

Polarization and Ambiguity*

Sandeep Baliga[†] Eran Hanany[‡] Peter Klibanoff[§]

February 1, 2013

Abstract

We offer a theory of polarization as an optimal response to ambiguity. Suppose individual A's beliefs first-order stochastically dominate individual B's. They observe a common signal. They exhibit polarization if A's posterior dominates her prior and B's prior dominates her posterior. Given agreement on conditional signal likelihoods, we show that polarization is impossible under Bayesian updating or after observing extreme signals. However, we also show that polarization can arise after intermediate signals as ambiguity averse individuals implement their optimal prediction strategies. We explore when this polarization will occur and the logic underlying it.

*Some previous versions of this paper had the title "Ambiguity, Dynamic Consistency and Behavioral Phenomena." We thank Subir Bose, Jeff Ely, Itzhak Gilboa, Anna Gumen, Christian Kellner, Sujoy Mukerji, Marco Ottaviani, Emre Ozdenoren, Wolfgang Pesendorfer, Ben Polak, Jean-Marc Tallon, three anonymous referees, the coeditor and a number of seminar and conference audiences for helpful comments and discussion. We especially thank Eran Shmaya for his help with the proof of Theorem 2.1. Daniel Garrett and Luciano Pomatto provided excellent research assistance. This research was partially supported by Grant No. 2006264 to Hanany and Klibanoff from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. The authors have no financial or other material interests related to this research to disclose.

[†]Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA. E-mail: baliga@kellogg.northwestern.edu

[‡]Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel. Email: hananye@post.tau.ac.il

[§]Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA. E-mail: peterk@kellogg.northwestern.edu

1 Introduction

A number of voters are in a television studio before a U.S. Presidential debate. They are asked the likelihood that the Democratic candidate will cut the budget deficit, as he claims. Some think it is likely and others unlikely. The voters are asked the same question again after the debate. They become even more convinced that their initial inclination is correct. A similar phenomenon can arise in financial markets. “Bulls” and “bears” have different beliefs. On seeing the same macroeconomic forecasts, they become more bullish and bearish respectively. Individuals observe the same evidence, and yet their beliefs move in opposite directions and end up further apart.

Similar polarization of beliefs has been documented in experiments. For example, Darley and Gross (1983) randomize subjects into different groups. They show one group evidence suggesting a child is from a high socioeconomic background; another that she is from a low socioeconomic background. The former predict the child’s reading abilities are higher than the latter. The groups then watch a film of the child taking an oral test on which she answers some questions correctly and others incorrectly. Those who received the information that the child came from a high socioeconomic background, rate her abilities higher than before; those who received the information indicating she came from a low socioeconomic background rate her lower than before. Thus, the common evidence - the film - leads beliefs to polarize.

We follow Dixit and Weibull (2007) in defining polarization as follows: Suppose that, prior to observing a common signal, two individuals have different beliefs and individual A ’s belief first-order stochastically dominates individual B ’s. Their beliefs exhibit *polarization* if, after observing the common signal, individual A ’s posterior dominates his prior and individual B ’s prior dominates his posterior.

Consider two individuals who agree on the probability of each signal conditional on an underlying parameter and use Bayes’ rule to update their beliefs. We show polarization cannot occur. As individuals share the same theory connecting parameters to signals, a given signal increases one individual’s belief

if and only if it increases the other's (Theorem 2.1).

In contrast, we show that polarization can occur as an optimal response to ambiguity aversion (i.e., aversion to subjective uncertainty about probabilities). We do this within a simple prediction model: An individual must predict the value of a parameter that determines the distribution of a random variable. The individual views the parameter as ambiguous and is ambiguity averse. He observes a signal that can inform his predictions. His payoff is decreasing in the squared difference between his prediction and the parameter. This is a standard model apart from ambiguity aversion.

How does ambiguity aversion affect behavior? An individual is exposed to ambiguity when the expected payoff to his strategy varies with probabilities over which he is uncertain. Different strategies may involve different exposure to ambiguity. Suppose there are just two possible parameter values, 0 and 1. If the individual predicts $\frac{1}{2}$, then the difference between the prediction and the parameter is the same no matter what the parameter value. Thus, this prediction strategy completely hedges against ambiguity, i.e., removes any exposure to the ambiguity about the parameter. However, if the individual predicts 1, the squared difference is much higher when the parameter is 0 than when it is 1, exposing the individual to ambiguity. An ambiguity averse individual will tend to favor strategies that reduce exposure. Reducing exposure is not the only concern of such an individual – for example, the more weight his beliefs place on the parameter equaling 1, the higher his optimal prediction.

Simple strategies such as “predict $\frac{1}{2}$ ” or “predict 1” are generally not optimal when a signal is forthcoming, as these strategies fail to condition on anticipated information. An optimal contingent strategy will make the prediction an increasing function of the observed signal likelihood ratio. We focus on individuals who form an ex-ante optimal contingent strategy (i.e., optimal assuming full commitment to that strategy once chosen) and who are indeed willing to carry it out after each possible contingency. Such an individual is said to be *dynamically consistent*.¹ Dynamic consistency is assumed in almost

¹See e.g., Hanany and Klibanoff (2007, 2009) for such an approach to modeling ambiguity averse individuals and for discussion and references to alternative approaches.

all standard economic analysis. Any theory of updating that is not dynamically consistent will lead to worse outcomes as evaluated by ex-ante welfare. Also, studying the dynamically consistent case identifies the key effects leading to polarization that apply even when substantial dynamic inconsistency may be present (see the Concluding Remarks).

Suppose the individual concludes that “predict $\frac{3}{4}$ if the signal is high, $\frac{1}{2}$ if the signal is medium, and $\frac{1}{4}$ if the signal is low” is his optimal strategy. This strategy leaves him partially exposed to the ambiguity about the parameter. Notice, however, that this exposure varies with the signal received. He is not exposed to ambiguity after a medium signal but is exposed after a high or low signal. Under ambiguity aversion, the greater exposure to ambiguity after seeing a high or low signal may lead to an increased desire to hedge against this ambiguity, while the lack of exposure after seeing a medium signal may diminish the value of hedging. These changed hedging motives, *ceteris paribus*, could lead the individual to want to depart from the ex-ante optimal strategy. We call this the *hedging effect*. There is also a more standard effect having nothing to do with ambiguity attitude. After a signal is realized, the likelihoods of this signal are no longer relevant for optimality going forward. We call this the *likelihood effect*. Because of this effect, after seeing the signal, if beliefs were not updated to incorporate that signal’s likelihoods, the individual might want to depart from the ex-ante optimal strategy. Dynamically consistent updating must neutralize both the hedging and the likelihood effects of the signal on the incentives of an ambiguity averse individual. Bayesian updating counterbalances only the likelihood effect. The presence of the hedging effect leads dynamically consistent updating to necessarily depart from Bayes’ rule under ambiguity aversion.² Importantly, the hedging effect may alter the direction of updating. Moreover, the hedging effect (but not the likelihood effect) depends on the ex-ante optimal strategy which, in turn, is influenced by the individual’s beliefs before observing the signal. Through this chain of reasoning, beliefs can influence the direction of updating. This allows for the

²In contrast, for an expected utility maximizer, dynamic consistency requires that subjective beliefs are updated using Bayes’ rule, thus ruling out polarization.

possibility of polarization. We study when polarization does and does not occur.

Even under ambiguity aversion, we show that polarization does not occur after observing the highest or the lowest signals (Theorem 3.1). Thus, polarization is a possibility only at signals with an intermediate likelihood ratio. We can offer a particularly clean result if the intermediate signal is neutral (i.e., has equal probability under both parameter values). Then the hedging effect is the only reason to update beliefs – there is no likelihood effect. We show that individuals with sufficiently extreme and opposite beliefs display polarization after observing a common neutral intermediate signal (Theorem 3.2). When there are exactly three possible signals and there is constant relative ambiguity aversion, we provide necessary and sufficient conditions for polarization (Theorem 3.3) that apply even when the intermediate signal is not neutral.

All of the above results apply whether or not the two individuals have the same degree of ambiguity aversion as long as initial beliefs differ. Finally, even if individuals have the same beliefs, as long as several conditionally independent observations are available then if they observe different private signals before observing a common signal, they can have different beliefs by the time they see the common signal and, at that point, our above results apply. We now turn to our model. Related literature is discussed at the end of the paper.

2 The Model and Benchmark Result

Consider an individual who is concerned with the value of a parameter $\theta \in \Theta \subset \mathbb{R}$. His beliefs are given by a full-support prior μ . To help inform the individual about θ , conditionally independent observations from a random variable X given θ may be available. This random variable has distribution π_θ and takes values in a finite set \mathcal{X} such that each $x \in \mathcal{X}$ has $\pi_\theta(x) > 0$ for some $\theta \in \Theta$. For example, θ might indicate a child’s reading ability, while π_θ might be the distribution of scores on a reading test for a child with that ability.

We assume that Θ is finite and, without loss of generality, index Θ so that

$\theta_1 < \theta_2 < \dots < \theta_{|\Theta|}$. A distribution $\hat{\eta}$ (first-order) stochastically dominates $\check{\eta}$ if

$$\sum_{i=1}^k \check{\eta}(\theta_i) \geq \sum_{i=1}^k \hat{\eta}(\theta_i) \text{ for all } k \in \{1, 2, \dots, |\Theta|\}.$$

The dominance is strict if at least one of these inequalities is strict. We adopt the following definition of polarization due to Dixit and Weibull (2007):

Definition 2.1. Fix two individuals with beliefs $\check{\eta}$ and $\hat{\eta}$ over Θ and with common support such that $\hat{\eta}$ stochastically dominates $\check{\eta}$. After they both observe a signal $x \in \mathcal{X}$ whose likelihood given $\theta \in \Theta$ is $\pi_\theta(x)$, we say that *polarization* occurs if and only if the resulting posterior beliefs, $\check{\nu}$ and $\hat{\nu}$ respectively, lie further apart, i.e., $\check{\eta}$ stochastically dominates $\check{\nu}$ and $\hat{\nu}$ stochastically dominates $\hat{\eta}$ with at least one dominance strict.

