Inattentive Valuation and Belief Polarization*

Savitar Sundaresan[†] Sébastien Turban[‡]

March 12, 2014

Abstract

Why do people disagree? Using the recent literature on inattention, we construct a simple setup that allows a priori identical agents, being shown the same set of signals from the same objective state of the world, to permanently come to different conclusions in their posteriors. The inattentive framework allows for two effects, which we call confirmation and confidence effects. The former states that agents who are biased in a particular direction will arrange their attention to perceive signals that agree with their bias. The latter states that the larger the degree of the bias, the less attention an agent will pay, until the agent stops paying attention altogether. These effects in combination allow for divergence of opinions and permanent polarization, even on issues with an objective truth.

*For helpful comments, we thank participants at seminars at Caltech, Columbia University and NYU, Alessandra Casella, Thomas Palfrey, Ricardo Reis and Michael Woodford.

[†]Columbia University, svs2102@columbia.edu

[‡]Caltech, sturban@hss.caltech.edu

1 Introduction

Can two people come to permanently differ in their beliefs about an objective state of the world? Can one person be permanently wrong about that state? It would appear that the simple answer to these questions is no. With Bayesian updating and informative signals, beliefs should converge to the truth irrespectively of the starting point. However, in reality, we do observe persistently heterogeneous beliefs in a variety of fields on issues of common values or objective truth from politics, to sports, to economics and finance. How can these divergences occur? In this paper, we show how the simple assumption that agents are attention-constrained implies the possibility of them coming to the wrong belief permanently. When an individual can choose the level of attention to pay to different states, he elects to obtain more precision in the state he judges more likely, generating a *confirmation bias*. The confirmation bias implies that when an agent perceives the wrong state of the world in one period, she is more likely to make a similar perceptual mistake in the next period. In addition, we show that when an agent's beliefs are biased enough, she stops paying attention to the problem whatsoever and beliefs are stuck, which we label the *confidence effect*. Eventually, an individual led to wrongly biased beliefs due to mistaken signals, a situation which is made more likely with confirmation bias, can thus be permanently wrong.

The impact of prior beliefs on the pattern of information collection, and especially in the direction of updating, has been observed in various settings. Polls by Pew Research Center show how Republicans' and Democrats' positions on government policies related to civil liberties evolved in opposite directions before and after the election of Barack Obama. In 2006, 75% of Republicans and 37% of Democrats viewed the "NSA surveillance programs" as "Acceptable"; In 2013, those numbers were 52% and 64%. In 2002, 53% of Republicans and 41% of Democrats agreed that the "government [should] be able to monitor emails"; in 2013 those numbers were 45% and 53%.

A divergence in the evolution of beliefs due to different priors has been evinced more systematically. Darley and Gross (1983) ask some experimental subjects to judge a child's ability based only on some information about their socio-economic background. They give the same socio-economic information to another set of subjects who are then shown a video of the child completing an academic test, after which the subjects are also asked to rate the child's ability. The viewing of the video increases the reported child's ability when the socio-economic information reveals the child to come from a high-income neighbourhood, the opposite occurs with the other treatment.

In this paper, we suggest that there can be polarization of beliefs on common value issues through the effect of priors on the collection and the interpretation of information. A biased agent pays more attention to a state she believes to be more likely to be true. More formally, with two possible states of the world, she *chooses* the signal to be more informative in the state which is more likely under her prior. Interpretation errors will thus be more numerous in the state that goes against her beliefs. Therefore, initial interpretation errors which bias the agents' belief against the true state of the world will beget new interpretation errors.

Our framework, while intuitively simple, actually enables us to achieve divergence. It is quite straightforward to achieve permanent divergence, and the divergence often occurs very quickly. However, the convenience of having such strong divergence results, is twofold. The first is that it allows us to provide an micro foundation option for heterogeneous beliefs. The second is that it occupies a space on the opposite end of the spectrum from Blackwell. Having two results under two different, but related frameworks, allows for the possibility of hybrid models that could lead to a more sophisticated understanding of disagreement.

Blackwell and Dubins (1962) showed that increasing information among agents with absolute continuity in priors, must lead to convergence (or as they termed it, 'merging') of opinions. Any Bayesian agent can observe a imperfect signal and be temporarily distracted, but in the limit, the truth must be reached. This powerful result has formed the bedrock of a strand of economic literature. In particular, how can we reconcile such a result with the many protracted disagreements we see around us? Acemoglu, Chernozhukov, and Yildiz (2006) and Acemoglu, Chernozhukov, and Yildiz (2008) use the idea of noise in the inference process of agents to allow for belief divergence. Other more direct critiques of Blackwell and Dubins have come from such papers as Freedman (1963) and Miller and Sanchirico (1999), which make a technical argument against the assumption of absolute continuity in their proof. Our approach is somewhat different - in that we allow for there to be noise in the process of a signal's perception, but there is no incomplete information, nor does our result rely on agent's placing zero weight on one state in their prior (as for example, is the case in Berk (1966)).

Our focus on the impact of prior beliefs on information interpretation is motivated by a long list of empirical findings. First, there has been substantial research uncovering the confirmation bias, and we refer the reader to the review by Nickerson (1998), who defines it as "the unwitting selectivity in the acquisition and use of evidence".

More particularly, various papers have tried to analyze how agents collect and interpret information. A first notion which has been studied in political science is that of selective exposure: agents choose the sources of information and their potential bias. Lord, Ross, and Lepper (1979), and Baumeister and Newman (1994), provided evidence for agents suggesting that people pay less attention to information confirming their prior and evaluate "disconfirming evidence" more thoroughly. In finance, Huberman and Regev (2001) narrates how a front-page story in the New York Times about a cancer-curing drug reported months earlier, publicly, by Nature, led to a dramatic jump in the stock of a company linked to the drug.[1]. In political economy, Redlawsk (2008) shows that voters are looking for information about preferred candidates and avoid information about candidates they dislike. Other references for similar results include Lodge, Steenbergen, and Brau (1995), Taber and Lodge (2006).

A second strand has looked at information misinterpretation, or how two agents can look at the *same* information differently. In addition to Darley and Gross (1983) mentioned above, Taber and Lodge (2006) show in the political realm that arguments in favor of the prior are considered stronger than arguments against. In accounting, Hirshleifer and Teoh (2003) demonstrate how the presentation of information affects its interpretation, a literature to which Hirst and Hopkins (1998) contribute as well. Other examples can be

1. See also Lipe (1998)

found in, for instance, Fryer, Harms, and Jackson (2013).

The overwhelming evidence for the existence of a bias in the collection of information has led to a flourishing theoretical literature to understand its impact on the path of beliefs, to which our work is closely related. The necessity of a new framework to understand how beliefs are formed has been made even more salient by a recent paper by Baliga, Hanany, and Klibanoff (2013) where they show that divergence cannot occur in a Bayesian updating framework. Various models which discuss the potential impact of priors on information collection have been proposed. Both Suen (2004) and Cukierman and Tommasi (1998) show how it might be rational to choose an information source biased towards one's prior because only such a source can credibly reveal information against the prior, which would make it worth it to incur the cost of updating. Rabin and Schrag (1999) model confirmation bias as an exogenous probability of misinterpretation of incongruent signals. Koszegi and Rabin (2006) assume that the utility function is a increasing function of the difference between actual consumption and a reference point so that prior expectations play a role. Fryer, Harms, and Jackson (2013), in the paper most closely related to our work, assume that agents receive ambiguous signals which they interpret as signals in favor of their prior, and keep only this interpretation in memory.

Instead, we adopt the Inattentive Valuation framework laid out in Woodford (2012) where the precision of information is an endogenous choice of the individual. Agents are motivated by simply being correct, but the accuracy of their perception – the precision of their signals – is limited by an attention cost. This attention cost is a fixed marginal cost imposed on the amount of information conveyed by a choice of precision levels. This information quantity is measured using tools from information theory developed by Shannon (1948) for communication systems, and also discussed by Sims (2003).

Importantly, agents' utility function is simply minimizing squared error terms. They are fully rational and try to maximize utility subject to certain constraints or costs, and further they are Bayesian, and use Bayes' Rule to update their beliefs given their perceived signals.

The paper proceeds as follows. Section 2 will explain Woodford (2012) and show that the naive approach will be insufficient to resolve the question we're trying to answer. Section 3 sets up a static problem where there is uncertainty about a binary variable, and the precision on the observation of the actual value of the variable is affected by the prior over the two alternatives. Section 4 extends the analysis to a dynamic setting where an agent is allowed to perform multiple (even infinite) observations about the variable. Section 6 concludes.

2 Setup

Suppose that there are two possible states of the world, A and B. An agent has a prior that the state is A which we denote by $\pi = P(A)$. We will henceforth write $P(E)$ for the probability of a given event E. The agent's decision is to choose the level of attention to pay to each possible state, i.e. the probability of observing it correctly, in order to maximize accuracy. The "perceptions" of the states $r \in \{a, b\}$ are assimilable to a typical "noisy signal". Formally, the attention that the agent pays to the states can be characterized by $\alpha = P(r = a|x = A)$ for state A and $\beta = P(r = b|x = B)$ for state B.

In standard models of this form with private information, α and β are usually given exogenously and signals are assumed to be informative, i.e. $\alpha, \beta > \frac{1}{2}$. Under these setups, agents who update their beliefs using Bayes Rule, as ours do, will always come to learn the truth in the limit. The reasoning behind this is simple: if the true state is A it must be the case that in the limit, the proportion of "a" signals that the agents sees is exactly α; if the true state is B, this fraction will be $1-\beta$. $\alpha = 1-\beta$ would imply that the signals are uninformative about the state.

In our model, α and β are decision variables: the signals' precisions are endogenous. Typically, higher values of α and β will be more desirable for the agent. If information was not costly, he would choose $\alpha = \beta = 1$. If we assume, however, that attention has a price, this will usually not be the case. Assume that the quantity of information allowed by a choice (α, β) can be written as a function $\Gamma(\alpha, \beta; \gamma)$ which depends on the attention choices, and perhaps some additional parameter(s) γ. Intuitively, a standard condition on Γ would be that it is increasing in both α and β. Given this quantity, we can also quantify the attention cost as a function of $\Gamma(.)$. We will turn to information theory to find a cost function that is well-founded in the literature.

We quantify information by considering a model of the perceptual system as a communication system, and use the tools developed by Shannon (1948). The decision frameworks developed by Sims (2003) and Woodford (2012) employ Shannon's work in the economic world in quantifying information.

We quantify the information conveyed by a choice of attention in the same way, and assume that there exists a fixed marginal cost incurred by an agent in increasing this quantity.

The information conveyed by an attentional choice (α, β) is quantified by a measure of the difference in uncertainty about the state of the world before and after the perceived signal.

The amount of uncertainty contained in a distribution can be quantified by its entropy. The entropy measures the minimum expected number of bits necessary to encode the information contained in one realization from this distribution. Formally, for a pdf f, the entropy is given by $H(f) = E_f[\log_2(f)]$. The higher the value, the more bits required, and hence the less informative the distribution.

The prior π yields the entropy of the prior belief. The choice of attention along with the prior yields the entropy of the posterior belief. The expected amount of information conveyed by the attention choice, from the point of view of the agent, is the difference in entropy between the prior and the posterior, which is called the mutual information between the two random variables. The larger the gap between the distributions, the more informative the attention choice.

In our model, this mutual information is thus given by:[2]

$$I((\alpha, \beta); \pi) = -E_{P(x,r)}\left[\ln(P(x)) - \ln(P(x|r))\right]$$

We then have:

2. We use the natural logarithm here instead of the logarithm in base 2, but this has no influence on the results.

$$
\begin{aligned}
I((\alpha, \beta); \pi) = {} & \alpha\pi \ln(\alpha) + (1-\alpha)\pi \ln(1-\alpha) \\
& + (1-\beta)(1-\pi) \ln(1-\beta) + \beta(1-\pi) \ln(\beta) \\
& - [\alpha\pi + (1-\beta)(1-\pi)] \ln(\alpha\pi + (1-\beta)(1-\pi)) \\
& - [(1-\alpha)\pi + \beta(1-\pi)] \ln((1-\alpha)\pi + \beta(1-\pi))
\end{aligned} \tag{1}
$$

The agent's problem is to maximize accuracy, or minimize errors, under an attention cost based on $I((\alpha, \beta); \pi)$. This cost can either be a pure marginal cost, or a shadow cost of relaxing a fixed constraint on the overall attention level. Given the simple setting, the maximization of accuracy is equivalent to minimizing squared error (a more conventional objective) considering a situation where the utility of being correct is 1 and the utility of being incorrect is 0.

The framework of rational inattention proposed by Sims (2003) uses mutual information directly as its measure of the quantity of information contained in an attention choice. If attention has a fixed marginal cost $\theta > 0$, the cost of attention is given by $\Gamma(\alpha, \beta; \gamma) = \theta I((\alpha, \beta); \pi)$ so that $\gamma \equiv \pi$ and the rational inattention model can be written as

$$
\underset{\alpha,\beta}{\text{Max}}\; \pi\alpha + (1-\pi)\beta - \theta I((\alpha, \beta); \pi) \tag{2}
$$

There are various way of interpreting θ. One important interpretation is that it represents an opportunity cost in terms of attention paid to other attention problems: θ would increase in the number or complexity of other decision problems that the agent faces.

Rational inattention induces a constraint on the expected amount of information conveyed by an attention choice, meaning that it can be cheap to invest attention on a low-probability state. In particular, Woodford (2012) shows that rational inattention cannot explain the higher attention paid to more likely states in the experiment of Shaw and Shaw (1977).

Instead, we assume that the choice of attention is costly because of the *information capacity* it generates. The capacity of a communication system, say, an undersea cable, is the maximum rate of information that can be transmitted without error (Shannon (1948)): if a source signal has an entropy higher than the capacity, the transmission *must* result in some errors, if it is lower, it can be transmitted perfectly. In our case, the capacity is then a measure of the amount of information that a choice of attention *can* transmit, as opposed to what it is *expected to* transmit: the cost of attention will thus be on the potential information it can convey rather than the actual information it transmits.

Formally, the constraint will not depend on the discrepancy of the posterior to the actual prior π, but on the maximum possible discrepancy of the posterior to a *potential* prior.

