be the i-th elements of $\mathbf{g}^n$ and $F(\mathbf{g}^n)$, respectively. Clearly, h_i is the flip of g_i, and $P(g_i|q, \theta) = P(h_i|q, -\theta)$, as both equal p_q if $g_i = \theta$ and $1-p_q$ if $g_i = -\theta$. Therefore, $\sum_\theta \left(\prod_{i=1}^m P(g_i|q, \theta)\right)\omega_0^\theta = \sum_\theta \left(\prod_{i=1}^m P(h_i|q, \theta)\right)\omega_0^\theta$ for any m due to the summation over both values of θ. Applying this to Equation 5, $b_q(\mathbf{g}^n) = b_q(F(\mathbf{g}^n))$. From Equation 4, $P^b(\theta = A|\mathbf{g}^n) = P^b(\theta = B|F(\mathbf{g}^n)) = 1 - P^b(\theta = A|F(\mathbf{g}^n))$, and the second property follows.

We claim that $E_0[D(\mathbf{s}^n)|q = \overline{q}] = 0$ for any $\overline{q} \in (0, 1)$. To be clear, this conditional expectation is taken over the econometrician's information set, but the true q remains unknown to the Bayesian and pre-screener. The proposition then follows due to the tower property of conditional expectations.

Let $\overline{q}$ be given. Observe that:

$$\begin{aligned} E[D(\mathbf{s}^n)|q = \overline{q}] = \omega_0^A &\left(\sum_{\{\mathbf{g}^n\}} P(\mathbf{g}^n|\overline{q}, A)D(\mathbf{g}^n) + \sum_{\{\mathbf{h}^n\}} P(\mathbf{h}^n|\overline{q}, A)D(\mathbf{h}^n) \right) \\ &+ \omega_0^B \left(\sum_{\{\mathbf{g}^n\}} P(\mathbf{g}^n|\overline{q}, B)D(\mathbf{g}^n) + \sum_{\{\mathbf{h}^n\}} P(\mathbf{h}^n|\overline{q}, B)D(\mathbf{h}^n) \right). \end{aligned}$$

The two properties, along the fact that F is one-to-one and onto, imply:

$$\sum_{\{\mathbf{h}^n\}} P(\mathbf{h}^n|\overline{q}, A)D(\mathbf{h}^n) = -\sum_{\{\mathbf{g}^n\}} P(\mathbf{g}^n|\overline{q}, B)D(\mathbf{g}^n)$$
$$\sum_{\{\mathbf{h}^n\}} P(\mathbf{h}^n|\overline{q}, B)D(\mathbf{h}^n) = -\sum_{\{\mathbf{g}^n\}} P(\mathbf{g}^n|\overline{q}, A)D(\mathbf{g}^n).$$

With $\omega_0^A = \omega_0^B$, the claim follows. The corollary $Var_0[D(\mathbf{s}^n)] > 0$ follows because $D(F(\mathbf{g}^n))^2 = D(\mathbf{g}^n)^2$.

A.3 Proof of Proposition 2

Lemma 2 *For all $\omega_0^\theta \in (0,1)$ and $\omega_0^q \in (0,1)$, $\kappa_H(\mathbf{s}^n) < w_0^H$ if and only if $b_H(\mathbf{s}^n) < b_L(\mathbf{s}^n)$. Likewise, $\kappa_H(\mathbf{s}^n) > w_0^H$ if and only if $b_H(\mathbf{s}^n) > b_L(\mathbf{s}^n)$. $\kappa_H(\mathbf{s}^n) = w_0^H$ if and only if $b_H(\mathbf{s}^n) = b_L(\mathbf{s}^n)$.*

Proof. For any given sequence of signals $\mathbf{s}^n = (s_1, s_2, \ldots, s_n)$, $\kappa_q(\mathbf{s}^n)$ can be re-written as

$$\begin{aligned}
\kappa_q(\mathbf{s}^n) &= \frac{b_q(\mathbf{s}^{n-1})\omega_0^q \sum_\theta \left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^\theta}{\sum_q b_q(\mathbf{s}^{n-1}) \sum_\theta \left(\prod_{t=1}^n P(s_n|q,\theta)\right)\omega_0^\theta \omega_0^q} \\
&= \frac{\left(\prod_{m=1}^{n-1}\left(\sum_\theta \left(\prod_{t=1}^m P(s_t|q,\theta)\right)\omega_0^\theta\right)\right)\omega_0^q \sum_\theta \left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^\theta}{\sum_q \left(\prod_{m=1}^{n-1}\left(\sum_\theta \left(\prod_{t=1}^m P(s_t|q,\theta)\right)\omega_0^\theta\right)\right)\sum_\theta \left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^\theta \omega_0^q} \\
&= \frac{b_q(\mathbf{s}^n)\omega_0^q}{\sum_q b_q(\mathbf{s}^n)\omega_0^q}.
\end{aligned}$$

Thus, the statement is shown for $\mathbf{s}^n = (s_1, s_2, \ldots, s_t)$.[3] ■

From Equation (4), the biased agent's posterior that the expert is high quality is lower than the Bayesian's if and only if $\kappa_H(\mathbf{s}^n) < w_0^H$. Lemma 2 shows that this is only the case if and only if $b_H(\mathbf{s}^n) < b_L(\mathbf{s}^n)$. Thus, $b_H(\mathbf{s}^n) < b_L(\mathbf{s}^n)$ if and only if $P^b(q = H|\mathbf{s}^n) < P^u(q = H|\mathbf{s}^n)$.

Consider $P^b(\theta = A|\mathbf{s}^n) < P^u(\theta = A|\mathbf{s}^n)$:

$$P^b(\theta = A|\mathbf{s}^n) < P^u(\theta = A|\mathbf{s}^n)$$
$$\frac{\omega_0^A \sum_q b_q(\mathbf{s}^n)\left(\prod_{t=1}^n P(s_t|q,A)\right)\omega_0^q}{\sum_q b_q(\mathbf{s}^n)\sum_\theta \left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^\theta \omega_0^q} < \frac{\omega_0^A \sum_q \left(\prod_{t=1}^n P(s_t|q,A)\right)\omega_0^q}{\sum_q \sum_\theta \left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^\theta \omega_0^q},$$

[3] If the signals are observed simultaneously (e.g., in period 1), then the above argument applies analogously, where $b_q(\mathbf{s}^{t-1}) = b_q(\emptyset) = 1$ instead. Thus, $\kappa_H(\mathbf{s}^n) < w_0^H$ implies $b_H(\mathbf{s}^n) < b_L(\mathbf{s}^n)$ and vice versa. Likewise when the inequality reverses or when the equality holds.

which is true if and only if

$$0 < \omega_0^A(1-\omega_0^A)\omega_0^H(1-\omega_0^H)(b_L(\mathbf{s}^n) - b_H(\mathbf{s}^n))\left(\left(\prod_{t=1}^n P(s_t|H,A)\right)\left(\prod_{t=1}^n P(s_t|L,B)\right) - \left(\prod_{t=1}^n P(s_t|H,B)\right)\left(\prod_{t=1}^n P(s_t|L,A)\right)\right)$$
$$0 < \omega_0^A(1-\omega_0^A)\omega_0^H(1-\omega_0^H)(b_L(\mathbf{s}^n) - b_H(\mathbf{s}^n))\left(p_H^{n_a}(1-p_H)^{n_b}p_L^{n_b}(1-p_L)^{n_a} - p_H^{n_b}(1-p_H)^{n_a}p_L^{n_a}(1-p_L)_b^n\right)$$
$$0 < \omega_0^A(1-\omega_0^A)\omega_0^H(1-\omega_0^H)(b_L(\mathbf{s}^n) - b_H(\mathbf{s}^n))\left(p_H^{n_a}(1-p_H)^{n_b}p_L^{n_b}(1-p_L)^{n_a} - p_H^{n_b}(1-p_H)^{n_a}p_L^{n_a}(1-p_L)_b^n\right)$$
$$0 < \omega_0^A(1-\omega_0^A)\omega_0^H(1-\omega_0^H)(b_L(\mathbf{s}^n) - b_H(\mathbf{s}^n))\,(p_H(1-p_H)p_L(1-p_L))^{n_b}\left((p_H(1-p_L))^{n_a-n_b} - ((1-p_H)p_L)^{n_a-n_b}\right),$$

which is true when $n_a > n_b$ since $p_H > p_L$. Clearly, $P^u(A|\mathbf{s}^n) > \frac{1}{2}$ only if $n_a > n_b$, so A is the (objectively) more likely state. Note that if $n_a = n_b$, then $P^b(\theta = A|\mathbf{s}^n) = Pr^u(\theta = A|\mathbf{s}^n)$ regardless of the biased agent's beliefs on the expert's quality. Thus, for any $n_a > n_b$ set of signals and for all $\omega_0^\theta \in (0,1)$, under-trust in expert quality implies pessimism in beliefs about the more likely state: If $P^b(q = H|\mathbf{s}^n) < P^u(q = H|\mathbf{s}^n)$, then $P^b(\theta = A|\mathbf{s}^n) < P^u(\theta = A|\mathbf{s}^n)$. Likewise, $P^b(\theta = A|\mathbf{s}^n) < P^u(\theta = A|\mathbf{s}^n)$ if and only if $b_H(\mathbf{s}^n) < b_L(\mathbf{s}^n)$ when $n_a > n_b$, which implies that $P^b(q = H|\mathbf{s}^n) < P^u(q = H|\mathbf{s}^n)$. Reversing the inequalities yields that overtrust in expert quality implies optimism in beliefs about the more likely state, and vice versa.

A.4 Proof of Proposition 3

Let n_a^s be the number of a signals and n_b^s be the number of b signals in sequence s. Consider any two sequences $\mathbf{x}_n$ and $\mathbf{y}_n$ with identical information content ($n_a^x = n_a^y$ and $n_b^x = n_b^y$). Let b_q^s correspond to sequence $s \in \{\mathbf{x}_n, \mathbf{y}_n\}$ and $q \in \{L, H\}$. Without loss of generality, let $n_a > n_b$.

1. Correlated disagreement

   Proof. Let $\mathbf{s}_J^n = \mathbf{x}_n$ and $\mathbf{s}_M^n = \mathbf{y}_n$. By direct comparison of the posteriors on expert quality, a necessary and sufficient condition for sequence x to generate more trust than sequence y (i.e., $P^b(q = H|\mathbf{x}^n) > P^b(q = H|\mathbf{y}^n)$) is $b_H^x b_L^y - b_L^x b_H^y > 0$. By direct comparison of the posteriors on the most likely state (which is A because $n_a > n_b$), a necessary and sufficient condition for the belief in A to be greater after observing sequence x than after observing sequence y (i.e., $P^b(\theta = A|\mathbf{x}^n) > P^b(\theta = A|\mathbf{y}^n)$) is $b_H^x b_L^y - b_L^x b_H^y > 0$. Since the same condition $b_H^x b_L^y - b_L^x b_H^y > 0$ is required for both $P^b(q = H|\mathbf{x}^n) > P^b(q = H|\mathbf{y}^n)$ and $P^b(\theta = A|\mathbf{x}^n) > P^b(\theta = A|\mathbf{y}^n)$, then disagreement between biased agents is correlated. That is, $P^b(q = H|\mathbf{x}^n) > P^b(q = H|\mathbf{y}^n)$ if and only if $P^b(\theta = A|\mathbf{x}^n) > P^b(\theta = A|\mathbf{y}^n)$. Clearly reversing all the inequalities applies as well.