Starting from two beliefs, one higher than the other, polarization occurs when the observation of a common signal leads the higher belief to move higher and the lower belief to move lower. The requirement that one initial belief be higher than the other is essential to rule out the possibility that posteriors move in opposite directions but toward each other. If initial beliefs were not ranked, one were updated upward and the other downward, it would not be clear whether beliefs had moved toward or away from one another. Even if the requirement that initial beliefs be ranked were somehow relaxed, the results of this paper would continue to hold. Our benchmark impossibility result (Theorem 2.1) does not rely on the initial beliefs being ranked. Additionally, in the setting we use in the next section to show polarization under ambiguity aversion, beliefs can always be ranked by dominance. In more complex settings, this aspect of defining polarization may matter. We favor Definition 2.1 because whenever it identifies polarization, there is no doubt that beliefs are moving away from one another.

The following result shows that, in this setting, under Bayesian updating, irrespective of any non-belief aspect of preference, polarization cannot occur. The theorem and proof formalize the intuitive statement that, as long as their priors share the same support, if two individuals who use Bayes' rule see the

same observation and agree on the probability of each observation conditional on the parameter, it is impossible for them to update in opposite directions in the sense of first-order stochastic dominance. All proofs are in the online Appendix.

Theorem 2.1. *Fix two individuals as in Definition 2.1. Polarization cannot occur if they use Bayesian updating.*

3 Polarization and Ambiguity

For the remainder of the paper, suppose the individual's goal is to predict the value of the parameter θ . For tractability, we assume $\Theta = \{0, 1\}$ so there are just two possible parameter values. Accordingly, throughout this section let μ denote $\mu(\theta = 1)$. For example, $\theta = 1$ might indicate a child has high reading ability and the prediction $\alpha \in [0, 1]$ might be interpreted as a probability that the child has high reading ability. We make the standard assumption that the payoff to a prediction α is given by quadratic loss; i.e., $-(\alpha - \theta)^2$. To avoid tedious corner cases, we assume π_θ has full support for each θ . The individual can condition his prediction on an observation $x \in \mathcal{X}$ drawn from the random variable X .³ Hence, his strategy is a function $\alpha : \mathcal{X} \rightarrow \mathbb{R}$. We use uppercase X and lowercase x to distinguish between the random variable and its realization, respectively.⁴ We assume the individual views θ as ambiguous, is risk neutral and evaluates prediction strategies according to ambiguity averse smooth ambiguity preferences (Klibanoff, Marinacci and Mukerji (2005)). Specifically, any prediction strategy is evaluated according to the concave objective function

$$E_{(\mu, 1-\mu)\phi} [E_{\pi_\theta} (-(\alpha(X) - \theta)^2)],$$

³Extensions of our results to the case of multiple conditionally independent observations may be found in the working paper version (Baliga, Hanany and Klibanoff (2013)).

⁴For example, $\alpha(X)$ is a random (since the observation is yet-to-be-realized) prediction, while $\alpha(x)$ is the realized prediction.

where ϕ is increasing, concave and continuously differentiable, E is the expectation operator (with respect to the subscripted distribution), $(\mu, 1 - \mu)$ is the belief about θ and π_θ is the distribution of X given θ . Observe that if ϕ is linear (i.e., ambiguity neutrality), the objective function reduces to expected quadratic loss. The concavity of ϕ reflects ambiguity aversion. We will sometimes additionally assume constant relative ambiguity aversion $\gamma \geq 0$, in which case $\phi(u) = -\frac{(-u)^{1+\gamma}}{1+\gamma}$ for $u \leq 0$.

The optimal strategy $\alpha^*(x)$ is the unique solution to the first-order conditions:

$$\frac{\alpha^*(x)}{1 - \alpha^*(x)} \frac{\phi'[E_{\pi_0}(-(\alpha^*(X))^2)]}{\phi'[E_{\pi_1}(-(1 - \alpha^*(X))^2)]} \frac{\pi_0(x)}{\pi_1(x)} = \frac{\mu}{1 - \mu} \quad (3.1)$$

for each $x \in \mathcal{X}$.

The difference from the usual prediction problem with ambiguity neutrality is the presence of the term $\frac{\phi'[E_{\pi_0}(-(\alpha^*(X))^2)]}{\phi'[E_{\pi_1}(-(1 - \alpha^*(X))^2)]}$ on the left-hand side of (3.1). Under ambiguity aversion, ϕ is concave, and this term reflects the desire to hedge or reduce the variation in expected payoffs as a function of the ambiguous parameter θ . Ambiguity aversion ensures that when expected payoffs across the θ 's differ, the ϕ' ratio pushes the optimal prediction strategy in the direction of equalizing them by moving the predictions toward the θ with the lower expected payoff. This is the manifestation of the value that ambiguity averse individuals place on hedging against ambiguity. For this reason, we call this ϕ' ratio the *hedging motive*. It compares the marginal value of an extra unit of expected utility when $\theta = 0$ to the marginal value when $\theta = 1$. When these expected payoffs are equal (i.e., a perfect hedge) the hedging motive equals one. Values above (below) one reflect a stronger (weaker) desire to shift expected payoff from $\theta = 1$ to $\theta = 0$, i.e., to hedge by adjusting the prediction strategy α^* downward.

We have the following useful implication of (3.1) that is true independent of ambiguity attitude: for any $x, y \in \mathcal{X}$,

$$\frac{\alpha^*(x)}{1 - \alpha^*(x)} \frac{\pi_0(x)}{\pi_1(x)} = \frac{\alpha^*(y)}{1 - \alpha^*(y)} \frac{\pi_0(y)}{\pi_1(y)}. \quad (3.2)$$

The intuition for this equality is the standard one concerning equating marginal rates of substitution across signal realizations.

In a prediction problem, *updating* maps beliefs about θ and new observations to posterior beliefs about θ . *Dynamically consistent updating* is updating that preserves the optimality of the contingent strategy $\alpha^*(x)$ after the observation is realized (i.e., ex-ante optimal updating).⁵ Let ν_x denote the posterior probability of $\theta = 1$ after observing x . Dynamically consistent updating is equivalent to these posteriors ν_x satisfying

$$\frac{\alpha^*(x)}{1 - \alpha^*(x)} \frac{\phi'[-(\alpha^*(x))^2]}{\phi'[-(1 - \alpha^*(x))^2]} = \frac{\nu_x}{1 - \nu_x} \quad (3.3)$$

for all $x \in \mathcal{X}$. Note that (3.3) is simply the first-order condition of the continuation prediction problem, evaluated at the ex-ante optimal strategy $\alpha^*(x)$, after x has been realized and assuming beliefs at that point are ν_x . It therefore guarantees that $\alpha^*(x)$ remains optimal upon seeing the observation. After the next result, we describe the difference, under ambiguity aversion, between dynamically consistent updating and Bayesian updating and show how the former allows polarization. First we show that several natural properties that hold under ambiguity neutrality continue to hold under ambiguity aversion:

Proposition 3.1. (i) *For each realization $x \in \mathcal{X}$, the optimal prediction $\alpha^*(x)$ is an increasing function of μ (the prior probability of $\theta = 1$) and of the likelihood ratio $\frac{\pi_1(x)}{\pi_0(x)}$. The best constant (i.e., not varying with x) prediction is also an increasing function of μ ;*

(ii) *The optimal prediction after observing x is an increasing function of the posterior ν_x . Furthermore, $\nu_x \gtrless \mu$ if and only if the optimal prediction after observing x and the (ex-ante) best constant prediction are similarly ordered. Under dynamically consistent updating, the optimal prediction after observing x is $\alpha^*(x)$.*

Proposition 3.1 implies that polarization as defined in terms of beliefs is equivalent to polarization in actions (here, predictions). A common signal

⁵For a thorough discussion and analysis of dynamically consistent updating under ambiguity aversion see Hanany and Klibanoff (2009).

moves optimal actions further apart and in opposite directions exactly when that signal moves beliefs further apart and in opposite directions. To see this, observe from the first part of the proposition that prior beliefs have the same order as the respective best predictions made without using the observation. From the second part of the proposition, posterior beliefs after a common signal compare to the prior beliefs in the same way as the optimal predictions after a common signal compare to the best predictions made without using the signal. Combining these yields the equivalence.

For the remainder of this section we assume dynamically consistent updating. Suppose a signal x is observed. Dynamic consistency requires that the optimal prediction strategy after this observation also be the optimal prediction strategy ex ante contingent on observing x . As emphasized above, under ambiguity aversion, the optimal prediction strategy is partly driven by the desire to hedge. Before the signal is realized, the hedging motive is as in (3.1). However, after observing x , the interim hedging motive is as in (3.3). These hedging motives are typically *not* equal. As we highlighted in the Introduction, the individual's hedging motive changes since he no longer needs to account for variation in his expected payoffs induced by the signal realization. To carry out the optimal prediction strategy, dynamically consistent updating departs from Bayesian updating in a way that exactly offsets this hedging effect. We use this to offer a characterization of the direction of dynamically consistent updating:

Proposition 3.2. *The posterior ν_x is above/equal to/below the prior μ if and only if*

$$\frac{\phi'[-(\alpha^*(x))^2]}{\phi'[-(1-\alpha^*(x))^2]} \frac{\pi_1(x)}{\pi_0(x)} \begin{matrix} \geq \\ < \end{matrix} \frac{\phi'[E_{\pi_0}(-(\alpha^*(X))^2)]}{\phi'[E_{\pi_1}(-(1-\alpha^*(X))^2)]}. \quad (3.4)$$

Notice that in addition to the hedging motive terms, (3.4) includes the likelihood ratio for the newly observed signal. This term reflects the likelihood effect referred to in the Introduction. Thus (3.4) formalizes the statement that dynamically consistent updating offsets both the hedging and likelihood effects (while Bayesian updating offsets only the likelihood effect). The condition in (3.4) is not always easy to apply, as it involves the endogenously determined

optimal strategy α^* . Nevertheless, we can prove some general properties of updating directly from this inequality: Observing a signal x^H with the highest likelihood ratio (i.e., $x^H \in \arg \max_{x \in \mathcal{X}} \frac{\pi_1(x)}{\pi_0(x)}$) always leads to updating upwards and observing a signal, x^L , with the lowest likelihood ratio always leads to updating downwards. One implication is that polarization cannot occur after “extreme” common signals.

