# Polarization and Ambiguity*

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

November 2011

**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. We show a sense in which polarization is impossible under Bayesian updating or after observing extreme signals. However, we also show that polarization after intermediate signals can arise from the efforts of ambiguity averse individuals to implement their optimal prediction strategies. We explore when polarization of this kind will occur and the logic underlying it.

*Some previous versions of this paper had the title "Ambiguity, Dynamic Consistency and Behavioral Phenomena." We thank Jeff Ely, Itzhak Gilboa, Anna Gumen, Christian Kellner, Sujoy Mukerji, Marco Ottaviani, Emre Ozdenoren, Wolfgang Pesendorfer, Ben Polak, Jean-Marc Tallon, audiences at Oxford, Northwestern, the University of British Columbia, the University of Texas, Austin and several conferences for helpful discussion. 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.

[†]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 really 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 priors. 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.

These examples are natural, and similar *polarization* of beliefs has been documented in experiments. In a seminal contribution, Lord, Ross and Lepper [1979] show that attitudes toward capital punishment polarize: Initial proponents of the death penalty support it even more strongly after observing additional evidence for and against the deterrent effect of capital punishment. But initial opponents of capital punishment support it even less after observing exactly the same evidence. Batson [1975] finds polarization of religious beliefs in believers and non-believers after subjects receive evidence purporting to show that the New Testament is fraudulent.

In these studies, there is ex ante heterogeneity in the subject pool. Polarization also occurs in an ex ante homogeneous subject pool where differences in initial beliefs are generated by randomly assigning individuals into treatment groups that are shown different preliminary evidence. When subjects are later shown common additional evidence at a second stage, their beliefs polarize. For example, Darley and Gross [1983] randomized subjects into different groups to ensure ex ante homogeneity. One group was shown evidence suggesting a child was from a high socioeconomic background; another that she was from a low socioeconomic background. The former predicted the child's reading abilities to be higher than the latter. The groups were then shown a film of the child taking an oral test on which she answers some questions correctly and others incorrectly. Those who had received the preliminary information that the child came from a high socioeconomic background, rated her abilities higher than before; those who had received the information indicating she came from a low socioeconomic background rated her lower than before. Thus, the common evidence in the form of the film led beliefs to polarize.

Polarization is incompatible with standard economic agents who understand the signal structure they are facing and use Bayes' rule to update their beliefs.[1] We offer a theory of polarization as an optimal response to ambiguity, i.e., subjective uncertainty about probabilities. An individual is exposed to ambiguity when the expected payoff to his strategy varies with the probabilities over which he is uncertain. An ambiguity averse individual dislikes this variation. Ambiguity aversion is our only departure from the standard modeling. In particular, agreement on the signal structure and the consistent updating of beliefs are maintained.

Why focus on ambiguity and ambiguity aversion? It seems to us that polarization is fundamentally related to uncertainty – if there is no uncertainty, it is hard to imagine much disagreement, let alone polarization. A large and growing branch of the modern literature on decisions under uncertainty has focused on ambiguity and ambiguity attitude as some of the more interesting aspects to model (both from descriptive and normative points of view).[2] At a very loose level, intuition suggests that situations where information is lacking might be good candidates for both ambiguity and polarization to be relevant. In fact, as we show in this paper, there is a formal connection and a sense in which ambiguity and ambiguity aversion can lead to polarization. Moreover, this theory leads to novel and appealing results, for example that polarization can be triggered by common observation of intermediate signals, while never being triggered by the most informative signals. This seems broadly consistent with the polarization experiments mentioned earlier, in that the polarizing observations there have elements favoring each direction.

How does ambiguity aversion affect behavior? Different strategies may

[1] As we discuss later in the introduction, one of our contributions is to prove this claim (Theorem 2.1).

[2] For a survey including experimental/descriptive work see e.g., Camerer and Weber [1992] and for a more normatively oriented survey see e.g., Gilboa and Marinacci [2011].

involve different exposure to ambiguity. This exposure is taken into consideration by an ambiguity averse individual in choosing an optimal strategy. For example, suppose an individual is subjectively uncertain about the probability that it will rain tomorrow. He thinks that with probability $\frac{1}{2}$ there is a 40% chance of rain, while with the remaining probability there is a 60% chance of rain. He considers two possible strategies: carrying an umbrella to work or leaving it at home. Utility payoffs are as follows:

| | Rain | No rain |
|---|---|---|
| Umbrella | -1 | -1 |
| No umbrella | -2 | 0 |

The expected utility payoffs under a 40% and a 60% chance of rain are therefore:

| | 40% | 60% |
|---|---|---|
| Umbrella | -1 | -1 |
| No umbrella | -0.8 | -1.2 |

Notice that taking the umbrella to work completely hedges against the ambiguity – the expected utility is the same whether the chance of rain is 40% or 60%. However, not taking the umbrella results in greater exposure to the ambiguity – the payoff is better under a 40% chance than under a 60% chance of rain. An ambiguity neutral individual (e.g., an expected utility maximizer) assigning probability $\frac{1}{2}$ to each scenario is indifferent between the two strategies. However, any degree of ambiguity aversion leads the individual to strictly prefer taking the umbrella. In general, *ceteris paribus*, ambiguity aversion disadvantages strategies that result in more exposure to ambiguity.

Polarization fundamentally concerns reaction to new information. Whenever new information is anticipated, optimal strategies will condition on this information. Continuing the above example, if the individual can look out the window before leaving the house, contingent strategies such as "take the umbrella if it looks cloudy and leave it at home otherwise" become possible. 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*.[3]

Why do we focus on dynamic consistency? First, normatively there is a strong case for dynamic consistency. Specifically, any theory of updating that is not dynamically consistent will lead to worse outcomes as evaluated by ex-ante welfare. Second, studying the dynamically consistent case identifies the key effects leading to polarization that apply even when substantial dynamic inconsistency may be present. For example, consider a multi-selves model, where the later selves differ in updated beliefs from the earlier selves. Suppose, however, as in the style of Gul and Pesendorfer [2001], the early self can (possibly at some cost) exert self-control on the decisions faced by the later selves. As long as the cost of self-control is not infinite, the effects we identify in the dynamically consistent theory will survive and will continue to drive polarization.

How does dynamic consistency relate to reaction to new information? When an individual observes new information, he updates his subjective beliefs to incorporate the information. Thus, the manner in which subjective beliefs are updated is crucial in determining whether the individual is dynamically consistent. For an ambiguity neutral individual, dynamic consistency requires that the individual update subjective beliefs using Bayes' rule. In this sense, dynamic consistency is *the* justification for Bayesian updating of subjective beliefs under expected utility. For an ambiguity averse individual, dynamic consistency also delivers a method of updating subjective beliefs, but this generally differs from Bayes' rule. In this sense, non-Bayesian updating is optimal for an ambiguity averse individual.

Returning to our earlier example, suppose an individual concludes that "take the umbrella if it looks cloudy and leave it at home otherwise" is his optimal strategy. This strategy leaves him partially exposed to the ambiguity about rain – his expected utility depends on the chance of rain, but not as

[3] See e.g., Hanany and Klibanoff [2009] for such an approach to modeling ambiguity averse individuals. Modeling dynamic preferences under ambiguity is not straightforward, and the literature has pursued several approaches. For an approach relaxing dynamic consistency see e.g., Siniscalchi [2011].

much as it would if he never carried an umbrella. Notice, however, that this exposure is not the same across the different possible views he may see in the morning. He is less exposed to the ambiguity about rain if it is cloudy in the morning than if it is sunny: if he takes the umbrella his expected utility will not be impacted by the probability of rain, whereas if he has no umbrella his expected utility is much lower if the probability of rain is high than if it is low. This variation in exposure is irrelevant if the individual is ambiguity neutral, but not if he is ambiguity averse. Under ambiguity aversion, the greater exposure to ambiguity after seeing it is sunny may lead to an increased desire to hedge against this ambiguity, while the lack of exposure after seeing it is cloudy 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 – only likelihoods of future signals matter at that point. Because of this change in which likelihoods are relevant, *ceteris paribus*, after seeing the signal, the individual might want to depart from what the ex-ante optimal strategy prescribes after that signal. We call this the *likelihood effect*. 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. We study when the resulting updating leads to polarization of beliefs.

We use a simple prediction model as a vehicle to study these issues: An individual must predict a parameter that determines the distribution of a random variable. The individual has a prior distribution over the possible parameters and is ambiguity averse. He observes a number of conditionally independent signals that can inform his predictions. The individual's payoff is decreasing in the quadratic difference between his prediction and the parameter. This is a standard model apart from ambiguity aversion.

We define polarization as follows: Suppose two individuals have different priors and individual $A$'s belief first-order stochastically dominates individual $B$'s. The individuals observe a common signal. Their beliefs exhibit *polarization* if individual $A$'s posterior dominates his prior and individual $B$'s prior dominates his posterior.

Polarization is hard to reconcile with standard models of Bayesian updating. We establish that there is a sense in which reconciliation is impossible. Consider two individuals who are Bayesian and agree on the probability of each signal conditional on the parameter. We show polarization cannot occur. As individuals share the same theory connecting parameters to signals, if one individual's belief increases after seeing a given signal, so does the other's (Theorem 2.1).

This simple picture changes completely when we allow individuals to be ambiguity averse. As dynamically consistent updating is not Bayesian in this case, our benchmark result does not apply and there is room for polarization. We study when polarization remains impossible and when it occurs.

Suppose there are just two possible parameters, 0 and 1. Signals can then be ordered by the ratio of their likelihood under 1 to their likelihood under 0. As a preliminary result, we show (Proposition 3.1) that our definition of polarization can also be thought of as comparing optimal actions (here, predictions) with and without observing a common signal. Given that no polarization occurs under Bayesian updating, there will continue to be no polarization if the hedging effect (present because of ambiguity aversion) simply reinforces the likelihood effect. We show that this is exactly what happens after observing the highest or the lowest signals. We conclude that polarization cannot occur at extreme signals even under ambiguity aversion (Theorem 3.1). Moreover, the same result shows that, after extreme signals, dynamically consistent updating overshoots the Bayesian update since the hedging effect reinforces the likelihood effect.

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). For a neutral signal, 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).

