INATTENTION AND BELIEF POLARIZATION
PRELIMINARY AND INCOMPLETE

KRISTOFFER P. NIMARK AND SAVITAR SUNDARESAN

Abstract. Disagreement persists over issues that have objective truths. In the presence of increasing amounts of data, such disagreement should vanish, but it is nonetheless observable. This paper studies persistent disagreement in a model where Bayesian rational agents learn about an unobserved state through noisy signals. We show that agents (i) choose signal structures that are more likely to reinforce their prior beliefs and (ii) choose signal precisions that are inversely related to the precision of their prior beliefs. We call the former the *confirmation effect* and the latter the *complacency effect*. The complacency effect may lead agents to stop updating their beliefs entirely, leading to permanent disagreement among agents about the true state of nature. Taken together, the confirmation and complacency effect implies that the beliefs of ex ante identical agents over time tend to cluster in two distinct groups near the boundaries of the belief space. The complacency effect is stronger and more general when information cost is proportional to channel capacity, compared to when it is proportional to reduction in entropy. We argue that in some settings, it may be more natural to model the cost of attention as proportional to channel capacity, rather than the common practice of specifying information cost as proportional to entropy reduction.

1. Introduction

There is ample evidence that economic agents often disagree about the probability of a given event occurring, and about whether a particular stated fact is true or not. For instance, there appears to be vastly divergent opinions among the general population about whether climate change is real and caused by human activities. One possible reason for such disagreement is that different agents have access to different information. However, most existing models suggest that over time as more signals are observed, the beliefs of different agents should converge even if they observe different signals as long as the signals are not biased. But that then raises the question, why would agents choose to observe biased information? In this paper we propose a theory that can explain why rational Bayesian agents may disagree permanently in a model where the agents can choose information channels that have different precisions in different states of the world. This result is driven by two effects. The first, which we call the *confirmation effect*, causes agents to choose to observe signals that are more precise in a priori more likely states and therefor also more likely to confirm their

Date: February 15, 2018, Nimark: Economics Department, Cornell University. *e-mail*: pkn8@cornell.edu *webpage*: www.kris-nimark.net. Sundaresan: Imperial College London. *e-mail*: s.sundaresan@ic.ac.uk, *webpage*: http://www.savitars.com. The authors thank Alessandra Casella, Navin Karthik, Alex Kohlhas, Thomas Palfrey, Ricardo Reis, Tomasz Strzalecki, Sebastien Turban, Laura Veldkamp, Michael Woodford, participants at seminars at NYU, Columbia, and Caltech, and conferences at Stanford and Oxford, as well as an anonymous referee, for helpful comments and suggestions.

prior beliefs. If two agents initially observe different realizations of signals drawn from the same distribution, the confirmation effect then makes it less likely that agents beliefs will converge over time. The second effect, which we call the *complacency effect*, makes agents' choose less precise channels as their uncertainty decreases. This effect causes agents to choose completely uninformative signals when their posterior beliefs are sufficiently precise. Combined, these two effects imply that the beliefs of different agents do not converge to a degenerate distribution.

A basic premise of this paper is that paying more attention allows agents to perceive reality more accurately. In our model, there is an objective, underlying reality that determines what is true and what is not. However, agents cannot observe this reality directly and have to choose channels that provide signals about the underlying true state of the world. More precise signals are more costly and agents can choose channels that will provide signals of different precision in different states of the world. One way to interpret this setup is that agents choose the medium through which a fact will be channeled. For example, there may be certain facts about the world that we are all trying to verify, such as understanding whether global warming exists. Informative sources require a lot of attention, while uninformative sources require very little: It may be harder to extract a signal from an article in *Nature* than from a sound byte on cable television. Some channels may also be more accurate in one state of the world, while other channels may be more accurate in other states. If precision is costly, agents want to allocate more precision to those states they find a priori more likely. That means that agents endogenously choose channels that are less likely to prove their priors wrong. The endogenous allocation of precision may then perpetuate disparate beliefs and lead to permanent disagreement about the true state of the world.