Theorem 3.1. *After observing x^H (resp. x^L), the posterior ν_x is above (resp. below) the prior μ . It is also above (resp. below) the Bayesian update of μ given x^H (resp. x^L).*

Remark 3.1. *If the signal is informative, so that $\frac{\pi_1(x^H)}{\pi_0(x^H)} > 1 > \frac{\pi_1(x^L)}{\pi_0(x^L)}$, then above (resp. below) in the statement of the theorem may be replaced by strictly above (resp. strictly below).*

We turn to our main positive results. Suppose there are two individuals with beliefs $\hat{\eta} > \check{\eta}$ (we continue our abuse of notation and denote $\hat{\eta}(\theta = 1)$ and $\check{\eta}(\theta = 1)$ by $\hat{\eta}$ and $\check{\eta}$ respectively). In all other respects, the individuals are equivalent. If they are ambiguity neutral, we know that polarization is impossible, from Theorem 2.1. If they both observe the same extreme signal, we know they will update in the same direction and thus fail to polarize, from Theorem 3.1. So, assume the individuals are ambiguity averse and there are at least three signals with distinct likelihood ratios. Thus, there is at least one intermediate (i.e., non-extreme) signal and these are the only signals after which individuals’ beliefs can possibly exhibit polarization. When and why can polarization occur?

Suppose an individual observes a signal, x^M , with intermediate likelihood ratio. Substituting for predictions $\alpha^*(x)$, $x \neq x^M$ using (3.2), inequality (3.4)

becomes:

$$\begin{aligned}
& \frac{\phi'[-(\alpha^*(x^M))^2] \pi_1(x^M)}{\phi'[-(1-\alpha^*(x^M))^2] \pi_0(x^M)} \\
& \geq \frac{\phi'[-(\alpha^*(x^M))^2 \sum_{y \in \mathcal{X}} \pi_0(y) \left(\frac{\frac{\pi_1(y)}{\pi_0(y)}}{\alpha^*(x^M) \frac{\pi_1(y)}{\pi_0(y)} + (1-\alpha^*(x^M)) \frac{\pi_1(x^M)}{\pi_0(x^M)}} \right)^2]}{\phi'[-(1-\alpha^*(x^M))^2 \sum_{y \in \mathcal{X}} \pi_1(y) \left(\frac{\frac{\pi_1(x^M)}{\pi_0(x^M)}}{\alpha^*(x^M) \frac{\pi_1(y)}{\pi_0(y)} + (1-\alpha^*(x^M)) \frac{\pi_1(x^M)}{\pi_0(x^M)}} \right)^2]}
\end{aligned} \tag{3.5}$$

The direction of this inequality determines the direction of updating. The connection between $\alpha^*(x^M)$ (and thus beliefs, since $\alpha^*(x^M)$ is increasing in beliefs by Proposition 3.1) and the direction of this inequality may be quite complex. It is simplest in the case where the signal is not only intermediate but also *neutral* (i.e., $\pi_0(x) = \pi_1(x)$). In the theorem below, we show that when $\alpha^*(x^M)$ (and thus belief) for one individual is close to 0 and for another is close to 1, polarization occurs after they commonly observe a neutral signal.

Theorem 3.2. *Polarization and Ambiguity: Assume there is a neutral signal, at least one informative signal and twice continuously differentiable ϕ with $\phi'' < 0 < \phi'$. Polarization occurs after commonly observing a neutral signal if belief $\hat{\eta}$ is sufficiently close to 1 and belief $\check{\eta}$ is sufficiently close to 0.*

Sketch of proof (for the full proof see the online Appendix): When $\alpha^*(x^M)$ is close to 0, if $\theta = 0$ then predictions will be close to perfect, both interim and ex-ante. Since payoffs are relatively insensitive to small changes in predictions in the neighborhood of perfection, any differences in the interim and ex-ante expected payoffs when $\theta = 0$ (i.e., any differences in the arguments of ϕ' in the numerators on each side of (3.5)) will be very small and will have minimal influence on updating (since $-\frac{\phi''}{\phi'}$ is finite).⁶ In contrast, if $\theta = 1$, predictions close to 0 will be very costly and small improvements in those predictions would be valuable. Therefore, (since $-\frac{\phi''}{\phi'}$ is non-zero) it is the

⁶The role of assuming $\phi'' < 0 < \phi'$ is to ensure that both the hedging motive, $\frac{\phi'[-\alpha^2]}{\phi'[-(1-\alpha)^2]}$, and ambiguity aversion (as measured by $-\frac{\phi''}{\phi'}$, the coefficient of (absolute) ambiguity aversion (see Klibanoff, Marinacci and Mukerji (2005))), are bounded away from zero and infinity.

differences in interim and ex-ante expected payoffs when $\theta = 1$ that drive the comparison of hedging motives when predictions are close to 0. Differentiating the arguments of the ϕ' terms in the denominators with respect to $\alpha^*(x^M)$ and evaluating at $\alpha^*(x^M) = 0$, yields that the ex-ante expected payoff when $\theta = 1$ is higher than the interim payoff when $\theta = 1$ if and only if the expected likelihood, $\sum_{y \in \mathcal{X}} \pi_1(y) \frac{\pi_1(y)}{\pi_0(y)}$, is higher than the realized likelihood, $\frac{\pi_1(x^M)}{\pi_0(x^M)}$. This comparison reflects the fact that the predictions $\alpha^*(y)$ optimally move toward 1 by an amount proportional to the likelihood $\frac{\pi_1(y)}{\pi_0(y)}$ so the expected or realized likelihoods reflect the expected or realized improvements in the prediction when $\theta = 1$. Notice that this expected likelihood is always larger than 1 because of the complementarity between the π_1 terms, so that if x^M is a neutral signal this condition will be satisfied.

As a result, when the signal likelihood $\frac{\pi_1(x^M)}{\pi_0(x^M)}$ is below $\sum_{y \in \mathcal{X}} \pi_1(y) \frac{\pi_1(y)}{\pi_0(y)}$, for all sufficiently low beliefs η (so that $\alpha^*(x^M)$ is sufficiently close to 0), the hedging motive is bigger ex-ante than after seeing the signal and so updating will be shaded downward compared to Bayesian updating. Similar reasoning for $\alpha^*(x^M)$ close to 1 shows that when $\frac{\pi_1(x^M)}{\pi_0(x^M)}$ lies above $\frac{1}{\sum_{y \in \mathcal{X}} \pi_0(y) \frac{\pi_0(y)}{\pi_1(y)}}$, for sufficiently high η , updating will be shaded upward compared to Bayesian updating. When the signal is neutral, Bayesian updating is flat, so these arguments imply updating will be downward when belief is sufficiently low and upward when belief is sufficiently high, generating polarization.

Remark 3.2. In this argument, the important aspect of quadratic loss is that the marginal payoff to improving a prediction is diminishing in the quality (i.e., closeness to the truth) of the prediction and vanishes at perfection. Any payoff function of the form $\psi(|\alpha - \theta|)$ where $\psi : [0, 1] \rightarrow R$ is a twice continuously differentiable function satisfying $\psi'(0) = 0$ and $\psi'' < 0$ will yield a similar result.

3.1 Threshold Rules for Updating

To further investigate when polarization occurs, we turn to a particularly clean structure for determining the direction of updating. Given an observation x ,

say that updating follows a *threshold rule* if there is a threshold $\tau \in [0, 1]$ such that all beliefs above the threshold are updated upward and those below the threshold are updated downward. Under ambiguity neutrality, the threshold is always degenerate – a given observation x either leads all priors to be updated upward or all priors to be updated downward depending on how the likelihood ratio $\frac{\pi_1(x)}{\pi_0(x)}$ compares to 1. In contrast, under ambiguity aversion, updating may follow a non-trivial threshold rule. In the online Appendix, we provide a characterization of when updating follows a threshold rule (Proposition A.2). To provide an explicit description, for the remainder of this section, we specialize by assuming constant relative ambiguity aversion and that there are exactly three distinct likelihood ratios associated with signals. Under these conditions, we show that updating always follows a threshold rule and we can explicitly derive the thresholds. By Theorem 3.1, all beliefs are updated in the same direction after extreme signals. Hence, we study thresholds given the intermediate signal.

Polarization is obviously impossible if two individuals have the same beliefs and have the same degree of ambiguity aversion. If, however, there is heterogeneity on either dimension, individuals may exhibit polarization when they observe a common signal. Theorem 3.3 and Proposition 3.3 characterize the conditions for a signal to lead to polarization when there is heterogeneity across individuals in beliefs and/or ambiguity aversion.