Denote $\pi^*(\alpha,\beta) = \underset{\pi}{argmax}\ I((\alpha,\beta);\pi)$ and $I^*(\alpha,\beta)$ the mutual information evaluated at $\pi^*(\alpha,\beta)$. $I^*(.)$ is the channel capacity of the perceptual system defined by the attention choice (α,β). The "inattentive valuation" problem then differs from the rational inattention model by assuming that the attention cost is based on $I^*(.)$. Assuming a fixed marginal cost $\theta > 0$, the cost function becomes $\Gamma(\alpha,\beta;\gamma) = \theta \cdot I^*((\alpha,\beta))$ so that the problem becomes:

$$\underset{\alpha,\beta}{\text{Max}}\ \pi\alpha + (1-\pi)\beta - \theta I^{*}(\alpha,\beta) \tag{3}$$

Note here that $\gamma \equiv \emptyset$. This is crucial - as it is the key differentiator of our Inattentive Valuation framework from Rational Inattention. The entire cost function can be written in terms of the choice variables, and does not depend at all on the prior distribution of the agent.

The use of the capacity measure is attractive for various reasons. First, the capacity of a communication channel has a strong importance in the information theory literature, and measures the amount of information a system with a given error rate (in our case, $(1-\alpha,\, 1-\beta)$ is able to transmit accurately. The assumption in the analogy with the perceptual system is that attention allows the reception of a given quantity of information, so that it should not depend on π, a measure of the quantity of information *for the current problem* directly. More practically, the independence of the capacity on π implies that the drawback of the mutual information quantity disappear: paying attention in unlikely states is now expensive because the attention choice would allow more information to come in.

One final assumption is in order. In the rest of the paper, we will focus on a choice of attention (α, β) such that $\alpha + \beta \geq 1$. Remember this condition's importance from our discussion of exogenously determined errors. This assumption is without loss of generality given the symmetry of the problem. The assumption simply states that a perception "x" is indeed perceived as favoring state X, but as is standard in those types of models, a simple relabelling of the signals could imply the opposite. We will write

$\Delta = \{(\alpha, \beta) \in [0,1]^2 | \alpha + \beta \geq 1\}$ and[3] $\mathring{\Delta} = \Delta \setminus \{(0,1),(1,0)\}$

3 The myopic problem

Let us first consider a myopic agent - who cares only about maximizing his single period utility, thus solving Problem (3).

3.1 Properties of I^*

In order to properly understand the results of this section, we first analyze the properties of the attention function $I^*(.)$.

First, for any choice of attention (α, β) we can find the prior which maximizes the quantity of information provided by this choice, denoted $\pi^*(\alpha, \beta)$. Using the first order condition with respect to π in Equation (1), we obtain:[4]

$$\forall \alpha, \beta \in (0,1)^2, \pi^*(\alpha, \beta) = \frac{f(\alpha, \beta)(1-\beta) - \beta}{(1-\alpha-\beta)(1+f(\alpha,\beta))}$$

where $f(\alpha, \beta) = \left[\frac{\alpha^\alpha(1-\alpha)^{1-\alpha}}{\beta^\beta(1-\beta)^{1-\beta}}\right]^{\frac{1}{1-\alpha-\beta}}$.

Given $\pi^*(.)$, we can find simple formulas for $I^*(.)$. In particular, if one denotes $q^*(\alpha, \beta) = \pi^*(\alpha, \beta) \cdot \alpha + (1 - \pi^*(\alpha, \beta)) \cdot (1 - \beta)$ the average probability

3. $\mathring{\Delta}$ is defined because all the quantities discussed in the paper, including partial derivatives of $I^*(.)$, for instance, are well defined on the interior of Δ and can be extended by continuity on $\mathring{\Delta}$

4. We write down the explicit formulas in the calculations for $\alpha, \beta \in (0,1)^2$. Although most of the quantities here and below are not defined on the boundary of Δ, they can be extended by continuity. The expressions for all α, β are available upon request.

of observing signal a:

$$\begin{aligned} I^*(\alpha,\beta) =& (1-\beta)\ln(1-\beta) + \beta\ln(\beta) - (1-\beta)\ln(q^*) - \beta\ln(1-q^*) \\ I^*(\alpha,\beta) =& (1-\alpha)\ln(1-\alpha) + \alpha\ln(\alpha) - \alpha\ln(q^*) - (1-\alpha)\ln(1-q^*) \\ I^*(1,1) =& \ln(2) \\ I^*(\alpha,1-\alpha) =& 0 \end{aligned}$$

where the arguments of $q^*(.)$ are implicit.

The specific quantities shown above are intuitive. $I^*(1,1)$ is the information conveyed by full attention. In base 2, full attention would imply that one bit of information can be seen perfectly: $\log_2(2) = 1$. In the case of the natural logarithm, a digit can be observed perfectly, which corresponds to $\ln(2)$ bits. On the other hand, $I^*(\alpha, 1-\alpha)$ represents the information conveyed by a choice of attention which is completely uninformative: each signal has the same probability in each state.

Figure 1 shows some qualitative properties of I^* – darker shades represent higher values. As expected, I^* is higher as α and β increases. Figure 2 displays the isoquants for I^*, with the diagonal corresponding to attention choices which convey no information ($I^* = 0$), and the curves shifting upwards with higher attention levels. Several properties are already apparent, notably, the concavity of the isoquants.

Proposition 1 (Convexity of information capacity)**.** *I^* is infinitely differentiable on $\mathring{\Delta}$ and is strictly convex in* (α, β)

As an agent wishes to increase precision in one state, she must sacrifice precision in the other to maintain the same level of informativeness, and increasingly so as the prior changes. Namely, the marginal rate of substitution of precision on state A for precision on state B is increasing: as the precision on one state becomes weaker, it becomes more "valuable" or "informative" at the margin, relatively to the other state. Proposition 1 shows that this pattern is the result of I^* being a convex function of the information structure (α, β).

Figure 2 also shows that if an agent tries to become certain about one state, the marginal increase in the information quantity increases to become infinite – so that with a fixed marginal cost, certainty would be infinitely costly. This means that for any fixed marginal cost, the agent will never choose fully informative signals. The marginal rate of substitution described above becomes unboundedly large at extreme priors. Proposition 2 shows formally that when α is close to 1, the marginal increase in $I^*()$ from a marginal increase in α is infinite and the marginal increase in $I^*()$ from a marginal increase in β is bounded away from 0.

In addition, the proposition shows that marginally increasing precision from a state where attention is completely uninformative bears no cost *as long as both precisions can be increased*, i.e. $(\alpha, \beta) \in \mathring{\Delta}$. This has important implications: it implies that there are only two possibilities: either an agent pays attention to the problem by setting α, β such that $I^*(.)$ is strictly positive, or one of the state's precision is set to 1 and the other to 0.

Proposition 2 (Marginal informativeness on the boundaries of $\mathring{\Delta}$)**.**

$\forall (\alpha, \beta) \in \mathring{\Delta}$,

- $\frac{\partial I^*(1,\beta)}{\partial \alpha} = \infty$

- $\frac{\partial I^*(1,\beta)}{\partial \beta} \in (0, \infty)$
- $\frac{\partial I^*(\alpha,\beta)}{\partial \alpha} = 0 \Leftrightarrow \alpha + \beta = 1$

The properties of $I^*(.)$ have important implications for the pattern of attention choices which result from solving Problem (3). First, the assumption that the cost of attention is independent of the prior generates a confirmation bias in that more attention is paid to the most likely state, given its higher marginal return (weighted by the prior) in terms of accuracy. As a consequence, attention is allocated evenly when the prior is unbiased ($\pi = .5$) and is increasingly unbalanced as π is further from .5. Because $I^*(.)$ is convex (Proposition 1) and because the marginal cost of attention is fixed, the confirmation bias implies that the total amount of attention allocated to the problem, endogenously given by $I^*(\alpha, \beta)$, decreases with the prior. Finally, because the marginal cost of certainty on a state is infinite (Proposition 2, there will be a level of bias after which it is optimal to pay no attention whatsoever to the problem at stake. Those results are shown formally below.

3.2 Attention choices

We first show in Proposition 3 that the choice of attention is strictly monotonous in the prior when some attention is actually paid (i.e. $I^*(.) > 0$) and the solution is interior. Because the optimal attention choices are continuous functions of the prior and that the solution is trivially interior at $\pi = .5$, the proposition shows that as the prior is more biased, more attention is invested in the more likely state . We label this effect the *confirmation bias.* In the following, we denote $\alpha(\theta; \pi)$ and $\beta(\theta; \pi)$ the solutions of Problem (3), and

$C(\theta;\pi) = I^*(\alpha(\theta;\pi), \beta(\theta;\pi))$

Proposition 3 (Confirmation bias). *$\forall\theta > 0$, $\forall\pi | (\alpha(\theta;\pi), \beta(\theta;\pi)) \in \mathring{\Delta}$, $\frac{\partial\alpha(\theta;\pi)}{\partial\pi} > 0$, $\frac{\partial\beta(\theta;\pi)}{\partial\pi} < 0$.*

$\forall\theta > 0$, $\forall\pi | (\alpha(\theta;\pi), \beta(\theta;\pi)) \in \Delta$, $\frac{\partial\alpha(\theta;\pi)}{\partial\pi} \geq 0$, $\frac{\partial\beta(\theta;\pi)}{\partial\pi} \leq 0$.

The condition that $(\alpha(\theta;\pi), \beta(\theta;\pi)) \in \mathring{\Delta}$, as we will see in the next propositions, is valid for priors π which are not "extreme", i.e. far away from $\frac{1}{2}$.

The confirmation bias has one immediate, important consequence: in a myopic (as opposed to forward-looking), dynamic model, an agent biased in favor of the wrong state is more likely to perceive the wrong state afterwards. Perceptual mistakes beget perceptual mistakes; and beliefs depend not only on the content of the perceptions but also on their order, as opposed to a standard Bayesian model.

We then show in Proposition 4 that this confirmation bias and the unequal attention to the two states imply that as an agent becomes more sure of her opinion, she lowers the overall attention paid to the problem. Eventually, she pays no attention to the problem at all for extreme priors: in particular, $I^*(\alpha, \beta)$ converges to 0 when π approaches unity, and attention level approaches certainty. We label this effect the *confidence effect*.

Proposition 4 (Confidence effect). *$C(\theta;\pi) \underset{\pi\to 1}{\to} 0$ and $C(\theta;\pi) \underset{\pi\to 0}{\to} 0$. Further, $\alpha(\theta;\pi) \underset{\pi\to 1}{\to} 1$ and $\beta(\theta;\pi) \underset{\pi\to 1}{\to} 0$; $\alpha(\theta;\pi) \underset{\pi\to 0}{\to} 0$ and $\beta(\theta;\pi) \underset{\pi\to 0}{\to} 1$.*

The combinations of the two previous results have important implications. One can easily see that when $\pi = .5$, the choice of precision in each state is equal, and signals are informative: $\forall\theta > 0, \alpha(\theta;.5) > \frac{1}{2}, \beta(\theta;.5) > \frac{1}{2}$.

However, because the solutions are continuous there exist a bias after which perceptions in *both* states are more likely to favor the same state. At the limit, perceptions in *both* states always favor the same state.

We show in Proposition 5, that this shutdown occurs before the agent is certain about the state, i.e. the optimal level of $I^*(\alpha, \beta)$ is nil for large priors: the agent chooses an information structure which conveys no information at all about the state by focusing their attention fully on the most likely state.

Proposition 5 (Information shutdown at extreme priors). *$\forall \theta, \exists \overline{\pi}(\theta) | \forall \pi > \overline{\pi}(\theta)$, $\beta(\theta; \pi) < 0.5$ and $\alpha(\theta; \pi) > 0.5$. $\forall \theta, \exists \bar{\bar{\pi}}(\theta) | \forall \pi > \bar{\bar{\pi}}(\theta)$, $\beta(\theta; \pi) = 0$ and $\alpha(\theta; \pi) = 1$*

The confidence effect reflected in a decrease in the amount of attention $I^*(.)$ for a given prior can be seen in Figure 3, where we set $\theta = 1$: the amount of information conveyed by the choice of attention falls increasingly quickly as the prior moves away from $\frac{1}{2}$ and hits 0 before $\pi = 1$. We show the corresponding choices of attention Figure 4: the attention paid to state A increases as the belief increases away from $\frac{1}{2}$, so that for a high level of the prior the agent only pays attention to state A.

The fact that an agent can stop thinking about the problem altogether when the prior is strictly below 1 makes it intuitively possible that at the end of time in a repeated game, her prior will be wrong in the sense that she will be stuck at a belief putting more weight on the wrong state. This would imply that a permanent polarization of beliefs is possible.

The next question is whether a forward-looking agent, who takes into account the consequences of a contemporaneous choice of biased attention on future mistakes, can diverge in that way.

4 The dynamic problem

In the previous section we have been comparing the attention choices of an agent when he observes a single signal, which can also be interpreted in a dynamic setting as the choices made by a myopic agent. We have stressed, in particular, the impact of the prior, given exogenously, on attention choices. This leads to another question: how are those priors formed? What is the impact of limited attention on the formation of the prior? In this section, we analyze this question by considering rational agents with the *same* initial belief and analyzing how the cost of attention affects their choices on the attention paid to a stream of signal. Because priors affect contemporaneous attention choices through the confirmation bias, for instance, a forward-looking agent must take into account the fact that a perceptual mistake made in one period makes future mistakes more likely. Because of the confidence effect, the forward-looking agent must also consider that the change in her beliefs' bias will affect the total amount of attention paid to future problems. The main question of interest, eventually, is whether the last result of the previous section holds in this dynamic case: can an agent's belief be permanently stuck against the actual state of the world? More generally, can two agents with the same initial prior diverge permanently in their beliefs?

4.1 Statement of the Dynamic Problem

As before, we will assume that there is a state of the world, either A or B. We assume that it remains constant over time, and that it sends a stream of unambiguous[5] signals to the agents in question. For example, if the state were A (as we will assume wlog) the signals would be “a”,“a”,“a”, ... The probabilities on the perceptions of those signals are dictated by their choices of attention at each period, so that they would typically change over time. The marginal cost of attention, θ, is fixed over time.