The early literature on costly information acquisition treated information as the scarce resource, e.g. Grossman and Stiglitz (1980). The key conceptual shift introduced by Sims (1998) and Sims (2003) was that information may be plentiful, but people's *attention* is a scarce resource. Beyond this basic change of perspective, the Rational Inattention literature has made specific assumptions about how to model the cost of attention. The most common formulations either put a hard constraint on the reduction in entropy that agents experience by observing a signal, or a utility or pecuniary cost that is increasing in the reduction in entropy, e.g. Sims (1998), Sims (2003), Maćkowiak and Wiederholt (2009), Woodford (2009), Matějka (2015) and Matejka and McKay (2015).[1] In our set up, as in the Rational Inattention literature, agents choose how precise their information is in different states of the world. However, we study to different specifications for the cost-of-information functions that both build on Shannon's (1948) entropy concept. We contrast the standard approach of modeling cost as proportional to the reduction in entropy as measured by the mutual information between the state and the signal to an alternative cost function that is proportional to channel capacity.[2] Channel capacity can be interpreted as the maximum possible entropy

[1]It is common in the Rational Inattention literature to not model the signals explicitly but instead to focus on the mutual information between actions and latent states. This is without loss of generality and is inconsequential for the argument here.

[2]While most of the existing literature has used modeled cost as being driven by the reduction in entropy (i.e. the mutual information between signals and states), one exception is Woodford (2012) who specifies a cost function that depends on channel capacity.

reduction that any agent could achieve by observing a given signal and using it to model information cost leads to different predictions about behavior.

One way to intuitively understand how measuring the quantity of information in terms channel capacity differs from mutual information is as follows. Consider the cost of watching a TV show about the causes of global warming. If we model the cost of watching the show in terms of mutual information, the cost would be very low for a climate scientist who already knows a lot about the topic. However, watching the same TV show would be very costly for a person without any schooling at all. This seems reasonable: In many situations a person who is already better informed will have a lower cost of absorbing information. However, in some situations it may be more natural to think of the costs as being determined by the maximum any person could learn from a text or broadcast. For instance, if we think of the attention cost as corresponding to time, it is less clear that it is less costly for an informed person to watch a broadcast than it is for an uninformed person. A longer broadcast that may contain more information is then more costly to watch, but also potentially more informative. It may be the case that some information is already known by an informed agent, but unless the agent knows exactly at which points in the broadcast already known information will be revealed, it is as costly for the well-informed as for the uninformed to watch the broadcast. This latter scenario corresponds to modeling the cost of information in terms of channel capacity.

Using mutual information as a measure of information may thus be more reasonable in settings where agents can design their signal structure to completely avoid repetition of already known information. However, using channel capacity may be more reasonable in settings where agents can choose the precision of information but cannot avoid repetition of already known information. Both approaches has appealing features and in this paper we do not take a stand on when one should use which approach. Instead, our focus is on demonstrating the different predictions that the two approaches imply about how agents form beliefs and how these evolve over time. In particular, we demonstrate that modeling information cost in terms of channel capacity imply that when agents' priors are sufficiently precise, they are more likely to choose completely uninformative signals and thus stop updating their beliefs.

To understand why the two cost functions imply different behavior, consider the decision problem of an agent who wants to decide whether a given signal is worth paying attention to or not. When the precision of the agent's prior is higher, the marginal value of observing an additional signal decreases since the agent is already pretty sure about the state of the world. The expected reduction in uncertainty from observing the signal is then small. But if the cost of the signal is measured in terms of entropy reduction, a given signal also becomes cheaper as the prior precision of an agent's beliefs increases. In the limit with infinitely precise priors, any signal regardless of its precision can be observed for free. On the other hand, if the cost of paying attention to a given signal depends only on the precision of the signal, agents demand less and less precise signals as the precision of their beliefs increase since the marginal usefulness of the signal then decreases. This implies that when an agent's beliefs are precise enough, he will choose to observe completely uninformative signals. The agent's beliefs are then not updated further. Importantly, an agent may stop updating his beliefs before they become degenerate on the true state of the world. In fact, even an agent that attaches a higher probability to the incorrect state than to the correct state of the world

may stop updating his beliefs. Because agents only stop updating their beliefs when they are sufficiently precise, beliefs of different agents may permanently cluster in two distinct groups, where one group is almost certain that one state has occurred and the other group is almost certain that it has not.