To gain intuition, suppose the individual believes parameter 1 is highly likely. Then, the optimal strategy will specify relatively high predictions. This leads to small prediction errors and high expected payoff at parameter 1. Marginal changes in prediction errors do not add or subtract much to payoffs when predictions are already very good. Thus the expected payoff given parameter 1 remains almost the same after observing the neutral signal as it was ex-ante. In contrast, at parameter 0, since the strategy is specifying high predictions, prediction errors are large and the marginal effect of changes in these predictions is large as well. Expected payoffs at parameter 0 place more weight on signals below the neutral signal, and hence on predictions closer to 0. This implies that, at parameter 0, the ex-ante expected payoff is higher than the payoff after observing the neutral signal. Thus there is a relative shifting of expected utility away from parameter 0 and this, all else equal, increases the hedging motive to push the prediction toward 0. However, all else is not equal. Specifically, dynamically consistent updating will counterbalance this change in the hedging effect by updating beliefs toward 1. A similar argument establishes that an individual who believes parameter 0 is highly likely will update toward 0 after observing a neutral signal. Thus the observation of a neutral signal results in polarization.

Theorem 3.2 is important mainly because it establishes that the hedging effect can lead to polarization. In general (i.e., for non-neutral intermediate signals and non-extreme beliefs), determining when the combination of the likelihood and hedging effects will lead to polarization is complex. However, we obtain a complete characterization of when polarization occurs under the additional assumptions that individuals display constant relative ambiguity aversion and that there are exactly three possible signals. In this case we establish a threshold result: if belief before seeing a signal is above a threshold, it is updated upwards after the intermediate signal and, if it is below a threshold, it is updated downwards after the intermediate signal (Proposition 3.3). Theorem 3.3 uses this to provide necessary and sufficient conditions for polarization.

All of the above results apply whether or not the two individuals have the same degree of ambiguity aversion. Finally, even if individuals are homogeneous (in all respects) ex ante, if they observe different private signals before observing a common signal, they will generally have different beliefs by the time they see the common signal and, at that point, our results concerning polarization with different beliefs apply (Theorem 3.4).

Before presenting our analysis, we discuss related theory literature. There are a number of approaches to modeling polarization-like phenomena.[4] 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 [2009] shows that polarization can be generated when individuals see different private signals that are correlated with a common public signal. Andreoni and Mylovanov [2010] provide theory along lines similar to Kondor and conduct a related experiment. 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 their model, like those above, can be interpreted as one where individuals sometimes disagree about the likelihood of the observed signal conditional on the parameter. 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.

Dixit and Weibull [2007] propose the same definition of polarization that

[4] We should also mention that there is a literature on political polarization which is not related to our work. For example, Esteban and Ray [2011] study whether conflict can be connected to the level of polarization in society. There is also a literature examining the polarizing or moderating effects of group versus individual decision making. See Adams and Ferreira [2010] for a study of the moderating effects of group decision-making. Eliaz, Ray and Razin [2006] build a theory of the polarizing effects of group decision-making with regard to risky decisions based on individual Allais paradox behavior.

we use. They 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 as defined. In fact, our Theorem 2.1 shows that, quite generally, polarization is impossible, irrespective of whether MLRP holds, under Bayesian updating. Instead, in their examples, while the means or the medians of two individuals move in opposite directions after observing a common signal, their beliefs cannot be ranked by stochastic dominance.

Zimper and Ludwig [2009] study particular forms of dynamically inconsistent updating in a model where agents are ambiguity sensitive and have non-additive beliefs. They examine conditions under which "expected" posterior signal probabilities may diverge in the limit as the number of observations grows where the expectation is taken in the sense of Choquet (see Schmeidler [1989]). In our model, 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.

Wilson [2004] studies a model with bounded memory where an individual chooses which signals to remember in order to economize on finite memory. If two individuals begin with different beliefs, the memory constraint can lead them to optimally allocate memory to different subsets of signals, and thereby possibly polarize.

# 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 about $\theta$ are given by a full-support prior $\mu$. To help inform the individual about $\theta$, conditionally independent observations from a random variable $X$ given $\theta$ may be available. This random variable has distribution $\pi_\theta$ 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, $\theta$ might indicate a child's reading ability, while $\pi_\theta$ might be the distribution of scores on a reading test for a child with that ability.

We assume that $\Theta$ is finite and, without loss of generality, index $\Theta$ so that $\theta_1 < \theta_2 < \ldots < \theta_{|\Theta|}$. A distribution $\hat{\eta}$ (first-order) stochastically dominates $\check{\eta}$ if

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

The dominance is strict if at least one of these inequalities is strict. We define polarization as follows:

**Definition 2.1** *Fix two individuals with beliefs $\check{\eta}$ and $\hat{\eta}$ over $\Theta$ 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.*

The following result shows that under Bayesian updating, irrespective of any non-belief aspect of preference, polarization cannot occur after any signal given positive probability. In fact, we show that if one posterior belief stochastically dominates the prior, so must the other, *without* invoking any dominance between the priors. 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 it is impossible for them to update in opposite directions in the sense of first-order stochastic dominance.[5]

**Theorem 2.1** *Polarization cannot occur if the two individuals use Bayesian updating.*

**Proof.** Bayesian updating is only well-defined following positive probability signals. Therefore, assume $\sum_i \check{\eta}\left(\theta_i\right) \pi_{\theta_i}(x) > 0$ and $\sum_i \hat{\eta}\left(\theta_i\right) \pi_{\theta_i}(x) > 0$. We use proof by contradiction. Suppose two individuals use Bayesian updating and that $\check{\eta}$ stochastically dominates $\check{\nu}$ and $\hat{\nu}$ stochastically dominates $\hat{\eta}$

[5] We thank Eran Shmaya for his help with the proof.

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 $\check{\eta}\left(\theta_{|\Theta|}\right) \geq \check{\nu}\left(\theta_{|\Theta|}\right) = \frac{\check{\eta}\left(\theta_{|\Theta|}\right)\pi_{\theta_{|\Theta|}}(x)}{\sum_i \check{\eta}(\theta_i)\pi_{\theta_i}(x)}$. Simplifying, this implies

$$\pi_{\theta_1}(x) \geq \sum_i \check{\eta}(\theta_i)\pi_{\theta_i}(x) \geq \pi_{\theta_{|\Theta|}}(x). \tag{2.1}$$

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 $\hat{\eta}\left(\theta_{|\Theta|}\right) \leq \hat{\nu}\left(\theta_{|\Theta|}\right) = \frac{\hat{\eta}\left(\theta_{|\Theta|}\right)\pi_{\theta_{|\Theta|}}(x)}{\sum_i \hat{\eta}(\theta_i)\pi_{\theta_i}(x)}$. Simplifying, this implies

$$\pi_{\theta_1}(x) \leq \sum_i \hat{\eta}(\theta_i)\pi_{\theta_i}(x) \leq \pi_{\theta_{|\Theta|}}(x). \tag{2.2}$$

The only way for (2.1) and (2.2) to be satisfied simultaneously is when

$$\pi_{\theta_1}(x) = \sum_i \check{\eta}(\theta_i)\pi_{\theta_i}(x) = \sum_i \hat{\eta}(\theta_i)\pi_{\theta_i}(x) = \pi_{\theta_{|\Theta|}}(x). \tag{2.3}$$

Notice that under (2.3) $\hat{\eta}(\theta_1) = \hat{\nu}(\theta_1)$, $\hat{\eta}\left(\theta_{|\Theta|}\right) = \hat{\nu}\left(\theta_{|\Theta|}\right)$, $\check{\eta}(\theta_1) = \check{\nu}(\theta_1)$ and $\check{\eta}\left(\theta_{|\Theta|}\right) = \check{\nu}\left(\theta_{|\Theta|}\right)$. 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

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

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 $\hat{\eta}(\theta_{n+1}) \geq \hat{\nu}(\theta_{n+1}) = \frac{\hat{\eta}(\theta_{n+1})\pi_{\theta_{n+1}}(x)}{\sum_i \hat{\eta}(\theta_i)\pi_{\theta_i}(x)} = \frac{\hat{\eta}(\theta_{n+1})\pi_{\theta_{n+1}}(x)}{\sum_i \check{\eta}(\theta_i)\pi_{\theta_i}(x)}$. 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. ■

**Remark 2.1** *For convenience alone, we assume the set of observations is fi-*

*nite. Theorem 2.1 can easily be extended to the case where there is a continuum of feasible observations. Also, we conjecture that a similar argument holds, under appropriate regularity conditions, for an infinite set of possible parameters $\Theta$.*

**Remark 2.2** *The common support condition cannot be relaxed to simply overlapping supports. When $|\Theta| \geq 3$, it is easy to construct examples where each individual has non-degenerate support, the supports overlap and each strictly moves in the opposite direction following some signal $x$. In particular, one individual can think $x$ has likelihood strictly increasing in $\theta$ for the $\theta$'s in his support while the other individual can think $x$ has likelihood strictly decreasing in $\theta$ for the $\theta$'s in his support.*

# 3 Polarization and Ambiguity

For the remainder of the paper, suppose the individual's goal is to predict the value of the parameter $\theta$. For tractability, we assume $\Theta = \{0, 1\}$ so there are just two possible parameter values. We make the standard assumption that the payoff to a prediction $\alpha \in [0, 1]$ is given by quadratic loss (i.e., $-(\alpha - \theta)^2$). To avoid tedious corner cases that do not change any of our conclusions, we assume $\pi_\theta$ has full support for each $\theta$. A strategy for the prediction problem when the individual is allowed to condition his prediction on $n \geq 0$ observations is a function $\alpha : \mathcal{X}^n \to \mathbb{R}$. We assume the individual views $\theta$ 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

$$E_\mu \phi[E_{\pi_\theta \ldots \pi_\theta}(-(\alpha\,(X_1, \ldots, X_n) - \theta)^2)],$$

where $\phi$ is increasing, concave and continuously differentiable. Observe that if $\phi$ is linear (i.e., ambiguity neutrality), the objective function reduces to expected quadratic loss. 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 $\alpha^*(x_1, \ldots, x_n)$ is therefore the unique solution to the first-order conditions:

$$E_\mu \left[ \phi'[E_{\pi_\theta \ldots \pi_\theta}(-(\alpha^*(X_1, \ldots, X_n) - \theta)^2)](\alpha^*(x_1, \ldots, x_n) - \theta) \prod_{i=1}^{n} \pi_\theta(x_i) \right] = 0 \tag{3.1}$$

for each $(x_1, \ldots, x_n) \in \mathcal{X}^n$. When no observations are available ($n = 0$), we sometimes denote the optimal prediction strategy by $\alpha^*(\emptyset)$ to emphasize that there are no observations and in this case (3.1) reduces to

$$E_\mu \left[ \phi'[-(\alpha^*(\emptyset) - \theta)^2](\alpha^*(\emptyset) - \theta) \right] = 0.$$