Consider a model with an infinite number of periods, in discrete time starting at date 0: each period is denoted $t \in \mathbb{N}$. At each period, the agent's instantaneous payoff is the benefit to accuracy net of the information cost, equivalent to the one-period objective of Problem (3). Future periods are discounted by $\delta < 1$. The decision maker chooses the attention level for every period to maximize the discounted expected value of those instantaneous payoffs. Formally, the dynamic problem can be written given an initial prior belief π_0 as

$$V(\pi_0) = \underset{\{\alpha_t,\beta_t\}_{t\in\mathbb{N}}}{\text{Max}} \mathbb{E}\left[\sum_{t=0}^{\infty} \left(u(x,r_t) - \theta I^*(\alpha_t,\beta_t)\right) |\pi_0\right] \tag{4}$$

where x is the fixed state of the world drawn at time 0, r_t is the perception made at time t, and

$$u(x,r_t) = \begin{cases} 1 & \text{if } r_t = x \\ 0 & \text{if } r_t \neq x \end{cases}$$

This particular structure of the new problem requires some explanation. In the static version of the game, the agent selected accuracies α and β in order to maximize accuracy subject to a constraint. The implicit assumption of that structure was that the agent wanted to maximize accuracy because

5. This assumption is relaxed in the next section.

there was a second decision, selecting whether the state was A or B after observing the signal, and that the payoffs were such that he would receive 1 for correct identification, and 0 for incorrect identification. The minimization of the squared error led to the objective function as stated. In the dynamic case, we cannot have sequential assessment of the agent's decisions. Under the assumption of perfectly informative signals (no noise in the data generating process), any assessment would be fully informative of the state. Our assumption here therefore, is that after each signal the agent takes an action (selecting A or B), which will be evaluated after all signals have been viewed. We believe that this is a better way to think about protracted disagreements. For example, consider the case of Global Warming - as the environment sends us signals which we try to attend to and understand, we are actively taking decisions: to limit carbon emissions, to try to find alternative energy sources etc. The value of these decisions will not be apparent to us for a long time. Another example is the determination of the long run riskiness of a particular market. As research is conducted, we may buy or sell assets as we refine our opinions, but our actions will not be assessed for some time. Matters in which our actions can be quickly evaluated, are much less likely to result in protracted or permanent disagreement, as the evaluation should give concrete evidence to support one opinion over another.

Problem 4 can be written in the form of a Bellman equation, where the state variable in question is the prior π_t with which an agent enters period t. This yields:

$$
\begin{aligned}
V(\pi_t) = \max_{\alpha_t, \beta_t} \pi_t \alpha_t + (1 - \pi_t)\beta_t - \theta I^*(\alpha_t, \beta_t) \\
+ \delta \Big[(\alpha_t \pi_t + (1 - \beta_t)(1 - \pi_t)) V(\pi_t^a) + ((1 - \alpha_t)\pi_t \\
+ \beta_t (1 - \pi_t)) V(\pi_t^b) \Big]
\end{aligned} \tag{5}
$$

where the priors at the start of period $t+1$ after perceiving signal "a" and "b" respectively, π_t^a and π_t^b are given by the posteriors at the end of period t:

$$
\pi_t^a = \frac{\alpha_t \pi_t}{\alpha_t \pi_t + (1 - \beta_t)(1 - \pi_t)}
$$
$$
\pi_t^b = \frac{(1 - \alpha_t)\pi_t}{(1 - \alpha_t)\pi_t + \beta_t(1 - \pi_t)}
$$

It is simple to show that the Blackwell conditions for the contraction defined by the Bellman equation are satisfied, so that the value function exists and is the fixed point of that contraction. We denote $\alpha_d(\theta; \pi)$ and $\beta_d(\theta; \pi)$ the attention choices made by the forward-looking agent with initial prior π. We omit the d subscript when there is no confusion possible with the static solutions.

4.2 Will Agents Disagree Forever?

Permanent divergence requires that an agent eventually stops paying attention to the problem. If that is not the case, the share of perceptions "a" and "b" will differ between states and all individuals will converge to the correct belief ($\pi = 1$ if the state is A, for instance). We show in Proposition 6 that when beliefs become extreme, such an information shutdown occurs, as we found in Proposition 5 in the myopic problem. For values of π sufficiently far from 0.5, the agent chooses signal precisions which prevent him from distinguishing between the two states at all: perceptions are uninformative. Importantly, the shutdown point is identical to the one found in the myopic case.

Proposition 6. $\forall\theta > 0, \exists\tilde{\pi}(\theta) > \frac{1}{2} | \left|\pi - \frac{1}{2}\right| > \left|\tilde{\pi}(\theta) - \frac{1}{2}\right| \Rightarrow \alpha_d(\theta;\pi) + \beta_d(\theta;\pi) = 1$. *Moreover,* $\forall\theta > 0, \tilde{\pi}(\theta) = \bar{\bar{\pi}}(\theta)$

Proposition 6 implies that if any agent starts a period with a prior that lies above $\tilde{\pi}$ or below $1 - \tilde{\pi}$ where $\tilde{\pi} \in (0.5, 1)$, then he will not pay any attention to signals from that or any subsequent period. Importantly here, $\tilde{\pi}(\theta) = \bar{\bar{\pi}}(\theta)$, which is to say that the point of certainty at which the myopic and forward-looking agents choose to stop paying attention to signals is the same. As we will see later, there is strong evidence to suggest that the amount of attention paid in the forward-looking case is weakly larger than the myopic. The reason for this change is that attention in the forward looking case affects not only signals at present, but the future, and the agent internalizes this future effect. One could think of this as a multiplier greater than one, that maps myopic attention to forward looking attention. However, no matter what the multiplier, under the priors for which no attention is being paid in the static case, no attention is paid either in the dynamic case.

Extreme beliefs are thus persistent: $\forall\pi' \in [0, 1 - \tilde{\pi}) \cup (\tilde{\pi}, 1]$, $\pi_t = \pi' \Rightarrow \forall u > t, \pi_u = \pi'$. Even mild disagreement can then persist forever, when θ is large: clearly, $\frac{\partial\tilde{\pi}(\theta)}{\partial\theta} < 0$ and $\tilde{\pi}(\theta) \underset{\theta\to\infty}{\to} \frac{1}{2}$. The question is then whether individuals with the same initial belief $\pi_0 \in (1 - \tilde{\pi}(\theta), \tilde{\pi}(\theta))$ can end up in either of these regions with positive probability.

In order to answer this, we need to consider the pattern of attention choices in the forward looking problem. We showed that the shutdown points are identical in the myopic and the dynamic problems. However, this does not mean that the amount of attention paid in the static and dynamic cases are the same for all values of the prior. Consider in particular the problem when $\pi_0 = 0.5$. We show in Proposition 7 that in that case, the attention devoted to both states by the forward-looking agent is higher than in the one-shot problem.

Proposition 7. $\forall \theta > 0$, $\alpha(\theta; .5) < \alpha_d(\theta; .5)$ *and* $\beta(\theta; .5) < \beta_d(\theta; .5)$.

The intuition for this result is simple. Agents in the forward-looking model not only understand the direct effect that their attention has on accuracy (as is the case in the static model) but also the indirect effect that their perceptions will have on future decisions. As such, they will choose to pay more attention based on this forward-looking effect than they would without it.

We can now show that permanent divergence of agents with the same initial prior can occur. For this purpose, we show that when the prior is uninformative, the decision makers consider only the first period signal and stop collecting any information after that

Proposition 8. *If* $\pi_t = 0.5$, *then* $I^*(\alpha_{t+1}, \beta_{t+1}) = 0$. *The agent will observe one signal, update, and shutdown.*

This is a strong result, and one that depends, in large part, on the strength of the signals. That is to say, the signals from each potential state of the world are perfect and unadulterated. Additionally, at the point $\pi = 0.5$, I^* is at its highest point - the most attention is being paid. Therefore, perceptions will be very accurate. Very accurate perceptions of perfectly accurate signals, mean that the perceptions are therefore additionally, very informative. As a result, large updates in the agents beliefs will take place. Large enough, in fact, to make them sure of their opinion. The agent prefers to front load the attentional cost, in order to forego paying it in any subsequent period.

Consider Figure 6. For different values of θ, agents will update to differing degrees away from 0.5 and then stop acquiring new information. Although we are unable to provide analytic results for other starting values of π, we can observe some simulations to get a sense of the dynamics. In Figure 5, we see agents starting at values of $\pi \neq 0.5$. It appears that agents who receive confirmatory signals (that is, agents who start with a prior $\pi > 0.5$ and see an "a" signal, or agents who start with a prior $\pi < 0.5$ and see a "b" signal), update once and then shut down. However, agents who receive contradictory signals (agents who start with a prior $\pi > 0.5$ and see a "b" signal, or agents who start with a prior $\pi < 0.5$ and see an "a" signal), will update and then continue to look for signals.

5 Extensions and Robustness

We consider here potential extensions and applications of the model to show the potential uses of the inattentive valuation framework and how the results developed in the two previous sections could possibly be extended quickly to more general settings and related questions.

First, we discuss the assumption that all the noise in the perceptions of the states in our model are endogenous. Second, we adapt the problem to cases where the states are not weighted identically, i.e. errors in the two states generate different disutility. Third, we discover what we can say about the intensive and extensive margins of disagreement. Finally, we summarize the properties of the information quantity which were sufficient to generate the results described in the previous sections.

5.1 Adding exogenous noise

5.1.1 The static case

In the framework described in previous sections, we assumed that the noise in the signals perceived by the agent were purely endogenous, the result of a choice of attention due to costly information acquisition. The assumption means that there exists an objective state of the world which sends perfectly informative signals. It is beyond the scope of this paper to discuss what we usually understand by the term "signal", but it seems intuitive that the information generated by the objective state of the word can actually be distorted even before their perception by an agent. This distortion will not be a choice, either conscious or unconscious, of the agent.

How can we understand this potential for an exogenous distortion of the signals? Formally, assume that in state A, the signal sent is a with probability $1 > q_a > \frac{1}{2}$ and b with probability $1 - q_a$, and define q_b similarly in state B.

The distinction between the exogenous and the endogenous distortion is important. An agent can only make an endogenous attention choice towards the “signals” derived from the state of the world and its inherent exogenous noise. Although in the original framework, the state and the exogenous signal were identical because we assumed no exogenous noise, we need to make the distinction here. Therefore, let us introduce the following notations. The state of the world is $x \in \{A, B\}$. State A sends signal $s = a$ with probability q_a, State B sends signal $s = b$ with probability q_b. The agent can only pay attention to those signals, so that the attention choice has to be rewritten as

$$P(r = a|s = a) = \alpha$$

$$P(r = b|s = b) = \beta$$

The latter notation requires a short discussion. As we alluded to in the setup of the original problem, the cost of attention is based on the reduction in uncertainty of an input generated by a given attention choice o this input. In the new frame with exogenous noise, the input is the imperfect signal. Given π, q_a and q_b, this input has a Bernoulli distribution with parameter $p = \pi q_a + (1-\pi)(1-q_b)$, where p is the probability of $s = a$. The expected quantity of information conveyed by the choice of attention, defined by the mutual information between the posterior and the prior distribution of the exogenous signal, is thus

$$I((\alpha, \beta); p) = -E_{P(s,r)}\left[\ln(P(s)) - \ln(P(s|r))\right]$$

The channel capacity can then be reformulated as

$$I^*_{exog}(\alpha, \beta) = \max_{p \in \mathcal{P}} I((\alpha, \beta); p)$$

Where $\mathcal{P} = [1 - q_b, q_a]$ is the image of $[0, 1]$ via the increasing function $p(.)$:

$$p : [0,1] \rightarrow [0,1]$$
$$\pi \rightarrow \pi q_a + (1-\pi)(1-q_b)$$

Therefore, if $\pi^*(\alpha, \beta) \in \mathcal{P}$ then

$$I^*_{exog}(\alpha, \beta) = I^*(\alpha, \beta)$$

Because $\pi^*(\alpha, \beta)$ is bounded away from 0 and 1[6], we have that the set $\mathcal{Q} = \{(q_a, q_b) | \forall(\alpha, \beta) \in (0, 1) I^*_{exog}(\alpha, \beta) = I^*(\alpha, \beta)\} \neq \emptyset$. Although we have not been able to prove it, we conjecture that the bounds on $\pi^*(\alpha, \beta)$ can be computed and that $\left|\pi^*(\alpha, \beta) - \frac{1}{2}\right| < \left|\frac{1}{2} - exp(-1)\right|$[7]. In particular, $\pi^*(\alpha, \beta)$ is maximal and equal to $1 - exp(-1)$ at $\alpha = 1, \beta \rightarrow 0$. One can then characterize $\mathcal{Q}$ explicitly: $\mathcal{Q} = \{(q_a, q_b) | min(q_a, q_b) > exp(-1)\}$.

We thus know that provided the exogenous signals are relatively informative ($(q_a, q_b) \in \mathcal{Q}$) the choice of attention in the new framework is the solution to the problem

$$\underset{\alpha,\beta}{\text{Max}}\ \pi\left[q_a\alpha + (1-q_a)(1-\beta)\right] + (1-\pi)\left[q_b\beta + (1-q_b)(1-\alpha)\right] - \theta I^*(\alpha, \beta) \tag{6}$$

6. For any attention choices, the mutual information would be 0 in these cases, while it is strictly positive at $\pi = \frac{1}{2}$

7. We can show that $\pi^*(1, \beta)$ is decreasing in β, but it remains to be shown that the maximum of $\pi^*(1, \beta)$ is indeed achieved at the border.

with $q_a = q_b = 1$ yielding the original problem. With simple manipulations, the problem is equivalent to:

$$\underset{\alpha,\beta}{\text{Max}}\ \alpha\left[q_a\pi - (1-q_b)(1-\pi)\right] + \beta\left[q_b(1-\pi) - \pi(1-q_a)\right] - \theta I^*(\alpha,\beta)$$

or writing $\lambda(q_a, q_b, \pi) = \frac{q_a\pi-(1-q_b)(1-\pi)}{(2q_a-1)\pi+(2q_b-1)(1-\pi)}$

$$\underset{\alpha,\beta}{\text{Max}}\ \alpha\lambda(q_a, q_b, \pi) + \beta(1-\lambda(q_a, q_b, \pi)) - \frac{\theta}{2q_a\pi + 2q_b(1-\pi) - 1} I^*(\alpha,\beta)$$

This reformulation means that we can directly find the attention choices as a function of the prior and the exogenous precisions, as we describe in Proposition 9. If we note the solutions to Problem (6) $\alpha_e(\theta, q_a, q_b; \pi), \beta_e(\theta, q_a, q_b; \pi)$, we have:

Proposition 9 (Attention Choices with Exogenous Noise)**.**

$$\alpha_e(\theta, q_a, q_b; \pi) = \begin{cases} \alpha\left(\frac{\theta}{2q_a\pi+2q_b(1-\pi)-1}; \lambda(q_a, q_b, \pi)\right) & \text{if } \lambda(q_a, q_b, \pi) \in [0,1] \\ 0 & \text{if } \lambda(q_a, q_b, \pi) < 0 \\ 1 & \text{if } \lambda(q_a, q_b, \pi) > 0 \end{cases}$$

and

$$\beta_e(\theta, q_a, q_b; \pi) = \begin{cases} \beta\left(\frac{\theta}{2q_a\pi+2q_b(1-\pi)-1}; \lambda(q_a, q_b, \pi))\right) & \text{if } \lambda(q_a, q_b, \pi) \in [0,1] \\ 1 & \text{if } \lambda(q_a, q_b, \pi) < 0 \\ 0 & \text{if } \lambda(q_a, q_b, \pi) > 0 \end{cases}$$

Given the similarities, most qualitative properties of the solutions are preserved. First, note that given $q_a, q_b > \frac{1}{2}$, $\frac{\partial \lambda(q_a,q_b,\pi)}{\partial \pi} > 0$. Second, the new marginal cost $\tilde{\theta}(\pi) = \frac{\theta}{2q_a\pi + 2q_b(1-\pi) - 1}$ depends on π, and $\tilde{\theta}'(\pi)$ is of the sign of $q_b - q_a$. Finally, $\tilde{\theta}(\pi)$ is bounded. The dependence of the effective marginal cost on π makes the confirmation bias results less clear-cut. However, given that it is monotonous, it is clear that $q_a > q_b \Rightarrow \frac{\partial \alpha_e(\theta,q_a,q_b;\pi)}{\partial \pi} > 0$, since as π increases, the decrease of the marginal cost complements the confirmation bias; a similar result is obtained on $\beta_e(.)$ when $q_b > q_a$. Finally, the existence of a shutdown point is still valid since $\tilde{\theta}(.)$ is bounded below.