Theorem 3.3. *Assume constant relative ambiguity aversion and exactly three distinct likelihood ratios. There exist $\hat{\tau}, \check{\tau} \in [0, 1]$ such that polarization occurs after commonly observing a signal with the non-extreme likelihood ratio if and only if belief $\hat{\eta} \geq \hat{\tau}$ and belief $\check{\eta} \leq \check{\tau}$ with at least one inequality strict.*

Notice that whenever the thresholds satisfy $\hat{\tau} < 1$ or $\check{\tau} > 0$, there exist beliefs that generate polarization. The theorem relies on the following proposition establishing that updating follows a threshold rule. The proposition is proved by explicitly constructing the threshold.

Proposition 3.3. *Assume constant relative ambiguity aversion $\gamma > 0$ and exactly three distinct likelihood ratios. The posterior ν_{x^M} is above/equal to/below*

the prior μ when the latter is above/equal to/below a threshold $\tau(\gamma, \pi_0, \pi_1)$ that is independent of μ .

Theorem 3.3 and Proposition 3.3 specialize immediately for the cases where heterogeneity is either only in beliefs or only in ambiguity aversion.

Corollary 3.1. *Assume exactly three distinct likelihood ratios. Then,*

(a) *Polarization with Homogeneous Beliefs: Two individuals with beliefs η and constant relative ambiguity aversions $\hat{\gamma}$ and $\check{\gamma}$ exhibit polarization after observing the intermediate signal if and only if*

$$\tau(\hat{\gamma}, \pi_0, \pi_1) \leq \eta \leq \tau(\check{\gamma}, \pi_0, \pi_1)$$

with at least one inequality strict; and

(b) *Polarization with Heterogeneous Beliefs: Two individuals with constant relative ambiguity aversion γ and beliefs $\hat{\eta}$ and $\check{\eta}$ exhibit polarization after observing the intermediate signal if and only if*

$$\hat{\eta} \geq \tau(\gamma, \pi_0, \pi_1) \geq \check{\eta}$$

with at least one inequality strict.

When the intermediate signal is a neutral signal, thresholds always lie strictly between 0 and 1, and take a particularly simple form:

Corollary 3.2. *Assume constant relative ambiguity aversion $\gamma > 0$ and exactly three distinct likelihood ratios. If x^M is a neutral signal, the threshold $\tau(\gamma, \pi_0, \pi_1)$ is*

$$\frac{1}{1 + \left(\frac{\pi_1(x^H)}{\pi_0(x^H)} \frac{\pi_1(x^L)}{\pi_0(x^L)} \right)^{\gamma + \frac{1}{2}}} \in (0, 1).$$

Heterogeneous tastes or beliefs are the source of polarization under ambiguity in Theorem 3.3. But this cannot explain the polarization observed by Darley and Gross (1983), where the groups exhibiting polarization were homogeneous. In their study, heterogeneity was induced across groups at the

interim stage by showing them different initial evidence. Our previous results, together with their extension to prediction after observing multiple signals as described in the working paper version (Baliga, Hanany and Klibanoff (2013)), imply that exactly this device can generate polarization in an ex ante homogeneous prediction problem. For example, suppose there are two individuals with a common coefficient of relative ambiguity aversion $\gamma > 0$, a common prior $\mu = \frac{1}{2}$ and signals with three distinct likelihood ratios and symmetric likelihoods (i.e., $\pi_0(x^L) = \pi_1(x^H)$, $\pi_0(x^M) = \pi_1(x^M)$ and $\pi_0(x^H) = \pi_1(x^L)$). The individuals are allowed to condition their prediction on two conditionally independent observations. Suppose one individual observes the sequence $\{x^L, x^M\}$, while the other observes the sequence $\{x^H, x^M\}$. The multi-signal version of Theorem 3.1 implies that after one observation the first individual will have updated beliefs $\tilde{\eta} < \frac{1}{2}$ and the second individual will have updated beliefs $\hat{\eta} > \frac{1}{2}$. From Corollary 3.2 and symmetry of the likelihoods, the threshold for updating upon observing x^M is $\tau(\gamma, \pi_0, \pi_1) = \frac{1}{2}$. Since the beliefs $\tilde{\eta}$ and $\hat{\eta}$ are on opposite sides of this threshold, Theorem 3.3 implies that polarization will occur after the second observation, x^M .

More generally, as long as the threshold is interior and enough observations are available, polarization is possible after an intermediate signal. This follows since if one individual observes a long sequence of high signals and another observes a long sequence of low signals, their posteriors will end up on different sides of this threshold. If they then observe a common intermediate signal, they will update in opposite directions and polarize. A formal statement of this result on polarization in a homogeneous environment may be found in Baliga, Hanany and Klibanoff (2013, Theorem 3.4).

4 Related Literature and Concluding Remarks

4.1 Related Literature

Dixit and Weibull (2007) show that polarization cannot occur under Bayesian updating in the standard linear-normal model where individuals' (different)

priors and (common) noise are normally distributed. Signals in this model satisfy the monotone likelihood ratio property (MLRP). They also argue via example that polarization can occur if signals do not satisfy MLRP. On closer inspection, however, their examples violating MLRP do not display polarization. In fact, our Theorem 2.1 shows that polarization cannot occur under Bayesian updating, whether MLRP or normality holds or not. Instead, in their examples, while the means or the medians of two individuals' beliefs move further apart after observing a common signal, their beliefs are not further apart according to stochastic dominance.

Acemoglu, Chernozhukov and Yildiz (2009) study asymptotic disagreement in a model where individuals have different priors on parameters and also different distributions on signals conditional on the parameter. They show that posteriors on parameters can diverge. Kondor (2012) shows that polarization can be generated when individuals see different private signals that are correlated with a common public signal. Andreoni and Mylovanov (2012) provide a similar theory and test their model experimentally. Rabin and Schrag (1999) study a model of confirmatory bias where agents ignore signals that do not conform with their first impressions, and thus updating is simply assumed to be biased in the direction of current beliefs, directly generating polarization. Notice that in all of these papers, individuals may disagree about the likelihood of a publicly observed signal conditional on the parameter. This is likely to be a common, if unsurprising, source of polarization. In contrast, in our model, conditional on the parameter, all individuals agree on the distribution over signals and their independence, and yet an interesting theory of polarization still emerges.

The only other paper we know of relating ambiguity to polarization is Zimper and Ludwig (2009). They study particular forms of dynamically inconsistent updating in a model where agents are Choquet expected utility maximizers (Schmeidler (1989)), and polarization is defined as divergence of “expected” signal probabilities as the number of observations goes to infinity. This contrasts with our model, where updating is optimal in the sense of dynamic consistency, beliefs have the standard additive form and polarization is

defined after any signal realization rather than as a limit phenomenon.

4.2 Concluding Remarks

The arrival of information changes the hedging motive of ambiguity averse individuals. Optimal (i.e., dynamically consistent) updating must counteract this hedging effect in addition to the more familiar likelihood effect. We show that this delivers a theory of polarization – describing when it can occur and when it cannot.

The model and theory can be extended in several ways. First, we have assumed the individual has perfect foresight of the number of observations that will be available before he needs to take an action and that there is only one action required in the problem. Suppose instead that foresight is limited and the individual believes that they must take an action after fewer observations than will, in reality, be available. This is a natural description of the approach plausibly taken by subjects in the experiments of Darley and Gross (1983). Suppose (1) the individual uses dynamically consistent updating in the part of the problem he foresees; and (2) when faced with the unforeseen continuation problem, he applies dynamically consistent updating to the continuation starting from the posterior beliefs inherited from the foreseen problem. Then, the possibility of polarization and the logic behind it described in our analysis continue to hold.

Second, our results are developed using the smooth ambiguity model of Klibanoff, Marinacci and Mukerji (2005). One benefit of using this model is that, like the standard Bayesian expected utility model, it allows us to describe beliefs, and thus polarization, through a probability measure. It is not obvious how to best define polarization for other models of ambiguity averse preferences. Putting that aside, the fundamental connection between ambiguity aversion and dynamically consistent updating that must counteract a hedging effect is present in any complete preference model of ambiguity aversion. Exactly when such a connection generates polarization is likely to vary with choice of model and spelling out these conditions for different models

is left for future research. Such an investigation may require additional tools, as the smooth ambiguity model allows us to characterize the unique solution of our prediction problem using first-order conditions.

Finally, we have assumed fully dynamically consistent updating. As was mentioned in the Introduction, the effects we identify continue to generate polarization even under substantially weaker assumptions. For example, after observing a signal, suppose the individual maximizes a weighted sum of utility under dynamically consistent updating and utility under Bayesian updating.⁷ As long as there is strictly positive weight on the former, all of our qualitative results on polarization under ambiguity aversion are preserved.

References

- Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz.** 2009. “Fragility of Asymptotic Agreement under Bayesian Learning.” <http://economics.mit.edu/files/3795>.
- Andreoni, James, and Timofiy Mylovanov.** 2012. “Diverging Opinions.” *American Economic Journal: Microeconomics*, 4(1): 209–232.
- Baliga, Sandeep, Eran Hanany, and Peter Klibanoff.** 2013. “Polarization and Ambiguity.” Northwestern University, Kellogg CMS-EMS Discussion Paper 1558, <http://www.kellogg.northwestern.edu/research/math/papers/1558.pdf>.
- Darley, John M., and Paget H. Gross.** 1983. “A Hypothesis-Confirming Bias in Labeling Effects.” *Journal of Personality and Social Psychology*, 44(1): 20–33.
- Dixit, Avinash K., and Jörgen W. Weibull.** 2007. “Political Polarization.” *Proceedings of the National Academy of Sciences*, 104(18): 7351–7356.

⁷One foundation for such a model is the temptation and costly self-control representation of Gul and Pesendorfer (2001) with the normative preference generated from dynamically consistent updating and the temptation preference generated from Bayesian updating.