Our paper is not the first to propose a theory that can explain why disagreement among agents may persist. The starting point of many studies is a well-known result by Savage (1954). He argued that repeated observation of signals will lead a Bayesian agent to assign probability 1 to a true event almost surely as his length of experience increases. Even if different agents start off with different priors, their beliefs should then converge over time. Underlying this result is an assumption that the true state of nature is assigned a positive probability a priori. Blackwell and Dubins (1962) built on Savage's result and showed that as long as the two agents priors are absolutely continuous, then their beliefs will also converge over time.

One way to break the result that beliefs converge over time is to assume that agents assign a zero probability to some states that may in fact occur. For instance, Freedman (1963) and Miller and Sanchirico (1999) show that relaxing the assumption of absolute continuity between the subjective and the true distribution weakens the result. In related work, Berk (1966) shows that if an agents use an incorrect model, his beliefs may not converge to single point. Our set-up generates persistent disagreement without such restrictions. Our agents are rational Bayesians and the support of their prior beliefs contain the true state of nature.

Gentzkow and Shapiro (2006) present a model that, like our model, is populated with agents that are rational Bayesians. They propose that if newspapers are rewarded for perceived accuracy of their reporting, they will bias their reporting to conform to agents' (possibly incorrect) priors. The mechanism behind this result is that the perceived accuracy of an information source is decreasing in the distance between signals and priors. Rational agents then perceive information sources that confirm their priors to be more accurate. In our model, agents know how precise their information is, and yet, they still choose channels that are more likely to confirm their prior beliefs.

The confirmation effect that make agents choose channels that are more likely to confirm their priors is related to other forms of confirmation biases that have been studied in the literature. Nickerson (1998) defines confirmation bias as "the unwitting selectivity in the acquisition and use of evidence". Models of confirmation bias include such as Suen (2004), Cukierman and Tommasi (1998), Rabin and Schrag (1999), Koszegi and Rabin (2006). Baliga et al. (2013) where they show that divergence cannot occur in a Bayesian updating framework. Fryer et al. (2013) assume that agents receive ambiguous signals which they interpret as signals in favor of their prior, and keep only this interpretation in memory. In this paper, the selectivity is not 'unwitting' but intentional and thus provides a theory to explain the observed behaviour without relying on behavioural biases.

One way to interpret the choice of channels in our model is to think o fit as a choice about which sources to get information from, or which news papers to read or what TV channels to watch. Gentzkow and Shapiro (2006) and Besley and Prat (2006) show competition among information providers makes it difficult to hide information, and Hong and Kacperczyk (2010) shows it can decrease reporting bias. Mullainathan and Shleifer (2005) show that if behavioral agents prefer news that favors their beliefs, media sources will tend to be biased.

This paper can deliver the same demand story while using rational agents. Additionally, this paper provides two dimension over which channels are graded: overall informativeness and bias.

Some papers such as Lord et al. (1979), Baumeister and Newman (1994), show that people pay less attention to information confirming their prior and evaluate "disconfirming evidence" more thoroughly. This paper's model finds that if agents receive disconfirming signals, they update their beliefs quite strongly, which is consistent with this result.

The next two sections present the basic set up and describes an agent's optimal channel choice. There, we formally derive the confirmation and the complacency effects and show that for sufficiently precise priors, agents will choose a completely uninformative channel. We then generalize these results to a dynamics setting where agents choose the precision of the current channel while taking into account that more precise information today will also increase future utility. We use the dynamic set up to show that the confirmation and complacency effects endogenously generate persistent disagreement in a population of ex ante identical agents. The dynamic model is extended to study belief dynamics when some of the channel noise is irreducible. We also show that if all agents care more about being right in one state rather than the other, then disagreement decreases. However, if different agents care more about being right in different states, then persistent disagreement becomes more likely.

2. A model of hidden states and costly channels

We start by presenting a simple set up in which a single agent wants to determine the state of the world. Different states of the world cannot be distinguished by direct observation and the agent instead needs to rely on noisy channels. The agent can choose how informative the channel is subject to a cost, and more informative channels are more costly.