In the simple case $q_a = q_b = q$, $\tilde{\theta}(\pi) \equiv \bar{\tilde{\theta}}$: the effective marginal cost is constant and we can apply all the results we found for Problem (3), including the confirmation bias. The confidence effect is even stronger. Indeed, the weight on α is positive if and only if $\frac{1-q}{q} \leq \frac{\pi}{1-\pi}$ and likewise, the weight on β is positive if and only if $\frac{q}{1-q} \geq \frac{\pi}{1-\pi}$. Those two conditions are equivalent to $1 - q \leq \pi$ and $q \geq \pi$. This is intuitive, as under these conditions the exogenous noises are larger than the uncertainty in the prior.

If the first condition is not satisfied, i.e. for low π, the weight on α is negative and α is trivially set to 0. Likewise, if the second condition is not satisfied, i.e. for high π, the weight on β is negative and β is set to 0.

5.1.2 The dynamic case

One of the results of the initial framework with no exogenous noise that is at first blush perplexing is that at the point of maximal uncertainty ($\pi = 0.5$), only one signal is required to make the agent shut off. It appears that the reason for this result is the strength of the signals from the state. If $\theta = 0$, an agent would require only one signal to have perfect information. On the other hand, in the exogenously noisy case, even when $\theta = 0$, the agent would require infinitely many signals to achieve perfect certainty on the value of the state. Hence, the updating process is slowed.

As we see in simulations of the case when $\theta > 0$, this carries over to our setup. Although there are still beliefs for which agents shut down, they approach those beliefs more slowly, as each signal carries two sources of noise (perceptual and exogenous), and thus less information.

Consider figure 8. The three panels show, in order, the evolution of 50 agents' priors. The agents are ex-ante identical, and observe 20 identical signals in these simulations. The cost of attention in all cases is given by $\theta = 10$. These signals are drawn from state A, which shows signal "a" with probability q - known to the agents. Here again, agents eventually shut down, after becoming sufficiently sure of their opinions, but this certainty takes time. In fact, some agents in the sample did not converge at all after receiving several conflicting signals.

In conclusion, in the case where $q_a = q_b = 1$, we have seen in previous sections that shutdown occurs (usually) after one signal. Here, we found suggestive evidence that when $1 > q_a = q_b > 0.5$, shutdown can take many periods. In the other limit case, where $q_a = q_b = 0.5$, all signals are perfectly uninformative, so no amount of attention will be able to distinguish them. Therefore, agents will set $\alpha + \beta = 1$, and will never move from their original priors.

5.2 Different weighing of states

We now consider an extension of the problem of inattentive valuation to the case where the accuracy in each state is not valued identically. The typical example is the one of Feddersen and Pesendorfer (1998) where a jury makes a decision on whether an individual is guilty or innocent. The jurors might not put the same weight on one error against the other. More generally, we can think of the traditional arguments underlying a different weighting for errors of type I and errors of type II.

We can modify the model in the simplest way by adding a parameter $\gamma > 0$ measuring the weight put on being correct in state B relative to state A. The problem is then:

$$\operatorname*{Max}_{\alpha,\beta} \pi\alpha + \gamma(1-\pi)\beta - \theta I^{*}(\alpha,\beta)$$

This setting is equivalent to assuming that the utility of being correct in state B is γ and the utility of being correct in state A is 1, while the utility of being incorrect in each state is 0. As in the case of exogenous noise, we can rewrite the problem in the shape of Problem (3)

$$\operatorname*{Max}_{\alpha,\beta} \frac{\pi}{\pi+\gamma(1-\pi)}\alpha + \frac{\gamma(1-\pi)}{\pi+\gamma(1-\pi)}\beta - \frac{\theta}{\pi+\gamma(1-\pi)} I^{*}(\alpha,\beta)$$

As with exogenous noise, we can thus find the solutions $\alpha_{rw}(\theta,\gamma;\pi), \beta_{rw}(\theta,\gamma;\pi)$ from the solutions to Problem (3). Likewise, the confirmation bias is weaker because of the dependence of the effective marginal cost on the prior, but because it is monotonous in π for a given γ, we have that for instance

$\gamma < 1 \Rightarrow \frac{\partial \alpha_{rw}(\theta,\gamma;\pi)}{\partial \pi} > 0$. Finally, the result that no attention is paid when beliefs are extreme, the confidence effect, is also unchanged given that the cost is bounded. The main difference will be that the shutdown of information occurs asymmetrically: if being accurate in state B is more important ($\gamma > 1$), the bias in favor of state B which would lead an agent to stop paying attention is higher than the required bias in favor of state A.

5.3 Size of divergent cohorts

We showed in section 4 that permanent divergence can occur between two individuals with identical prior, based on their initial perceptions. Can we quantify the magnitude of this divergence? There are two ways to characterize this disagreement. The first is to talk about the size of the groups that disagree: is the minority a small percentage or almost half? The second is to talk about the degree of disagreement: are the groups relatively close in their beliefs, or is there a large gulf between them? We will term the former the extensive margin of disagreement, and the latter the intensive margin.

Consider some $0 < \theta < \infty$, and $\pi_0 = \frac{1}{2}$. We know that $\alpha_d(\theta, \pi_0) = \beta_d(\theta, \pi_0) = \zeta(\theta) \in \left(\frac{1}{2}, 1\right)$. Using the results developed in the proofs of Proposition 8, we know that for all $t > 1$ a share ζ of individuals will have a posterior $\pi_1 = \pi_t = \zeta$ and that a share $(1-\zeta)$ will have a posterior $\pi_1 = \pi_t = 1-\zeta$. Moreover, we know that $\zeta'(\theta) < 0$. Therefore, as the cost of attention increases, the intensive margin of disagreement decreases, but the extensive margin increases.

These results are displayed for different values of θ in Figure 7. In the figure, we show the final distribution of agents when a unit mass starts with a prior of $\pi_0 = .5$. The horizontal axis represents the final beliefs, while the height of the bars show the percentage of the unit mass ending with this belief.

Convergence Under Fixed Capacity It is a relevant question to think about whether or not the setup of attention in this problem is important to the result. Much of the literature on attention assumes, instead of a fixed cost of attention, a fixed capacity, or stock of attention. If we instead consider a problem of the following form:

$$\max_{\alpha,\beta} \alpha\pi + \beta(1-\pi)$$
$$s.t. I^*(\alpha, \beta) \leq C$$

where C is exogenously given, can the same results as above hold? The following proposition will show that they do not.

Proposition 10. *As $t \to \infty$, $\pi_t \to 1$ for a fixed C. That is, any agent will always converge to the truth in the limit.*

In both Rational Inattention and Inattentive Valuation, the intuition for this result is relatively straightforward. Under Rational Inattention, there is only one unique fixed point - and that is $\pi = 1$. Therefore, no divergence to any other point is possible. Under Inattentive Valuation, there are actually two fixed points: $\pi = 1$ and $\pi = 0$. However, intuitively, agents under fixed capacity are prevented from paying less attention as their certainty increases. Therefore, as an agent becomes convinced of the wrong belief, he must still observe signals with the same net level of accuracy. Therefore, if the agent sees a string of 'b's at first, a subsequent 'a' will actually be quite informative as α and β are necessarily bounded away from 0. Therefore, no matter how close the agent gets to $\pi = 0$, observing an 'a'' at that point would move him quite far in the opposite direction.

5.4 Attention cost function

We showed in the previous sections that our model can be used to solve more general problems by manipulating the reward for being correct. The final question is to understand the influence of the choice of the channel capacity as our measure of attention for the results described in the paper. Indeed, the use of information theory, mutual information, and channel capacity to quantify information, and the inclusion of a fixed marginal cost for this quantity, can be seen as specific cases of a general model with costly attention. Therefore, it is important to understand which conditions on $I^*(.)$ were necessary to produce the results described in Section 3 and 4. Here, we described conditions on the information quantity which are sufficient to generate the confirmation bias and the confidence effect in the static case.

We consider here classes of functions which do not depend on the prior. We explained in the discussion of rational inattention that one drawback in using the actual prior is that unlikely states are overweighted by definition. In this case, more attention can be invested towards those states, which would negate a confirmation bias.

Several other restrictions are more intuitive. In particular, we restrict our discussion to functions which are symmetric in the states – the states are "anonymous". We also assume that the information quantity is strictly increasing in its arguments. To prevent some potential problems linked to the relabelling of the signals, we thus restrict the arguments' space to the upper quadrant of the unit square – one can extend the function to the lower quadrant by symmetry.

Therefore, remember that $\Delta = \{(x, y) \in [0, 1] | x + y \geq 1\}$ and denote $\mathcal{F} = \{f : \Delta \to \mathbb{R}^+ | f(x, y) = f(y, x), \forall y \in [0, 1] x > x' \geq 1 - y \Rightarrow f(x, y) > f(x', y)\}$

We want to find the properties of a function $f \in \mathcal{F}$ yielding the results from the previous section when the agent's problem is

$$\underset{\alpha,\beta}{\text{Max}}\ \pi\alpha + (1 - \pi)\beta - \theta f(\alpha, \beta) \tag{7}$$

Let us denote the solutions to Problem (7) as $(\alpha_f(\theta; \pi), \beta_f(\theta; \pi))$

The proof of Proposition 3 shows that the confirmation bias can be extended to all cases where f is "smooth", concave and the states' attention levels are complementary. We formalize this result in Proposition

Proposition 11 (Confirmation bias)**.** *Assume that f is twice differentiable on $\mathring{\Delta}$, and strictly convex. In addition, assume that $\forall (\alpha, \beta) \in \mathring{\Delta}$, $\frac{\partial^2 f}{\partial \alpha \beta} > 0$. Then*

$\forall \theta > 0$, $\forall \pi | (\alpha_f(\theta; \pi), \beta_f(\theta; \pi)) \in \mathring{\Delta}$, $\frac{\partial \alpha_f(\theta;\pi)}{\partial \pi} > 0$, $\frac{\partial \beta_f(\theta;\pi)}{\partial \pi} < 0$.

$\forall \theta > 0$, $\forall \pi | (\alpha_f(\theta; \pi), \beta_f(\theta; \pi)) \in \Delta$, $\frac{\partial \alpha_f(\theta;\pi)}{\partial \pi} \geq 0$, $\frac{\partial \beta_f(\theta;\pi)}{\partial \pi} \leq 0$.

The confidence effect as described in Proposition 4 requires slightly stronger assumptions. Intuitively, the pattern of collection of information at extreme priors depend on the value of marginal precisions when those precisions are already extreme. The properties of $I^*(.)$ described in Proposition 2, where we listed those values on the boundary of Δ, were sufficient to generate the confidence effect.

The assumption on f that we require for the effect to be obtained is the combination of the strict convexity of f and the properties of the marginal informativeness of attention described in Proposition 2: when precisions can be increased and no information is currently collected, increasing attention is costless; certainty on one state is infinitely costly.

The intuition is clear. The second part of Proposition 2 is required since as the prior increases, the marginal return to precision in the least likely state becomes negligible. If the f satisfies the condition, the attention choice must converge towards the anti-diagonal of the unit square, reflecting uninformativeness. At the same time, the marginal return on the most likely state remains positive, so that the only solution at extreme priors is the corner solution, where all the precision is oriented towards the most likely state. The first part of Proposition 2 is required since otherwise, it would be possible for an agent to have full precision on one of the state while maintaining information on the other, when the marginal cost θ is small.

Formally, define $C_f(\theta;\pi) = f((\alpha_f(\theta;\pi), \beta_f(\theta;\pi)))$.

Proposition 12 (Confidence effect)**.** *Assume f satistifies the properties described in Proposition 11.*

In addition, assume $\forall(\alpha,\beta) \in \mathring{\Delta}$

- $\frac{\partial f(1,\beta)}{\partial \alpha} = \infty$
- *,* $\frac{\partial f(\alpha,\beta)}{\partial \alpha} = 0 \Leftrightarrow \alpha + \beta = 1.$

Then,

$C_f(\theta;\pi) \underset{\pi\to 1}{\to} 0$ *and* $C_f(\theta;\pi) \underset{\pi\to 0}{\to} 0$. *Further,* $\alpha_f(\theta;\pi) \underset{\pi\to 1}{\to} 1$ *and* $\beta_f(\theta;\pi) \underset{\pi\to 1}{\to} 0$*;* $\alpha_f(\theta;\pi) \underset{\pi\to 0}{\to} 0$ *and* $\beta_f(\theta;\pi) \underset{\pi\to 0}{\to} 1$.

Finally, it is also worth emphasizing that as Shannon (1948) has shown, the use of the entropy measure to quantify the amount of information in a random variable relies on really weak and intuitive assumptions. In particular, it is the only function satisfying three properties: it is continuous in the distribution of interest, it increases with the size of the support when the distribution is uniform, and is independent of the decomposition of choices.