Abusing notation to let $\mu$ denote $\mu(\theta = 1)$, the first-order conditions for $\alpha^*(x_1, \ldots, x_n)$ and $\alpha^*(\emptyset)$ simplify to the following:

$$\frac{\alpha^*(x_1, \ldots, x_n)}{1 - \alpha^*(x_1, \ldots, x_n)} \frac{\phi'[E_{\pi_0 \ldots \pi_0}(-(\alpha^*(X_1, \ldots, X_n))^2)]}{\phi'[E_{\pi_1 \ldots \pi_1}(-(1 - \alpha^*(X_1, \ldots, X_n))^2)]} \prod_{i=1}^{n} \frac{\pi_0(x_i)}{\pi_1(x_i)} = \frac{\mu}{1 - \mu} \tag{3.2}$$

for each $(x_1, \ldots, x_n) \in \mathcal{X}^n$, and

$$\frac{\alpha^*(\emptyset)}{1 - \alpha^*(\emptyset)} \frac{\phi'[-(\alpha^*(\emptyset))^2]}{\phi'[-(1 - \alpha^*(\emptyset))^2]} = \frac{\mu}{1 - \mu}. \tag{3.3}$$

When no observations are available, the difference from the usual prediction problem with ambiguity neutrality is the presence of the term $\frac{\phi'[-(\alpha^*(\emptyset))^2]}{\phi'[-(1-\alpha^*(\emptyset))^2]}$ on the left-hand side of (3.3). Under ambiguity aversion, $\phi$ is concave, and this term reflects the desire to hedge or reduce the variation in payoffs as a function of the ambiguous parameter $\theta$. To see this, note that the argument of $\phi'$ in the numerator is the payoff when $\theta = 0$ and the argument in the denominator is the payoff when $\theta = 1$. The $\phi'$ ratio 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 ratio equals 1. Increases above 1 reflect a stronger desire to shift expected payoff from $\theta = 1$ to $\theta = 0$, i.e., to hedge by adjusting the prediction strategy

$\alpha^*$ downward. Similarly, decreases below 1 reflect a weaker desire to hedge by adjusting $\alpha^*$ downward. In general, ambiguity aversion ensures that when expected payoffs across the $\theta$'s differ, the $\phi'$ ratio pushes the optimal prediction in the direction of equalizing them by moving the prediction toward the $\theta$ 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 use the term *hedging motive* to describe the expression $\frac{\phi'[-(\alpha^*(\emptyset))^2]}{\phi'[-(1-\alpha^*(\emptyset))^2]}$ and its generalizations to cases when observations are available. The larger this hedging motive, the more hedging concerns push in the direction of strategies that perform well when $\theta = 0$.

An additional consideration enters when at least one observation is available ($n \geq 1$). From (3.2) it is clear that $\alpha^*(x_1, \ldots, x_n)$ depends on the optimal prediction strategy for other possible realizations of the $n$ observations. This is a fundamental non-separability introduced by ambiguity aversion – when the individual is ambiguity neutral ($\phi$ affine) this interdependence disappears.

Why does ambiguity aversion lead to such interdependence? The individual cares about the ambiguity concerning $\theta$ only to the extent that the expected payoff of the individual's strategy varies with $\theta$. An ambiguity averse individual, as explained above, dislikes this variation. In the prediction setting, the mapping from $\theta$ to expected payoffs depends on the *whole* prediction strategy $\alpha$. Therefore, an ambiguity averse individual will take the whole strategy into account when determining any part of that strategy. This is the non-separability reflected in (3.2).

Despite this interdependence, we have the following useful implication of (3.2) that is true independent of ambiguity attitude: for any $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathcal{X}^n$,

$$\frac{\alpha^*(x_1, \ldots, x_n)}{1 - \alpha^*(x_1, \ldots, x_n)} \prod_{i=1}^{n} \frac{\pi_0(x_i)}{\pi_1(x_i)} = \frac{\alpha^*(y_1, \ldots, y_n)}{1 - \alpha^*(y_1, \ldots, y_n)} \prod_{i=1}^{n} \frac{\pi_0(y_i)}{\pi_1(y_i)}. \tag{3.4}$$

The intuition for this equality is the standard one concerning equating marginal rates of substitution across signal realizations. An implication of (3.4) is that all optimal contingent predictions move up or down together. Given this, the interpretation and implications of the $\phi'$ ratio described above in the context of (3.3) apply equally well to the more general case, given by (3.2), where the optimal strategy is contingent on observations. This interpretation proves very helpful when examining updating.

In a prediction problem, *updating* maps beliefs about $\theta$ and new observations to posterior beliefs about $\theta$. *Dynamically consistent updating* is updating that preserves the optimality of the contingent strategy $\alpha^*(x_1, \ldots, x_n)$ as observations are realized (i.e., ex-ante optimal updating).[6] Let $\nu_\ell$ denote the posterior probability of $\theta = 1$ after observing $x_1, \ldots, x_\ell$ in a prediction problem with $n \geq \ell$ observations available. Dynamically consistent updating is equivalent to these posteriors $\nu_\ell$ satisfying

$$\frac{\alpha^*(x_1, \ldots, x_n)}{1 - \alpha^*(x_1, \ldots, x_n)} \frac{\phi'[E_{\pi_0 \ldots \pi_0}(-(\alpha^*(x_1, \ldots, x_\ell, X_{\ell+1,\ldots,}X_n))^2)]}{\phi'[E_{\pi_1 \ldots \pi_1}(-(1 - \alpha^*(x_1, \ldots, x_\ell, X_{\ell+1,\ldots,}X_n))^2)]} \prod_{i=\ell+1}^{n} \frac{\pi_0(x_i)}{\pi_1(x_i)} = \frac{\nu_\ell}{1 - \nu_\ell} \tag{3.5}$$

for all $0 \leq \ell \leq n$ and all $(x_1, \ldots, x_n) \in \mathcal{X}^n$. Note that (3.5) is simply the first-order condition of the continuation prediction problem, evaluated at the ex-ante optimal strategy $\alpha^*(x_1, \ldots, x_n)$, after $x_1, \ldots, x_\ell$ have been realized and assuming beliefs at that point are $\nu_\ell$. It therefore guarantees that $\alpha^*(x_1, \ldots, x_n)$ remains optimal as observations accumulate. After the next result, we will expand on the difference, under ambiguity aversion, between dynamically consistent updating and Bayesian updating and show how the former allows polarization. We begin by showing that several natural properties that hold under ambiguity neutrality continue to hold under ambiguity aversion.

Two such properties are identified in the next Proposition:

**Proposition 3.1** *(i) For all $n \geq 0$, with $n$ observations available, for each possible realization $(x_1, \ldots, x_n) \in \mathcal{X}^n$ of these observations, the optimal prediction $\alpha^*(x_1, \ldots, x_n)$ is an increasing function of $\mu$ (the prior probability of $\theta = 1$) and of the likelihood ratio $\prod_{i=1}^{n} \frac{\pi_1(x_i)}{\pi_0(x_i)}$;*

[6] For a thorough discussion and analysis of dynamically consistent updating under ambiguity aversion see Hanany and Klibanoff [2009].

*(ii) The posterior probability of $\theta = 1$ after observing $x_1, \ldots, x_n$ in the prediction problem with $n \geq 0$ observations available is above/equal to/below the posterior probability of $\theta = 1$ after observing $y_1, \ldots, y_m$ in the prediction problem with $m \geq 0$ observations available if and only if the optimal predictions in the respective continuation prediction problems are similarly ordered. Under dynamically consistent updating, the same is true of the ex-ante optimal contingent predictions (i.e., $\alpha^*(x_1, \ldots, x_n) \gtreqless \alpha^*(y_1, \ldots, y_m)$).*

Proposition 3.1 implies that polarization as we have defined it in terms of beliefs is equivalent to polarization in actions (here, predictions). A common signal 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 optimal predictions with no observations available. From the second part of the proposition with $m = 0$, 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 optimal predictions with no observations available. Combining these yields the equivalence.

Suppose signals $x_1, \ldots, x_\ell$ are observed. Dynamic consistency requires that the optimal prediction strategy after these observations also be the optimal prediction strategy ex ante contingent on observing $x_1, \ldots, x_\ell$. As emphasized above, under ambiguity aversion, the optimal prediction strategy is partly driven by the desire to hedge (i.e., to reduce the sensitivity of expected payoff to the ambiguous parameter $\theta$). Ex ante, before signals are realized, this hedging motive is captured by the ratio

$$\frac{\phi'[E_{\pi_0 \ldots \pi_0}(-(\alpha^*(X_1, \ldots, X_n))^2)]}{\phi'[E_{\pi_1 \ldots \pi_1}(-(1 - \alpha^*(X_1, \ldots, X_n))^2)]}. \tag{3.6}$$

Recall that, the larger this term is, the more hedging pushes in the direction of strategies that are expected to perform well when $\theta = 0$. However, after observing $x_1, \ldots, x_\ell$, the expected payoffs that determine the hedging motive involve only the uncertainty about the remaining observations. This interim hedging motive is captured by the ratio

$$\frac{\phi'[E_{\pi_0 \dots \pi_0}(-(\alpha^*(x_1,\dots,x_\ell,X_{\ell+1,\dots,}X_n))^2)]}{\phi'[E_{\pi_1 \dots \pi_1}(-(1-\alpha^*(x_1,\dots,x_\ell,X_{\ell+1,\dots,}X_n))^2)]}. \tag{3.7}$$

If the individual is ambiguity neutral, (3.6) and (3.7) are equal. In this case, by comparing (3.5) and (3.2) we see that the posterior is proportional to the prior times the likelihood,

$$\frac{\nu_\ell}{1-\nu_\ell} = \frac{\mu}{1-\mu}\prod_{i=1}^{\ell}\frac{\pi_1(x_i)}{\pi_0(x_i)}, \tag{3.8}$$

and thus Bayesian updating guarantees dynamic consistency. However, if the individual is ambiguity averse, (3.6) and (3.7) are typically *not* equal. The individual's hedging motive changes, as he no longer needs to account for variation in his expected payoffs induced by the first $\ell$ realizations. This change in the hedging motive is the hedging effect highlighted in the Introduction. To carry out the optimal prediction strategy, dynamically consistent updating departs from Bayesian updating in a way that exactly offsets this hedging effect.