- Gul, Faruk, and Wolfgang Pesendorfer.** 2001. "Temptation and Self Control." *Econometrica*, 69(6): 1403–1435.
- Hanany, Eran, and Peter Klibanoff.** 2007. "Updating Preferences with Multiple Priors." *Theoretical Economics*, 2: 261–298.
- Hanany, Eran, and Peter Klibanoff.** 2009. "Updating Ambiguity Averse Preferences." *B.E. Journal of Theoretical Economics*, 9 (Advances): Article 37.
- Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji.** 2005. "A Smooth Model of Decision Making under Ambiguity." *Econometrica*, 73(6): 1849–1892.
- Kondor, Péter.** 2012. "The More We Know about the Fundamental, the Less We Agree on the Price." *The Review of Economic Studies*, 79(3): 1175–1207.
- Rabin, Matthew, and Joel L. Schrag.** 1999. "First Impressions Matter: A Model of Confirmatory Bias." *The Quarterly Journal of Economics*, 114(1): 37–82.
- Schmeidler, David.** 1989. "Subjective Probability and Expected Utility without Additivity." *Econometrica*, 57(3): 571–587.
- Zimper, Alexander, and Alexander Ludwig.** 2009. "On Attitude Polarization under Bayesian Learning with Non-Additive Beliefs." *Journal of Risk and Uncertainty*, 39(2): 181–212.

A Online Appendix to accompany Baliga, Hanany and Klibanoff, “Polarization and Ambiguity”, *American Economic Review*

This Appendix contains all proofs not included in the main text and some further results on the direction of updating.

A.1 Proofs not in the Main Text

Proof. [Proof of Theorem 2.1] Bayesian updating is only well-defined following positive probability signals. Therefore, assume $\sum_i \check{\eta}(\theta_i) \pi_{\theta_i}(x) > 0$ and $\sum_i \hat{\eta}(\theta_i) \pi_{\theta_i}(x) > 0$. We use proof by contradiction. Suppose two individuals use Bayesian updating and that $\check{\eta}$ stochastically dominates $\check{\nu}$ and $\hat{\nu}$ stochastically dominates $\hat{\eta}$ with at least one dominance strict (i.e., that polarization occurs). Observe that $\check{\eta}$ stochastically dominates $\check{\nu}$ implies $\check{\eta}(\theta_1) \leq \check{\nu}(\theta_1) = \frac{\check{\eta}(\theta_1)\pi_{\theta_1}(x)}{\sum_i \check{\eta}(\theta_i)\pi_{\theta_i}(x)}$ and $\check{\eta}(\theta_{|\Theta|}) \geq \check{\nu}(\theta_{|\Theta|}) = \frac{\check{\eta}(\theta_{|\Theta|})\pi_{\theta_{|\Theta|}}(x)}{\sum_i \check{\eta}(\theta_i)\pi_{\theta_i}(x)}$. Simplifying, this implies

$$\pi_{\theta_1}(x) \geq \sum_i \check{\eta}(\theta_i) \pi_{\theta_i}(x) \geq \pi_{\theta_{|\Theta|}}(x). \quad (\text{A.1})$$

Similarly, observe that $\hat{\nu}$ stochastically dominates $\hat{\eta}$ implies $\hat{\eta}(\theta_1) \geq \hat{\nu}(\theta_1) = \frac{\hat{\eta}(\theta_1)\pi_{\theta_1}(x)}{\sum_i \hat{\eta}(\theta_i)\pi_{\theta_i}(x)}$ and $\hat{\eta}(\theta_{|\Theta|}) \leq \hat{\nu}(\theta_{|\Theta|}) = \frac{\hat{\eta}(\theta_{|\Theta|})\pi_{\theta_{|\Theta|}}(x)}{\sum_i \hat{\eta}(\theta_i)\pi_{\theta_i}(x)}$. Simplifying, this implies

$$\pi_{\theta_1}(x) \leq \sum_i \hat{\eta}(\theta_i) \pi_{\theta_i}(x) \leq \pi_{\theta_{|\Theta|}}(x). \quad (\text{A.2})$$

The only way for (A.1) and (A.2) to be satisfied simultaneously is when

$$\pi_{\theta_1}(x) = \sum_i \check{\eta}(\theta_i) \pi_{\theta_i}(x) = \sum_i \hat{\eta}(\theta_i) \pi_{\theta_i}(x) = \pi_{\theta_{|\Theta|}}(x). \quad (\text{A.3})$$

Notice that under (A.3) $\hat{\eta}(\theta_1) = \hat{\nu}(\theta_1)$, $\hat{\eta}(\theta_{|\Theta|}) = \hat{\nu}(\theta_{|\Theta|})$, $\check{\eta}(\theta_1) = \check{\nu}(\theta_1)$ and $\check{\eta}(\theta_{|\Theta|}) = \check{\nu}(\theta_{|\Theta|})$. Given $\sum_i \check{\eta}(\theta_i) \pi_{\theta_i}(x) = \sum_i \hat{\eta}(\theta_i) \pi_{\theta_i}(x)$, consider the

induction hypothesis that, for some $1 \leq n < |\Theta|$,

$$\hat{\eta}(\theta_i) = \hat{\nu}(\theta_i) \text{ and } \check{\eta}(\theta_i) = \check{\nu}(\theta_i) \text{ for } i = 1, \dots, n.$$

Under this hypothesis, $\check{\eta}$ stochastically dominates $\check{\nu}$ implies $\check{\eta}(\theta_{n+1}) \leq \check{\nu}(\theta_{n+1}) = \frac{\check{\eta}(\theta_{n+1})\pi_{\theta_{n+1}}(x)}{\sum_i \check{\eta}(\theta_i)\pi_{\theta_i}(x)}$ and $\hat{\nu}$ stochastically dominates $\hat{\eta}$ implies $\hat{\eta}(\theta_{n+1}) \geq \hat{\nu}(\theta_{n+1}) = \frac{\hat{\eta}(\theta_{n+1})\pi_{\theta_{n+1}}(x)}{\sum_i \hat{\eta}(\theta_i)\pi_{\theta_i}(x)}$. Therefore,

$$\hat{\eta}(\theta_{n+1}) = \hat{\nu}(\theta_{n+1}) \text{ and } \check{\eta}(\theta_{n+1}) = \check{\nu}(\theta_{n+1}).$$

Since we showed above that the induction hypothesis holds for $n = 1$, we conclude that $\check{\eta}$ stochastically dominates $\check{\nu}$ and $\hat{\nu}$ stochastically dominates $\hat{\eta}$ implies $\check{\eta} = \check{\nu}$ and $\hat{\eta} = \hat{\nu}$. This contradicts our supposition of polarization. \square

Proof. [Proof of Proposition 3.1] It is immediate from (3.1) that $\alpha^*(x) \in (0, 1)$ since $\mu \in (0, 1)$ and $\phi' > 0$. To prove (i), fix any $x \in \mathcal{X}$ and, from (3.2), observe that for any $y \in \mathcal{X}$, $\alpha^*(y)$ is a strictly increasing function of $\alpha^*(x)$ in any solution of the system of first-order conditions. This and the fact that ϕ is concave implies that the left-hand side of (3.1) is strictly increasing in $\alpha^*(x)$ and decreasing in $\frac{\pi_1(x)}{\pi_0(x)}$. The right-hand side of (3.1) is strictly increasing in μ and constant in $\alpha^*(x)$. Therefore, $\alpha^*(x)$ is well-defined and strictly increasing in μ and $\frac{\pi_1(x)}{\pi_0(x)}$. The first-order condition describing the best constant prediction, which we denote here by $\bar{\alpha}$, is

$$\frac{\bar{\alpha}}{1 - \bar{\alpha}} \frac{\phi'[-(\bar{\alpha})^2]}{\phi'[-(1 - \bar{\alpha})^2]} = \frac{\mu}{1 - \mu}. \quad (\text{A.4})$$

Again, concavity of ϕ implies that the left-hand side is strictly increasing in $\bar{\alpha}$ and thus the best constant prediction is strictly increasing in μ .

To prove (ii), let $\beta^*(x)$ denote the optimal prediction after observing x . By the first-order conditions for optimality, these predictions must satisfy

$$\frac{\beta^*(x)}{1 - \beta^*(x)} \frac{\phi'[-(\beta^*(x))^2]}{\phi'[-(1 - \beta^*(x))^2]} = \frac{\nu_x}{1 - \nu_x}. \quad (\text{A.5})$$

Therefore, by the same reasoning as in (i), $\beta^*(x)$ is strictly increasing in the posterior ν_x . Comparing (A.4) and (A.5) and using concavity of ϕ yields

$$\nu_x \underset{\leq}{\overset{\geq}{\approx}} \mu$$

if and only if

$$\beta^*(x) \underset{\leq}{\overset{\geq}{\approx}} \bar{\alpha}.$$

Finally, under dynamically consistent updating, from (3.3), the posteriors must satisfy

$$\frac{\alpha^*(x)}{1 - \alpha^*(x)} \frac{\phi'[-(\alpha^*(x))^2]}{\phi'[-(1 - \alpha^*(x))^2]} = \frac{\nu_x}{1 - \nu_x}.$$

Therefore, $\beta^*(x) = \alpha^*(x)$. □

Proof. [Proof of Proposition 3.2] Dynamically consistent updating implies that (3.3) is satisfied in addition to (3.1). Combining the two equalities yields,

$$\nu_x \underset{\leq}{\overset{\geq}{\approx}} \mu$$

if and only if

$$\frac{\phi'[-(\alpha^*(x))^2]}{\phi'[-(1 - \alpha^*(x))^2]} \underset{\leq}{\overset{\geq}{\approx}} \frac{\phi'[E_{\pi_0}(-(\alpha^*(X))^2)]}{\phi'[E_{\pi_1}(-(1 - \alpha^*(X))^2)]} \frac{\pi_0(x)}{\pi_1(x)}.$$