Therefore, we consider the use of the tools of information theory to be useful and relevant to understand the role of limited attention on individual decisions. However, it is clear that one main future area of research will be to know more about the attention cost function, as Caplin and Dean (2014) do for instance. We hope that by providing simple properties which would generate confirmation bias and permanent disagreement, we can provide intuitions on the properties we would expect from such a function.

6 Conclusion

We have shown in this paper that polarization of beliefs on common value issues can occur under the simple assumption that people are motivated by belief accuracy but are constrained in the attention they can pay to the issue. Costly information, as measured by the amount of information that a choice of attention can potentially convey, leads agents to select to obtain more precise signals about the states of the world they consider likely, generating a confirmation bias. When this information has a fixed marginal cost, a highly biased agent will decide to pay no attention whatsoever to figuring out the state of the world, so that his beliefs are permanently stuck. As a consequence, an agent who makes perceptual mistakes at the start of the information collection process can be permanently wrong, his beliefs being biased against the actual state of the world.

The assumption that attention is limited should be relatively uncontroversial. The fact that this simple limitation can cause rational individuals to eventually disagree completely on an objective issue, where they both aim to understand correctly, is humbling. Indeed, our results show that our disagreements can simply be the consequence of random, initial mistakes in the perception of signals. Long-term disagreements can also come from initial, random shocks to prior beliefs. Kaplan and Mukand (2011) show that the increase in Republican registration after September 11th, 2001, persists after several years. Mullainathan and Washington (2009) find that two years after a presidential election, the degree of polarization differs between cohorts around the age of voting eligibility two years before, with much more polarization in the eligible one. Outside the world of politics, it is easy to see how these results could affect, for instance, the interpretation of information in financial markets: Hong and Stein (2007) provides a review of recent financial literature, and conclude that behavioral techniques, including that of heterogeneous priors, are required in asset pricing.

More work should be done in order to understand the consequences of limited attention on information patterns and economic choices. To answer those issues, the most important avenue of investigation is the analysis of the properties of the attention cost function. Here, we considered the model proposed by Woodford (2012), and we discussed in the final section which properties of the cost function seemed necessary for our findings. We consider the use of entropy and its derivatives as a measure of information to be the most promising, for its reliance on a small and intuitive set of assumptions. However, more can be learnt empirically about the properties of the function theorists should use. In particular, Caplin and Dean (2014) provides interesting experimental findings such as the increase in total attention as stakes increase and that more attention is devoted to discriminate states with higher stakes. Higher stakes, in our framework, are isomorphic to a lower relative cost of attention, and so this arises endogenously. More experimental work is needed to uncover additional comparative statics, or test the properties of existing proposals – in our case, it would be important to test the concavity of $I^*(.)$ and its behavior at high levels of certainty.

A second key element of future research is to consider is the impact of limited attention in settings where individuals can interact. For instance, given that the probability distribution of perceived signals is history dependent, one immediate interrogation is the potential impact of limited attention on herding behavior. More generally, how does limited attention influence information aggregation? For instance, if one person is unaware of another's initial bias and thus cannot correct for it, a perceptual mistake by the second person might be communicated to the first, and separate groups could diverge faster than individuals.

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz. 2006. "Learning and Disagreement in an Uncertain World." *SSRN Electronic Journal.*

———. 2008. "Fragility of Asymptotic Agreement Under Bayesian Learning." *SSRN Electronic Journal,* no. February.

Baliga, Sandeep, Eran Hanany, and P Klibanoff. 2013. "Polarization and ambiguity." *. . . , Kellogg CMS-EMS . . .*

Baumeister, R. F., and L. S. Newman. 1994. "Self-Regulation of Cognitive Inference and Decision Processes." *Personality and Social Psychology Bulletin* 20, no. 1 (February): 3–19.

Berk, Robert H. 1966. "Limiting Behavior of Posterior Distributions when the Model is Incorrect." *The Annals of Mathematical Statistics* 37, no. 1 (February): 51–58.

Blackwell, David, and Lester Dubins. 1962. "Merging of Opinions with Increasing Information." *The Annals of Mathematical Statistics* 33, no. 3 (September): 882–886.

Caplin, A, and M Dean. 2014. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *NBER Working Paper.*

Cukierman, Alex, and M. Tommasi. 1998. "When does it take a Nixon to go to China?" *American Economic Review* 88 (1): 180–197.

Darley, John M., and Paget H. Gross. 1983. "A hypothesis-confirming bias in labeling effects." *Journal of Personality and Social Psychology* 44 (1): 20–33.

Feddersen, Timothy J., and W. Pesendorfer. 1998. "Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting." *The American Political Science Review* 92 (1): 23–35.

Freedman, David A. 1963. "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case." *The Annals of Mathematical Statistics* 34, no. 4 (December): 1386–1403.

Fryer, R, P Harms, and M Jackson. 2013. "Updating Beliefs with Ambiguous Evidence: Implications for Polarization." *Available at SSRN 2263504.*

Hirshleifer, David, and Siew Hong Teoh. 2003. "Limited attention, information disclosure, and financial reporting." *Journal of Accounting and Economics* 36, nos. 1-3 (December): 337–386.

Hirst, D Eric, and Patrick E Hopkins. 1998. "Comprehensive Income Reporting and Analysts' Valuation Judgments." *Journal of Accounting Research* 36 (1998): 47.

Hong, Harrison, and Jeremy C Stein. 2007. "Disagreement and the Stock Market." *Journal of Economic Perspectives* 21, no. 2 (May): 109–128.

Huberman, Gur, and Tomer Regev. 2001. "Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar." *The Journal of Finance* 56, no. 1 (February): 387–396.

Kaplan, Ethan, and Sharun Mukand. 2011. “The Persistence of Political Partisanship: Evidence from 9/11.”

Koszegi, B., and M. Rabin. 2006. “A model of reference-dependent preferences.” *The Quarterly Journal of Economics* 121 (4): 1133.

Lipe, M.G. 1998. “Individual investors’ risk judgments and investment decisions: The impact of accounting and market data.” *Accounting, Organizations and Society* 23, no. 7 (October): 625–640.

Lodge, Milton, Marco R. Steenbergen, and Shawn Brau. 1995. “The Responsive Voter: Campaign Information and the Dynamics of Candidate Evaluation.” *The American Political Science Review* 89, no. 2 (June): 309.

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

Miller, RI, and CW Sanchirico. 1999. “The Role of Absolute Continuity in Merging of Opinions and Rational Learning.” *Games and Economic Behavior* 29 (1-2): 170–190.

Mullainathan, Sendhil, and Ebonya Washington. 2009. “Sticking with Your Vote: Cognitive Dissonance and Political Attitudes.” *American Economic Journal: Applied Economics* 1, no. 1 (January): 86–111.

Nickerson, Raymond S. 1998. "Confirmation bias: A ubiquitous phenomenon in many guises." *Review of General Psychology* 2 (2): 175–220.

Rabin, M, and J. Schrag. 1999. "First impressions matter: A model of confirmatory bias." *The Quarterly Journal of Economics,* no. February.

Redlawsk, David P. 2008. "Hot Cognition or Cool Consideration? Testing the Effects of Motivated Reasoning on Political Decision Making." *The Journal of Politics* 64, no. 04 (July).

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27, no. 4 (October): 623–656.

Shaw, M L, and P Shaw. 1977. "Optimal allocation of cognitive resources to spatial locations." *Journal of experimental psychology. Human perception and performance* 3, no. 2 (May): 201–11.

Sims, Christopher a. 2003. "Implications of rational inattention." *Journal of Monetary Economics* 50, no. 3 (April): 665–690.

Suen, Wing. 2004. "The Self Perpetuation of Biased Beliefs." *The Economic Journal* 114:377–396.

Taber, Charles S, and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50, no. 3 (July): 755–769.

Woodford, Michael. 2012. "Inattentive Valuation and Reference-Dependent Choice." *Working Paper.*

A Proofs

Proof of proposition 1

We will use the following lemma

Lemma 1 (Convexity of the maximum of convex functions). *Let $f_i : X \to Y$, $i \in I$ be a family of convex functions and $f = max_i f_i$. Then f is convex.*

Proof of lemma 1. Consider $f = max(f_1, f_2)$ with both f_i convex. Then for any $x, y, \lambda \in (0,1)$, there exists $i \in \{1,2\}$ such that $f(\lambda x + (1-\lambda)y) = f_i(\lambda x + (1-\lambda)y)$. By convexity of f_i, $f(\lambda x+(1-\lambda)y) \leq \lambda f_i(x)+(1-\lambda)f_i(y) \leq \lambda f(x) + (1-\lambda)f(y)$

□

Proposition 1 (Convexity of information capacity). *I^* is infinitely differentiable on $\mathring{\Delta}$ and is strictly convex in (α, β)*

Proof of proposition 1. To show that I^* is convex, we show that $I_\pi(\alpha, \beta)$ is convex for any π. We then use lemma 1 to conclude. Let us denote $q = \alpha\pi + (1-\beta)(1-\pi)$. The mutual information given a prior π is

$$\begin{aligned} I_\pi(\alpha,\beta) = {} & \alpha\pi\ln(\alpha) + (1-\alpha)\pi\ln(1-\alpha) \\ & + (1-\beta)(1-\pi)\ln(1-\beta) + \beta(1-\pi)\ln(\beta) \\ & - q\ln(q) - (1-q)\ln(1-q) \end{aligned}$$

Then for all π, $I_\pi(.)$ is twice continuously differentiable in (α, β) and

$$
\begin{aligned}
\frac{\partial I_\pi(\alpha,\beta)}{\partial\alpha} =& \pi\ln(\alpha)-\pi\ln(1-\alpha)-\pi\ln(q)+\pi\ln(1-q)\\
\frac{\partial I_\pi(\alpha,\beta)}{\partial\beta} =& -(1-\pi)\ln(\beta)-(1-\pi)\ln(1-\beta)+(1-\pi)\ln(q)-(1-\pi)\ln(1-q)
\end{aligned}
$$

and the second derivatives are given by

$$
\begin{aligned}
\frac{\partial^2 I_\pi(\alpha,\beta)}{\partial\alpha^2} =& \frac{\pi}{\alpha}+\frac{\pi}{1-\alpha}-\frac{\pi^2}{q}-\frac{\pi^2}{1-q}\\
\frac{\partial^2 I_\pi(\alpha,\beta)}{\partial\beta^2} =& \frac{1-\pi}{\beta}+\frac{1-\pi}{1-\beta}-\frac{(1-\pi)^2}{q}-\frac{(1-\pi)^2}{1-q}\\
\frac{\partial^2 I_\pi(\alpha,\beta)}{\partial\beta\partial\alpha} =& \frac{\pi(1-\pi)}{q}+\frac{\pi(1-\pi)}{1-q}
\end{aligned}
$$

Given the Hessian of $I_\pi(.)$, we can directly show that for any π, $I_\pi(.)$ is convex, by showing that (denoting $I \equiv I_\pi(\alpha,\beta)$)

$$
\Delta = \frac{\partial^2 I}{\partial\alpha^2}\cdot\frac{\partial^2 I}{\partial\beta^2}-\frac{\partial^2 I}{\partial\beta\partial\alpha} > 0
$$

Let us first rewrite the terms of the Hessian:

$$
\begin{aligned}
\frac{\partial^2 I}{\partial\alpha^2} =& \frac{\pi q(1-q)-\pi^2\alpha(1-\alpha)}{\alpha(1-\alpha)q(1-q)}\\
\frac{\partial^2 I}{\partial\beta^2} =& \frac{(1-\pi)q(1-q)-(1-\pi)^2\beta(1-\beta)}{\beta(1-\beta)q(1-q)}\\
\frac{\partial^2 I}{\partial\beta\partial\alpha} =& \frac{\pi(1-\pi)}{q(1-q)}
\end{aligned}
$$

Hence,

$$
\Delta > 0 \Leftrightarrow \pi[q(1-q)-\pi\alpha(1-\alpha)](1-\pi)[q(1-q)-(1-\pi)\beta(1-\beta)]-\alpha(1-\alpha)\beta(1-\beta)\pi^2(1-\pi)^2 > 0
$$

or

$$\Delta > 0 \Leftrightarrow [q(1-q)-\pi\alpha(1-\alpha)][q(1-q)-(1-\pi)\beta(1-\beta)]-\alpha(1-\alpha)\beta(1-\beta)\pi(1-\pi) > 0$$

Finally,

$$\Delta > 0 \Leftrightarrow [q^2(1-q)^2 - \pi\alpha(1-\alpha)q(1-q) - (1-\pi)\beta(1-\beta)q(1-q)] > 0$$

which is equivalent to

$$\Delta > 0 \Leftrightarrow [q(1-q) - \pi\alpha(1-\alpha) - (1-\pi)\beta(1-\beta)] > 0$$

Plugging in $q = \alpha\pi + (1-\pi)(1-\beta)$ and dividing by $\pi(1-\pi)$ yields the condition

$$\Delta > 0 \Leftrightarrow (1 - (\alpha+\beta)^2) > 0$$

Hence, $I_\pi(.)$ is convex for all π.

I^* is therefore the pointwise maximum of convex functions, and is thus convex by Lemma 1.

□

Proof of Proposition 2

Proposition 2 (Marginal informativeness on the boundaries of $\mathring{\Delta}$).

$\forall(\alpha,\beta) \in \mathring{\Delta}$,

- $\frac{\partial I^*(1,\beta)}{\partial\alpha} = \infty$
- $\frac{\partial I^*(1,\beta)}{\partial\beta} \in (0,\infty)$

- $\frac{\partial I^*(\alpha,\beta)}{\partial \alpha} = 0 \Leftrightarrow \alpha + \beta = 1$

Proof of Proposition 2. By the envelope theorem, we know that $\frac{\partial I^*}{\partial \alpha} = \frac{\partial I(\alpha,\beta;\pi^*)}{\partial \alpha}$ where $\pi^*(\alpha, \beta)$ is the prior maximizing I for a given information structure. But we know that

$$\begin{aligned}
I((\alpha,\beta);(\pi,1-\pi)) =& \alpha\pi\ln(\alpha) + (1-\alpha)\pi\ln(1-\alpha) \\
&+ (1-\beta)(1-\pi)\ln(1-\beta) + \beta(1-\pi)\ln(\beta) \\
&- [\alpha\pi + (1-\beta)(1-\pi)]\ln(\alpha\pi + (1-\beta)(1-\pi)) \\
&- [(1-\alpha)\pi + \beta(1-\pi)]\ln((1-\alpha)\pi + \beta(1-\pi))
\end{aligned}$$

So that

$$\frac{\partial I(\alpha,\beta;\pi^*)}{\partial \alpha} = \pi^*\left[\ln\frac{\alpha}{1-\alpha} - \ln\frac{q^*}{1-q^*}\right] \tag{8}$$

$$\frac{\partial I(\alpha,\beta;\pi^*)}{\partial \beta} = (1-\pi^*)\left[\ln\frac{\beta}{1-\beta} - \ln\frac{1-q^*}{q^*}\right] \tag{9}$$

where $q^* = \alpha\pi^* + (1-\beta)(1-\pi^*)$.