Costly information acquisition can be modeled in several ways. We use the set up presented here to study two cost of information functions that both build on Shannon's 1948 entropy concept. *Mutual information* is a measure of how much information an agent with a specific prior on average learns from observing a given signal. *Channel capacity* is a measure of how much any agent can learn from a given signal. It is defined as the maximum mutual information of a given signal over all priors. Stated differently, channel capacity measures how much a hypothetical minimally informed agent on average learns from a given signal.

To date, most of the literature has followed Sims (2003) and used mutual information to specify the cost of information, e.g. Mackowiak and Wiederholt (2009), Woodford (2009), Matejka (2015, 2016) and McKay and Matejka (2016). An exception is Woodford (2012) who uses channel capacity to model consumer choices. In this section, we formally define mutual information and channel capacity and discuss some of their properties that are important for understanding the results in the sections that follow.

2.1. **States and agent utility.**

Nature draws the state of the world $\omega \in \{0,1\} \equiv \Omega$ from the initial distribution $p(\omega)$. An agent cannot observe the state of nature directly and we denote an agent's prior that the fact is true as $\pi(\omega)$.

An agent can pay a cost to observe the signal s about the state of the world. We will refer to the information structure defined by the error probabilities ϵ_0 and ϵ_1 that determine the

probability of observing a signal $s \neq \omega$ as a *channel* through which the agent gets information about the true state of nature.

Definition 1. *A channel* $S \in (0,1) \times (0,1)$ *is defined by the error probabilities* $\epsilon_0 \equiv p(s = 1 \mid \omega = 0)$ *and* $\epsilon_1 \equiv p(s = 0 \mid \omega = 1)$.

Conditional on the signal s agents use Bayes rule to update their beliefs. We denote the posterior belief of an agent as $\pi(\omega \mid s)$. We assume, without loss of generality, that $\epsilon_0 + \epsilon_0 \leq 1$ so that $p(s = \omega) \geq \frac{1}{2}$.

Agents choose error probabilities ϵ_0 and ϵ_1 in order to maximize the objective function Φ

$$\Phi(S, \pi) = U(S, \pi) - \theta \Gamma(S, \pi) \tag{2.1}$$

where U is the utility an agent gets from having accurate beliefs about the state ω

$$U(S, \pi) \equiv 1 - E\left[(\omega - p(\omega \mid s))^2 \mid S, \pi\right].$$

An agent thus gets one unit of utility for correctly assessing the state, and nothing for being completely wrong. Γ is the cost associated with the channel S and θ is a parameter that scales the cost function Γ. Below, we will analyze the implications of two different cost functions that both build on Shannon's (1948) concept of entropy to quantify information.

2.2. **Information, entropy and channel capacity.** Shannon (1948) proposed to define the amount of information I contained in a message that is observed with probability p as

$$I \equiv \log \frac{1}{p} = -\log p. \tag{2.2}$$

This definition is appealing: Being told that something unexpected happened is more informative than being told about an event occurring that you already assigned a high probability to. Entropy $H(\Omega)$ of a source (or random variable) $\omega \in \Omega$ is defined as the expected amount of information

$$H(\Omega) = \sum_{\omega \in \Omega} p(\omega) \log \frac{1}{p(\omega)}. \tag{2.3}$$

In a binary state model, entropy is maximized when both states are equally likely, i.e. when $p(\omega) = \frac{1}{2}$. Entropy is minimized when one state is certain, i.e when $p(\omega)$ equals 0 or 1.

Conditional entropy $H(\Omega \mid S)$ quantifies how much uncertainty remains about Ω after observing S. It is defined as

$$H(\Omega \mid S) = \sum_{s \in S} \sum_{\omega \in \Omega} p(\omega, s) \log \frac{1}{p(\omega \mid s)}. \tag{2.4}$$

Conditional entropy (2.4) can be used to compute the mutual information $I(\Omega, S)$

$$I(\Omega, S) = H(\Omega) - H(\Omega \mid S) \tag{2.5}$$

which quantifies how much information is revealed about Ω by observing S. In a binary state model, mutual information is maximized when $p(\omega) = \frac{1}{2}$ and $p(\omega \mid s)$ equals 0 or 1 in all states.