□

Proof. [Proof of Theorem 3.1] By Proposition 3.2,

$$\nu_{x^H} \geq \mu$$

if and only if

$$\frac{\phi'[-(\alpha^*(x^H))^2]}{\phi'[-(1 - \alpha^*(x^H))^2]} \geq \frac{\phi'[E_{\pi_0}(-(\alpha^*(X))^2)]}{\phi'[E_{\pi_1}(-(1 - \alpha^*(X))^2)]} \frac{\pi_0(x^H)}{\pi_1(x^H)}. \quad (\text{A.6})$$

For all $y \in \mathcal{X}$, since $\frac{\pi_1(x^H)}{\pi_0(x^H)} \geq \frac{\pi_1(y)}{\pi_0(y)}$, it follows from part (i) of Proposition 3.1 that

$$\alpha^*(x^H) \geq \alpha^*(y).$$

Therefore $(\alpha^*(x^H))^2 \geq E_{\pi_0}(\alpha^*(X))^2$ and $(1 - \alpha^*(x^H))^2 \leq E_{\pi_1}(1 - \alpha^*(X))^2$. As ϕ is concave, this implies

$$\frac{\phi'[-(\alpha^*(x^H))^2]}{\phi'[-(1 - \alpha^*(x^H))^2]} \geq \frac{\phi'[E_{\pi_0}(-(\alpha^*(X))^2)]}{\phi'[E_{\pi_1}(-(1 - \alpha^*(X))^2)]}. \quad (\text{A.7})$$

Since $\frac{\pi_1(x^H)}{\pi_0(x^H)} \geq 1$, (A.6) follows. Thus $\nu_{x^H} \geq \mu$.

Furthermore, (3.3), (A.7) and (3.1) imply

$$\begin{aligned} \frac{\nu_{x^H}}{1 - \nu_{x^H}} &= \frac{\alpha^*(x^H)}{1 - \alpha^*(x^H)} \frac{\phi'[-(\alpha^*(x^H))^2]}{\phi'[-(1 - \alpha^*(x^H))^2]} \\ &\geq \frac{\alpha^*(x^H)}{1 - \alpha^*(x^H)} \frac{\phi'[E_{\pi_0}(-(\alpha^*(X))^2)]}{\phi'[E_{\pi_1}(-(1 - \alpha^*(X))^2)]} \\ &= \frac{\mu}{1 - \mu} \frac{\pi_1(x^H)}{\pi_0(x^H)}, \end{aligned}$$

where the last expression is the posterior ratio generated by Bayesian updating of μ after observing x^H .

An analogous argument shows $\mu \geq \nu_{x^L}$ and

$$\frac{\nu_{x^L}}{1 - \nu_{x^L}} \leq \frac{\mu}{1 - \mu} \frac{\pi_1(x^L)}{\pi_0(x^L)}.$$

□

Proof. [Proof of Theorem 3.2] Recall that the optimal prediction $\alpha^*(x)$ is continuous and increasing in the prior probability of $\theta = 1$. Denote this probability by η . As the optimal prediction is 0 if $\eta = 0$ and 1 if $\eta = 1$, considering η close enough to 0 or η close enough to 1 is equivalent to considering $\alpha^*(x)$ close enough to 0 or 1 respectively. The proof strategy for determining updating for sufficiently extreme beliefs will be to consider updating for sufficiently extreme predictions.

Observe, by applying (3.1) and (3.3), that dynamically consistent updating of η after seeing $x \in \mathcal{X}$ will be shaded upward/equal to/shaded downward compared to Bayesian updating if and only if

$$\begin{aligned} & \phi'[-(\alpha^*(x))^2] \phi'[-\sum_{y \in \mathcal{X}} \pi_1(y)(1 - \alpha^*(y))^2] \\ & \geq \phi'[-(1 - \alpha^*(x))^2] \phi'[-\sum_{y \in \mathcal{X}} \pi_0(y)(\alpha^*(y))^2]. \end{aligned} \quad (\text{A.8})$$

From (3.2), $\alpha^*(y) = \beta_{\pi_1, \pi_0}(\alpha^*(x); y)$ where $\beta_{\pi_1, \pi_0} : [0, 1] \times \mathcal{X} \rightarrow [0, 1]$ is defined by $\beta_{\pi_1, \pi_0}(z; y) = \frac{z \frac{\pi_1(y)}{\pi_0(y)}}{z \frac{\pi_1(y)}{\pi_0(y)} + (1-z) \frac{\pi_1(x)}{\pi_0(x)}}$ for all $z \in [0, 1]$ and $y \in \mathcal{X}$. Define the function $f : [0, 1] \rightarrow \mathbb{R}$ such that

$$f(z) = \frac{\phi'[-\sum_{y \in \mathcal{X}} \pi_1(y)(1 - \beta_{\pi_1, \pi_0}(z; y))^2]}{\phi'[-(1 - z)^2]} - \frac{\phi'[-\sum_{y \in \mathcal{X}} \pi_0(y)(\beta_{\pi_1, \pi_0}(z; y))^2]}{\phi'(-z^2)}.$$

Under our assumptions, f is continuous and differentiable. By comparing f with (A.8), observe that when $z = \alpha^*(x) \in (0, 1)$, the direction in which updating is shaded relative to Bayesian updating is determined by the sign of f . Therefore we want to determine the sign of $f(z)$ when z is close 0 and when it is close to 1. By the assumptions in the statement of the theorem, $0 < \phi'(0) < \phi'(-1) < \infty$ where the last inequality comes from the fact that ϕ' is continuous on $[-1, 0]$ and thus bounded. Then $f(0) = f(1) = 0$. Therefore, the sign of $f(z)$ when z is close 0 and when it is close to 1 is determined by the sign of $f'(z)$ at 0 and 1 respectively. Differentiating f (and denoting the

derivative of β_{π_1, π_0} with respect to z evaluated at $(z; y)$ by $\beta'_{\pi_1, \pi_0}(z; y)$ yields,

$$\begin{aligned}
f'(z) &= \frac{2\phi''[-\sum_{y \in \mathcal{X}} \pi_1(y)(1 - \beta_{\pi_1, \pi_0}(z; y))^2] \sum_{y \in \mathcal{X}} \pi_1(y)(1 - \beta_{\pi_1, \pi_0}(z; y))\beta'_{\pi_1, \pi_0}(z; y)}{\phi'[-(1 - z)^2]} \\
&\quad - \frac{2\phi'[-\sum_{y \in \mathcal{X}} \pi_1(y)(1 - \beta_{\pi_1, \pi_0}(z; y))^2]\phi''[-(1 - z)^2](1 - z)}{(\phi'[-(1 - z)^2])^2} \\
&\quad + \frac{2\phi''[-\sum_{y \in \mathcal{X}} \pi_0(y)(\beta_{\pi_1, \pi_0}(z; y))^2] \sum_{y \in \mathcal{X}} \pi_0(y)(\beta_{\pi_1, \pi_0}(z; y))\beta'_{\pi_1, \pi_0}(z; y)}{\phi'(-z^2)} \\
&\quad - \frac{2\phi''(-z^2)(z)\phi'[-\sum_{y \in \mathcal{X}} \pi_0(y)(\beta_{\pi_1, \pi_0}(z; y))^2]}{(\phi'(-z^2))^2}.
\end{aligned}$$

Thus,

$$f'(0) = 2 \left(-\frac{\phi''(-1)}{\phi'(-1)} \right) \left[1 - \sum_{y \in \mathcal{X}} \pi_1(y)\beta'_{\pi_1, \pi_0}(0; y) \right] + (0) \left(-\frac{\phi''(0)}{\phi'(0)} \right) \left[1 - \sum_{y \in \mathcal{X}} \pi_0(y)\beta'_{\pi_1, \pi_0}(0; y) \right]$$

and

$$f'(1) = (0) \left(-\frac{\phi''(0)}{\phi'(0)} \right) \left[1 - \sum_{y \in \mathcal{X}} \pi_1(y)\beta'_{\pi_1, \pi_0}(1; y) \right] + 2 \left(-\frac{\phi''(-1)}{\phi'(-1)} \right) \left[1 - \sum_{y \in \mathcal{X}} \pi_0(y)\beta'_{\pi_1, \pi_0}(1; y) \right].$$

Since ϕ'' is negative and finite (since ϕ'' is continuous on a bounded interval), the coefficient of ambiguity aversion, $-\frac{\phi''}{\phi'}$, is everywhere positive and finite. This allows us to conclude that the sign of $f'(0)$ is the same as the sign of $1 - \sum_{y \in \mathcal{X}} \pi_1(y)\beta'_{\pi_1, \pi_0}(0; y)$, while the sign of $f'(1)$ is the sign of $1 - \sum_{y \in \mathcal{X}} \pi_0(y)\beta'_{\pi_1, \pi_0}(1; y)$. Differentiating $\beta_{\pi_1, \pi_0}(z; y)$ shows that $\beta'_{\pi_1, \pi_0}(0; y) = \frac{\pi_1(y)/\pi_1(x)}{\pi_0(y)/\pi_0(x)}$ and $\beta'_{\pi_1, \pi_0}(1; y) = \frac{\pi_1(x)/\pi_1(y)}{\pi_0(x)/\pi_0(y)}$. Thus $f'(0) < 0$ and $f'(1) < 0$ if and only if

$$\frac{1}{\sum_{y \in \mathcal{X}} \pi_0(y) \frac{\pi_0(y)}{\pi_1(y)}} < \frac{\pi_1(x)}{\pi_0(x)} < \sum_{y \in \mathcal{X}} \pi_1(y) \frac{\pi_1(y)}{\pi_0(y)}. \quad (\text{A.9})$$

Summarizing, we have shown that f is negative for values sufficiently close to 0 and positive for values sufficiently close to 1 if and only if (A.9) is satisfied. Therefore, it is exactly under these conditions that updating will be shaded downward compared to Bayesian updating for beliefs sufficiently close to 0 and

shaded upward compared to Bayesian updating for beliefs sufficiently close to 1.