   ■

2. Expected disagreement

   Proof. Let $\mathbf{s}_J^n = \mathbf{x}_n$ and $\mathbf{s}_M^n = \mathbf{y}_n$. It is sufficient to show that $P^b(\theta = A|\mathbf{x}^n) \neq P^b(\theta = A|\mathbf{y}^n)$ for at least two sequences $\mathbf{x}_n$ and $\mathbf{y}_n$ with identical information content ($n_a^x = n_a^y = n_a$ and $n_b^x = n_b^y = n_b$). Consider two sequences such that $n_a \geq n_b$, where the first $j = n_a + n_b$

signals are the same and $n-2 \geq j \geq 1$, the two sequences differ in the $j+1$ and $j+2$ signals, and then all subsequent signals are identical (i.e., terms $j+3$ through n). Let $x_{j+1} = a$, $x_{j+2} = b$, $y_{j+1} = b$, and $y_{j+2} = a$. As already shown in the proof of Lemma 1, $P^b(q = H|\mathbf{x}^n) > P^b(q = H|\mathbf{y}^n)$ when $n_a > n_b$. As shown in the preceding proof of correlated disagreement between two prescreeners, this implies that $P^b(\theta = A|\mathbf{x}^n) > P^b(\theta = A|\mathbf{y}^n)$. Thus, $P^b(\theta = A|\mathbf{x}^n) \neq P^b(\theta = A|\mathbf{y}^n)$ for at least two sequences $\mathbf{x}_n$ and $\mathbf{y}_n$ with identical information content ($n_a^x = n_a^y = n_a$ and $n_b^x = n_b^y = n_b$).

■

A.5 Proof of Lemma 1

Let n_a^s be the number of a signals and n_b^s be the number of b signals in sequence s. Given any two sequences $\mathbf{x}_n$ and $\mathbf{y}_n$ with identical information content ($n_a^x = n_a^y = n_a$ and $n_b^x = n_b^y = n_b$), by direct comparison of the posteriors on expert quality, a necessary and sufficient condition for sequence x to generate more trust than sequence y (i.e., $P^b(q = H|\mathbf{x}^n) > P^b(q = H|\mathbf{y}^n)$) is $b_H^x b_L^y - b_L^x b_H^y > 0$, where b_q^s corresponds to sequence $s \in \{\mathbf{x}_n, \mathbf{y}_n\}$ and $q \in \{L, H\}$. Consider two sequences such that $n_a \geq n_b$, where the first $j = n_a + n_b$ signals are the same and $n-2 \geq j \geq 1$, the two sequences differ in the $j+1$ and $j+2$ signals, and then all subsequent signals are identical (i.e., terms $j+3$ through n). Let $x_{j+1} = a$, $x_{j+2} = b$, $y_{j+1} = b$, and $y_{j+2} = a$. (For example, sequence 1 could be *aababaa* and sequence 2 could be *aabbaaa* - here $j = 3$, $n_a = 2$, $n_b = 1$.) Then $b_H^x b_L^y - b_L^x b_H^y > 0$ whenever $n_a > n_b$ and $b_H^x b_L^y - b_L^x b_H^y = 0$ whenever $n_a = n_b$. To see this, note that, given the general expression for b_q^s, all of the terms are identical for b_q^x and b_q^y except term $j+1$. This implies that when $\omega_0^\theta = 0.5$, then $b_H^x b_L^y - b_L^x b_H^y \geq 0$ if

$$\left(p_H^{n_a+1}(1-p_H)^{n_b} + (1-p_H)^{n_a+1}p_H^{n_b}\right)\left(p_L^{n_a}(1-p_L)^{n_b+1} + (1-p_L)^{n_a}p_L^{n_b+1}\right) -$$
$$\left(p_L^{n_a+1}(1-p_L)^{n_b} + (1-p_L)^{n_a+1}p_L^{n_b}\right)\left(p_H^{n_a}(1-p_H)^{n_b+1} + (1-p_H)^{n_a}p_H^{n_b+1}\right) \geq 0$$
$$(p_H - p_L)\left(p_H^{n_a}(1-p_H)^{n_b}p_L^{n_a}(1-p_L)^{n_b} - p_H^{n_b}(1-p_H)^{n_a}p_L^{n_b}(1-p_L)^{n_a}\right)$$
$$+\left(p_H p_L - (1-p_H)(1-p_L)\right)\left(p_H^{n_a}(1-p_H)^{n_b}p_L^{n_b}(1-p_L)^{n_a} - p_H^{n_b}(1-p_H)^{n_a}p_L^{n_a}(1-p_L)^{n_b}\right) \geq 0.$$

We can verify that both terms are positive when $n_a > n_b$ and zero when $n_a = n_b$. (Note that we can easily verify that this order effect holds for any $\omega_0^\theta \geq 0.5$.) Thus, $P^b(H|\mathbf{x}^n) > P^b(H|\mathbf{y}^n)$ when $n_a > n_b$.

Using Proposition 1, we can basically iteratively apply this fact to order sequences of fixed composition in decreasing trust by starting with the sequence with the least reversals (all a's followed by all b's), and iteratively switching the first b and last a to generate sequences where the first b moves forward. E.g., *aaaabb* generates more trust than *aaabab*, which generates more trust than *aabaab* which generates more trust than *abaaab*. Then, *aaabba* generates more trust than *aababa* than *abaaba*, where *aaabab* generates more trust than *aaabba* and *abaaab* generates more trust than *abaaba*. We can keep doing this (and applying Proposition 1) to establish that *aaaabb* generates the most trust and *ababaa* generates the least trust.

A.6 Proof of Proposition 4

1. Positive first impressions

   Proof. Suppose the agent observes $n_a \geq 1$ consecutive a signals, followed by m pairs of (b, a) signals: $\mathbf{s}^n = (a, a, a, \ldots, b, a, b, a)$. This sequence generates:

$$b_q(\mathbf{s}^n) = \left(\frac{1}{2}\right)^{n_a+m} [p_q(1-p_q)]^{m(m+1)} \left([p_q^{n_a-1} + (1-p_q)^{n_a-1}][p_q^{n_a} + (1-p_q)^{n_a}]\right)^m \left(\prod_{i=1}^{n_a}(p_q^i + (1-p_q)^i)\right). \tag{14}$$

   Further,

$$\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} = \left(\frac{1}{2}\right)^{n_a+m} [p_q(1-p_q)]^{m(m+1)} \left([p_q^{n_a-1} + (1-p_q)^{n_a-1}][p_q^{n_a} + (1-p_q)^{n_a}]\right)^{m-1} \left(\prod_{i=1}^{n_a}(p_q^i(1-p_q)^i)\right)$$
$$\left(mp_q(1-p_q)\left((n_a-1)(p_q^{n_a-2} - (1-p_q)^{n_a-2})(p_q^{n_a} + (1-p_q)^{n_a}) + n_a(p_q^{n_a-1} + (1-p_q)^{n_a-1})(p_q^{n_a-1} - (1-p_q)^{n_a-1})\right)\right.$$
$$\left. + (p_q^{n_a-1} + (1-p_q)^{n_a-1})(p_q^{n_a} + (1-p_q)^{n_a}) \left(m(m+1)(1-2p_q) + p_q(1-p_q)\sum_{i=1}^{n_a} \frac{i(p_q^{i-1} - (1-p_q)^{i-1})}{p_q^i + (1-p_q)^i}\right)\right). \tag{15}$$

   Since b_q is symmetric about $p_q = 1/2$, we focus on the case of $p_q \geq 1/2$ but the conclusion extends to $1/2 > p_L > 1 - p_H$. From Equations (14) and (15), we can see that

   - $\frac{\partial}{\partial p_q}(b_q(\mathbf{s}^n)) = 0$ when $p_q \in \{\frac{1}{2}, 1\}$,
   - $b_q(\mathbf{s}^n)) > 0$ when $p_q = \frac{1}{2}$,
   - $b_q(\mathbf{s}^n)) = 0$ when $p_q = 1$.

   Moreover, using the fact that $b_q(\mathbf{s}^n) = 0$ when $p_q = 1/2$, then

$$\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}} = b_q(\mathbf{s}^n)[p_q(1-p_q)(p_q^{n_a} + (1-p_q)^{n_a})(p_q^{n_a-1} + (1-p_q)^{n_a-1})]^{-1}$$
$$\left(\frac{1}{2}\right)^{2n_a-3} \left(2m[(n_a-1)^2 - (m+1)] + \frac{1}{3}n_a(n_a-1)(n_a+1)\right) \tag{16}$$

   Note that the last term of Equation (16) increases in n_a for all $n_a > 1$, and that it is positive for all $m > m'$ where

$$2m'[(n_a-1)^2 - (m'+1)] + \frac{1}{3}n_a(n_a-1)(n_a+1) = 0.$$

By direct computation, we can see that $\frac{\partial b_q}{\partial p_q} < 0$ for all $m \geq 1$ and $p_q \in (1/2, 1)$ for $n_a \in \{1, 2\}$. This implies that the agent under-trusts and is pessimistic about the most likely state for all $m \geq 1$ when $n_a \leq 2$.

Consider the case of $n_a \geq 3$. Since $b_q(\mathbf{s}^n) > 0$ when $p_q = 1/2$, $b_q(\mathbf{s}^n) = 0$ when $p_q = 1$, $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q}\big|_{p_q=1/2} = 0$, and $b_q(\mathbf{s}^n) \geq 0$ for any $p_q \in [0, 1]$, then there exists some threshold $\frac{1}{2} < p' < 1$ such that $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} > 0$ for all $p_q < p'$ when $\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}} > 0$, which holds when $m < m'$. This implies that the agent overtrusts and is optimistic about the most likely state when $m < m'$ and $p_L < p_H \leq p'$. Since the last term of Equation (16) increases in n_a for all $n_a > 1$ and $\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}} > 0$ for $m \leq 3$, then $m' > 3$ for all $n_a \geq 3$.