We can expand on this intuition to offer a characterization of the direction of dynamically consistent updating:

**Proposition 3.2** *With $n \geq 1$ observations and dynamically consistent updating, for $0 \leq k < m \leq n$, the posterior probability of $\theta = 1$ after observing $x_1,\dots,x_m$ is above/equal to/below the posterior probability of $\theta = 1$ after observing $x_1,\dots,x_k$ if and only if*

$$\begin{aligned} &\frac{\phi'[E_{\pi_0 \dots \pi_0}(-(\alpha^*(x_1,\dots,x_m,X_{m+1,\dots,}X_n))^2)]}{\phi'[E_{\pi_1 \dots \pi_1}(-(1-\alpha^*(x_1,\dots,x_m,X_{m+1,\dots,}X_n))^2)]}\prod_{i=k+1}^{m}\frac{\pi_1(x_i)}{\pi_0(x_i)} \\ \gtreqless\ &\frac{\phi'[E_{\pi_0 \dots \pi_0}(-(\alpha^*(x_1,\dots,x_k,X_{k+1,\dots,}X_n))^2)]}{\phi'[E_{\pi_1 \dots \pi_1}(-(1-\alpha^*(x_1,\dots,x_k,X_{k+1,\dots,}X_n))^2)]}. \end{aligned} \tag{3.9}$$

Under ambiguity neutrality, the hedging motive terms on each side cancel and (3.9) reduces to the familiar statement that updating is up/flat/down as the likelihood ratio of the newly observed signals is above/equal to/below 1.

Under ambiguity aversion, as we described earlier, the two hedging motives are typically not identical, and dynamic consistency requires updating to offset this change by pushing in the opposite direction. This explains why, for example, when the interim hedging term is larger than the ex-ante hedging term so that hedging pushes towards valuing performance when $\theta = 0$ more strongly at the interim stage, Proposition 3.2 tells us that updating will be shaded upward (thus pushing more toward valuing performance when $\theta = 1$) compared to Bayesian updating.

The condition in (3.9) is not always easy to apply, as it involves the endogenously determined optimal strategy $\alpha^*$. 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., $x^H \in \arg\max_{x \in \mathcal{X}} \frac{\pi_1(x)}{\pi_0(x)} \geq 1$) always leads to updating upwards and observing the signal with the lowest likelihood ratio always leads to updating downwards. One implication is that polarization, if it is to occur at all, can only occur after "intermediate" common signals.

**Theorem 3.1** *With $n \geq 1$ observations, for $0 \leq k < m \leq n$, after observing $x_1, \ldots, x_k, \underbrace{x^H, \ldots, x^H}_{m-k \text{ times}}$ (resp. $x_1, \ldots, x_k, \underbrace{x^L, \ldots, x^L}_{m-k \text{ times}}$) the posterior probability of $\theta = 1$ after the $m$ observations is above (resp. below) the posterior probability (denoted by $\nu_k$) of $\theta = 1$ after observing $x_1, \ldots, x_k$ . It is also above (resp. below) the Bayesian update of $\nu_k$ given $\underbrace{x^H, \ldots, x^H}_{m-k \text{ times}}$ (resp. $\underbrace{x^L, \ldots, x^L}_{m-k \text{ times}}$).*

**Remark 3.1** *If signals are 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 corollary may be replaced by strictly above (resp. strictly below).*

To understand this result, first note that if, *for both values of* $\theta$, the expected optimal prediction after seeing the additional $m - (k + 1)$ observations is above the same expectation taken before those observations then, since $\phi'$ is decreasing by ambiguity aversion, the $\phi'$ ratio on the left-hand side of (3.9) is greater than the $\phi'$ ratio on the right-hand side. If, additionally, $\prod_{i=k+1}^{m} \frac{\pi_1(x_i)}{\pi_0(x_i)}$ is at least 1, it follows from (3.9) that updating must be upwards. When the additional observations are all $x^H$, these properties are satisfied. To see this, note that by Proposition 3.1, optimal predictions are higher after observing $x^H$ than after any other observations, so the expected optimal prediction after seeing a string of $m-(k+1)$ signals $x^H$ is greater than the same expectation before those observations.

The above argument shows that after observing $x^H$, both the hedging and likelihood effects push beliefs upward. Since Bayesian updating offsets the likelihood effect only, it also follows that dynamically consistent updating overshoots the Bayesian update. All of the arguments just made for the case of observing $x^H$ hold with the inequalities reversed when observing a lowest likelihood ratio signal $x^L \in \arg\min_{x\in\mathcal{X}} \frac{\pi_1(x)}{\pi_0(x)} \leq 1$.

We turn to polarization and offer some intuition for our main positive results.[7] Suppose there are two individuals with beliefs $\hat{\eta}$ and $\check{\eta}$ and assume $\hat{\eta}(\theta = 1) > \check{\eta}(\theta = 1)$ so $\hat{\eta}$ stochastically dominates $\check{\eta}$ (we also continue our abuse of notation and denote $\hat{\eta}(\theta = 1)$ and $\check{\eta}(\theta = 1)$ by $\hat{\eta}$ and $\check{\eta}$ respectively). These beliefs could be the individuals' priors or posteriors after observing some sequence of signals. (For simplicity, we suppress the notation for the history of signals in the discussion below.) 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 a sequence of extreme signals, we know they will update in the same direction, from Theorem 3.1. So to have any chance of polarization, 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. Specializing to the case of one observation and substituting for predic-

[7] For the purposes of this section, it is sufficient to consider only polarization that occurs following the last observation before a prediction is to be made.

tions $\alpha^*(x)$, $x \neq x^M$ using (3.4), inequality (3.9) becomes:

$$\frac{\phi'[-(\alpha^*\left(x^M\right))^2]}{\phi'[-(1-\alpha^*\left(x^M\right))^2]}\frac{\pi_1(x^M)}{\pi_0(x^M)} \tag{3.10}$$

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

The direction of this inequality determines the direction of updating. The connection between $\alpha^*\left(x^M\right)$ (and thus $\mu$, since $\alpha^*\left(x^M\right)$ is increasing in $\mu$ by Proposition 3.1) and the direction of this inequality may be quite complex. It is simpler 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 $\alpha^*\left(x^M\right)$ (and thus $\mu$) 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 $x^N$, at least one informative signal and twice continuously differentiable $\phi$ with $\phi'' < 0 < \phi'$. Polarization occurs after commonly observing $x^N$ 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 Appendix): If $\alpha^*\left(x^M\right) = 0$ or $\alpha^*\left(x^M\right) = 1$, the hedging motive expressions are the same on both sides of (3.10). When $\alpha^*\left(x^M\right)$ 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 $\phi'$ in the numerators on each side of (3.10)) will be very small and will have minimal influence on updating (since $-\frac{\phi''}{\phi'}$ is finite).[8] In contrast, if $\theta = 1$, predictions close to 0 will be very costly and small improvements in

[8]The 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.

those predictions would be valuable. Therefore, (since $-\frac{\phi''}{\phi'}$ is non-zero) it is the 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 $\phi'$ terms in the denominators with respect to $\alpha^*\left(x^M\right)$ and evaluating at $\alpha^*\left(x^M\right) = 0$, 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 $\pi_1$ terms, so that if $x^M$ is a neutral signal this condition will be satisfied.

Similar reasoning applies when $\alpha^*\left(x^M\right)$ is close to 1. In this case, it is when $\theta = 0$ that predictions will be far from perfect and differences in interim and ex-ante expected payoffs will be valuable and thus drive the hedging comparison. Differentiating using the numerators and evaluating at $\alpha^*\left(x^M\right) = 1$ yields that the ex-ante expected payoff when $\theta = 0$ is higher than the interim payoff when $\theta = 1$ if and only if the expected inverse likelihood, $\sum_{y\in\mathcal{X}} \pi_0(y)\frac{\pi_0(y)}{\pi_1(y)}$, is higher than the realized inverse likelihood, $\frac{\pi_0(x^M)}{\pi_1(x^M)}$. Again a neutral realized signal will satisfy this condition. The inverse likelihood appears in this case because as the signal $y$ varies, a higher inverse likelihood, $\frac{\pi_0(y)}{\pi_1(y)}$, moves the prediction $\alpha^*(y)$ closer to 0, leading to more payoff when $\theta = 0$.

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 $\eta$ (so that $\alpha^*\left(x^M\right)$ 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. Similarly, 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 $\eta$, updating will be shaded upward compared to Bayesian updating. When the signal is neutral, so that Bayesian updating is flat, these imply updating will be downward when belief is sufficiently low and upward when belief is sufficiently high, generating polarization.

**Remark 3.2** As the sketch of the proof suggests, we show something more: Under the assumptions of the theorem, dynamically consistent updating is shaded downward compared to Bayesian updating for all beliefs sufficiently close to 0 if and only if the likelihood ratio $\frac{\pi_1(x_n)}{\pi_0(x_n)}$ of the observed signal lies below $\sum_{y\in\mathcal{X}} \pi_1(y)\frac{\pi_1(y)}{\pi_0(y)}$. This bound is above 1 and is the average signal likelihood ratio given $\theta = 1$. Similarly, dynamically consistent updating is shaded upward compared to Bayesian updating for all beliefs sufficiently close to 1 if and only if $\frac{\pi_1(x_n)}{\pi_0(x_n)}$ lies above $\frac{1}{\sum_{y\in\mathcal{X}} \pi_0(y)\frac{\pi_0(y)}{\pi_1(y)}}$. This bound is below 1 and is the inverse of the average signal likelihood ratio given $\theta = 0$.

**Remark 3.3** The same strategy used to prove Theorem 3.2 also may be used to show that quadratic loss payoffs are not crucial. Specifically, any payoff function of the form $\psi(|\alpha - \theta|)$ where $\psi : [0,1] \to R$ is a twice continuously differentiable function satisfying $\psi'(0) = 0$ and $\psi'' < 0$ will yield a similar result. Thus, the important aspect of quadratic loss ($\psi(d) = -d^2$) 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.