We now show that a neutral signal necessarily satisfies (A.9). Note that $\sum_{y \in \mathcal{X}} \pi_1(y) \frac{\pi_1(y)}{\pi_0(y)} \geq 1$ and $\sum_{y \in \mathcal{X}} \pi_0(y) \frac{\pi_0(y)}{\pi_1(y)} \geq 1$ because the strictly convex constrained minimization problem $\min_{w_1, \dots, w_{|\mathcal{X}|}} \sum_{i=1}^{|\mathcal{X}|} \frac{w_i^2}{v_i}$ subject to $\sum_{i=1}^{|\mathcal{X}|} w_i = 1$, assuming $\sum_{i=1}^{|\mathcal{X}|} v_i = 1$ and $v_i > 0$ for $i = 1, \dots, |\mathcal{X}|$, has first order conditions equivalent to $\frac{w_i}{v_i}$ constant in i , thus the minimum is achieved at $\frac{1}{\sum_{i=1}^{|\mathcal{X}|} v_i} = 1$ with $w_i = \frac{v_i}{\sum_{i=1}^{|\mathcal{X}|} v_i} = v_i$. Moreover, since there exists at least one informative signal, i.e., $y \in \mathcal{X}$ such that $\frac{\pi_1(y)}{\pi_0(y)} \neq 1$, the unique minimum is not attained and so $\sum_{y \in \mathcal{X}} \pi_1(y) \frac{\pi_1(y)}{\pi_0(y)} > 1$ and $\sum_{y \in \mathcal{X}} \pi_0(y) \frac{\pi_0(y)}{\pi_1(y)} > 1$. Thus, (A.9) is always satisfied if $\frac{\pi_1(x)}{\pi_0(x)} = 1$ (i.e., if x is a neutral signal).

Finally, observe that if x is a neutral signal, then, since Bayesian updating would be flat, updating shaded downward implies updating is downward and updating shaded upward implies updating is upward, generating polarization. \square

Remark A.1. The theorem remains true if $\phi'(0) = 0$ and the requirements of the theorem are otherwise satisfied. This case requires an argument based on second-order comparisons. Intuitively, second-order differences that were previously masked may now become important in the limit because the zero creates unboundedly large ambiguity aversion (as measured by $-\frac{\phi''}{\phi'}$) near perfect predictions. Specifically, one can show that, for beliefs close to θ , a second-order comparison yields that the payoff following a neutral signal is larger than the expected payoff before seeing the signal. This drives the comparison of ex-ante versus interim hedging effects and generates the polarization. Moreover, in this case, the polarization result may be extended beyond neutral signals to all signals having a likelihood ratio lying in an interval containing 1.

Proof. [Proof of Proposition 3.3] From Lemma A.1, $\nu_{x^M} \gtrless \mu$ if and only if

$$\sum_{y \in \mathcal{X}} \pi_1(y) \frac{\left(\frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^{\frac{1}{\gamma}+2} - \frac{\pi_1(y)}{\pi_0(y)}}{[\alpha^*(x^M) \frac{\pi_1(y)}{\pi_0(y)} + (1 - \alpha^*(x^M)) \frac{\pi_1(x^M)}{\pi_0(x^M)}]^2} \gtrless 0. \quad (\text{A.10})$$

We consider the following exhaustive list of possibilities:

(i) $\left(\frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^{\frac{1}{\gamma}+2} \geq \frac{\pi_1(x^H)}{\pi_0(x^H)}$. In this case, using $\frac{\pi_1(x^L)}{\pi_0(x^L)} < \frac{\pi_1(x^M)}{\pi_0(x^M)} < \frac{\pi_1(x^H)}{\pi_0(x^H)}$, the left-hand side of (A.10) is strictly positive, and therefore updating is always upward, so set $\tau(\gamma, \pi_0, \pi_1) = 0$. Note that a necessary condition for this case is that $\frac{\pi_1(x^M)}{\pi_0(x^M)} > 1$.

(ii) $\left(\frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^{\frac{1}{\gamma}+2} \leq \frac{\pi_1(x^L)}{\pi_0(x^L)}$. In this case, using $\frac{\pi_1(x^L)}{\pi_0(x^L)} < \frac{\pi_1(x^M)}{\pi_0(x^M)} < \frac{\pi_1(x^H)}{\pi_0(x^H)}$, the left-hand side of (A.10) is strictly negative, and therefore updating is always downward, so set $\tau(\gamma, \pi_0, \pi_1) = 1$. Note that a necessary condition for this case is that $\frac{\pi_1(x^M)}{\pi_0(x^M)} < 1$.

(iii) $\frac{\pi_1(x^H)}{\pi_0(x^H)} > \left(\frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^{\frac{1}{\gamma}+2} > \frac{\pi_1(x^L)}{\pi_0(x^L)}$. In this case, using $\frac{\pi_1(x^L)}{\pi_0(x^L)} < \frac{\pi_1(x^M)}{\pi_0(x^M)} < \frac{\pi_1(x^H)}{\pi_0(x^H)}$, in the left-hand side of (A.10), the term for $y = x^L$ is positive and has a denominator strictly decreasing in $\alpha^*(x^M)$, the term for $y = x^M$ is constant in $\alpha^*(x^M)$, and the term for $y = x^H$ is negative and has a denominator strictly increasing in $\alpha^*(x^M)$. Therefore the whole sum is strictly increasing in $\alpha^*(x^M)$ and thus can change signs at most once. Three sub-cases are relevant:

(iii)(a) the left-hand side of (A.10) is non-negative when 0 is plugged in for $\alpha^*(x^M)$. In this case, updating is always upward, so set $\tau(\gamma, \pi_0, \pi_1) = 0$.

(iii)(b) the left-hand side of (A.10) is non-positive when 1 is plugged in for $\alpha^*(x^M)$. In this case, updating is always downward, so set $\tau(\gamma, \pi_0, \pi_1) = 1$.

(iii)(c) otherwise. In this case, continuity and strict increasingness of the left-hand side of (A.10) in $\alpha^*(x^M)$ implies there exists a unique solution for a in $(0, 1)$ to

$$\sum_{y \in \mathcal{X}} \pi_1(y) \frac{\left(\frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^{\frac{1}{\gamma}+2} - \frac{\pi_1(y)}{\pi_0(y)}}{\left(a \frac{\pi_1(y)}{\pi_0(y)} + (1-a) \frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^2} = 0.$$

Since (A.11) holds with equality when $z = a$, using constant relative ambiguity aversion ($\phi'(z) = (-z)^\gamma$) and given the monotonicity of $\alpha^*(x^M)$ in μ , the associated threshold for μ may be found by substituting $z = a$ into (A.11) with equality and solving for $\mu = \tau(\gamma, \pi_0, \pi_1)$. Doing this yields

$$\frac{\tau(\gamma, \pi_0, \pi_1)}{1 - \tau(\gamma, \pi_0, \pi_1)} = \left(\frac{a}{1-a}\right)^{2\gamma+1}.$$

Therefore

$$\tau(\gamma, \pi_0, \pi_1) = \frac{a^{2\gamma+1}}{a^{2\gamma+1} + (1-a)^{2\gamma+1}}.$$

Collecting these results into an overall expression, the threshold is defined by:

$$\tau(\gamma, \pi_0, \pi_1) = \frac{b^{2\gamma+1}}{b^{2\gamma+1} + (1-b)^{2\gamma+1}},$$

where

$$b \equiv \begin{cases} 0 & \text{if } S(0) \geq 0 \\ a & \text{if } S(a) = 0 \text{ and } a \in (0, 1) \\ 1 & \text{if } S(1) \leq 0 \end{cases}$$

and

$$S(\lambda) \equiv \sum_{y \in \{x^L, x^M, x^H\}} \pi_1(y) \frac{\left(\frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^{\frac{1}{\gamma}+2} - \frac{\pi_1(y)}{\pi_0(y)}}{\left(\lambda \frac{\pi_1(y)}{\pi_0(y)} + (1-\lambda) \frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^2}.$$

□

Proof. [Proof of Theorem 3.3] Polarization is equivalent to $\hat{\nu} \geq \hat{\eta}$ and $\check{\nu} \leq \check{\eta}$ with at least one inequality strict. If $\gamma = 0$, updating is Bayesian and polarization is impossible by Theorem 2.1, so set $\hat{\tau} = 1$ and $\check{\tau} = 0$. By Proposition 3.3, if $\gamma > 0$ then polarization occurs if and only if $\hat{\eta} \geq \tau(\hat{\gamma}, \pi_0, \pi_1)$ and $\check{\eta} \leq \tau(\check{\gamma}, \pi_0, \pi_1)$ with at least one inequality strict, where the τ function is the one defined in that result. □

Proof. [Proof of Corollary 3.1] From Proposition 3.3, $\hat{\tau} = \tau(\hat{\gamma}, \pi_0, \pi_1)$ and $\check{\tau} = \tau(\check{\gamma}, \pi_0, \pi_1)$. The rest is immediate from Theorem 3.3. □

Proof. [Proof of Corollary 3.2] From Proposition 3.3, such a threshold exists. Since $\pi_0(x^M) = \pi_1(x^M)$ implies $\pi_0(x^L) - \pi_1(x^L) = \pi_1(x^H) - \pi_0(x^H) > 0$, calculation shows that the relevant case in the proof of Proposition 3.3 is case (iii)(c). Thus $\tau(\gamma, \pi_0, \pi_1) = \frac{a^{2\gamma+1}}{a^{2\gamma+1} + (1-a)^{2\gamma+1}} = \frac{1}{1 + \left(\frac{1-a}{a}\right)^{2\gamma+1}}$ where $a \in (0, 1)$ is

the unique solution of $S(a) = 0$. Simplifying yields

$$\frac{1-a}{a} = \sqrt{\frac{\pi_1(x^H) \pi_1(x^L)}{\pi_0(x^H) \pi_0(x^L)}}.$$

□

A.2 Further Results on the Direction of Updating

The next result combines Proposition 3.2 and equations (3.2) and (3.3) to show a general form relating fundamentals to the direction of updating.