An information channel $p(S \mid \Omega)$ is defined by a probability distribution of signals conditional on different states. The channel capacity C is the maximum reduction in entropy

that can be achieved by applying the channel to a (source) distribution $p(\omega)$. In our binary state setting, it thus defined as

$$C\left(p(S \mid \Omega)\right) = \max_{q(\omega)\in(0,1)} I(\Omega, S) \tag{2.6}$$

where the maximum is taken over all possible beliefs $q\left(\omega\right) \in (0,1)$. When computing the channel capacity, the precision of the channel, i.e. the distribution $p(S \mid \Omega)$ of signals conditional on the state, is taken as given. Below we will study the optimal channel choices under two different specification of the cost function Γ. To ease notation, we will denote this function as $\Gamma_I(\Omega, S, \pi) \equiv I(\Omega, S)$ when the channel cost is proportional to the mutual information of the signal and the state, and as $\Gamma_C(\Omega, S) \equiv I(\Omega, S)$ when the channel cost is proportional to the channel's channel capacity.

2.3. **Properties of mutual information and channel capacity.** Using mutual information or channel capacity to model the cost of information Γ in agents' utility function have different implications for agent's optimal actions. The properties of the two functions that are most important for our purposes are summarized by the following four lemmas.

Lemma 1. *Mutual information $I(\Omega, S)$ and channel capacity $C(\Omega, S)$ are both convex in the precision of signals $(1-\epsilon_0)$ and $(1-\epsilon_1)$.*

Proof. That mutual information is convex in the precision of signals is a standard result, e.g. Theorem 2.7.4 in Cover and Thomas (2006). That channel capacity is convex follows from that the (point-wise) maximum of convex functions is a convex function. □

Lemma 1 follows directly from the definitions of mutual information and channel capacity. It implies that regardless of whether we model the cost of information using mutual information or channel capacity, the marginal cost of information is increasing in the signal precisions $(1-\epsilon_0)$ and $(1-\epsilon_1)$. The lemma also implies that the marginal cost of precision in one state is increasing in the precision of the channel in the other state.

Mutual information $I(\Omega, S)$ and channel capacity $C(\Omega, S)$ thus share some properties. However, as information cost functions, they have different implications for what constitutes an optimal channel. Most importantly, this is so because mutual information depends on an agent's prior beliefs while channel capacity does not. The following lemma states a general property of the relationship between mutual information and the prior distribution that is important for understanding an agent's choice of channel precision.

Lemma 2. *The mutual information implied by a given channel S is concave in the prior $\pi(\omega)$ everywhere on $(\epsilon_0, \epsilon_1) \in (0,1) \times (0,1)$*

Proof. See Theorem 2.7.4 in Cover and Thomas (2006). □

Again, Lemma 2 is a standard result and follows directly from manipulation of the definition (2.5). Importantly, the lemma implies that the marginal cost of information is decreasing as an agent's prior $\pi(\omega)$ tends to either 0 or 1. As mentioned above, the same is not true for channel capacity. Since channel capacity is defined as the maximum mutual information that can be achieved between a given channel S for any prior π, channel capacity cost is by construction independent of the precision of an individual agent's prior.

Lemma 3. *The channel capacity of a given channel S is independent of an agent's prior.*

Proof. The lemma follows immediately from the definition (2.6) of channel capacity. □

The last properties of mutual information and channel capacity that we discuss here relates to how these functions behave at the boundaries of the precision of the channel S and prior beliefs $\pi(\omega)$.

Lemma 4. *Mutual information $I(\Omega, S)$ tends to zero when $\pi(\omega)$ tends to either 0 or 1 or when the channel S tends towards being completely uninformative. Channel capacity $C(\Omega, S)$ tend to zero only when the channel tend towards being completely uninformative.*

Proof. In the Appendix. □

Lemma 1 above is useful for ensuring that a solution to the agent's optimization problem exists and can be characterized by standard first order conditions (?). The properties described by Lemma 2, 3 and 4 are important for understanding why the two cost functions imply different predictions about an agent's channel choice, and in extension, his beliefs. In particular, the fact that information cost modeled as channel capacity is independent of an agent's prior while cost modeled as mutual information is decreasing in the precision of an agent's prior precision is important for the results that follow.