**Remark 3.4** 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 $\theta$, 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.

To further investigate when polarization occurs, we turn to a particularly clean structure for determining the direction of updating. 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 (i.e., $\tau \in (0,1)$). In the Appendix, we provide a general characterization of when updating follows a threshold rule (Proposition A.2). To provide an explicit description of the threshold rule in circumstances where it is guaranteed to exist, 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. Note that constant relative ambiguity aversion simplifies the determination of the direction of updating, as it makes the right-hand side of (3.10) multiplicatively separable. Under these conditions, we show that updating always follows a threshold rule and explicitly derive the threshold.

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 the 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. With* $n \geq 1$ *observations, the posterior* *probability of $\theta = 1$ after $x_1, \ldots, x_{n-1}, x^M$ is above/equal to/below the probability of $\theta = 1$ after $x_1, \ldots, x_{n-1}$ when the latter is above/equal to/below a threshold $\tau(\gamma, \pi_0, \pi_1)$ that is independent of $n$ and $x_1, \ldots, x_{n-1}$ and beliefs.*

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 $\eta$ and constant relative ambiguity aversions $\hat{\gamma}$ and $\check{\gamma}$ exhibit polarization after observing the intermediate signal if and only if*

$$\tau(\hat{\gamma}, \pi_0, \pi_1) \leq \eta \leq \tau(\check{\gamma}, \pi_0, \pi_1) \tag{3.11}$$

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

$$\hat{\eta} \geq \tau(\gamma, \pi_0, \pi_1) \geq \check{\eta} \tag{3.12}$$

*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 $\tau(\gamma, \pi_0, \pi_1) \in (0, 1)$ and equals*

$$\frac{1}{1 + C^{2\gamma+1}}$$

*where*

$$C \equiv \frac{\pi_1\left(x^L\right)\sqrt{\pi_1\left(x^H\right)\pi_0\left(x^H\right)} - \pi_1\left(x^H\right)\sqrt{\pi_1\left(x^L\right)\pi_0\left(x^L\right)}}{\pi_0\left(x^H\right)\sqrt{\pi_1\left(x^L\right)\pi_0\left(x^L\right)} - \pi_0\left(x^L\right)\sqrt{\pi_1\left(x^H\right)\pi_0\left(x^H\right)}}.$$

Ambiguity neutral individuals with different levels of risk aversion or different beliefs cannot exhibit polarization if they use Bayesian updating. But dynamically consistent ambiguity averse individuals with heterogeneous ambiguity aversion or beliefs can exhibit polarization. This theory offers one plausible interpretation of polarization. Bulls and bears have different beliefs and are ambiguity averse. Polarization is a manifestation of their effort to implement their optimal plans. Death penalty opponents and proponents have different preferences and respond to intermediate signals by updating in opposite directions.

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. We next show that our previous results 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., $\pi_0\left(x^L\right) = \pi_1\left(x^H\right)$, $\pi_0\left(x^M\right) = \pi_1\left(x^M\right)$ and $\pi_0\left(x^H\right) = \pi_1\left(x^L\right)$). The individuals are allowed to condition their prediction on two observations. Suppose one individual observes the sequence $\{x^L, x^M\}$, while the other observes the sequence $\{x^H, x^M\}$. Applying Theorem 3.1, after one observation the first individual will have updated beliefs $\check{\eta} < \frac{1}{2}$ 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 $\check{\eta}$ 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:

**Theorem 3.4** Polarization in a Homogeneous Environment: *Assume common constant relative ambiguity aversion* $\gamma > 0$*, common prior* $\mu \in (0,1)$ *and exactly three distinct likelihood ratios. If* $\tau(\gamma, \pi_0, \pi_1) \in (0,1)$ *and the number of observations* $n$ *is sufficiently large, polarization occurs after observing the intermediate signal* $x^M$ *if up to that point one individual observes* $\underbrace{x^H, \ldots, x^H}_{n-1 \text{ times}}$ *while the other observes* $\underbrace{x^L, \ldots, x^L}_{n-1 \text{ times}}$.

# 4 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 we use to motivate our study (e.g., 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, we have assumed fully dynamically consistent updating. As was mentioned in the introduction, the effects we identify will survive and will continue to generate polarization even under substantially weaker assumptions. This could be studied through a multi-selves model with temptation and costly self-control in the style of Gul and Pesendorfer [2001] where the temptation is to use Bayesian updating. In such a setting, as long as self-control is not infinitely costly, an ambiguity averse individual will end up acting as if he uses an update rule in between Bayesian and dynamically consistent updating. This preserves a role for the hedging effect which was key to our results.

# A Appendix (For Online Publication)

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 of Proposition 3.1.** It is immediate from (3.2) and (3.3) that $\alpha^*(x_1,\ldots,x_n)\in(0,1)$ since $\mu\in(0,1)$ and $\phi'>0$. To prove (i), fix any $n\geq 0$ and $(x_1,\ldots,x_n)\in\mathcal{X}^n$ and, from (3.4), observe that for any $(y_1,\ldots,y_n)\in\mathcal{X}^n$, $\alpha^*(y_1,\ldots,y_n)$ is a strictly increasing function of $\alpha^*(x_1,\ldots,x_n)$ in any solution of the system of first-order conditions. This and the fact that $\phi$ is concave implies that the left-hand side of the corresponding first-order condition is strictly increasing in $\alpha^*(x_1,\ldots,x_n)$ and decreasing in $\prod_{i=1}^{n}\frac{\pi_1(x_i)}{\pi_0(x_i)}$. The right-hand side of (3.2) (or, when $n=0$, (3.3)) is strictly increasing in $\mu$ and constant in $\alpha^*(x_1,\ldots,x_n)$. Therefore, $\alpha^*(x_1,\ldots,x_n)$ is well-defined and strictly increasing in $\mu$ and $\prod_{i=1}^{n}\frac{\pi_1(x_i)}{\pi_0(x_i)}$.

To prove (ii), let $\nu_n$ (resp. $\nu_m$) denote the the posterior probability of $\theta=1$ after observing $x_1,\ldots,x_n$ (resp. $y_1,\ldots,y_m$) in the prediction problem with $n\geq 0$ (resp. $m\geq 0$) observations available. Let $\beta^*(x_1,\ldots,x_n)$ (resp. $\beta^*(y_1,\ldots,y_m)$) denote the optimal prediction in the continuation problem given that posterior.

By the first-order conditions for optimality, these predictions and posteriors must satisfy

$$\frac{\beta^*(x_1,\ldots,x_n)}{1-\beta^*(x_1,\ldots,x_n)}\frac{\phi'[-(\beta^*(x_1,\ldots,x_n))^2]}{\phi'[-(1-\beta^*(x_1,\ldots,x_n))^2]}=\frac{\nu_n}{1-\nu_n}$$

and

$$\frac{\beta^*(y_1,\ldots,y_m)}{1-\beta^*(y_1,\ldots,y_m)}\frac{\phi'[-(\beta^*(y_1,\ldots,y_m))^2]}{\phi'[-(1-\beta^*(y_1,\ldots,y_m))^2]}=\frac{\nu_m}{1-\nu_m}.$$

Therefore,

$$\nu_n \gtreqless \nu_m$$

if and only if

$$\frac{\beta^*(x_1,\ldots,x_n)}{1-\beta^*(x_1,\ldots,x_n)}\frac{\phi'[-(\beta^*(x_1,\ldots,x_n))^2]}{\phi'[-(1-\beta^*(x_1,\ldots,x_n))^2]} \gtreqless \frac{\beta^*(y_1,\ldots,y_m)}{1-\beta^*(y_1,\ldots,y_m)}\frac{\phi'[-(\beta^*(y_1,\ldots,y_m))^2]}{\phi'[-(1-\beta^*(y_1,\ldots,y_m))^2]}.$$

Since $\frac{z}{1-z}\frac{\phi'[-(z)^2]}{\phi'[-(1-z)^2]}$ is strictly increasing in $z$ on $(0,1)$, this is equivalent to

$$\beta^*(x_1,\ldots,x_n) \gtreqless \beta^*(y_1,\ldots,y_m).$$

Finally, under dynamically consistent updating, from (3.5), the posteriors must satisfy

$$\frac{\alpha^*(x_1,\ldots,x_n)}{1-\alpha^*(x_1,\ldots,x_n)}\frac{\phi'[-(\alpha^*(x_1,\ldots,x_n))^2]}{\phi'[-(1-\alpha^*(x_1,\ldots,x_n))^2]}=\frac{\nu_n}{1-\nu_n}$$

and

$$\frac{\alpha^*(y_1,\ldots,y_m)}{1-\alpha^*(y_1,\ldots,y_m)}\frac{\phi'[-(\alpha^*(y_1,\ldots,y_m))^2]}{\phi'[-(1-\alpha^*(y_1,\ldots,y_m))^2]}=\frac{\nu_m}{1-\nu_m}.$$

Therefore, $\alpha^*(x_1,\ldots,x_n)=\beta^*(x_1,\ldots,x_n)$ and $\alpha^*(y_1,\ldots,y_m)=\beta^*(y_1,\ldots,y_m)$, so that the above argument yields

$$\nu_n \gtreqless \nu_m$$

if and only if

$$\alpha^*(x_1,\ldots,x_n) \gtreqless \alpha^*(y_1,\ldots,y_m).$$

■

**Proof of Proposition 3.2.** Let $\nu_m$ (resp. $\nu_k$) denote the the posterior probability of $\theta = 1$ after observing $x_1, \ldots, x_m$ (resp. $x_1, \ldots, x_k$) in the prediction problem with $n \geq 1$ observations available. Dynamically consistent updating implies that (3.5) is satisfied for $\ell = m$ and $\ell = k$. Therefore,

$$\frac{\alpha^*(x_1,\ldots,x_n)}{1-\alpha^*(x_1,\ldots,x_n)} \frac{\phi'[E_{\pi_0\ldots\pi_0}(-(\alpha^*(x_1,\ldots,x_m,X_{m+1,\ldots,}X_n))^2)]}{\phi'[E_{\pi_1\ldots\pi_1}(-(1-\alpha^*(x_1,\ldots,x_m,X_{m+1,\ldots,}X_n))^2)]} \prod_{i=m+1}^{n} \frac{\pi_0(x_i)}{\pi_1(x_i)} = \frac{\nu_m}{1-\nu_m}$$

and

$$\frac{\alpha^*(x_1,\ldots,x_n)}{1-\alpha^*(x_1,\ldots,x_n)} \frac{\phi'[E_{\pi_0\ldots\pi_0}(-(\alpha^*(x_1,\ldots,x_k,X_{k+1,\ldots,}X_n))^2)]}{\phi'[E_{\pi_1\ldots\pi_1}(-(1-\alpha^*(x_1,\ldots,x_k,X_{k+1,\ldots,}X_n))^2)]} \prod_{i=k+1}^{n} \frac{\pi_0(x_i)}{\pi_1(x_i)} = \frac{\nu_k}{1-\nu_k}.$$