Proposition A.1. *The posterior ν_x is above/equal to/below the prior μ if and only if the fundamentals $(\mu, \phi, \pi_1, \pi_0)$ are such that*

$$\frac{z}{1-z} \frac{\phi'[-z^2]}{\phi'[-(1-z)^2]} \begin{matrix} \geq \\ < \end{matrix} \frac{\mu}{1-\mu}, \quad (\text{A.11})$$

for the unique $z \in (0, 1)$ solving

$$\begin{aligned} & \frac{z}{1-z} \frac{\phi' \left[-z^2 \sum_{y \in \mathcal{X}} \frac{\pi_0(y) \left(\frac{\pi_1(y)}{\pi_0(y)} \right)^2}{\left(z \frac{\pi_1(y)}{\pi_0(y)} + (1-z) \frac{\pi_1(x)}{\pi_0(x)} \right)^2} \right]}{\phi' \left[-(1-z)^2 \sum_{y \in \mathcal{X}} \frac{\pi_1(y) \left(\frac{\pi_1(x)}{\pi_0(x)} \right)^2}{\left(z \frac{\pi_1(y)}{\pi_0(y)} + (1-z) \frac{\pi_1(x)}{\pi_0(x)} \right)^2} \right]} \\ &= \frac{\mu}{1-\mu} \frac{\pi_1(x)}{\pi_0(x)}. \end{aligned} \quad (\text{A.12})$$

Proof. Substituting (3.3) into (3.4) and rearranging yields

$$\frac{\alpha^*(x)}{1-\alpha^*(x)} \frac{\phi'[-\alpha^*(x)^2]}{\phi'[-(1-\alpha^*(x))^2]} \begin{matrix} \geq \\ < \end{matrix} \frac{\mu}{1-\mu}.$$

From (3.2), we obtain for all $y \in \mathcal{X}$,

$$\alpha^*(y) = \frac{\alpha^*(x) \frac{\pi_1(y)}{\pi_0(y)}}{\alpha^*(x) \frac{\pi_1(y)}{\pi_0(y)} + (1-\alpha^*(x)) \frac{\pi_1(x)}{\pi_0(x)}}.$$

Using this together with (3.3), $\alpha^*(x)$ is the unique solution to

$$\begin{aligned} & \frac{\alpha^*(x)}{1 - \alpha^*(x)} \frac{\phi' \left[-\alpha^*(x)^2 \sum_{y \in \mathcal{X}} \frac{\pi_0(y) \left(\frac{\pi_1(y)}{\pi_0(y)} \right)^2}{\left(\alpha^*(x) \frac{\pi_1(y)}{\pi_0(y)} + (1 - \alpha^*(x)) \frac{\pi_1(x)}{\pi_0(x)} \right)^2} \right]}{\phi' \left[-(1 - \alpha^*(x))^2 \sum_{y \in \mathcal{X}} \frac{\pi_1(y) \left(\frac{\pi_1(x)}{\pi_0(x)} \right)^2}{\left(\alpha^*(x) \frac{\pi_1(y)}{\pi_0(y)} + (1 - \alpha^*(x)) \frac{\pi_1(x)}{\pi_0(x)} \right)^2} \right]} \\ &= \frac{\mu}{1 - \mu} \frac{\pi_1(x)}{\pi_0(x)}. \end{aligned}$$

□

In interpreting inequality (A.11), it is important to realize that z is an increasing function of beliefs μ (as follows from the argument used in proving part (i) of Proposition 3.1 with z playing the role of $\alpha^*(x)$). In fact, (A.12) combines (3.2) and (3.3). This implies that $z = \alpha^*(x)$, the optimal prediction given the observation x . From (A.12), in the case of ambiguity neutrality (ϕ affine) $\frac{z}{1-z}$ is simply a multiple of $\frac{\mu}{1-\mu}$ so that updating is either always upward (if $\frac{\pi_1(x)}{\pi_0(x)} \geq 1$) or always downward (if $\frac{\pi_1(x)}{\pi_0(x)} \leq 1$). Similarly, we see that under ambiguity aversion, $\frac{z}{1-z}$ is generally a non-linear function of $\frac{\mu}{1-\mu}$ (reflecting the balancing of the desire to hedge with the likelihood based motivation from the ambiguity neutral case) which creates the possibility that inequality (A.11) may change direction as beliefs μ change. In general, the regions where it goes one way and where it goes the other may be very complex. We now offer a characterization of when updating follows a threshold rule so that A.11 changes direction at most once.

Proposition A.2. *There is a threshold rule for updating μ after observing x*

if and only if

$$\frac{\phi'[-z^2] \frac{\pi_1(x)}{\pi_0(x)}}{\phi'[-(1-z)^2] \frac{\pi_1(x)}{\pi_0(x)}} \quad (\text{A.13})$$

$$= \frac{\phi' \left[-z^2 \sum_{y \in \mathcal{X}} \frac{\pi_0(y) \left(\frac{\pi_1(y)}{\pi_0(y)} \right)^2}{\left(z \frac{\pi_1(y)}{\pi_0(y)} + (1-z) \frac{\pi_1(x)}{\pi_0(x)} \right)^2} \right]}{\phi' \left[-(1-z)^2 \sum_{y \in \mathcal{X}} \frac{\pi_1(y) \left(\frac{\pi_1(x)}{\pi_0(x)} \right)^2}{\left(z \frac{\pi_1(y)}{\pi_0(y)} + (1-z) \frac{\pi_1(x)}{\pi_0(x)} \right)^2} \right]}$$

as a function of z has at most one zero in $(0, 1)$ and, if a zero exists, (A.13) is increasing at that zero.

Proof. The result follows by combining the definition of a threshold updating rule with the characterization of the direction of updating given by Proposition A.1. \square

Finally, we present a lemma showing how inequality (3.4), which identifies the direction of updating after observing a signal, simplifies under the assumption of constant relative ambiguity aversion. In proving Theorem 3.3, we use this inequality to help establish and calculate the threshold rule.

Lemma A.1. *With constant relative ambiguity aversion $\gamma > 0$, the posterior ν_x is above/equal to/below the prior μ if and only if*

$$\sum_{y \in \mathcal{X}} \pi_1(y) \frac{\left(\frac{\pi_1(x)}{\pi_0(x)} \right)^{\frac{1}{\gamma} + 2} - \frac{\pi_1(y)}{\pi_0(y)}}{\left(\alpha^*(x) \frac{\pi_1(y)}{\pi_0(y)} + (1 - \alpha^*(x)) \frac{\pi_1(x)}{\pi_0(x)} \right)^2} \geq 0. \quad (\text{A.14})$$

Proof. From inequality (A.11) and equation (A.12), $\nu_x \geq \mu$ if and only if

$$\frac{\phi'[-(\alpha^*(x))^2] \frac{\pi_1(x)}{\pi_0(x)}}{\phi'[-(1 - \alpha^*(x))^2] \frac{\pi_1(x)}{\pi_0(x)}} \geq \frac{\phi' \left[-\alpha^*(x)^2 \sum_{y \in \mathcal{X}} \frac{\pi_0(y) \left(\frac{\pi_1(y)}{\pi_0(y)} \right)^2}{\left(\alpha^*(x) \frac{\pi_1(y)}{\pi_0(y)} + (1 - \alpha^*(x)) \frac{\pi_1(x)}{\pi_0(x)} \right)^2} \right]}{\phi' \left[-(1 - \alpha^*(x))^2 \sum_{y \in \mathcal{X}} \frac{\pi_1(y) \left(\frac{\pi_1(x)}{\pi_0(x)} \right)^2}{\left(\alpha^*(x) \frac{\pi_1(y)}{\pi_0(y)} + (1 - \alpha^*(x)) \frac{\pi_1(x)}{\pi_0(x)} \right)^2} \right]}. \quad (\text{A.15})$$

Under constant relative ambiguity aversion, $\phi'(z) = (-z)^\gamma$ and therefore (A.15) is equivalent to

$$\left(\frac{\pi_1(x)}{\pi_0(x)}\right)^{\frac{1}{\gamma}} \geq \frac{\sum_{y \in \mathcal{X}} \frac{\pi_0(y) \left(\frac{\pi_1(y)}{\pi_0(y)}\right)^2}{\left(\alpha^*(x) \frac{\pi_1(y)}{\pi_0(y)} + (1-\alpha^*(x)) \frac{\pi_1(x)}{\pi_0(x)}\right)^2}}{\sum_{y \in \mathcal{X}} \frac{\pi_1(y) \left(\frac{\pi_1(x)}{\pi_0(x)}\right)^2}{\left(\alpha^*(x) \frac{\pi_1(y)}{\pi_0(y)} + (1-\alpha^*(x)) \frac{\pi_1(x)}{\pi_0(x)}\right)^2}}.$$

Simplifying yields inequality (A.14). □