■

2. Negative first impressions

   Proof. Suppose the agent observes $n_b \geq 1$ pairs of (a, b) signals, followed by $m \geq 1$ consecutive a signals, where $m \geq 1$: $\mathbf{s}^n = (a, b, a, b, \ldots, a, a, a)$. This sequence generates:

$$b_q(\mathbf{s}^n) = \left(\frac{1}{2}\right)^{n_b} (p_q(1-p_q))^{n_b^2} \left(\prod_{i=1}^{m} \frac{1}{2}\left(p_q^{i+n_b}(1-p_q)^{n_b} + p_q^{n_b}(1-p_q)^{i+n_b}\right)\right) \tag{17}$$

$$= \left(\frac{1}{2}\right)^{n_b+m} (p_q(1-p_q))^{n_b(n_b+m)} \left(\prod_{i=1}^{m} (p_q^i + (1-p_q)^i)\right). \tag{18}$$

Further,

$$\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} = \left(\frac{1}{2}\right)^{n_b+m} (p_q(1-p_q))^{n_b(n_b+m)-1} \left(\prod_{i=1}^{m} (p_q^i + (1-p_q)^i)\right) \left(n_b(n_b+m)(1-2p_q) + p_q(1-p_q)\sum_{i=1}^{m}\left(\frac{i(p_q^{i-1} - (1-p_q)^{i-1}}{p_q^i + (1-p_q)^i}\right)\right) \tag{19}$$

Since b_q is symmetric about $p_q = 1/2$, we focus on the case of $p_q \geq 1/2$ but the conclusion extends to $1/2 > p_L > 1 - p_H$. Evaluating Equation 20 when $m = 1$ (i.e., $n_a = n_b + 1$ where $n_b \geq 1$), $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q}|_{m=1} < 0$. Thus, by Proposition 2, the pre-screener under-trusts and is pessimistic about the mostly likely state, A, when he observes a sequence $\mathbf{s}^n = (a, b, a, b, \ldots, a, b, a)$ where $n_a = n_b = 1$. Further, evaluating Equation 20 when $m = 2$ (i.e., $n_a = n_b + 2$ where $n_b \geq 1$), $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q}|_{m=1} < 0$. Thus, the pre-screener still under-trusts and is pessimistic about the most likely state, A, when $\mathbf{s}^n = (a, b, a, b, \ldots, a, a)$ where $n_a = n_b + 2$ for all $n_b \geq 1$. Further, evaluating Equation 20 when $m = 3$ (i.e., $n_a = n_b + 3$ where $n_b \geq 1$), $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q}|_{m=3} \leq 0$ with equality at $p_q = \frac{1}{2}$ only if $n_b = 1$. Since the third term of Equation (20) is decreasing in n_b for all $p_q \in (\frac{1}{2}, 1]$, then $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q}|_{m=3} < 0$ for all $n_b > 1$. Thus, the pre-screener still under-trusts and is pessimistic about the most likely state, A, when $\mathbf{s}^n = (a, b, a, b, \ldots, a, a)$ where

$n_a = n_b + 3$ for all $n_b \geq 1$. Therefore, there exists some $m^* > 3$ such that $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} < 0$ for all $m < m^*$, which implies that the pre-screener will under-trust for $m < m^*$. Moreover, since the third term of Equation (20) is decreasing in n_b for all $p_q \in (\frac{1}{2}, 1]$, then m^* is increasing in n_b. ■

A.7 Proof of Proposition 5

1. **Proof.** Since b_q is symmetric about $p_q = 1/2$, we focus on the case of $p_q \geq 1/2$ but the conclusion extends to $1/2 > p_L > 1 - p_H$. Note that Equation (15) can be re-written as

$$\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} = \left(\frac{1}{2}\right)^{n_a+m} m[p_q(1-p_q)]^{m(m+1)} \left([p_q^{n_a-1} + (1-p_q)^{n_a-1}][p_q^{n_a} + (1-p_q)^{n_a}]\right)^{m-1} \left(\prod_{i=1}^{n_a}(p_q^i(1-p_q)^i)\right) \Bigg($$

where $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q}$ is negative whenever Z is negative and $p_q \in (\frac{1}{2}, 1)$, and

$$Z = p_q(1-p_q)\left((n_a-1)(p_q^{n_a-2} - (1-p_q)^{n_a-2})(p_q^{n_a} + (1-p_q)^{n_a}) + n_a(p_q^{n_a-1} + (1-p_q)^{n_a-1})(p_q^{n_a-1} - (1-p_q)^{n_a-1})\right)$$
$$+ (p_q^{n_a-1} + (1-p_q)^{n_a-1})(p_q^{n_a} + (1-p_q)^{n_a}) \left((m+1)(1-2p_q) + \left(\frac{1}{m}\right) p_q(1-p_q) \sum_{i=1}^{n_a} \frac{i(p_q^{i-1} - (1-p_q)^{i-1})}{p_q^i + (1-p_q)^i}\right).$$

For given n_a, Z is more than linearly decreasing in m. Thus, there exists $\hat{m}$, defined by $Z(\hat{m}) = 0$, such that $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} < 0$ for all $p_q \in (\frac{1}{2}, 1)$ when $m > \hat{m}$. Thus for any given n_a, there exists $\hat{m}$ such that when $m > \hat{m}$, the pre-screener under-trusts and is pessimistic about the most likely state for any (p_L, p_H). ■

2. **Proof.** Since b_q is symmetric about $p_q = 1/2$, we focus on the case of $p_q \geq 1/2$ but the conclusion extends to $1/2 > p_L > 1 - p_H$. From Equations (18) and (20), we can see that

   - $\frac{\partial}{\partial p_q}(b_q(\mathbf{s}^n)) = 0$ when $p_q \in \{\frac{1}{2}, 1\}$,
   - $b_q(\mathbf{s}^n)) > 0$ when $p_q = \frac{1}{2}$,
   - $b_q(\mathbf{s}^n)) = 0$ when $p_q = 1$.

   Since $b_q(\mathbf{s}^n) = 0$ when $p_q = 1$, $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q}\Big|_{p_q=1} = 0$, and $b_q(\mathbf{s}^n) \geq 0$ for any $p_q \in [0, 1]$, then there exists some threshold $\hat{p} < 1$ such that $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} < 0$ and $b_q(\mathbf{s}^n) < b_q(\mathbf{s}^n)\Big|_{p_q=\frac{1}{2}}$ for all $p_q > \hat{p}$. Therefore, $b_L(\mathbf{s}^n) > b_H(\mathbf{s}^n)$ so the pre-screener under-trusts and is pessimistic about the most likely state if $\hat{p} \leq p_L < p_H$.

Moreover, using the fact that $b_q(\mathbf{s}^n) = 0$ when $p_q = 1/2$, then

$$\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}} = b_q(\mathbf{s}^n)\left(-8n_b(n_b+m) + \sum_{i=1}^{m} 4i(i-1)\right)$$
$$= b_q(\mathbf{s}^n)\left(-8n_b(n_b+m) + \frac{4}{3}m(m-1)(m+1)\right). \tag{20}$$

Thus there exists some threshold $\frac{1}{2} < \check{p} < 1$ such that $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} > 0$ for all $p_q < \check{p}$ when $\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}} > 0$. Note that $\check{p} > \frac{1}{2}$ for any given n_b if m is sufficiently large that $\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}} > 0$. Thus the pre-screener also under-trusts and is pessimistic about the most likely state if $p_L \leq \check{p}$ and $p_H > \hat{p}$ where $\check{p} \geq \frac{1}{2}$. Note that we have already shown directly (in the preceding proof) that the pre-screener under-trusts and is pessimistic for $m = 1, 2, 3$ regardless of p_L, p_H, and n_b. ■

A.8 Proof of Proposition 6

Shown in Proof of Proposition 2.

A.9 Proof of Proposition 7

To show the results when agents receive signals from multiple experts, note that Equation (9) can also be re-written as

$$P^b(q_1, q_2, \theta|\mathbf{s}^{n_1,n_2}) = \frac{\left(\prod_{t=n_1+1}^{n_1+n_2} P(s_{t2}|q_2,\theta)\right)\left(\prod_{t=1}^{n_1} P(s_{t1}|q_1,\theta)\right)\omega_0^{q_1}\omega_0^{q_2}\omega_0^{\theta} b_{q_1}(\mathbf{s}^{n_1}) b_{q_2q_1}(\mathbf{s}^{n_1,n_2})}{\sum_{q_2}\sum_{q_1}\sum_{\theta}\left(\prod_{t=n_1+1}^{n_1+n_2} P(s_{t2}|q_2,\theta)\right)\left(\prod_{t=1}^{n_1} P(s_{t1}|q_1,\theta)\right)\omega_0^{q_1}\omega_0^{q_2}\omega_0^{\theta} b_{q_1}(\mathbf{s}^{n_1}) b_{q_2q_1}(\mathbf{s}^{n_1,n_2})}, \tag{21}$$

where the functions $b_{q_1}(\mathbf{s}^{n_1})$ and $b_{q_1q_2}(\mathbf{s}^{n_1,n_2})$ reflect the path dependency of the biased agent's beliefs and $b_{q_1}(\emptyset) = 1$:

$$b_{q_1}(\mathbf{s}^{n_1}) = \prod_{m=1}^{n_1}\left(\sum_{\theta}\left(\prod_{t=1}^{m} P(s_{t1}|q_1,\theta)\right)\omega_0^{\theta}\right)$$
$$b_{q_2q_1}(\mathbf{s}^{n_1,n_2}) = \prod_{m=n_1+1}^{n_1+n_2}\left(\sum_{\theta}\left(\prod_{t=n_1+1}^{m} P(s_{t2}|q_2,\theta)\right)\left(\prod_{t=1}^{n_1} P(s_{t1}|q_1,\theta)\right)\omega_0^{\theta}\right).$$

Consider a sequence of signals such that the agent observes k a signals from expert 1, followed by k b signals from expert 2: $\mathbf{s}^{n_1} = (a, \ldots, a)$ and $\mathbf{s}^{n_2} = (b, \ldots, b)$ where $n_1 = n_2 = k$.

To show this, note that the following properties hold when $\omega_0^{\theta} = 1/2$ and the two experts send either (1) an equal number k of opposing signals, or (2) an equal number of completely mixed signals:

$\prod_{i=k+1}^{2k} P(s_{i2}|H_2, A) = \prod_{i=1}^{k} P(s_{i1}|H_1, B)$

$\prod_{i=k+1}^{2k} P(s_{i2}|L_2, A) = \prod_{i=1}^{k} P(s_{i1}|L_1, B)$

$\prod_{i=k+1}^{2k} P(s_{i2}|H_2, B) = \prod_{i=1}^{k} P(s_{i1}|H_1, A)$

$\prod_{i=k+1}^{2k} P(s_{i2}|L_2, B) = \prod_{i=1}^{k} P(s_{i1}|L_1, A)$

For all $\omega_0^\theta \in (0, 1)$, then $P^b(\theta|\mathbf{s}^{n_1,n_2}) > 1/2$ only if

$$\omega_0^H(1-\omega_0^H)\left(\left(\prod_{i=k+1}^{2k} P(s_{i2}|H_2,A)\right)\left(\prod_{i=k+1}^{2k} P(s_{m2}|L_2,B)\right) - \left(\prod_{i=k+1}^{2k} P(s_{i2}|L_2,A)\right)\left(\prod_{i=2}^{2k} P(s_{i2}|H_2,B)\right)\right)$$
$$(b_{L_1}(\mathbf{s}^n)b_{H_2L_1}(\mathbf{s}^n) - b_{H_1}(\mathbf{s}^n)b_{L_2H_1}(\mathbf{s}^n)) > 0. \tag{22}$$

When the two experts send an equal number of opposing signals in sequence (and suppressing the arguments of $b_{q_1}(\mathbf{s}^n)$ and $b_{q_1q_2}(\mathbf{s}^n)$ for brevity of exposition), we also know

$\prod_{i=k+1}^{2k} P(s_{i2}|H_2, A) = \prod_{i=1}^{k} P(s_{i1}|H_1, B) = (1-p_H)^k$