3. Optimal channels and beliefs

In this section we study the optimal choice of channel while taking the prior distribution $\pi(\omega)$ as given. We show that agents will choose channels that: (i) confirm her existing beliefs in expectation (ii) are less informative the stronger her beliefs and (iii) at some sufficiently strong beliefs, are completely uninformative. In the next section we generalize these results to a dynamic setting where an agent's prior in period $t+1$ is his posterior beliefs from period t. In what follows, we denote the parameters of an optimally chosen channel S^* as ϵ_0^* and ϵ_1^*.

3.1. Priors and optimal channels.

An agent choose the parameters ϵ_0^* and ϵ_1^* that determine the precision of the channel. An optimally chosen channel equates the expected marginal benefit of increasing the precision in given state with the marginal cost of doing so. To achieve this, an agent must trade off a lower error probability in one state against a lower error probability in the other state. The optimality conditions can be expressed as

$$\frac{\partial U\left(\epsilon_0, \epsilon_1, \pi\right)}{\partial \epsilon_0} - \theta \frac{\partial \Gamma\left(\epsilon_0, \epsilon_1, \pi\right)}{\partial \epsilon_0} = 0 \tag{3.1}$$

and

$$\frac{\partial U\left(\epsilon_0, \epsilon_1, \pi\right)}{\partial \epsilon_1} - \theta \frac{\partial \Gamma\left(\epsilon_0, \epsilon_1, \pi\right)}{\partial \epsilon_1} = 0. \tag{3.2}$$

The expected marginal benefit of a more precise signal is higher in the state that the agent believes is more likely to occur. We can state this result more formally.

Proposition 1. *(Confirmation effect) The optimal informative channel is more precise in the state that is a priori more likely, i.e. $\epsilon_0^* > \epsilon_1^*$ if $\pi(\omega) > \frac{1}{2}$ and $\epsilon_0^* < \epsilon_1^*$ if $\pi(\omega) < \frac{1}{2}$.*

Proof. In the Appendix. □

The expected benefit U of a more precise signal is independent of the cost function Γ and the confirmation effect holds regardless of whether the information cost is given by Γ_I or Γ_C. However, the effect is stronger under Γ_I since a signal that is expected to confirm an agent's prior is then cheaper than a signal that is more likely to under mutual information, while the cost is independent of the agent's prior under channel capacity. Agents thus choose channels that make it more likely that they observe a signal that confirms their prior beliefs.

The expected benefit of a given channel decreases as the precision of an agent's prior is increased. However, the cost Γ of a more precise channel depends on whether cost is modeled as mutual information or channel capacity. Under Γ_I, cost is proportional to the reduction in entropy. A channel of a given precision is then cheaper for an agent that already has a relatively precise prior. Under Γ_C, the cost of a channel is independent of an agent's prior. Because the benefit of a given channel is decreasing in the precision the prior, while the cost is not, agents will display what we call a *complacency effect*. We state this result more formally as follows.

Proposition 2. *(Complacency effect) Under the cost function Γ_C, an agent will choose a less precise channel with lower capacity as the precision of his prior is increased so that*

$$\frac{\partial \Gamma_C\left(\Omega, S^*\right)}{\partial \left|\pi(\omega) - \frac{1}{2}\right|} < 0. \tag{3.3}$$

Proof. In the Appendix. □

The complacency effect states that an agent will choose a less precise channel as he becomes better informed and this holds under general conditions when information cost is proportional to channel capacity. When cost is instead proportional to entropy reduction, the complacency effect applies to some intervals of an agent's prior, but it does not hold generally and limit case illustrates why this is the case. Under Γ_I, the cost of any channel tend to zero as the precision of the prior increases. Since the marginal usefulness of the channel also tend to zero, in the limit, an agent is then indifferent between different channels. Unlike under Γ_C, the precision of a channel is thus not a monotonically decreasing function of the precision of the prior under Γ_I.

The complacency effect is sufficiently strong under Γ_C so that for some sufficiently precise prior, an agent will choose a completely uninformative channel.