Combining the above,

$$\nu_m \gtreqless \nu_k$$

if and only if

$$\begin{aligned} &\frac{\phi'[E_{\pi_0\ldots\pi_0}(-(\alpha^*(x_1,\ldots,x_m,X_{m+1,\ldots,}X_n))^2)]}{\phi'[E_{\pi_1\ldots\pi_1}(-(1-\alpha^*(x_1,\ldots,x_m,X_{m+1,\ldots,}X_n))^2)]} \\ \gtreqless\ &\frac{\phi'[E_{\pi_0\ldots\pi_0}(-(\alpha^*(x_1,\ldots,x_k,X_{k+1,\ldots,}X_n))^2)]}{\phi'[E_{\pi_1\ldots\pi_1}(-(1-\alpha^*(x_1,\ldots,x_k,X_{k+1,\ldots,}X_n))^2)]} \prod_{i=k+1}^{m} \frac{\pi_0(x_i)}{\pi_1(x_i)}. \end{aligned}$$

■

**Proof of Theorem 3.1.** Let $\nu_m^H$ (resp. $\nu_m^L$) denote the the posterior probability of $\theta = 1$ after observing $x_1, \ldots, x_k, \underbrace{x^H, \ldots, x^H}_{m-k \text{ times}}$ (resp. $x_1, \ldots, x_k, \underbrace{x^L, \ldots, x^L}_{m-k \text{ times}}$) in the prediction problem with $n \geq 1$ observations available. By Proposition 3.2,

$$\nu_m^H \geq \nu_k$$

if and only if

$$\begin{aligned} & \frac{\phi'[E_{\pi_0 \dots \pi_0}(-(\alpha^*\left(x_1, \dots, x_k, x^H, \dots, x^H, X_{m+1,\dots,}X_n\right))^2)]}{\phi'[E_{\pi_1 \dots \pi_1}(-(1-\alpha^*\left(x_1, \dots, x_k, x^H, \dots, x^H, X_{m+1,\dots,}X_n\right))^2)]} \qquad \text{(A.1)} \\ \geq \quad & \frac{\phi'[E_{\pi_0 \dots \pi_0}(-(\alpha^*\left(x_1, \dots, x_k, X_{k+1,\dots,}X_n\right))^2)]}{\phi'[E_{\pi_1 \dots \pi_1}(-(1-\alpha^*\left(x_1, \dots, x_k, X_{k+1,\dots,}X_n\right))^2)]} \prod_{i=k+1}^{m} \frac{\pi_0(x^H)}{\pi_1(x^H)}. \end{aligned}$$

For all $(y_{k+1}, \dots, y_n)$, since $\prod_{i=k+1}^{m} \frac{\pi_1(x^H)}{\pi_0(x^H)} \geq \prod_{i=k+1}^{m} \frac{\pi_1(y_i)}{\pi_0(y_i)}$, it follows from (3.4) that

$$\alpha^*\left(x_1, \dots, x_k, x^H, \dots, x^H, y_{m+1}, \dots, y_n\right) \geq \alpha^*\left(x_1, \dots, x_k, y_{k+1}, \dots, y_n\right).$$

Therefore $E_{\pi_0 \dots \pi_0}(\alpha^*\left(x_1, \dots, x_k, x^H, \dots, x^H, X_{m+1,\dots,}X_n\right))^2 \geq E_{\pi_0 \dots \pi_0}(\alpha^*(x_1, \dots, x_k, X_{k+1,\dots,}X_n))^2$ and $E_{\pi_1 \dots \pi_1}(1-\alpha^*\left(x_1, \dots, x_k, x^H, \dots, x^H, X_{m+1,\dots,}X_n\right))^2 \leq E_{\pi_1 \dots \pi_1}(1-\alpha^*\left(x_1, \dots, x_k, X_{k+1,\dots,}X_n\right))^2$. As $\phi$ is concave, this implies

$$\begin{aligned} & \frac{\phi'[E_{\pi_0 \dots \pi_0}(-(\alpha^*\left(x_1, \dots, x_k, x^H, \dots, x^H, X_{m+1,\dots,}X_n\right))^2)]}{\phi'[E_{\pi_1 \dots \pi_1}(-(1-\alpha^*\left(x_1, \dots, x_k, x^H, \dots, x^H, X_{m+1,\dots,}X_n\right))^2)]} \qquad \text{(A.2)} \\ \geq \quad & \frac{\phi'[E_{\pi_0 \dots \pi_0}(-(\alpha^*\left(x_1, \dots, x_k, X_{k+1,\dots,}X_n\right))^2)]}{\phi'[E_{\pi_1 \dots \pi_1}(-(1-\alpha^*\left(x_1, \dots, x_k, X_{k+1,\dots,}X_n\right))^2)]}. \end{aligned}$$

Since $\frac{\pi_1(x^H)}{\pi_0(x^H)} \geq 1$, (A.1) follows. Furthermore, (3.5) for $\ell = m$ and $\ell = k$ and

(A.2) imply

$$
\begin{aligned}
& \frac{\nu_m^H}{1-\nu_m^H} \\
= \; & \frac{\alpha^*\left(x_1,\ldots,x_k,x^H,\ldots,x^H,x_{m+1},\ldots,x_n\right)}{1-\alpha^*\left(x_1,\ldots,x_k,x^H,\ldots,x^H,x_{m+1},\ldots,x_n\right)} \times \\
& \frac{\phi'[E_{\pi_0\ldots\pi_0}(-(\alpha^*\left(x_1,\ldots,x_k,x^H,\ldots,x^H,X_{m+1,\ldots,}X_n\right))^2)]}{\phi'[E_{\pi_1\ldots\pi_1}(-(1-\alpha^*\left(x_1,\ldots,x_k,x^H,\ldots,x^H,X_{m+1,\ldots,}X_n\right))^2)]} \prod_{i=m+1}^{n} \frac{\pi_0(x_i)}{\pi_1(x_i)} \\
\geq \; & \frac{\alpha^*\left(x_1,\ldots,x_k,x^H,\ldots,x^H,x_{m+1},\ldots,x_n\right)}{1-\alpha^*\left(x_1,\ldots,x_k,x^H,\ldots,x^H,x_{m+1},\ldots,x_n\right)} \times \\
& \frac{\phi'[E_{\pi_0\ldots\pi_0}(-(\alpha^*\left(x_1,\ldots,x_k,X_{k+1,\ldots,}X_n\right))^2)]}{\phi'[E_{\pi_1\ldots\pi_1}(-(1-\alpha^*\left(x_1,\ldots,x_k,X_{k+1,\ldots,}X_n\right))^2)]} \prod_{i=m+1}^{n} \frac{\pi_0(x_i)}{\pi_1(x_i)} \\
= \; & \frac{\nu_k}{1-\nu_k}\left(\frac{\pi_1(x^H)}{\pi_0(x^H)}\right)^{m-(k+1)}.
\end{aligned}
$$

Thus,

$$\frac{\nu_m^H}{1-\nu_m^H} \geq \frac{\nu_k}{1-\nu_k}\left(\frac{\pi_1(x^H)}{\pi_0(x^H)}\right)^{m-(k+1)}$$

where the right-hand side is the posterior ratio generated by Bayesian updating of $\nu_k$ after observing $\underbrace{x^H,\ldots,x^H}_{m-k \text{ times}}$.

An analogous argument shows $\nu_k \geq \nu_m^L$ and

$$\frac{\nu_m^L}{1-\nu_m^L} \leq \frac{\nu_k}{1-\nu_k}\left(\frac{\pi_1(x^L)}{\pi_0(x^L)}\right)^{m-(k+1)}.$$

■

**Proof of Theorem 3.2.** Recall that the optimal prediction $\alpha^*(x_1,\ldots,x_{n-1},x_n)$ is continuous and increasing in the posterior probability of $\theta=1$ after observing $x_1,\ldots,x_{n-1}$. Denote this posterior probability by $\eta$. As the optimal prediction is 0 if $\eta=0$ and 1 if $\eta=1$, considering $\eta$ close enough to 0 or $\eta$ close enough to 1 is equivalent to considering $\alpha^*(x_1,\ldots,x_{n-1},x_n)$ 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 (3.5), that updating $\eta$ after seeing $x_n$ will be shaded upward/equal to/shaded downward compared to Bayesian updating if and only if

$$\begin{aligned} & \phi'[-(\alpha^*(x_1,\ldots,x_{n-1},x_n))^2]\phi'[-\sum_{y\in\mathcal{X}}\pi_1(y)(1-\alpha^*(x_1,\ldots,x_{n-1},y))^2] \\ \gtreqless\; & \phi'[-(1-\alpha^*(x_1,\ldots,x_{n-1},x_n))^2]\phi'[-\sum_{y\in\mathcal{X}}\pi_0(y)(\alpha^*(x_1,\ldots,x_{n-1},y))^2]. \end{aligned} \tag{A.3}$$

From (3.4), $\alpha^*(x_1,\ldots,x_{n-1},y)=\beta_{\pi_1,\pi_0}(\alpha^*(x_1,\ldots,x_{n-1},x_n);y)$ where $\beta_{\pi_1,\pi_0}:[0,1]\times\mathcal{X}\to[0,1]$ is defined by $\beta_{\pi_1,\pi_0}(z;y)=\frac{z\frac{\pi_1(y)}{\pi_0(y)}}{z\frac{\pi_1(y)}{\pi_0(y)}+(1-z)\frac{\pi_1(x_n)}{\pi_0(x_n)}}$ for all $z\in[0,1]$ and $y\in\mathcal{X}$. Define the function $f:[0,1]\to\mathbb{R}$ 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.3), observe that when $z=\alpha^*(x_1,\ldots,x_{n-1},x_n)\in(0,1)$, 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 $\phi'$ 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 $\beta_{\pi_1,\pi_0}$ with respect to $z$ evaluated at $(z;y)$ by $\beta'_{\pi_1,\pi_0}(z;y)$) yields,