$\prod_{i=k+1}^{2k} P(s_{i2}|L_2, A) = \prod_{i=1}^{k} P(s_{i1}|L_1, B) = (1-p_L)^k$

$\prod_{i=k+1}^{2k} P(s_{i2}|H_2, B) = \prod_{i=1}^{k} P(s_{i1}|H_1, A) = p_H^k$

$\prod_{i=k+1}^{2k} P(s_{i2}|L_2, B) = \prod_{i=1}^{k} P(s_{i1}|L_1, A) = p_L^k$

$b_{q_1} = \prod_{i=1}^{k}(\frac{1}{2})(p_{q_1}^i + (1-p_{q_1})^i)$, where we have previously shown that $b_{H_1} > b_{L_1}$

$b_{q_2q_1} = \prod_{i=1}^{k}(\frac{1}{2})\left((1-p_{q_2})^i p_{q_1}^k + p_{q_2}^i(1-p_{q_1})^k\right)$

Substituting all of these into the biased agent's posterior on the state, $P^b(\theta|\mathbf{s}^{n_1,n_2}) > 1/2$ only if

$$\omega_0^H(1-\omega_0^H)\left((1-p_H)^k p_L^k - (1-p_L)^k p_H^k\right)(b_{L_1}b_{H_2L_1} - b_{H_1}b_{L_2H_1}) > 0. \tag{23}$$

The first term of Equation (23) is positive and the second term is clearly negative, since $p_H > p_L$. Note that $b_L < b_H$ for $n_a > 1$ and $b_L = b_H$ for $n_a = 1$. Comparing a given mth term of $b_{L_2H_1} - b_{H_2L_1}$ yields

$$\begin{aligned}&(\frac{1}{2})\left((1-p_L)^m p_H^k + p_L^m(1-p_H)^k - (1-p_H)^m p_L^k - p_H^m(1-p_L)^k\right)\\ &\quad = (\frac{1}{2})\left(p_H^m(1-p_L)^m(p_H^{k-m} - (1-p_L)^{k-m}) + p_L^m(1-p_H)^m((1-p_H)^{k-m} - p_L^{k-m})\right),\end{aligned}$$

which is zero if $k = m$ and positive if $m < k$. Thus, each mth term of $b_{L_2H_1}$ is strictly greater than the mth term of $b_{H_2L_1}$ for $m < k$ and is equal when $m = k$, implying that $b_{L_2H_1} > b_{H_2L_1}$ if

$k > 1$ (and $b_{L_2 H_1} = b_{H_2 L_1}$ if $k = 1$) . This implies that the third term of Equation (23) is strictly negative when $k > 1$, so Equation (23) is satisfied. Thus, $P^b(\theta = A|\mathbf{s}^{n_1,n_2}) > 1/2$ when $\omega_0^A = 1/2$ and $k > 1$, and $P^b(\theta = A|\mathbf{s}^{n_1,n_2}) = 1/2$ when $\omega_0^A = 1/2$ and $k = 1$.

Substituting all of these into the biased agent's posteriors on expert qualities, we have that $P^b(q_1|\mathbf{s}^{n_1,n_2}) > P^b(q_2|\mathbf{s}^{n_1,n_2})$ only if

$$\omega_0^H(1-\omega_0^H)\left((1-p_H)^k p_L^k - (1-p_L)^k p_H^k\right)(b_{L_1}b_{H_2L_1} - b_{H_1}b_{L_2H_1}) > 0,$$

which is exactly Equation (23) again. Thus, the biased agent believes that the first expert is more likely to be high quality than the second expert: $P^b(q_1|\mathbf{s}^{n_1,n_2}) > P^b(q_2|\mathbf{s}^{n_1,n_2})$.

A.10 Proof of Proposition 8

Consider a sequence of signals such that the agent observes k a signals from expert 1, followed by k b signals from expert 2: $\mathbf{s}^{n_1} = (a, \ldots, a)$ and $\mathbf{s}^{n_2} = (b, \ldots, b)$ where $n_1 = n_2 = k$.

Letting $k \to \infty$ and factoring, we can re-write the terms a, b, c, and d as

$$\begin{aligned}
a &= 2(\omega_0^H)^2(1-p_H)^k p_H^{2k(k+1)}(\frac{1-p_H}{p_H})^{\frac{k(k+1)}{2}} \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right) \\
&= p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}} \left(2(\omega_0^H)^2(\frac{1-p_H}{1-p_L})^{k+\frac{k(k+1)}{2}} \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\right) \\
b &= \omega_0^H(1-\omega_0^H)\left((1-p_H)^k p_L^k + p_H^k(1-p_L)^k\right) p_L^{\frac{k(k+1)}{2}} p_H^{\frac{k(k+1)}{2}} p_L^{k^2}(\frac{1-p_H}{p_H})^{\frac{k(k+1)}{2}} \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_L}{p_L})^m)\right) \\
&= p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}} \left(\omega_0^H(1-\omega_0^H)\left(1+(\frac{p_L(1-p_H)}{p_H(1-p_L)})^k\right)\left(\frac{1-p_H}{1-p_L}\right)^{\frac{k(k+1)}{2}}(\frac{p_L}{p_H})^{k^2+\frac{k(k+1)}{2}}\left(\prod_{m=1}^{\infty}(1+(\frac{1-p_L}{p_L})^m)\right)\right) \\
c &= \omega_0^H(1-\omega_0^H)\left((1-p_L)^k p_H^k + p_L^k(1-p_H)^k\right) p_H^{\frac{k(k+1)}{2}} p_L^{\frac{k(k+1)}{2}} p_H^{k^2}(\frac{1-p_L}{p_L})^{\frac{k(k+1)}{2}} \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right) \\
&= p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}} \left(\omega_0^H(1-\omega_0^H)\left(1+\left(\frac{p_L(1-p_H)}{p_H(1-p_L)}\right)^k\right)\left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\right) \\
d &= 2(1-\omega_0^H)^2(1-p_L)^k p_L^{2k(k+1)}(\frac{1-p_L}{p_L})^{\frac{k(k+1)}{2}} \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_L}{p_L})^m)\right) \\
&= p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}} \left(2(1-\omega_0^H)^2\left(\frac{p_L}{p_H}\right)^{\frac{3k(k+1)}{2}}\left(\prod_{m=1}^{\infty}(1+(\frac{1-p_L}{p_L})^m)\right)\right)
\end{aligned}$$

Note that the term $p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}}$ drops out since it is in every term when calculating the joint posteriors. Also, note that a necessary and sufficient condition for $\prod_{m=1}^{\infty}(1+(\frac{1-p_q}{p_q})^m)$ to converge is that $\sum_{m=1}^{\infty}(\frac{1-p_q}{p_q})^m$ is absolutely convergent, which is clearly satisfied when

$p_q > \frac{1}{2}$.

Thus, when $1 > p_H > p_L > \frac{1}{2}$,

$$\begin{aligned}
&\lim_{k\to\infty} a = 0\\
&\lim_{k\to\infty} b = 0\\
&\lim_{k\to\infty} c = \omega_0^H(1-\omega_0^H)\left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\\
&\lim_{k\to\infty} d = 0.
\end{aligned}$$

This implies that when $1 > p_H > p_L > \frac{1}{2}$,

$$\begin{aligned}
&\lim_{k\to\infty} P^b(H_1, H_2|\mathbf{s}^{n_1,n_2}) = 0\\
&\lim_{k\to\infty} P^b(L_1, H_2|\mathbf{s}^{n_1,n_2}) = 0\\
&\lim_{k\to\infty} P^b(H_1, L_2|\mathbf{s}^{n_1,n_2}) = 1\\
&\lim_{k\to\infty} P^b(L_1, L_2|\mathbf{s}^{n_1,n_2}) = 0.
\end{aligned}$$

An extremely similar proof applies to show that $\lim_{k\to\infty} P^b(\theta = A|k\ a$'s from expert 1, $k\ b$'s from expert 2) 1 when $1 > p_H > p_L > \frac{1}{2}$. Note that if $\frac{1}{2} > p_L > 1 - p_H$, then we can instead factor out $1 - p_L$ instead of p_L, so all the p_L and $1 - p_L$ terms are exchanged and the proof applies.

The result still holds if $p_L = \frac{1}{2}$. Letting $k \to \infty$ and $p_L = \frac{1}{2}$ and factoring, we can re-write the terms a, b, c, and d as

$$\begin{aligned}
a &= 2(\omega_0^H)^2(1-p_H)^k p_H^{2k(k+1)}(\frac{1-p_H}{p_H})^{\frac{k(k+1)}{2}} \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\\
&= 2(\omega_0^H)^2(1-p_H)^{k+\frac{k(k+1)}{2}} p_H^{\frac{3k(k+1)}{2}} \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\\
&= p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}} \left(2(\omega_0^H)^2(\frac{1-p_H}{1-p_L})^{k+\frac{k(k+1)}{2}} \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\right)\\
b &= \omega_0^H(1-\omega_0^H)(\frac{1}{2})^k\left((1-p_H)^k+p_H^k\right)(\frac{1}{2})^{k^2+\frac{k(k+1)}{2}} p_H^{\frac{k(k+1)}{2}}(2)^k \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\\
&= \omega_0^H(1-\omega_0^H)(\frac{1}{2})^{k^2+\frac{k(k+1)}{2}}\left((1-p_H)^k+p_H^k\right) p_H^{\frac{k(k+1)}{2}} \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\\
&= p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}} \left(\omega_0^H(1-\omega_0^H)\left(\frac{2}{(2p_H)^k}\right)^k(1+(\frac{1-p_H}{p_H})^k)\left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\right)\\
c &= \omega_0^H(1-\omega_0^H)(\frac{1}{2})^{k+\frac{k(k+1)}{2}} p_H^{k^2+\frac{k(k+1)}{2}}(p_H^k+(1-p_H)^k)\left(1+(\frac{1-p_H}{p_H})^k\right)^k \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\\
&= p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}} \left(\omega_0^H(1-\omega_0^H)(1+(\frac{1-p_H}{p_H})^k)(1+(\frac{1-p_H}{p_H})^k)^k \left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)\right)\\
d &= 2(1-\omega_0^H)^2(\frac{1}{2})^{2k^2+3k}(2)^{2k}\\
&= 2(1-\omega_0^H)^2(\frac{1}{2})^{2k^2+k}\\
&= p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}} \left(2(1-\omega_0^H)^2(\frac{1}{2p_H})^{\frac{3k(k+1)}{2}}(\frac{1}{2})^{\frac{k(k-1)}{2}}\right)
\end{aligned}$$

Note that the term $p_H^{\frac{3k(k+1)}{2}}(1-p_L)^{k+\frac{k(k+1)}{2}}$ drops out since it is in every term when calculating the joint posteriors. Also, note that a necessary and sufficient condition for $\prod_{m=1}^{\infty}(1+(\frac{1-p_q}{p_q})^m)$ to converge is that $\sum_{m=1}^{\infty}(\frac{1-p_H}{p_H})^m$ is absolutely convergent, which is clearly satisfied when $p_H > \frac{1}{2}$.