From the optimization of I, we know that

$$\frac{q^*}{1-q^*} = \left[\frac{\alpha^\alpha(1-\alpha)^{1-\alpha}}{\beta^\beta(1-\beta)^{1-\beta}}\right]^{\frac{1}{\alpha+\beta-1}}$$

In particular, when $\alpha = 1$, for any $\beta > 0$, $q^* \in (0,1)$. Hence, $\ln\frac{q^*}{1-q^*}$ is finite.

From equation 8, this gives us that $\frac{\partial I(\alpha,\beta;\pi^*)}{\partial \alpha}$ evaluated at $\alpha = 1$ and $\beta > 0$ is infinite while equation 9 tells us that at the same point, $\frac{\partial I(\alpha,\beta;\pi^*)}{\partial \beta}$ is finite – and bounded away from 0[8].

8. Note that π^* and q^* are related and at the same α, β, we will have $\pi^* \in (0,1)$

Finally, assume $\alpha \in (0,1)$. By writing $\beta = 1-\alpha+u$ where $u \to 0$, one can show that $\frac{q^*}{1-q^*} = \frac{\alpha}{1-\alpha}\left(1 - \frac{u}{2\alpha(1-\alpha} + o(u)\right)$. This expansion implies that $\pi^*(\alpha, 1-\alpha) = \frac{1}{2}$ by continuity, and using Equation 8 and 9, we obtain that the partial derivatives of I^* are all null on the inverted diagonal of the square.

The discussion above shows that those partial derivatives are also strictly positive outside of this diagonal, so that the equivalence stated in the proposition is proven.

□

Proof of proposition 3

Proposition 3 (Confirmation bias). *$\forall \theta > 0$, $\forall \pi | (\alpha(\theta;\pi), \beta(\theta;\pi)) \in \mathring{\Delta}$, $\frac{\partial \alpha(\theta;\pi)}{\partial \pi} > 0$, $\frac{\partial \beta(\theta;\pi)}{\partial \pi} < 0$.*

$\forall \theta > 0$, $\forall \pi | (\alpha(\theta;\pi), \beta(\theta;\pi)) \in \Delta$, $\frac{\partial \alpha(\theta;\pi)}{\partial \pi} \geq 0$, $\frac{\partial \beta(\theta;\pi)}{\partial \pi} \leq 0$.

Proof. Proof of proposition 3

Assume that the solution to Problem (3) is interior and consider the system derived from the first order conditions:

$$\frac{\partial I^*(\alpha(\theta;\pi), \beta(\theta;\pi))}{\partial \alpha} = \frac{\pi}{\theta} \tag{10}$$

$$\frac{\partial I^*(\alpha(\theta;\pi), \beta(\theta;\pi))}{\partial \beta} = \frac{1-\pi}{\theta} \tag{11}$$

For clarity, we will denote the (hypothetical) solution to the system (which is unique, by the concavity of the problem) $\alpha^*(\pi,\theta), \beta^*(\pi,\theta)$.

Denote $\tilde{\beta}(\alpha|\theta,\pi)$ the only point in $(1-\alpha, 1)$ such that $\frac{\partial I^*(\alpha, \tilde{\beta}(\alpha|\theta,\pi))}{\partial \beta} = \frac{1-\pi}{\theta}$. This is possible because for all α

$$\frac{\partial I^*(\alpha,1)}{\partial \beta} = \infty$$
$$\frac{\partial I^*(\alpha,1-\alpha)}{\partial \beta} = 0$$

and $\frac{\partial I^{*2}(\alpha,\beta)}{\partial \beta^2} > 0$.

Likewise, we can define $\tilde{\alpha}(\beta|\theta,\pi)$ to be the only point in $(1-\beta,1)$ such that $\frac{\partial I^*(\tilde{\alpha}(\beta|\theta,\pi),\beta)}{\partial \alpha} = \frac{\pi}{\theta}$.

Moreover, because $\frac{\partial I^{*2}(\alpha,\beta)}{\partial \beta^2} > 0$ and $\frac{\partial I^{*2}(\alpha,\beta)}{\partial \alpha \partial \beta} > 0$, $\tilde{\beta}(.|\theta,\pi)$ is a decreasing function of α. Likewise, $\tilde{\alpha}$ is a decreasing function of β.

Consider now an increase in $\pi \in (0,1)$, to $\pi' > \pi \in (0,1)$ for which the solution is also interior.

Because the loci of $\tilde{\beta}(.)$ and $\tilde{\alpha}(.)$ cross only once, $\tilde{\alpha}(\beta) \underset{\beta \to 0}{\to} 1$ and $\tilde{\beta}(\alpha) \underset{\alpha \to 1}{\to} \bar{\beta} \in (0,1)$, it must be that:

$$\tilde{\beta}(\tilde{\alpha}(\beta|\pi,\theta)|\pi,\theta) > \beta \Leftrightarrow \beta < \beta^*(\pi,\theta) \tag{12}$$

We also know that $\forall \beta, \tilde{\alpha}(\beta|\pi') > \tilde{\alpha}(\beta|\pi)$ since $\frac{\partial I^*}{\partial \alpha}$ increases in α and equation (10) must hold.

Because $\tilde{\beta}$ is decreasing, we have that for all β

$$\tilde{\beta}(\tilde{\alpha}(\beta|\pi')|\pi') < \tilde{\beta}(\tilde{\alpha}(\beta|\pi)|\pi')$$

One can show as we did for $\tilde{\alpha}$ that $\forall \alpha, \tilde{\beta}(\alpha|\pi') < \tilde{\beta}(\alpha|\pi)$. This implies

$$\tilde{\beta}(\tilde{\alpha}(\beta|\pi)|\pi') < \tilde{\beta}(\tilde{\alpha}(\beta|\pi)|\pi)$$

Evaluating the latter two equations at $\beta = \beta^*(\pi)$, we get

$$\tilde{\beta}(\tilde{\alpha}(\beta^*(\pi)|\pi')|\pi') < \beta^*(\pi)$$

By relation (12), it must be that $\beta^*(\pi) > \beta^*(\pi')$

Likewise, $\alpha^*(\pi) < \alpha^*(\pi')$ $\square$

Proof of proposition 4

Proposition 4 (Confidence effect). *$C(\theta;\pi) \underset{\pi\to 1}{\to} 0$ and $C(\theta;\pi) \underset{\pi\to 0}{\to} 0$. Further, $\alpha(\theta;\pi) \underset{\pi\to 1}{\to} 1$ and $\beta(\theta;\pi) \underset{\pi\to 1}{\to} 0$; $\alpha(\theta;\pi) \underset{\pi\to 0}{\to} 0$ and $\beta(\theta;\pi) \underset{\pi\to 0}{\to} 1$.*

Proof of proposition 4. By symmetry, we will concern ourselves with the limits as $\pi \to 1$. As a reminder, we also only consider the case where $\alpha + \beta \geq 1$. Fixing $\theta > 0$, we write $\alpha(\pi)$ and $\beta(\pi)$ to be the solutions of Problem (3) with the dependence in θ implicitly understood.

Because when the solution is interior, α and β are monotonous in π and they are bounded, there exists $(\bar{\alpha}, \bar{\beta}) \in [0,1]^2$ such that $\alpha(\pi) \underset{\pi\to 1}{\to} \bar{\alpha}$ and $\beta(\pi) \underset{\pi\to 1}{\to} \bar{\beta}$

We show in the following that the only possibility is $\bar{\alpha} = 1 - \bar{\beta}$ so that $C \underset{\pi\to 1}{\to} 0$

When the solution is not interior at extreme priors Assume first the that there exists $\bar{\pi} > \frac{1}{2}$ such that $\forall \pi > \bar{\pi}$, the solution to the maximization problem is not interior to $\alpha + \beta \geq 1$.

If for any $1 \geq \pi > \bar{\pi}$, $\beta(\pi) = 1$ (resp. $\alpha(\pi) = 1$), it must be the case that $\alpha(\pi) = 0$ (resp. $\beta(\pi) = 0$) by Proposition 2 since otherwise, the marginal increase in the objective function in decreasing β (resp. α) would be infinite.

Given these results, the only case consistent with optimization at $\pi > \frac{1}{2}$ is $\alpha = 1$ and $\beta = 0$. This yields the desired results.

The only possible other boundary to check is $\alpha + \beta = 1$. In that case, $I^*(\alpha, \beta) = 0$ and the only choice consistent with optimization for $\pi > \frac{1}{2}$ is also $\alpha = 1$ and $\beta = 0$.

When the solution is always interior Assume now that the solution is interior at all $\pi's$. The first order conditions are given by Equations (10) and (11).

Equation (11) requires that $\frac{\partial I^*(\alpha(\pi),\beta(\pi))}{\partial \beta} \underset{\pi \to 1}{\to} 0$.

We know that $\frac{\partial I^*(\alpha,\beta)}{\partial \beta} = 0 \Leftrightarrow \alpha = 1 - \beta$

Hence, $\bar{\alpha} = 1 - \bar{\beta}$. This immediately yields that $C(\theta; \pi) \underset{\pi \to 1}{\to} 0$.

We can also show immediately that $\bar{\alpha} = 1$. Assume $\bar{\alpha} \in (0, 1)$, and consider the first order condition in Equation (10). If it holds for all π, then the RHS converges to 0 since $\bar{\alpha} = 1 - \bar{\beta}$ while the LHS converges to 1. This is a contradiction. Hence it must be that $\bar{\alpha} \in \{0, 1\}$. But $\bar{\alpha} = 0, \bar{\beta} = 1$ would be obviously inconsistent with maximization in a neighbourhood of $\pi = 1$.

□

Proof of Proposition 5

Proposition 5 (Information shutdown at extreme priors)**.** *$\forall \theta, \exists \overline{\pi}(\theta) | \forall \pi > \overline{\pi}(\theta)$, $\beta(\theta; \pi) < 0.5$ and $\alpha(\theta; \pi) > 0.5$. $\forall \theta, \exists \overline{\overline{\pi}}(\theta) | \forall \pi > \overline{\overline{\pi}}(\theta)$, $\beta(\theta; \pi) = 0$ and $\alpha(\theta; \pi) = 1$*

Proof of Proposition 5. The first part of the proposition follows immediately from 4, the continuity of the solutions and the fact that $\beta(0.5) = \alpha(0.5) > 0.5$.

For the second part, we show that in a rational inattention model, $\beta = 0$ for high priors. We then show that when $\beta = 0$ in a rational inattention model, it has to be 0 in the inattentive valuation model.

The rational inattention problem with a fixed cost of attention θ is described in Problem (2)

Denote $a = exp\left(\frac{1}{\theta}\right)$. The first order conditions, provided the solution is interior, yield

$$\frac{\alpha(1-q)}{q(1-\alpha)} = a$$
$$\frac{\beta q}{(1-\beta)(1-q)} = a$$

Equalizing the two expressions for $\frac{q}{1-q}$, we have that

$$\frac{\alpha}{1-\alpha} = a^2 \frac{1-\beta}{\beta}$$

Hence,

$$\alpha = \frac{a^2(1-\beta)}{\beta + a^2(1-\beta)}$$

Note that $\alpha = 1 \Leftrightarrow \beta = 0$. The previous expression implies that

$$q = \frac{a^2(1-\beta)\pi}{\beta + a^2(1-\beta)} + (1-\beta)(1-\pi)$$
$$1 - q = \frac{\beta\pi}{\beta + a^2(1-\beta)} + \beta(1-\pi)$$

From

$$\frac{\beta q}{(1-\beta)(1-q)} = a$$

and substituting for q and $1-q$ We have that

$$\frac{\beta}{1-\beta} \cdot \frac{a^2(1-\beta)\pi + (1-\beta)(1-\pi)(\beta + a^2(1-\beta))}{\beta(1-\pi)(\beta + a^2(1-\beta)) + \beta\pi} = a$$

Or

$$\frac{a^2\pi + (1-\pi)(\beta + a^2(1-\beta))}{(1-\pi)(\beta + a^2(1-\beta)) + \pi} = a$$

Thus

$$a^2 + \beta[1-\pi][1-a^2] = a[\pi + a^2(1-\pi)] + a\beta[1-\pi][1-a^2]$$

Or

$$\beta \cdot (1-\pi)(a^2-1)(a-1) = a[\pi + a^2(1-\pi) - a]$$

We also have that

$$a[\pi + a^2(1-\pi) - a] = a(a-1)(a(1-\pi) - \pi)$$

So that

$$\beta = \frac{a(a(1-\pi) - \pi)}{(1-\pi)(a^2-1)}$$

This expression is not negative only if $\frac{\pi}{1-\pi} \leq a$ or $\pi < \frac{a}{1+a}$. Therefore, for all $\pi \geq \frac{a}{1+a}$, the solution is not interior and $\beta = 0$

Now, let us write

$$
\begin{aligned}
V^{RI} =& \max_{\alpha,\beta} \tilde{\pi}\alpha + (1-\tilde{\pi})\beta - \theta I((\alpha,\beta);(\tilde{\pi},1-\tilde{\pi})) \\
=& \max_{\alpha,\beta} h(\alpha,\beta,\tilde{\pi}) \\
V^{IV} =& \max_{\alpha,\beta} \tilde{\pi}\alpha + (1-\tilde{\pi})\beta - \theta I^*(\alpha,\beta) \\
=& \max_{\alpha,\beta} g(\alpha,\beta,\tilde{\pi})
\end{aligned}
$$

By definition, $\forall \alpha,\beta,\pi,\ g(\alpha,\beta,\pi) \leq h(\alpha,\beta,\pi)$.