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

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 $\phi''$ is negative and finite (since $\phi''$ 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 $\beta'_{\pi_1,\pi_0}(0;y) = \frac{\pi_1(y)}{\pi_0(y)}/\frac{\pi_1(x_n)}{\pi_0(x_n)}$ and $\beta'_{\pi_1,\pi_0}(1;y) = \frac{\pi_1(x_n)}{\pi_0(x_n)}/\frac{\pi_1(y)}{\pi_0(y)}$. Thus $f'(0) < 0$ and $f'(1) < 0$ if and only if

$$\frac{1}{\sum_{y\in\mathcal{X}}\pi_0(y)\frac{\pi_0(y)}{\pi_1(y)}} < \frac{\pi_1(x_n)}{\pi_0(x_n)} < \sum_{y\in\mathcal{X}}\pi_1(y)\frac{\pi_1(y)}{\pi_0(y)}. \qquad \text{(A.4)}$$

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.4) 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.4). 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 $\min_{w_1,\ldots,w_{|\mathcal{X}|}} \sum_{i=1}^{|\mathcal{X}|} \frac{w_i^2}{v_i}$ subject to $\sum_{i=1}^{|\mathcal{X}|} w_i = 1$, assuming $\sum_{i=1}^{|\mathcal{X}|} v_i = 1$ and $v_i > 0$ for $i = 1, ..., |\mathcal{X}|$, 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.4) is always satisfied if $\frac{\pi_1(x_n)}{\pi_0(x_n)} = 1$ (i.e., if $x_n$ is a neutral signal).

Finally, observe that if $x_n$ 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. ■

**Proof of Proposition 3.3.** From Lemma A.1, $\nu_n \gtreqless \nu_{n-1}$ after observing $x^M$ if and only if

$$\sum_{y\in\mathcal{X}} \pi_1(y) \frac{\left(\frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^{\frac{1}{\gamma}+2} - \frac{\pi_1(y)}{\pi_0(y)}}{[\alpha^*(x_1,\ldots,x_{n-1},x^M)\frac{\pi_1(y)}{\pi_0(y)} + (1-\alpha^*(x_1,\ldots,x_{n-1},x^M))\frac{\pi_1(x^M)}{\pi_0(x^M)}]^2} \gtreqless 0. \tag{A.5}$$

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.5) 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.5) 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.5), the term for $y = x^L$ is positive and has a denominator strictly decreasing in $\alpha^*(x_1, \ldots, x_{n-1}, x^M)$, the term for $y = x^M$ is constant in $\alpha^*(x_1, \ldots, x_{n-1}, x^M)$, and the term for $y = x^H$ is negative and has a denominator strictly increasing in $\alpha^*(x_1, \ldots, x_{n-1}, x^M)$. Therefore the whole sum is strictly increasing in $\alpha^*(x_1, \ldots, x_{n-1}, x^M)$ and thus can change signs at most once. Three sub-cases are relevant:

$(iii)(a)$ the left-hand side of (A.5) is non-negative when 0 is plugged in for $\alpha^*(x_1, \ldots, x_{n-1}, x^M)$. In this case, updating is always upward, so set $\tau(\gamma, \pi_0, \pi_1) = 0$.

$(iii)(b)$ the left-hand side of (A.5) is non-positive when 1 is plugged in for $\alpha^*(x_1, \ldots, x_{n-1}, x^M)$. 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.5) in $\alpha^*(x_1, \ldots, x_{n-1}, x^M)$ implies there exists a unique solution for $a$ in $(0, 1)$ to

$$\sum_{y \in \mathcal{X}} \pi_1(y) \frac{\left(\frac{\pi_1(x^M)}{\pi_0(x^M)}\right)^{\frac{1}{\gamma}+2} - \frac{\pi_1(y)}{\pi_0(y)}}{(a\frac{\pi_1(y)}{\pi_0(y)} + (1-a)\frac{\pi_1(x^M)}{\pi_0(x^M)})^2} = 0. \tag{A.6}$$

Since (A.7) holds with equality when $z = a$, using constant relative ambiguity aversion ($\phi'(z) = (-z)^\gamma$) and given the monotonicity of $\alpha^*(x_1, \ldots, x_{n-1}, x^M)$ in $\nu_{n-1}$, the associated threshold for $\nu_{n-1}$ may be found by substituting $z = a$ into (A.7) with equality and solving for $\nu_{n-1} = \tau(\gamma, \pi_0, \pi_1)$. 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

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

and

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

■

**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 $\tau$ function is the one defined in that result. ■

**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 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 $S(a) = 0$. This simplifies to

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

■

**Proof of Theorem 3.4.** Under dynamically consistent updating starting from any prior $\mu \in (0,1)$, observing a sufficiently long string of $x^L$'s (resp.

$x^H$'s) results in a posterior close enough to 0 (resp. 1) so as to be on opposite sides of the threshold $\tau(\gamma, \pi_0, \pi_1)$. Bayesian updating displays this property. By Theorem 3.1, dynamically consistent updating is above (resp. below) the Bayesian update given $n-1$ observations of $x^H$ (resp. $x^L$) and must also display this property. By Theorem 3.3, after such strings of extreme observations, a common observation of $x_n = x^M$ will result in polarization – the smaller posterior will become even smaller, while the larger posterior will increase. Key to this is that $\tau(\gamma, \pi_0, \pi_1)$ was shown to be independent of $n$ and the history $x_1, \ldots, x_{n-1}$ (Proposition 3.3). ■

## A.2 Further Results on the Direction of Updating

The next result combines Proposition 3.2 and equations (3.4) and (3.5) to show a general form relating fundamentals to the direction of updating.

**Proposition A.1** *In the prediction problem with $n \geq 1$ observations and dynamically consistent updating, for $0 \leq m < n$, the posterior probability of $\pi_1$ after $x_1, \ldots, x_n$ (denoted by $\nu_n$) is above/equal to/below the posterior probability of $\pi_1$ after $x_1, \ldots, x_m$ (denoted by $\nu_m$) if and only if the fundamentals $(\nu_m, \phi, \pi_1, \pi_0)$ are such that*

$$\frac{z}{1-z}\frac{\phi'[-z^2]}{\phi'[-(1-z)^2]} \gtreqless \frac{\nu_m}{1-\nu_m}, \tag{A.7}$$

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

$$\begin{aligned} &\frac{z}{1-z}\frac{\phi'\left[-z^2\sum_{(y_{m+1},\ldots,y_n)\in\mathcal{X}^{n-m}}\frac{\prod_{i=m+1}^n \pi_0(y_i)\left(\frac{\pi_1(y_i)}{\pi_0(y_i)}\right)^2}{\left(z\prod_{i=m+1}^n\frac{\pi_1(y_i)}{\pi_0(y_i)}+(1-z)\prod_{i=m+1}^n\frac{\pi_1(x_i)}{\pi_0(x_i)}\right)^2}\right]}{\phi'\left[-(1-z)^2\sum_{(y_{m+1},\ldots,y_n)\in\mathcal{X}^{n-m}}\frac{\prod_{i=m+1}^n \pi_1(y_i)\left(\frac{\pi_1(x_i)}{\pi_0(x_i)}\right)^2}{\left(z\prod_{i=m+1}^n\frac{\pi_1(y_i)}{\pi_0(y_i)}+(1-z)\prod_{i=m+1}^n\frac{\pi_1(x_i)}{\pi_0(x_i)}\right)^2}\right]} \\ &= \frac{\nu_m}{1-\nu_m}\prod_{i=m+1}^n\frac{\pi_1(x_i)}{\pi_0(x_i)}. \end{aligned} \tag{A.8}$$

**Proof.** Substituting (3.5) into (3.9) and rearranging yields

$$\frac{\alpha^*(x_1,\ldots,x_n)}{1-\alpha^*(x_1,\ldots,x_n)}\frac{\phi'[-\alpha^*(x_1,\ldots,x_n)^2]}{\phi'[-(1-\alpha^*(x_1,\ldots,x_n))^2]} \gtreqless \frac{\nu_m}{1-\nu_m}.$$

From (3.4), we obtain for all $(y_{m+1},\ldots,y_n) \in \mathcal{X}^{n-m}$,

$$\begin{aligned}&\alpha^*(x_1,\ldots,x_m,y_{m+1},\ldots,y_n)\\ = &\frac{\alpha^*(x_1,\ldots,x_n)\prod_{i=m+1}^n \frac{\pi_1(y_i)}{\pi_0(y_i)}}{\alpha^*(x_1,\ldots,x_n)\prod_{i=m+1}^n \frac{\pi_1(y_i)}{\pi_0(y_i)} + (1-\alpha^*(x_1,\ldots,x_n))\prod_{i=m+1}^n \frac{\pi_1(x_i)}{\pi_0(x_i)}}.\end{aligned}$$

Using this together with (3.5), $\alpha^*(x_1,\ldots,x_n)$ is the unique solution to

$$\begin{aligned}&\frac{\alpha^*(x_1,\ldots,x_n)}{1-\alpha^*(x_1,\ldots,x_n)}\times\\ &\frac{\phi'\left[-\sum_{(y_{m+1},\ldots,y_n)\in\mathcal{X}^{n-m}} \frac{\alpha^*(x_1,\ldots,x_n)^2\prod_{i=m+1}^n \pi_0(y_i)\left(\frac{\pi_1(y_i)}{\pi_0(y_i)}\right)^2}{\left(\alpha^*(x_1,\ldots,x_n)\prod_{i=m+1}^n \frac{\pi_1(y_i)}{\pi_0(y_i)}+(1-\alpha^*(x_1,\ldots,x_n))\prod_{i=m+1}^n \frac{\pi_1(x_i)}{\pi_0(x_i)}\right)^2}\right]}{\phi'\left[-\sum_{(y_{m+1},\ldots,y_n)\in\mathcal{X}^{n-m}} \frac{(1-\alpha^*(x_1,\ldots,x_n))^2\prod_{i=m+1}^n \pi_1(y_i)\left(\frac{\pi_1(x_i)}{\pi_0(x_i)}\right)^2}{\left(\alpha^*(x_1,\ldots,x_n)\prod_{i=m+1}^n \frac{\pi_1(y_i)}{\pi_0(y_i)}+(1-\alpha^*(x_1,\ldots,x_n))\prod_{i=m+1}^n \frac{\pi_1(x_i)}{\pi_0(x_i)}\right)^2}\right]}\\ = &\frac{\nu_m}{1-\nu_m}\prod_{i=m+1}^n \frac{\pi_1(x_i)}{\pi_0(x_i)}.\end{aligned}$$