Terms a, b, and d converge to 0. Term c converges to $\omega_0^H(1-\omega_0^H)\left(\prod_{m=1}^{\infty}(1+(\frac{1-p_H}{p_H})^m)\right)$, which is a finite number, because $\lim_{k\to\infty}(1+(\frac{1-p_H}{p_H})^k)^k = 1$ (re-arranging and using L'Hopital's rule a couple of times):

$$
\begin{aligned}
\lim_{k\to\infty}(1+(\frac{1-p_H}{p_H})^k)^k &= \lim_{k\to\infty}\left(\exp\left(\ln(1+(\frac{1-p_H}{p_H})^k)\right)^k\right)\\
&= \lim_{k\to\infty}\exp\left(k\ln(1+(\frac{1-p_H}{p_H})^k)\right)\\
&= \exp\lim_{k\to\infty}\left(k\ln(1+(\frac{1-p_H}{p_H})^k)\right)\\
&= \exp\lim_{k\to\infty}\frac{\ln(1+(\frac{1-p_H}{p_H})^k)}{\frac{1}{k}}\\
&= \exp\lim_{k\to\infty}\frac{\frac{(\frac{1-p_H}{p_H})^k\ln(\frac{1-p_H}{p_H})}{1+(\frac{1-p_H}{p_H})^k}}{-(\frac{1}{k})^2}\\
&= \exp\lim_{k\to\infty}(\ln(\frac{1-p_H}{p_H}))\frac{-k^2}{\frac{1+(\frac{1-p_H}{p_H})^k}{(\frac{1-p_H}{p_H})^k}}\\
&= \exp\lim_{k\to\infty}(\ln(\frac{1-p_H}{p_H}))\frac{-2k}{-\frac{\ln(\frac{1-p_H}{p_H})}{(\frac{1-p_H}{p_H})^k}}\\
&= \exp\lim_{k\to\infty}2\left(\frac{k}{\frac{1}{(\frac{1-p_H}{p_H})^k}}\right)\\
&= \exp\lim_{k\to\infty}2\left(\frac{1}{-\frac{\ln(\frac{1-p_H}{p_H})}{(\frac{1-p_H}{p_H})^k}}\right)\\
&= \exp\lim_{k\to\infty}2\left(\frac{(\frac{1-p_H}{p_H})^k}{-\ln(\frac{1-p_H}{p_H})}\right)\\
\lim_{k\to\infty}(1+(\frac{1-p_H}{p_H})^k)^k &= \exp(0) = 1.
\end{aligned}
$$

This implies that when $1 > p_H > p_L = \frac{1}{2}$,

$$
\begin{aligned}
\lim_{k\to\infty}P^b(H_1,H_2|\mathbf{s}^{n_1,n_2}) &= 0\\
\lim_{k\to\infty}P^b(L_1,H_2|\mathbf{s}^{n_1,n_2}) &= 0\\
\lim_{k\to\infty}P^b(H_1,L_2|\mathbf{s}^{n_1,n_2}) &= 1\\
\lim_{k\to\infty}P^b(L_1,L_2|\mathbf{s}^{n_1,n_2}) &= 0.
\end{aligned}
$$

An extremely similar proof applies to show that $\lim_{k\to\infty} P^b(\theta = A|\mathbf{s}^{n_1,n_2}) = 1$ where $n_1 = n_2 = k$ when $1 > p_H > p_L = \frac{1}{2}$.

A.11 Proof of Proposition 9

Let s_{itj} be the ith signal, observed in period t, sent by expert $j \in \{1, 2\}$. As before, expert 1 reports first, and expert j sends n_j signals in total. In this notation, if the agent observes one signal per period from expert 1, followed by one signal per period from expert 2, then $\mathbf{s}^{n_1} = (s_{111}, s_{221}, \dots s_{n_1,n_1,1}$ and $\mathbf{s}^{n_2} = (s_{n_1+1,n_1+1,2}, s_{n_2+1,n_2+1,2}, \dots s_{n_1+n_2,n_1+n_2,2})$. If expert 2 sends all of his n_2 signals in the first period $m = k+1$ after expert 1 reports, then $\mathbf{s}^{n_2} = (s_{1m2}, s_{2m2}, \dots s_{n_2,m,2})$. Since the reliability of a signal i from expert j is independent of the period in which it is observed and the other expert's quality, note that $P(s_{itj}|q_j, q_k, \theta) = P(s_{ij}|q_j, \theta)$ for all t where $j \neq k$.

Then the biased agent's updating after observing generalize to Equations (8) and (9), where the functions $b_{q_1}(\mathbf{s}^{n_1})$ and $b_{q_1q_2}(\mathbf{s}^{n_1,n_2})$ reflect the path dependency of the biased agent's beliefs and will differ depending on both timing and order of signals. For example, if the agent observes one signal per period from expert 1, followed by one signal per period from expert 2, then $\mathbf{s}^{n_1} = (s_{111}, s_{221}, \dots s_{n_1,n_1,1})$ and $\mathbf{s}^{n_2} = (s_{n_1+1,n_1+1,2}, s_{n_2+1,n_2+1,2}, \dots s_{n_1+n_2,n_1+n_2,2})$. If expert 1 sends all of his n_1 signals in period 1, then $\mathbf{s}^{n_1} = (s_{111}, s_{211}, \dots s_{n_1,1,1})$.

1. If the biased agent receives n signals simultaneously (say, from expert 1 in period 1), then his posterior after observing $\mathbf{s}^{n_1} = (s_{111}, s_{211}, \dots s_{n_1,1,1})$ all together is still described by Equation 4, but his $b^{q_1}(\mathbf{s}^n)$ is instead given by

$$b_{q_1}(\mathbf{s}^n) = \left(\sum_\theta P(s_1|q,\theta)P(s_2|q,\theta)\dots P(s_n|q,\theta)\omega_0^\theta\right) \tag{24}$$

$$= \sum_\theta \left(\prod_{t=1}^{n} P(s_t|q,\theta)\right)\omega_0^\theta. \tag{25}$$

Let x be the event in which expert 1 sends k a's simultaneously: $\mathbf{s}_x^{n_1} = (s_{111}, s_{211}, \dots s_{n_1,1,1})$. Let y be the event in which expert 1 sends k a's sequentially $\mathbf{s}_y^{n_1} = (s_{111}, s_{221}, \dots s_{n_1,n_1,1})$. Then

(a) $b_{q_1}^x = (\frac{1}{2})(p_{q_1}^k + (1-p_{q_1})^k)$

(b) $b_{q_1}^y = \prod_{i=1}^k (\frac{1}{2})(p_{q_1}^i + (1-p_{q_1})^i)$

By direct comparison of the posteriors, $P^b(q_1 = H|\mathbf{s}_x^{n_1}) < P^b(q_1 = H|\mathbf{s}_y^{n_1})$ only if $b_H^x b_L^y - b_H^y b_L^x < 0$, which is satisfied:

$$\begin{aligned} b_H^x b_L^y - b_H^y b_L^x &= (\tfrac{1}{2})(p_H^k + (1-p_H)^k)\left(\prod_{i=1}^k (\tfrac{1}{2})(p_L^i + (1-p_L)^i)\right) - (\tfrac{1}{2})(p_L^k + (1-p_L)^k)\left(\prod_{i=1}^k (\tfrac{1}{2})(p_H^i + (1-p_H)^i)\right) \\ &= (\tfrac{1}{2})^{k+1}(p_L^k + (1-p_L)^k)(p_H^k + (1-p_H)^k)\left(\left(\prod_{i=1}^{k-1}(p_L^i + (1-p_L)^i)\right) - \left(\prod_{i=1}^{k-1}(p_H^i + (1-p_H)^i)\right)\right) \end{aligned}$$

since $\frac{\partial}{\partial p_q}(p_q^k + (1-p_q)^k) < 0$ for $k > 1$ and the relevant parameter restrictions on p_L and p_H. Thus, if expert 1 sends k identical signals, the pre-screener with a flat prior on the state trusts him more when they are sent sequentially than simultaneously (though there is still overtrust in both cases, which is straightforward to show given $b_{q_1}^x$ and $b_{q_1}^y$).

2. First, we show that the pre-screener still believes that state A is more likely than B. Note that Equation (22) must be satisfied for this to be true, whether either expert sends signals simultaneously or sequentially. What differs based on simultaneous versus sequential signals is the terms b_{q_1} and $b_{q_2q_1}$. Let W be the event in which expert 1 sends k a's simultaneously and expert 2 sends k b's simultaneously: $\mathbf{s}^{n_1} = (s_{111}, s_{211}, \dots s_{n_1,1,1})$, $\mathbf{s}^{n_2} = (s_{n_1+1,2,2}, s_{n_1+2,2,2}, \dots s_{n_1+n_2,2,2})$. Let X be the event in which expert 1 sends k a's simultaneously and expert 2 sends k b's sequentially: $\mathbf{s}^{n_1} = (s_{111}, s_{211}, \dots s_{n_1,1,1})$, $\mathbf{s}^{n_2} = (s_{n_1+1,n_1+1,2}, s_{n_1+2,n_1+2,2}, \dots s_{n_1+n_2,n_1+n_2,2})$. Let Y be the event in which expert 1 sends k a's sequentially and expert 2 sends k b's simultaneously: $\mathbf{s}^{n_1} = (s_{111}, s_{221}, \dots s_{n_1,n_1,1})$, $\mathbf{s}^{n_2} = (s_{n_1+1,n_1+1,2}, s_{n_1+2,n_1+1,2}, \dots s_{n_1+n_2,n_1+1,2})$. Let Z be the event in which expert 1 sends k a's sequentially and expert 2 sends k b's sequentially: $\mathbf{s}^{n_1} = (s_{111}, s_{221}, \dots s_{n_1,n_1,1})$ and $\mathbf{s}^{n_2} = (s_{n_1+1,n_1+1,2}, s_{n_2+1,n_2+1,2}, \dots s_{n_1+n_2,n_1+n_2,2})$. Note that in each of these events, expert 1's signals are sent strictly before expert 2's signals.

   When $n_1 = n_2 = k$ where expert 1's signals are all a's and expert 2's signals are all b's,

$$\begin{aligned} &b_{q_1}^W = b_X^{q_1} = (\tfrac{1}{2})(p_{q_1}^k + (1-p_{q_1})^k) \\ &b_{q_1}^Y = b_Z^{q_1} = \textstyle\prod_{i=1}^k (\tfrac{1}{2})(p_{q_1}^i + (1-p_{q_1})^i) \\ &b_{q_2q_1}^W = b_{q_2q_1}^Y = (\tfrac{1}{2})\left((1-p_{q_2})^k p_{q_1}^k + p_{q_2}^k(1-p_{q_1})^k\right) \\ &b_{q_2q_1}^X = b_{q_2q_1}^Z = \textstyle\prod_{i=1}^k (\tfrac{1}{2})\left((1-p_{q_2})^i p_{q_1}^k + p_{q_2}^i(1-p_{q_1})^k\right) \end{aligned}$$

   Also, note that $b_H^E > b_L^E$ for all $k > 1$ and $E \in \{W, X, Y, Z\}$. We have already shown previously that $b_{L_2H_1}^X > b_{H_2L_1}^X$ for $k > 1$, and obviously $b_{L_2H_1}^W = b_{H_2L_1}^W$.