Hence, by taking the max on both sides with respect to α,β, we have that for all π, $V^{IV} \leq V^{RI}$

We know that for $\tilde{\pi} \geq \bar{\pi} = \frac{a}{1+a}$, $V^{RI} = \pi$ since $\alpha = 1, \beta = 0$ and $I((\alpha,\beta);(\tilde{\pi},1-\tilde{\pi})) = 0$. Hence, for $\tilde{\pi} \geq \bar{\pi}$, $V^{IV} \leq \pi$. But setting $\alpha = 1$ and $\beta = 0$ in the IV problem yields a value of $\tilde{\pi}$. Hence, it realizes the maximum.

Therefore, we have that when under RI, $\alpha = 1$ and $\beta = 0$, then it must be the case under IV. In particular, for all $\pi \geq \frac{a}{1+a}$, the solution of the inattentive valuation problem is not interior and $\beta = 0, \alpha = 1$

□

Proof of Proposition 6

Proposition 6. $\forall \theta > 0, \exists \tilde{\pi}(\theta) > \frac{1}{2} | \left|\pi - \frac{1}{2}\right| > \left|\tilde{\pi}(\theta) - \frac{1}{2}\right| \Rightarrow \alpha_d(\theta;\pi) + \beta_d(\theta;\pi) = 1$. *Moreover,* $\forall \theta > 0, \tilde{\pi}(\theta) = \bar{\bar{\pi}}(\theta)$

Proof of Proposition 6. Consider the dynamic setting under Rational Inattention:

$$
V(\pi) = \max_{\alpha,\beta} \alpha\pi + \beta(1-\pi) + \delta\Big[qV\left(\frac{\alpha\pi}{q}\right) + (1-q)V\left(\frac{(1-\alpha)\pi}{1-q}\right)\Big] - \theta I(\alpha,\beta,\pi)
$$

Now consider V under no attention ($\alpha + \beta = 1$), (immediately we get $\pi_a = \pi_b = \pi$):

$$\begin{aligned}
V(\pi) &= \max_{\alpha,\beta} \pi\alpha + (1-\pi)\beta + \delta[qV(\pi) + (1-q)V(\pi)] \\
V(\pi)(1-\delta) &= \max_{\alpha,\beta} \pi\alpha + (1-\pi)\beta \\
V(\pi) &= \frac{\pi}{1-\delta}
\end{aligned}$$

Now, since V is continuous, and we are trying to show a discontinuity in the first order conditions (a point at which they stop holding), there must be a $\tilde{\pi}$ such that for all $\pi > \tilde{\pi}$, $I(\pi) = 0$. We know that $\tilde{\pi} = 1$ is a solution to this, but we want to check whether there is another solution where $\tilde{\pi} < 1$. Therefore, $V(\tilde{\pi})$ must satisfy the first order conditions *and* satisfy $V(\tilde{\pi}) = \frac{\tilde{\pi}}{1-\delta}$. If we can find this $\tilde{\pi}$ we are done. Importantly, at $\tilde{\pi}$, $V(\tilde{\pi}_a) = V(\tilde{\pi}_b) = \tilde{\pi}$ and $V'(\tilde{\pi}_a) = V'(\tilde{\pi}_b)$:

$$V(\tilde{\pi}) = \max_{\alpha,\beta} \alpha\tilde{\pi} + \beta(1-\tilde{\pi}) + \delta\Big[qV\left(\frac{\alpha\tilde{\pi}}{q}\right) + (1-q)V\left(\frac{(1-\alpha)\tilde{\pi}}{1-q}\right)\Big] - \theta I(\alpha, \beta, \tilde{\pi})$$

So the first order condition with respect to α becomes:

$$\begin{aligned}
\tilde{\pi} + \delta\tilde{\pi}\left[V(\tilde{\pi}_a) - V(\tilde{\pi}_b) + \frac{q-\alpha\tilde{\pi}}{q}V'(\tilde{\pi}_a) + \frac{(1-\alpha)\tilde{\pi}-(1-q)}{1-q}V'(\pi_b)\right] &= \theta I_\alpha \\
\tilde{\pi} + \delta\tilde{\pi}V'(\tilde{\pi})\left[\frac{q-\alpha\tilde{\pi}}{q} + \frac{(1-\alpha)\tilde{\pi}-(1-q)}{1-q}\right] &= \theta I_\alpha \\
\tilde{\pi} &= \theta\tilde{\pi}\ln\left(\frac{\alpha(1-q)}{q(1-\alpha)}\right) \\
1 &= \theta\ln\left(\frac{\alpha(1-q)}{q(1-\alpha)}\right)
\end{aligned}$$

The first order condition with respect to β becomes:

$$\begin{aligned}
(1-\tilde{\pi})+\delta(1-\tilde{\pi})\left[V(\tilde{\pi}_a)-V(\tilde{\pi}_b)+\frac{\alpha\tilde{\pi}}{q}V'(\tilde{\pi}_a)+\frac{(1-\alpha)\tilde{\pi}}{(1-q)}V'(\pi_b)\right] &= \theta I_\beta \\
(1-\tilde{\pi})+\delta(1-\tilde{\pi})V'(\tilde{\pi})\left[\frac{\alpha\tilde{\pi}}{q}+\frac{(1-\alpha)\tilde{\pi}}{(1-q)}\right] &= \theta I_\beta \\
1 &= \theta\ln\left(\frac{\beta q}{(1-q)(1-\beta)}\right)
\end{aligned}$$

This is identical to the static case and thus, the exact same proof will work. Namely, we know that there is a point $\tilde{\pi}$ such that the first order conditions above are both satisfied with equality, *and* $\alpha+\beta=1$ - it will be exactly the $\tilde{\pi}$ at which the first order conditions of the static case break down.

The same logic as the above for the Inattentive Value formulation will show that at a hypothetical shutdown point, the FOC for the dynamic and static case become the same. We know that under the FOC for the static case, that there is a shutdown point that occurs before that of Rational Inattention. Therefore, the same shutdown point must exist in the dynamic case.

Therefore, such a shutdown point is defined in the same way in the dynamic case as in the static for both rational inattention and inattentive valuation. □

Proof of Proposition 7

Proposition 7. *$\forall\theta>0$, $\alpha(\theta;.5)<\alpha_d(\theta;.5)$ and $\beta(\theta;.5)<\beta_d(\theta;.5)$.*

Proof of Proposition 7. Under static case we get $q=\pi=\pi^*=0.5$, $\alpha=\beta$

and FOC:

$$\begin{aligned}
1 &= \theta \ln\left(\frac{\alpha}{1-\alpha}\right) \\
e^{\frac{1}{\theta}} &= \frac{\alpha}{1-\alpha} \\
\alpha &= \frac{e^{\frac{1}{\theta}}}{1+e^{\frac{1}{\theta}}}
\end{aligned}$$

Under the dynamic case, we get that $q = \pi = \pi^* = 0.5$, $\alpha = \beta$, $V(\pi_a) = V(\pi_b)$ (due to symmetry) and $V'(\pi_a) = -V'(\pi_b)$ (again due to symmetry). The FOC is:

$$\begin{aligned}
\tilde{\pi} + \delta\tilde{\pi}\left[V(\tilde{\pi}_a) - V(\tilde{\pi}_b) + \frac{q-\alpha\tilde{\pi}}{q}V'(\tilde{\pi}_a) + \frac{(1-\alpha)\tilde{\pi}-(1-q)}{1-q}V'(\pi_b)\right] &= \theta\tilde{\pi}\ln\left(\frac{\alpha}{1-\alpha}\right) \\
1 + \delta\left[(1-\alpha)V'(\tilde{\pi}_a) - \alpha V'(\pi_b)\right] &= \theta \ln\left(\frac{\alpha}{1-\alpha}\right)
\end{aligned}$$

$V'(\tilde{\pi}_a) > 0$, as is shown below, and so the amount on the left hand side of the equation is larger that in the static case, hence requiring a larger value of α to satisfy equality.

To try to show that $V'(\pi) > 0$ when $\pi > 0.5$:

$$V'(\pi) = \alpha - \beta + \delta\left[(\alpha+\beta-1)V\left(\frac{\alpha\pi}{q}\right) + V'\left(\frac{\alpha\pi}{q}\right)\frac{\alpha q - \alpha\pi(\alpha+\beta-1)}{q} - (\alpha+\beta-1)V\left(\frac{(1-\alpha}{1-}\right.\right.$$

It is easy to see two things immediately. If π is past the shutdown point, then $V'(\pi) = 1$, because $\alpha + \beta = 1$ and $\alpha \neq \beta$. Second, $V'(0.5) = 0$. Now, we can get the rest simply showing that the Value function is convex.

We show that the value function of our problem is convex by the following reasoning:

- For any V convex, TV is convex

- Therefore, for any V convex, $T^{(n)}V$ is convex
- Because the set of convex functions on $[0,1]$ is closed, $\lim_{\pi\to 1} T^{(n)}V$ is convex
- Hence, the value function which solves our problem is convex

The main step to prove is the first one.

Let us thus show that if V is convex in π, then TV is also convex in π. To show that, we will first show that the function inside the maximum is convex in π for any α, β. As the maximum of a family of convex functions, TV will thus be convex in π

Because the first part of the function is linear in π, we can focus on the term in square brackets to show the convexity. Let us define, in particular

$$f(\pi) = \left[qV(\pi^a) + (1-q)V(\pi^b)\right]$$

And note that

$$\begin{aligned}
\frac{\partial q}{\partial \pi} &= \alpha + \beta - 1 \\
\frac{\partial \pi^a}{\partial \pi} &= \frac{\alpha q - \alpha\pi(\alpha+\beta-1)}{q^2} \\
\frac{\partial \pi^b}{\partial \pi} &= \frac{(1-\alpha)(1-q) + (1-\alpha)\pi(\alpha+\beta-1)}{(1-q)^2}
\end{aligned}$$

Hence,

$$\begin{aligned}
f'(\pi) = {} & (\alpha+\beta-1)\left[V(\pi^a) - V(\pi^b)\right] \\
& + V'(\pi^a)\frac{\alpha q - \alpha\pi(\alpha+\beta-1)}{q} + V'(\pi^b)\frac{(1-\alpha)(1-q) + (1-\alpha)\pi(\alpha+\beta-1)}{1-q}
\end{aligned}$$

and

$$f''(\pi) = (\alpha+\beta-1)\left[V'(\pi^a)\frac{\alpha q - \alpha\pi(\alpha+\beta-1)}{q^2} - V'(\pi^b)\frac{(1-\alpha)(1-q)+(1-\alpha)\pi(\alpha+\beta-1)}{(1-q)^2}\right]$$
$$- V'(\pi^a)\frac{(\alpha+\beta-1)(\alpha q - \alpha\pi(\alpha+\beta-1))}{q^2}$$
$$+ V'(\pi^b)\frac{(\alpha+\beta-1)((1-\alpha)(1-q)+(1-\alpha)\pi(\alpha+\beta-1))}{(1-q)^2}$$
$$+ V''(\pi^a)\frac{1}{q}\left(\frac{\partial\pi^a}{\partial\pi}\right)^2 + V''(\pi^b)\frac{1}{1-q}\left(\frac{\partial\pi^b}{\partial\pi}\right)^2$$

The terms in V' cancel out and we are only left with the terms in V'' weighted by positive numbers. Hence, f'' inherits the sign of V'' if it is constant. In particular, f is convex if V'' is convex. $\square$

Proof of Proposition 8

Proposition 6. $\forall\theta > 0, \exists\tilde{\pi}(\theta) > \frac{1}{2} \big| \left|\pi - \frac{1}{2}\right| > \left|\tilde{\pi}(\theta) - \frac{1}{2}\right| \Rightarrow \alpha_d(\theta;\pi) + \beta_d(\theta;\pi) = 1$. *Moreover,* $\forall\theta > 0, \tilde{\pi}(\theta) = \bar{\bar{\pi}}(\theta)$

Proof of Proposition 8. Take a unit interval of agents who all have a prior of $\pi_0 = 0.5$. Since we know that agents pay more attention to signals in the forward-looking problem than in the myopic, we have that:

$$\alpha = \beta \geq \frac{e^{\frac{1}{\theta}}}{1+e^{\frac{1}{\theta}}}$$

Further, we know that no matter which way the agents update (whether they see an $'a'$ or a $'b'$:

$$\pi_1 \geq \frac{\pi_0\alpha}{q_0} \quad \text{OR} \quad \pi_1 \leq \frac{(1-\alpha)\pi_0}{1-q_0}$$

Remember that $\pi_0 = q_0 = 0.5$, which means that the above equation simplifies to the condition that $\pi_1 \geq \alpha$ or $\pi_1 \leq 1-\alpha$ depending on what signal is observed. But remember also that the shutdown point is (at least) the same as $\alpha(0.5)$! Therefore, for agents starting at 0.5, they will observe one signal, either A or B, update, and then shut down. $\square$

B Figures

Figure 1: Values of $I^*(\alpha, \beta)$.