■

In interpreting inequality (A.7), it is important to realize that $z$ is an increasing function of beliefs $\nu_m$ (as follows from the argument used in proving part (i) of Proposition 3.1 with $z$ playing the role of $\alpha^*(x_1,\ldots,x_n)$ and $\nu_m$ playing the role of $\mu$). In fact, (A.8) combines the first-order conditions (3.4) and (3.5). This implies that $z = \alpha^*(x_1,\ldots,x_n)$, the optimal prediction given the observations. From (A.8), in the case of ambiguity neutrality ($\phi$ affine) $\frac{z}{1-z}$ is simply a multiple of $\frac{\nu_m}{1-\nu_m}$ so that updating is either always upward (if $\prod_{i=m+1}^n \frac{\pi_1(x_i)}{\pi_0(x_i)} \geq 1$) or always downward (if $\prod_{i=m+1}^n \frac{\pi_1(x_i)}{\pi_0(x_i)} \leq 1$). Similarly, we see that under ambiguity aversion, $\frac{z}{1-z}$ is generally a non-linear function of $\frac{\nu_m}{1-\nu_m}$ (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.7) may change direction as beliefs $\nu_m$ 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.7 changes direction at most once.

**Proposition A.2** *There is a threshold rule for updating $\nu_m$ after observing $x_{m+1}, \ldots, x_n$ if and only if*

$$\frac{\phi'[-z^2]}{\phi'[-(1-z)^2]} \prod_{i=m+1}^{n} \frac{\pi_1(x_i)}{\pi_0(x_i)} - \frac{\phi'\left[-z^2 \sum_{(y_{m+1},\ldots,y_n)\in\mathcal{X}^{n-m}} \frac{\prod_{i=m+1}^{n} \pi_0(y_i)\left(\frac{\pi_1(y_i)}{\pi_0(y_i)}\right)^2}{\left(z\prod_{i=m+1}^{n} \frac{\pi_1(y_i)}{\pi_0(y_i)} + (1-z)\prod_{i=m+1}^{n} \frac{\pi_1(x_i)}{\pi_0(x_i)}\right)^2}\right]}{\phi'\left[-(1-z)^2 \sum_{(y_{m+1},\ldots,y_n)\in\mathcal{X}^{n-m}} \frac{\prod_{i=m+1}^{n} \pi_1(y_i)\left(\frac{\pi_1(x_i)}{\pi_0(x_i)}\right)^2}{\left(z\prod_{i=m+1}^{n} \frac{\pi_1(y_i)}{\pi_0(y_i)} + (1-z)\prod_{i=m+1}^{n} \frac{\pi_1(x_i)}{\pi_0(x_i)}\right)^2}\right]} \quad \text{(A.9)}$$

*as a function of $z$ has at most one zero in $(0,1)$ and, if a zero exists, (A.9) 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. ■

Finally, we present a lemma showing how inequality (3.9) 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** *In the prediction problem with $n \geq 1$ observations, dynamically consistent updating and constant relative ambiguity aversion $\gamma > 0$, the posterior probability of $\pi_1$ after $x_1, \ldots, x_n$ is above/equal to/below the posterior probability of $\pi_1$ after $x_1, \ldots, x_{n-1}$ if and only if*

$$\sum_{y\in\mathcal{X}} \pi_1(y) \frac{\left(\frac{\pi_1(x_n)}{\pi_0(x_n)}\right)^{\frac{1}{\gamma}+2} - \frac{\pi_1(y)}{\pi_0(y)}}{\left(\alpha^*(x_1,\ldots,x_n)\frac{\pi_1(y)}{\pi_0(y)} + (1-\alpha^*(x_1,\ldots,x_n))\frac{\pi_1(x_n)}{\pi_0(x_n)}\right)^2} \gtreqless 0. \quad \text{(A.10)}$$

**Proof.** Let $\nu_n$ (resp. $\nu_{n-1}$) denote the posterior probability of $\pi_1$ after observing $x_1, \dots, x_n$ (resp. $x_1, \dots, x_{n-1}$) in the prediction problem with $n \geq 1$ observations available. From inequality (A.7) and equation (A.8), $\nu_n \gtreqless \nu_{n-1}$ if and only if

$$\begin{aligned} &\frac{\phi'[-(\alpha^*(x_1,\dots,x_n))^2]}{\phi'[-(1-\alpha^*(x_1,\dots,x_n))^2]}\frac{\pi_1(x_n)}{\pi_0(x_n)} \qquad \text{(A.11)} \\ \gtreqless\; &\frac{\phi'\left[-\alpha^*(x_1,\dots,x_n)^2 \sum_{y\in\mathcal{X}} \frac{\pi_0(y)\left(\frac{\pi_1(y)}{\pi_0(y)}\right)^2}{\left(\alpha^*(x_1,\dots,x_n)\frac{\pi_1(y)}{\pi_0(y)}+(1-\alpha^*(x_1,\dots,x_n))\frac{\pi_0(y)}{\pi_0(y)}\frac{\pi_1(x_n)}{\pi_0(x_n)}\right)^2}\right]}{\phi'\left[-(1-\alpha^*(x_1,\dots,x_n))^2 \sum_{y\in\mathcal{X}} \frac{\pi_1(y)\left(\frac{\pi_1(x_n)}{\pi_0(x_n)}\right)^2}{\left(\alpha^*(x_1,\dots,x_n)\frac{\pi_1(y)}{\pi_0(y)}+(1-\alpha^*(x_1,\dots,x_n))\frac{\pi_1(x_n)}{\pi_0(x_n)}\right)^2}\right]}. \end{aligned}$$

Under constant relative ambiguity aversion, $\phi'(z) = (-z)^\gamma$ and therefore (A.11) is equivalent to

$$\left(\frac{\pi_1(x_n)}{\pi_0(x_n)}\right)^{\frac{1}{\gamma}} \gtreqless \frac{\sum_{y\in\mathcal{X}} \frac{\pi_0(y)\left(\frac{\pi_1(y)}{\pi_0(y)}\right)^2}{\left(\alpha^*(x_1,\dots,x_n)\frac{\pi_1(y)}{\pi_0(y)}+(1-\alpha^*(x_1,\dots,x_n))\frac{\pi_1(x_n)}{\pi_0(x_n)}\right)^2}}{\sum_{y\in\mathcal{X}} \frac{\pi_1(y)\left(\frac{\pi_1(x_n)}{\pi_0(x_n)}\right)^2}{\left(\alpha^*(x_1,\dots,x_n)\frac{\pi_1(y)}{\pi_0(y)}+(1-\alpha^*(x_1,\dots,x_n))\frac{\pi_1(x_n)}{\pi_0(x_n)}\right)^2}}.$$

Simplifying yields inequality (A.10). ■

# References


D. Acemoglu, V. Chernozhukov and M. Yildiz (2009): "Fragility of Asymptotic Agreement under Bayesian Learning," mimeo., M.I.T.

R. Adams and D. Ferreira (2010): "Moderation in Groups: Evidence from Betting on Ice Break-ups in Alaska," *Review of Economic Studies*, 77, 882–913.

J. Andreoni and T. Mylovanov (2010): "Diverging Opinions," mimeo., University of Pennsylvania.

C. D. Batson (1975): "Rational Processing or Rationalization? The Effect of Disconfirming Evidence on a Stated Religious Belief," *Journal of Personality and Social Psychology*, 32, 176-184.

C. Camerer and M. Weber (1992): "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty*, 5, 325–370.

J. Darley and P. Gross (1983): "A Hypothesis-Confirming Bias in Labeling Effects," *Journal of Personality and Social Psychology*, 44(1), 20–33.

A. K. Dixit and J. W. Weibull (2007): "Political Polarization," *Proceedings of the National Academy of Sciences*, 104(18), 7351–7356.

K. Eliaz, D. Ray and R. Razin (2006): "Choice Shifts in Groups: A Decision-Theoretic Basis," *American Economic Review*, 96(4), 1321–1332.

J. Esteban and D. Ray (2011): "Linking Conflict to Inequality and Polarization," *American Economic Review*, 101(4), 1345–1374.

I. Gilboa and M. Marinacci (2011): "Ambiguity and the Bayesian Paradigm" in *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress of the Econometric Society*, forthcoming.

F. Gul and W. Pesendorfer (2001): "Temptation and Self Control," *Econometrica*, 69, 1403–1435.

E. Hanany and P. Klibanoff (2009): "Updating Ambiguity Averse Preferences," *B.E. Journal of Theoretical Economics*, 9 (Advances), Article 37.

P. Kondor (2009): "The More We Know, The Less We Agree: Higher-Order Expectations, Public Announcement and Rational Inattention," mimeo., Central European University.

P. Klibanoff, M. Marinacci, S. Mukerji (2005): "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73, 1849–1892.

C. G. Lord, L. Ross, and M. R. Lepper (1979): “Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence,” *Journal of Personality and Social Psychology*, 37, 2098–2109.

M. Rabin and J. L. Schrag (1999): “First Impressions Matter: A Model of Confirmatory Bias,” *The Quarterly Journal of Economics*, 114(1), 37–82.

D. Schmeidler (1989): “Subjective Probability and Expected Utility without Additivity,” *Econometrica*, 57(3), 571–587.

M. Siniscalchi (2011): “Dynamic Choice under Ambiguity,” *Theoretical Economics*, 6(3), 379–421.

A. Wilson (2004): “Bounded Memory and Biases in Information Processing,” mimeo., University of Chicago.

A. Zimper and A. Ludwig (2009): “On attitude polarization under Bayesian learning with non-additive beliefs,” *Journal of Risk and Uncertainty*, 39, 181–212.