   Using these properties in Equation (22), we can verify that $P_E^b(\theta = A|\mathbf{s}^{n_1,n_2}) > \frac{1}{2}$ when $\omega_0^\theta = \frac{1}{2}$, $k > 1$, and $E \in \{W, X, Y, Z\}$.

To show that sending the k b signals simultaneously rather than sequentially gives more credibility to expert 2, it is sufficient to show that $P^b_X(\theta = A|\mathbf{s}^{n_1,n_2}) > P^b_Y(\theta = A|\mathbf{s}^{n_1,n_2})$ and $P^b_W(\theta = A|\mathbf{s}^{n_1,n_2}) > P^b_X(\theta = A|\mathbf{s}^{n_1,n_2})$.

$P^b_Y(\theta = A|\mathbf{s}^{n_1,n_2}) > P^b_Z(\theta = A|\mathbf{s}^{n_1,n_2})$ is satisfied only if

$$\begin{aligned}\omega_0^H(1-\omega_0^H)[p_H^k(1-p_L)^k - p_L^k(1-p_H)^k]\left((\omega_0^H)^2 p_H^k(1-p_H)^k(b_H^Y)^2\left(b^Y_{H_2H_1}(b^Z_{L_2H_1}-b^Z_{H_2L_1}) - b^Z_{H_2H_1}(b^Y_{L_2H_1}-b^Y_{H_2L_1})\right)\right.\\ \left.+(1-\omega_0^H)^2 p_L^k(1-p_L)^k(b_L^Y)^2\left(b^Y_{L_2L_1}(b^Z_{L_2H_1}-b^Z_{H_2L_1}) - b^Z_{L_2L_1}(b^Y_{L_2H_1}-b^Y_{H_2L_1})\right)\right)\\ +(\omega_0^H)^2(1-\omega_0^H)^2[p_H^{2k}(1-p_L)^{2k} - p_L^{2k}(1-p_H)^{2k}]b_H^Y b_L^Y\left(b^Z_{L_2H_1}b^Y_{H_2L_1} - b^Y_{L_2H_1}b^Z_{H_2L_1}\right).\end{aligned}$$

Note that $b^Y_{L_2H_1} = b^Y_{H_2L_1}$ and $b^X_{L_2H_1} > b^X_{H_2L_1}$ for $k > 1$, so the third term is positive. For the first and second terms, $b^X_{L_2H_1} > b^X_{H_2L_1}$, and $b^Y_{L_2H_1} > b^Y_{H_2L_1}$, so it is sufficient to show that $b^Y_{qq} - b^X_{qq}$ for them to each be positive:

$$\begin{aligned}b^Y_{qq} - b^X_{qq} &= p_q^k(1-p_q)^k - \prod_{i=1}^{k}\frac{1}{2}\left((1-p_q)^i p_q^k + p_q^i(1-p_q)^k\right)\\ &= p_q^k(1-p_q)^k\left(1 - \prod_{i=1}^{k-1}\frac{1}{2}\left((1-p_q)^i p_q^k + p_q^i(1-p_q)^k\right)\right),\end{aligned}$$

where each term of $(1-p_q)^i p_q^k + p_q^i(1-p_q)^k$ is bounded above by $\frac{1}{2}$ for $k > 1$. Thus, $b^Y_{qq} - b^X_{qq} > 0$ for $k > 1$ which implies that the first and second terms are positive when $k > 1$.

The same argument applies for $P^b_W(\theta = A|\mathbf{s}^{n_1,n_2}) > P^b_X(\theta = A|\mathbf{s}^{n_1,n_2})$. Thus, sending the k b signals simultaneously rather than sequentially gives more credibility to expert 2, given expert 1's signals.

A.12 Proof of Proposition 10

1. **Proof.** Let $\mathbf{s}^n$ be a sequence of n observed signals with n_a a's and n_b b's, let s_{n+1} be the $(n+1)$th observed signal, and let ω_n^b equal the pre-screener's joint posterior after the sequence $\mathbf{s}^n$.

   First, note that each joint belief on the state and quality for the prior ω_n^b, denoted $\omega_n^{q\theta}$, is given by

$$\omega_n^{q\theta} \equiv P^b(q,\theta|\{\mathbf{s}^n\}) = \frac{\left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^\theta\omega_0^q b_q(\mathbf{s}^n)}{\sum_\theta\sum_q\left(\prod_{t=1}^n P(s_t|q,\theta)\right)\omega_0^\theta\omega_0^q b_q(\mathbf{s}^n)}, \quad (26)$$

where

$$b_q(\mathbf{s}^n) = \prod_{m=1}^{n} \left(\sum_{\theta} \left(\prod_{t=1}^{m} P(s_t|q,\theta) \right) \omega_0^\theta \right). \tag{27}$$

Thus, the Bayesian's posterior belief given the biased prior is

$$P^u(\theta = A|\text{prior} = \omega_n^b, \{s_{n+1}\}) = \frac{\omega_0^A \sum_q \left(\prod_{t=1}^{n+1} P(s_t|q,A)\right) \omega_0^q b_q(\mathbf{s}^n)}{\sum_\omega \omega_0^\theta \sum_q P(s_{n+1}|q,\theta) \left(\prod_{t=1}^{n} P(s_t|q,\theta)\right) \omega_0^q b_q(\mathbf{s}^n)}.$$

In contrast, the pre-screener's posterior belief after observing $\{\mathbf{s}^n, s_{n+1}\}$ is

$$P^b(\theta = A|\{\mathbf{s}^n, s_{n+1}\}) = \frac{\omega_0^A \sum_q \left(\prod_{t=1}^{n+1} P(s_t|q,A)\right) \omega_0^q b_q(\{\mathbf{s}^n, s_{n+1}\})}{\sum_\theta \omega_0^\theta \sum_q \left(\prod_{t=1}^{n+1} P(s_t|q,\theta)\right) \omega_0^q b_q(\{\mathbf{s}^n, s_{n+1}\})},$$

where

$$\begin{aligned} b_q(\{\mathbf{s}^n, s_{n+1}\}) &= \prod_{m=1}^{n+1} \left(\sum_{\theta} \left(\prod_{t=1}^{m} P(s_t|q,\theta) \right) \omega_0^\theta \right) & (28) \\ &= b_q(\mathbf{s}^n) \left(\sum_{\theta} \left(\prod_{t=1}^{n+1} P(s_t|q,\theta) \right) \omega_0^\theta \right) & (29) \end{aligned}$$

Substituting all of the preceding information into $P^b(\omega = A|\{\mathbf{s}^n, s_{n+1}\}) > P^u(\omega = A|prior = \omega_n^b, \{s_{n+1}\})$, the inequality is only satisfied if

$$\begin{aligned} &\omega_0^A(1-\omega_0^A)\omega_0^H(1-\omega_0^H) b_L(\mathbf{s}^n) b_H(\mathbf{s}^n) \underbrace{\left(\left(\sum_\theta \left(\prod_{t=1}^{n+1} P(s_t|H,\theta) \right) \omega_0^\theta \right) - \left(\sum_\theta \left(\prod_{t=1}^{n+1} P(s_t|L,\theta) \right) \omega_0^\theta \right) \right)}_{X} \\ &\underbrace{\left(\left(\prod_{t=1}^{n+1} P(s_t|H,A) \right) \left(\prod_{t=1}^{n+1} P(s_t|L,B) \right) - \left(\prod_{t=1}^{n+1} P(s_t|H,B) \right) \left(\prod_{t=1}^{n+1} P(s_t|L,A) \right) \right)}_{Y} > 0, \end{aligned} \tag{30}$$

Without loss of generality, suppose that $n_a \geq n_b$.

If $s_{n+1} = a$, then $\{\mathbf{s}^n, s_{n+1}\}$ has $n_a + 1$ a's and n_b b's. Then the term Y is given by

$$\begin{aligned} &p_H^{n_a+1}(1-p_H)^{n_b}(1-p_L)^{n_a+1}(p_L)^{n_b} - (1-p_H)^{n_a+1}(p_H)^{n_b}(p_L)^{n_a+1}(1-p_L)^{n_b} \\ &= [p_H p_L (1-p_H)(1-p_L)]^{n_b}[(p_H(1-p_L))^{n_a-n_b+1} - (p_L(1-p_H))^{n_a-n_b+1}] \end{aligned}$$

Thus if $s_{n+1} = a$, then

$$Y(s_{n+1} = a) \begin{cases} > 0 & \text{if } n_a + 1 > n_b \\ = 0 & \text{if } n_a + 1 = n_b \\ < 0 & \text{if } n_a + 1 < n_b. \end{cases}$$

If $s_{n+1} = b$, then $\{\mathbf{s}^n, s_{n+1}\}$ has n_a a's and $n_b + 1$ b's. Then the term Y is given by

$$p_H^{n_a}(1-p_H)^{n_b+1}(1-p_L)^{n_a}(p_L)^{n_b+1} - (1-p_H)^{n_a}(p_H)^{n_b+1}(p_L)^{n_a}(1-p_L)^{n_b+1}$$
$$= [p_H p_L(1-p_H)(1-p_L)]^{n_b}[(p_H(1-p_L))^{n_a-n_b} p_L(1-p_H) - ((1-p_H)p_L)^{n_a-n_b} p_H(1-p_L)]$$

Thus if $s_{n+1} = b$, then

$$Y(s_{n+1} = b) \begin{cases} > 0 & \text{if } n_a > n_b + 1 \\ = 0 & \text{if } n_a = n_b + 1 \\ < 0 & \text{if } n_a < n_b + 1. \end{cases}$$

Thus, Y is positive if $\{\mathbf{s}^n, s_{n+1}\}$ has more a's than b's, Y is negative if $\{\mathbf{s}^n, s_{n+1}\}$ has more b's than a's, and Y is zero if $\{\mathbf{s}^n, s_{n+1}\}$ has an equal number of a's and b's.

Moreover, note that $\kappa_{\mathbf{s}^n} \equiv \frac{b_q(\mathbf{s}^n)\omega_0^q}{\sum_q b_q(\mathbf{s}^n)\omega_0^q}$. Then Equation (29) implies that $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) > \kappa_H(\mathbf{s}^n)$ if and only if
$\omega_0^H(1-\omega_0^H)b_H(\mathbf{s}^n)b_L(\mathbf{s}^n)\left(\left(\sum_\theta \left(\prod_{t=1}^{n+1} P(s_t|H,\theta)\right)\omega_0^\theta\right) - \left(\sum_\theta \left(\prod_{t=1}^{n+1} P(s_t|L,\theta)\right)\omega_0^\theta\right)\right) > 0$, which is the requirement that $X > 0$.