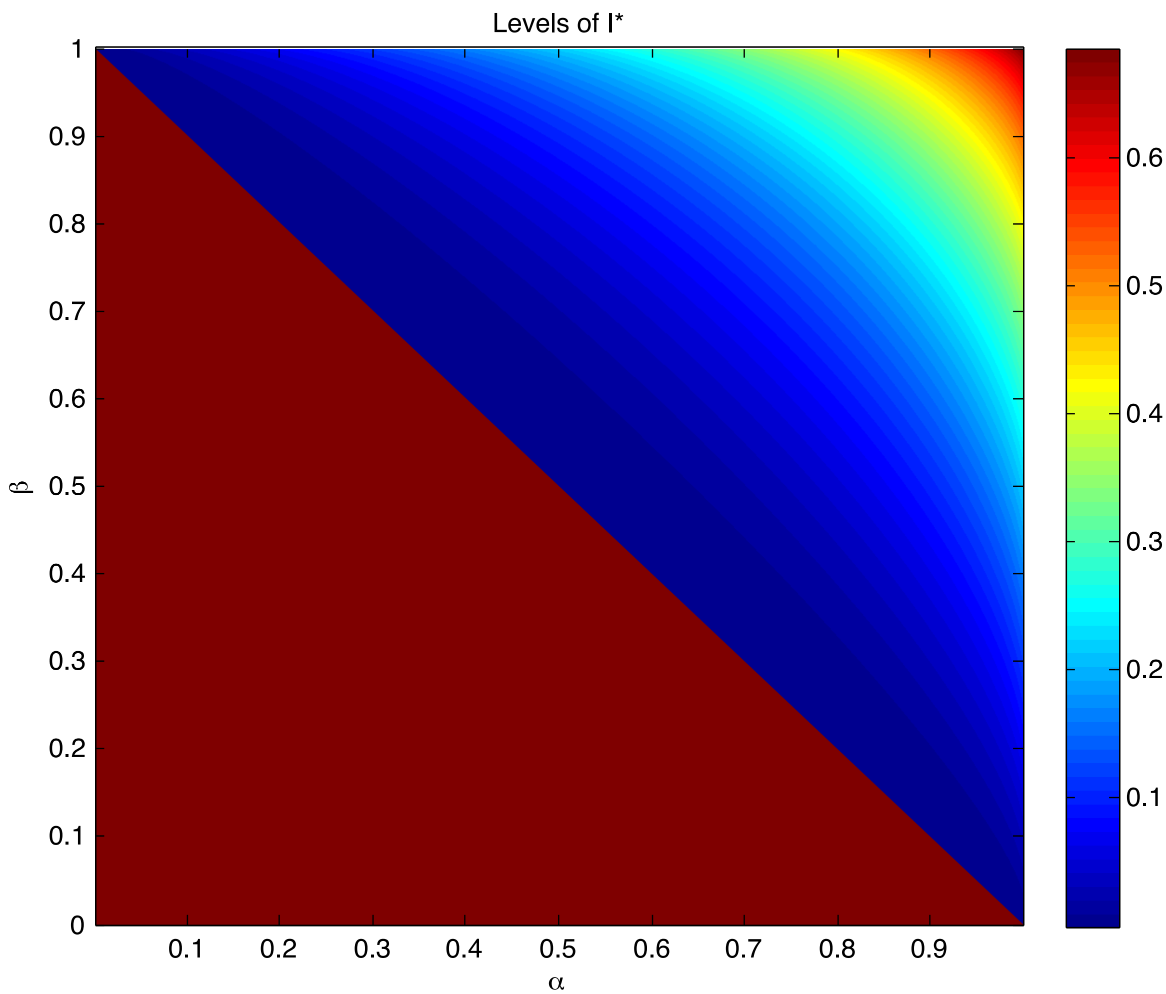
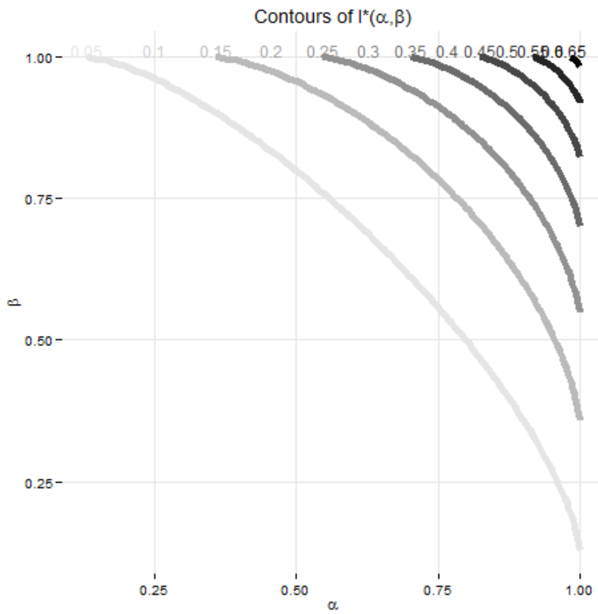


Figure 2: Isoquants for I^*



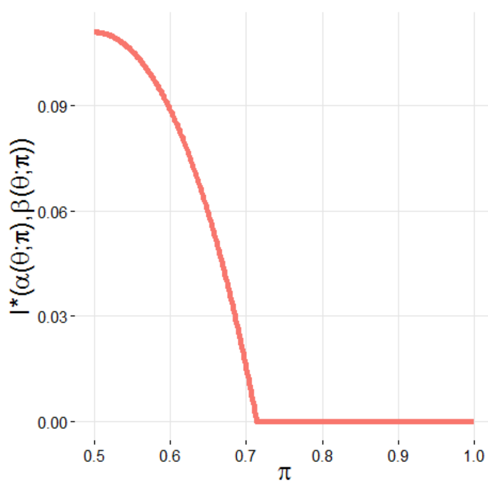


Figure 3: Information capacity allocated to the problem, as a function of the prior on state A for $\theta = 1$.

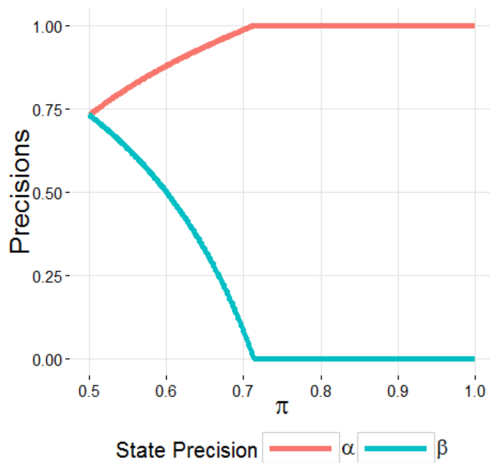
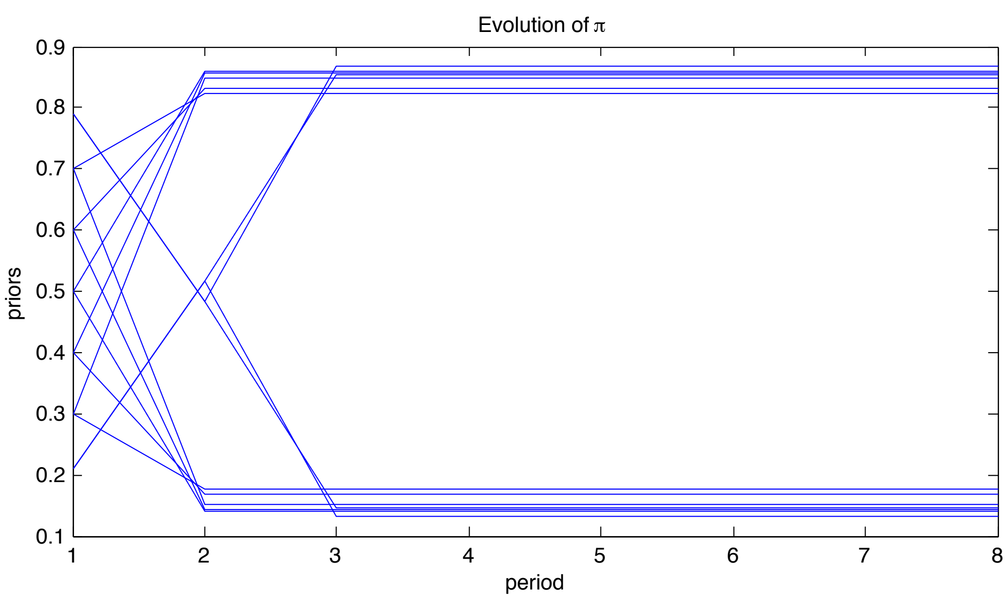


Figure 4: Optimal information structure (α, β) *as a function of the prior on state A for* $\theta = 1$.

Figure 5: Evolution of different starting values of π *over 8 periods where* $\theta = 10$



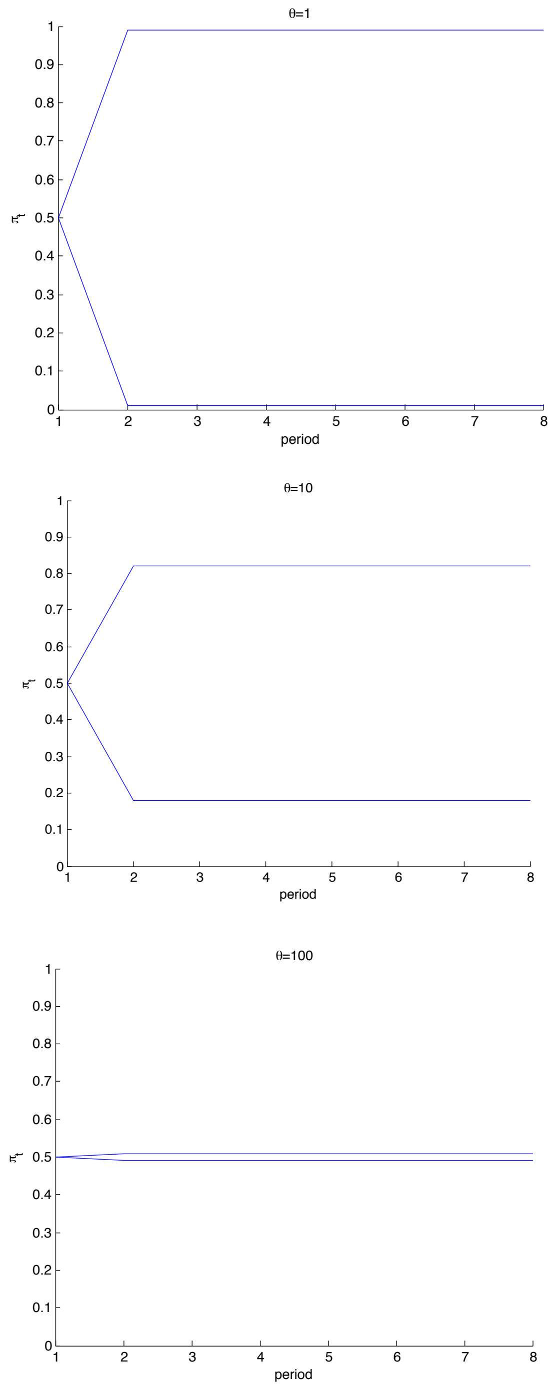


Figure 6: Evolution of π for agents under different values of θ.

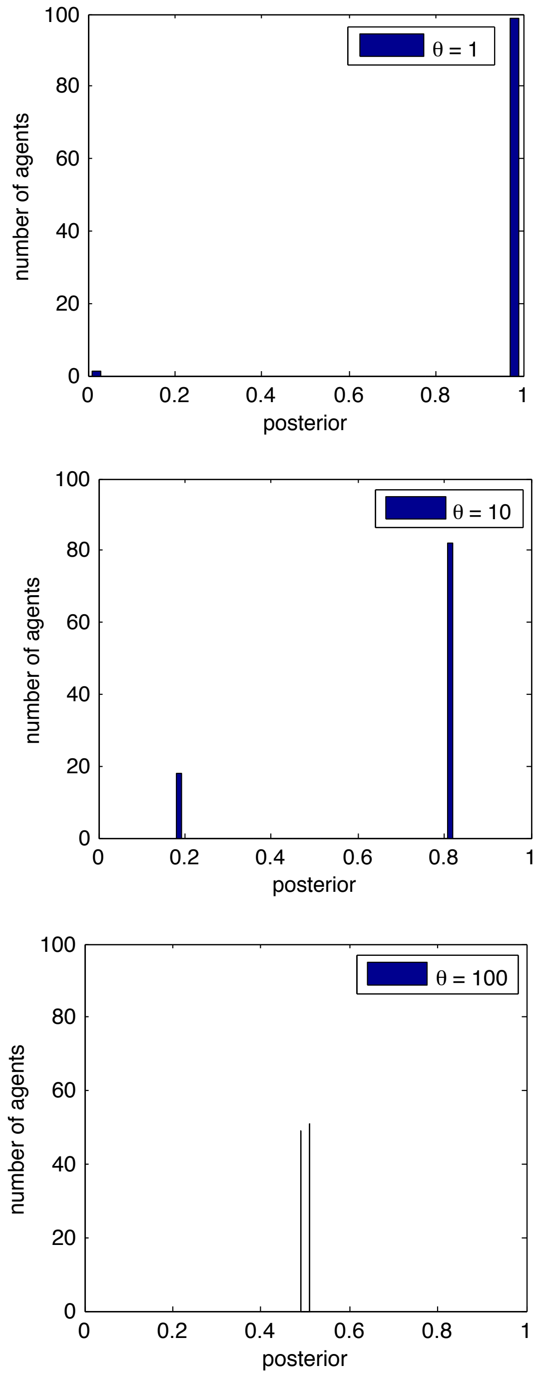


Figure 7: Finishing histograms for 100 agents starting at $\pi_0 = 0.5$ for different values of θ

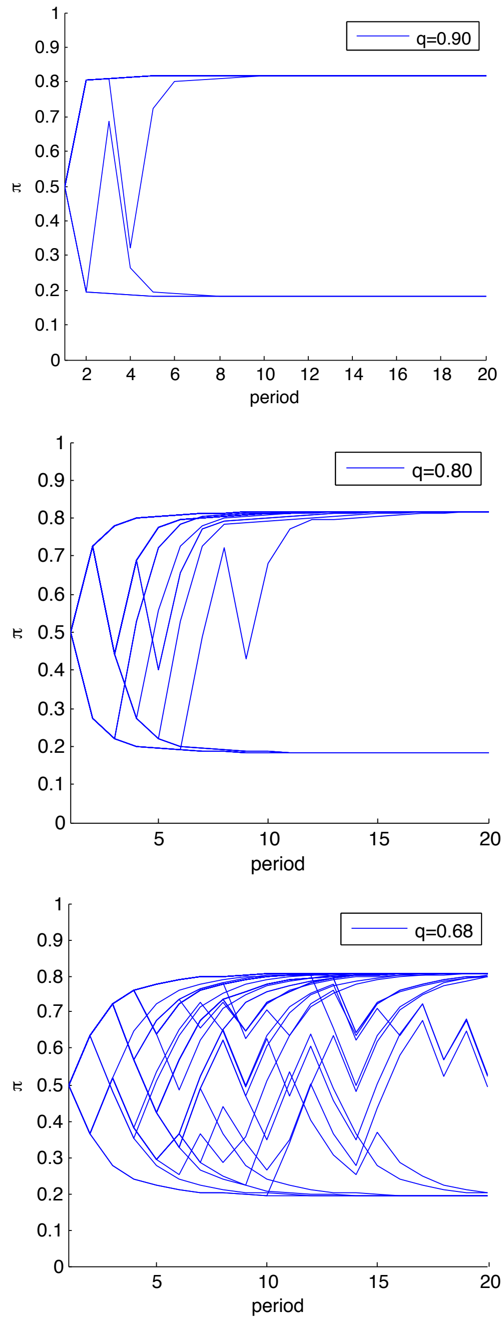


Figure 8: Evolution of priors for 50 agents starting at $\pi_0 = 0.5$ and $\theta = 10$ for different values of q.