In other words,

$$X \begin{cases} > 0 & \text{if and only if } \kappa_H(\{\mathbf{s}^n, s_{n+1}\}) > \kappa_H(\mathbf{s}^n) \\ = 0 & \text{if and only if } \kappa_H(\{\mathbf{s}^n, s_{n+1}\}) = \kappa_H(\mathbf{s}^n) \\ < 0 & \text{if and only if } \kappa_H(\{\mathbf{s}^n, s_{n+1}\}) < \kappa_H(\mathbf{s}^n). \end{cases}$$

From above, we can see that the sign of Equation (30) depends on the sign of XY. Putting everything together, then

$P^b(\theta = A|\{\mathbf{s}^n, s_{n+1}\}) = P^u(\theta = A|prior = \omega_n^b, \{s_{n+1}\})$ if either (1) $\{\mathbf{s}^n, s_{n+1}\}$ has an equal number of a's and b's or (2) $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) = \kappa_H(\mathbf{s}^n)$

$P^b(\theta = A|\{\mathbf{s}^n, s_{n+1}\}) > P^u(\theta = A|prior = \omega_n^b, \{s_{n+1}\})$ if (3) $\{\mathbf{s}^n, s_{n+1}\}$ has more a's than b's and $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) > \kappa_H(\mathbf{s}^n)$ or (4) $\{\mathbf{s}^n, s_{n+1}\}$ has more b's than a's and

$\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) < \kappa_H(\mathbf{s}^n)$

$P^b(\theta = A|\{\mathbf{s}^n, s_{n+1}\}) < P^u(\theta = A|prior = \omega_n^b, \{s_{n+1}\})$ if (5) $\{\mathbf{s}^n, s_{n+1}\}$ has more a's than b's and $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) < \kappa_H(\mathbf{s}^n)$ or (6) $\{\mathbf{s}^n, s_{n+1}\}$ has more b's than a's and $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) > \kappa_H(\mathbf{s}^n)$

Note that the statement $P^b(\theta = A|\{\mathbf{s}^n, s_{n+1}\}) > P^u(\theta = A|prior = \omega_n^b, \{s_{n+1}\})$ if $\{\mathbf{s}^n, s_{n+1}\}$ has more a's than b's and $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) > \kappa_H(\mathbf{s}^n)$ is equivalent to the statement $P^b(\theta = A|\{\mathbf{s}^n, s_{n+1}\}) < P^u(\theta = A|prior = \omega_n^b, \{s_{n+1}\})$ if $\{\mathbf{s}^n, s_{n+1}\}$ has more b's than a's and $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) > \kappa_H(\mathbf{s}^n)$. They both say that if the first-stage updated belief in the expert's high quality after $\{\mathbf{s}^n, s_{n+1}\}$ is that the agent is higher than the first-stage updated belief in the expert's high quality after $\mathbf{s}^n$, then the pre-screener over-updates toward the most likely state on the last signal.

Likewise, the statement $P^b(\theta = A|\{\mathbf{s}^n, s_{n+1}\}) > P^u(\theta = A|prior = \omega_n^b, \{s_{n+1}\})$ if $\{\mathbf{s}^n, s_{n+1}\}$ has more b's than a's and $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) < \kappa_H(\mathbf{s}^n)$ is equivalent to the statement $P^b(\theta = A|\{\mathbf{s}^n, s_{n+1}\}) < P^u(\theta = A|prior = \omega_n^b, \{s_{n+1}\})$ if $\{\mathbf{s}^n, s_{n+1}\}$ has more a's than b's and $\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) < \kappa_H(\mathbf{s}^n)$. They both say that if the first-stage updated belief in the expert's high quality after $\{\mathbf{s}^n, s_{n+1}\}$ is that the agent is lower than the first-stage updated belief in the expert's high quality after $\mathbf{s}^n$, then the pre-screener under-updates toward the most likely state on the last signal.

Therefore, we can state the proposition assuming that the number of a's be greater than or equal to the number of b's in $\{\mathbf{s}^n, s_{n+1}\}$ without loss of generality.

Moreover, note that $Pr^u(q = H|\{\mathbf{s}^n, s_{n+1}\}) = \frac{\omega_0^H \sum_\theta \left(\prod_{t=1}^{n+1} P(s_t|q,\theta)\right)\omega_0^\theta}{\sum_q \omega_0^q \sum_\theta \left(\prod_{t=1}^{n+1} P(s_t|q,\theta)\right)\omega_0^\theta}$. From this definition, we know that $Pr^u(q = H|\{\mathbf{s}^n, s_{n+1}\}) > \omega_0^H$ if and only if $\left(\left(\sum_\theta \left(\prod_{t=1}^{n+1} P(s_t|H,\theta)\right)\omega_0^\theta\right) - \left(\sum_\theta \left(\prod_{t=1}^{n+1} P(s_t|L,\theta)\right)\omega_0^\theta\right)\right) > 0$, which is the requirement that $X > 0$. Thus,

$$\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) > \kappa_H(\mathbf{s}^n) \text{ if and only if } Pr^u(q = H|\{\mathbf{s}^n, s_{n+1}\}) > \omega_0^H$$
$$\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) = \kappa_H(\mathbf{s}^n) \text{ if and only if } Pr^u(q = H|\{\mathbf{s}^n, s_{n+1}\}) = \omega_0^H$$
$$\kappa_H(\{\mathbf{s}^n, s_{n+1}\}) < \kappa_H(\mathbf{s}^n) \text{ if and only if } \; Pr^u(q = H|\{\mathbf{s}^n, s_{n+1}\}) < \omega_0^H.$$

■

2. **Proof.** First, note that $P^b[q, \theta | \{\mathbf{s}^n, s_{n+1}\}]$ is equal to

$$P^b[q, \theta | \{\mathbf{s}^n, s_{n+1}\}] = \frac{b_q(\{\mathbf{s}^n, s_{n+1}\}) \left(\prod_{t=1}^{n+1} P(s_t | q, \theta) \right) \omega_0^\theta \omega_0^q}{\sum_q b_q(\{\mathbf{s}^n, s_{n+1}\}) \sum_\theta \left(\prod_{t=1}^{n+1} P(s_t | q, \theta) \right) \omega_0^\theta \omega_0^q} \tag{31}$$

where $b_q(\{\mathbf{s}^n, s_{n+1}\})$ is described by Equations (28) or (29).

Second, applying the generalized pre-screening described in A.1, $P^b[q, \theta | \text{prior} = \omega_n^b, \{s_{n+1}\}]$ is equal to

$$P^b[q, \theta | \text{prior} = \omega_n^b, \{s_{n+1}\}] = \frac{b_{q\theta}(s_{n+1}) \left(\frac{1}{\sum_\theta \omega_n^{q\theta}} \right) P(s_{t+1} | q, \theta) \omega_n^{q\theta}}{\sum_q \sum_\theta b_{q\theta}(s_{n+1}) \left(\frac{1}{\sum_\theta \omega_n^{q\theta}} \right) P(s_{t+1} | q, \theta) \omega_n^{q\theta}}, \tag{32}$$

where

$$b_{q\theta}(s_{n+1}) = \sum_\theta P(s_{n+1} | q, \theta) \omega_n^{q\theta},$$

and $\omega_n^{q\theta}$ is described by Equation (26) and $b_q(\mathbf{s}^n)$ is described by Equation (27). Substituting this into $P^b[q, \theta | \text{prior} = \omega_n^b, \{s_{n+1}\}]$ yields:

$$\begin{aligned} P^b[q, \theta | \text{prior} = \omega_n^b, \{s_{n+1}\}] &= \frac{b_{q\theta}(s_{n+1}) \left(\frac{1}{\sum_\theta \omega_n^{q\theta}} \right) P(s_{t+1} | q,\theta) b_q(\mathbf{s}^n) \left(\prod_{t=1}^{n} P(s_t | q,\theta) \right) \omega_0^\theta \omega_0^q}{\sum_q \sum_\theta b_{q\theta}(s_{n+1}) \left(\frac{1}{\sum_\theta \omega_n^{q\theta}} \right) P(s_{t+1} | q,\theta) b_q(\mathbf{s}^n) \left(\prod_{t=1}^{n} P(s_t | q,\theta) \right) \omega_0^\theta \omega_0^q} \\ &= \frac{b_{q\theta}(s_{n+1}) \left(\frac{1}{\sum_\theta \omega_n^{q\theta}} \right) b_q(\mathbf{s}^n) \left(\prod_{t=1}^{n+1} P(s_t | q, \theta) \right) \omega_0^\theta \omega_0^q}{\sum_q \sum_\theta b_{q\theta}(s_{n+1}) \left(\frac{1}{\sum_\theta \omega_n^{q\theta}} \right) b_q(\mathbf{s}^n) \left(\prod_{t=1}^{n+1} P(s_t | q, \theta) \right) \omega_0^\theta \omega_0^q}, \end{aligned} \tag{33}$$

where

$$\begin{aligned} b_{q\theta}(s_{n+1}) \left(\frac{1}{\sum_\theta \omega_n^{q\theta}} \right) &= \sum_\theta \left(\prod_{i=1}^{n+1} P(s_t | q, \theta) \right) b_q(\mathbf{s}^n) \omega_0^\theta \left(\frac{\omega_0^q}{\sum_\theta \omega_n^{q\theta}} \right) \\ &= b_q(\mathbf{s}^n) \left(\sum_\theta \left(\prod_{t=1}^{n+1} P(s_t | q, \theta) \right) \omega_0^\theta \right) \left(\frac{\omega_0^q}{\sum_\theta \omega_n^{q\theta}} \right). \end{aligned}$$

Equation (29) implies that Equation (33) equals Equation (31) if and only if $\omega_0^q = \sum_\theta \omega_n^{q\theta}$. Since $\omega_n^{q\theta} \equiv P^b(q, \theta | \{\mathbf{s}^n\})$, then $P^b[q, \theta | \{\mathbf{s}^n, s_{n+1}\}] \neq P^b[q, \theta | \text{prior} = \omega_n^b, \{s_{n+1}\}]$ if $P^b[q | \mathbf{s}^n] \neq \omega_0^q$ and $P^b[q, \theta | \{\mathbf{s}^n, s_{n+1}\}] = P^b[q, \theta | \text{prior} = \omega_n^b, \{s_{n+1}\}]$ if $P^b[q | \mathbf{s}^n] = \omega_0^q$. ■

A.13 Proof of Proposition 11

1. **Lemma 3** *Suppose the agent observes $n_a = n_b$ signals of* a*'s and* b*'s in alternating order: $\mathbf{s}^n = (a, b, \ldots a, b)$ where $n_a = n_b = k$. Then the biased agent is always underconfident that the agent is high quality.*

   Proof. An alternating sequence of $n_a = n_b = k$ signals of a's and b's generates:

   $$b_q(\mathbf{s}^n) = \left(\frac{1}{2}\right)^k \left(\prod_{i=1}^{k-1} (p_q(1-p_q))^{2i}\right) (p_q(1-p_q))^k = \left(\frac{1}{2}\right)^k (p_q(1-p_q))^{k^2}.$$

   This implies that $\frac{\partial}{\partial p_q}(b_q(\mathbf{s}^n)) < 0$ for all p_q:

   $$\frac{\partial}{\partial p_q}(b_q(\mathbf{s}^n)) = \left(\frac{1}{2}\right)^k k^2 \left(p_q(1-p_q)\right)^{k^2-1} (1-2p_q),$$

   Since $(p_H(1-p_H))^{k^2} < (p_L(1-p_L))^{k^2}$ whenever $p_H > p_L \geq \frac{1}{2}$ or $\frac{1}{2} \geq p_L > 1 - p_H$, then $b_H(\mathbf{s}^n) < b_L(\mathbf{s}^n)$ for all $p_H > p_L \geq \frac{1}{2}$, which implies that the biased agent's belief that the expert is high quality is underconfident relative to the Bayesian: $Pr^b(H|\mathbf{s}^n) < Pr^u(H|\mathbf{s}^n)$. ■

   Suppose the agent observes $n_a > n_b$ signals, where n_b a's and n_b b's alternate followed by the remaining $m \equiv n_a - n_b$ a's where $m \geq 1$: $\mathbf{s}^n = (a, b, a, b, \ldots, a, a, a)$. Since b_q is symmetric about $p_q = 1/2$, we focus on the case of $p_q \geq 1/2$ but the conclusion extends to $1/2 > p_L > 1 - p_H$. From Equations (18) and (20), we can see that

   - $\frac{\partial}{\partial p_q}(b_q(\mathbf{s}^n)) = 0$ when $p_q \in \{\frac{1}{2}, 1\}$,
   - $b_q(\mathbf{s}^n)) > 0$ when $p_q = \frac{1}{2}$,
   - $b_q(\mathbf{s}^n)) = 0$ when $p_q = 1$.

   Moreover, using the fact that $b_q(\mathbf{s}^n) = 0$ when $p_q = 1/2$, then

   $$\begin{aligned}\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}} &= b_q(\mathbf{s}^n)\left(-8n_b(n_b+m) + \sum_{i=1}^{m} 4i(i-1)\right) \\ &= b_q(\mathbf{s}^n)\left(-8n_b(n_b+m) + \frac{4}{3}m(m-1)(m+1)\right).\end{aligned}$$

   Thus for any given n_b, there exists some threshold $\frac{1}{2} < \check{p} < 1$ whenever $m > m'$, where $-8n_b(n_b + m') + \frac{4}{3}m'(m'-1)(m'+1) = 0$. Let $n_a^* = n_b + m^*$. Then for n_a, n_b where $0 \geq n_b < n_a^* < n_a$ and $p_L < p_H \leq \check{p}$, the agent overtrusts and is optimistic that the state is A. Since this is the sequence that generates the least trust by Proposition 1, then if it results in overtrust then all other sequences of such a combination must generate overtrust and optimism as well.

2. **Lemma 4** *After observing $n_a > 1$ and $n_b = 0$ signals in sequence or simultaneously, the pre-screener overtrusts and is overoptimistic about the reported state.*

   Proof. Without loss of generality, suppose the sequence is n_a a's: $\mathbf{s}^n = (a, a, \ldots, a)$ where $n_a = n$ and $n_b = 0$. Then

$$b_q(\mathbf{s}^n) = \prod_{i=1}^{n_a} (\frac{1}{2})(p_q^i + (1 - p_q)^i).$$

   Considering each ith component of $b_q(\mathbf{s}^n)$, $p_H^i + (1-p_H)^i > p_L^i + (1-p_L)^i$ is positive for $i > 0$ when $p_H > p_L \geq \frac{1}{2}$ or when $\frac{1}{2} \geq p_L > 1-p_H$, which implies that $b_H(s_{11} = a, s_{22} = a, \ldots s_{n_a,n_a} = a) > b_L(s_{11} = a, s_{22} = a, \ldots s_{n_a,n_a} = a)$. Thus, applying Proposition 2, the pre-screener overtrusts and is overoptimistic about the reported state when he observes $n_a > 1$ and $n_b = 0$ signals in sequence. Since the simultaneous case implies $b_q(\mathbf{s}^n) = p_q^{n_a} + (1 - p_q)^{n_a}$, then this argument also shows the result when the biased agent observes $n_a > 1$ signals simultaneously. ■

Lemma 5 *Consider a sequence of signals such that the first k observed signals are a, followed by k b signals: $\mathbf{s}^n = (a, a, \ldots, a, b, b, \ldots, b)$ where $n_a = n_b = k$. There exists some $\underline{p} > \frac{1}{2}$ and $\overline{p} < 1$ such that the pre-screener under-trusts if (1) $k \in \{1, 2, 3\}$, (2) if $\overline{p} \leq p_L < p_H$, or (3) if $p_L \leq \underline{p}$ and $p_H \geq \overline{p}$.*

Proof. WLOG, suppose the sequence is n_a a's, then n_b b's:. Then

$$\begin{aligned} b_q(\mathbf{s}^n) &= \left(\prod_{i=1}^{n_a} (\frac{1}{2})(p_q^i + (1 - p_q)^i)\right) \left(\prod_{i=1}^{n_b} (\frac{1}{2})(p_q^{n_a}(1 - p_q)^i + p_q^i(1 - p_q)^{n_a})\right) \\ &= (\frac{1}{2})^{n_a+n_b} \left(\prod_{i=1}^{n_a} (p_q^i + (1 - p_q)^i)\right) \left(\prod_{i=1}^{n_b} (p_q^{n_a}(1 - p_q)^i + p_q^i(1 - p_q)^{n_a})\right). \end{aligned}$$

In particular, if the sequence is $n_a = n_b = k$, then:

$$b_q(\mathbf{s}^n) = (\frac{1}{2})^{2k} \prod_{i=1}^{k} \left(p_q^i + (1 - p_q)^i\right) \left(p_q^k(1 - p_q)^i + p_q^i(1 - p_q)^k\right) \tag{34}$$

Characterizing $\frac{\partial}{\partial p_q}(b_q(\mathbf{s}^n))$ when $n_a = n_b = k$, we can use the property that a product of multiple factors is given by:

$$\frac{d}{dx}\left(\prod_{i=1}^{k} f_i(x)\right) = \left(\prod_{i=1}^{k} f_i(x)\right)\left(\sum_{i=1}^{k} \frac{f_i'(x)}{f_i(x)}\right).$$

Applying this to Equation (34) yields

$$\frac{\partial}{\partial p_q}(b_q(\mathbf{s}^n)) = (\frac{1}{2})^{2k}\left(\prod_{i=1}^{k}\left(p_q^i + (1-p_q)^i\right)\left(p_q^k(1-p_q)^i + p_q^i(1-p_q)^k\right)\right)$$
$$\left(\sum_{i=1}^{k} \frac{i(p_q^{i-1}-(1-p_q)^{i-1})(p_q^k(1-p_q)^i+p_q^i(1-p_q)^k)+\left(p_q^i+(1-p_q)^i\right)\left(k\left(p_q^{k-1}(1-p_q)^i-p_q^i(1-p_q)^{k-1}\right)+i\left(p_q^{i-1}(1-p_q)^k-p_q^k(1-p_q)^{i-1}\right)\right)}{\left(p_q^i+(1-p_q)^i\right)\left(p_q^k(1-p_q)^i+p_q^i(1-p_q)^k\right)}\right) \tag{35}$$

Since b_q is symmetric about $p_q = 1/2$, we focus on the case of $p_q \geq 1/2$ but the conclusion extends to $1/2 > p_L > 1 - p_H$. From Equation (34) we can see that

- $\frac{\partial}{\partial p_q}(b_q(\mathbf{s}^n)) = 0$ when $p_q \in \{\frac{1}{2}, 1\}$,
- $b_q(\mathbf{s}^n)) > 0$ when $p_q = \frac{1}{2}$,
- $b_q(\mathbf{s}^n)) = 0$ when $p_q = 1$.

Moreover, using the fact that $b_q(\mathbf{s}^n) = 0$ when $p_q = 1/2$, then

$$\begin{aligned}\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}} &= b_q(\mathbf{s}^n)\left(\sum_{i=1}^{k} 4(2i(i-1) + k(k-1) - 2ki)\right)\\ &= b_q(\mathbf{s}^n)\left(\frac{8}{3}k(-3k + k^2 - 1)\right),\end{aligned}$$

so $\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}}$ is negative when $k < \frac{3+\sqrt{13}}{2} \approx 3.3028$ and positive when $k > \frac{3+\sqrt{13}}{2}$.

Since $b_q(\mathbf{s}^n) = 0$ when $p_q = 1$, $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q}\Big|_{p_q=1} = 0$, and $b_q(\mathbf{s}^n) \geq 0$ for any $p_q \in [0, 1]$, then there exists some threshold $\overline{p} < 1$ such that $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} < 0$ and $b_q(\mathbf{s}^n) < b_q(\mathbf{s}^n)\Big|_{p_q=\frac{1}{2}}$ for all $p_q > \overline{p}$.

Since $b_q(\mathbf{s}^n) > 0$ when $p_q = \frac{1}{2}$, $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q}\Big|_{p_q=\frac{1}{2}} = 0$, and $\frac{\partial^2 b_q(\mathbf{s}^n)}{\partial p_q^2}\Big|_{p_q=\frac{1}{2}} > 0$ when $k > \frac{3+\sqrt{13}}{2}$, then there exists some threshold $\underline{p} > \frac{1}{2}$ such that $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} > 0$ and $b_q(\mathbf{s}^n) > b_q(\mathbf{s}^n)\Big|_{p_q=\frac{1}{2}}$ for all $p_q < \underline{p}$ when $k > \frac{3+\sqrt{13}}{2}$.

When $k \leq \frac{3+\sqrt{13}}{2}$, we can show by direct computation of $b_q(\mathbf{s}^n)$ that $\frac{\partial b_q(\mathbf{s}^n)}{\partial p_q} < 0$ for all $p_q \in (\frac{1}{2}, 1)$ when $k \in \{1, 2, 3\}$.

This implies that the pre-screener under-trusts for all values of $p_L < p_H$ whenever $k \leq 3$, since $b_H(\mathbf{s}^n) < b_L(\mathbf{s}^n)$. When $k > 3$, there are two other sufficient conditions for the pre-screener to under-trust: (1) if $\overline{p} \leq p_L < p_H$, or (2) if $p_L \leq \underline{p}$ and $p_H \geq \overline{p}$ where $\underline{p} > \frac{1}{2}$ and $\overline{p} < 1$. If either of these sufficient conditions is met, then $b_H(\mathbf{s}^n) < b_L(\mathbf{s}^n)$ for $k > 3$. $\blacksquare$

Lemma 4 shows that the agent overtrusts and is overoptimistic about the reported state for a given $n_a > 1$ and $n_b = 0$. Clearly, the agent's degree of overtrust is monotonically decreasing as n_b increases. Lemma 5 shows that there exists some $\underline{p} > \frac{1}{2}$ and $\overline{p} < 1$ such that the pre-screener under-trusts if (1) $k \in \{1, 2, 3\}$, (2) if $\overline{p} \leq p_L < p_H$, or (3) if $p_L \leq \underline{p}$ and $p_H \geq \overline{p}$. By the intermediate value theorem, there exists some $\hat{n}_b$ such that the agent under-trusts when $\mathbf{s}^n = (a, a, \ldots, a, b, b, \ldots, b)$ where $0 \leq \hat{n_b} \leq n_b < n_a$. By Proposition 1, this is the sequence most likely to generate overtrust, so *all other sequences* of such a fixed combination (n_a, n_b) will also result in under-trust and pessimism about the mostly likely state. Thus, if one of the last two sufficient conditions for Lemma 5 is satisfied, then there exists some $\hat{n}_b$ such that the agent under-trusts when $\mathbf{s}^n = (a, a, \ldots, a, b, b, \ldots, b)$ where $0 \leq \hat{n_b} \leq n_b < n_a$.