Proposition 3. *(Information shutdown) Under Γ_C there exists sufficiently precise priors $\overline{\pi}(\omega) \in [\frac{1}{2}, 1)$ and $\underline{\pi}(\omega) \in (0, \frac{1}{2}]$ so that an agent will choose a completely uninformative channel.*

Proof. In the Appendix. □

Proposition 3 states that there exists prior beliefs, $\overline{\pi}$ and $\underline{\pi}$ away from the boundary set $(0, 1)$ such that an agent will choose a completely uninformative channel. Clearly, at a boundary, where an agent has a perfectly precise prior, the expected marginal usefulness of any channel regardless of precision, is zero. To prove the proposition, it is sufficient to show that the marginal benefit of increasing precision at the boundary is lower than the marginal cost of doing so for all values of $\epsilon_0, \epsilon_1 < 1$. The result then follows from the continuity of both U and Γ in ϵ_0 and ϵ_1.

The optimal channel choices are illustrated in Figure 1 where we have plotted the optimal error probabilities ϵ_0^* and ϵ_1^* as functions of the prior $\pi(\omega)$. An agent choose the most precise channel when his prior is most uncertain, i.e. when $\pi(\omega) = 1/2$. When an agent becomes more certain about the state, he will increase the precision of his channel in that state and decrease the precision in the other. In the figure, this can be seen from the fact that the optimal error probability ϵ_1^* (blue line) is decreasing in $\pi(\omega)$ and that the optimal error probability ϵ_0^* (red line) is increasing in $\pi(\omega)$. When the prior is sufficiently precise, i.e. when $\pi(\omega) < 0.15$ or when $\pi(\omega) > 0.85$, the agent chooses a completely uninformative channel and $\epsilon_0^* = \epsilon_1^* = 1/2$. Here, we have now shown that regardless of cost function, an agent will

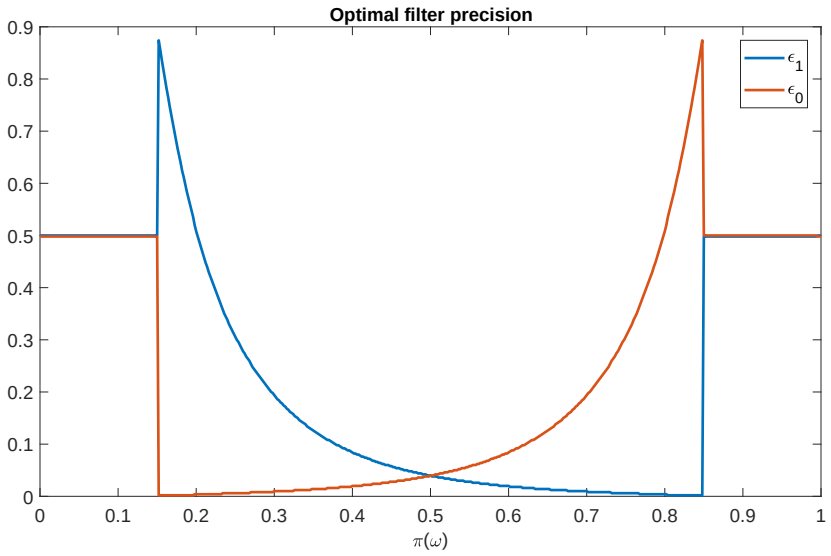


FIGURE 1. Optimal channel precision as a function of prior beliefs for $\theta = 0.2$.

choose a channel that is more precise in the state of the world that is a priori more likely. As the precision of the prior is increased, an agent who's information cost is proportional to the channel capacity of his channel will always choose a less informative channel. We call this the complacency effect, and we showed that it leads an agent to choose a completely uninformative channel when his prior becomes sufficiently precise. In the next section, where we study a dynamic setting, we will show that the confirmation and complacency effects will make agents choose channels that lead to permanent disagreement about he state ω.

4. Optimal channels and divergence of beliefs

The analysis above treated the agent's prior $\pi(\omega)$ as given. Here, we study how an agent's beliefs evolve over time as he recursively choose the precision of his channel. We also study how the distribution of beliefs across a population changes over time. In particular, we show that beliefs of different agents will endogenously diverge to a time-invariant distribution in which some agents permanently attach a higher probability to $\omega = 1$ while other agents permanently attach a higher probability to $\omega = 0$.

4.1. **Dynamic set up.** Time is discrete an indexed by $t = 0, 1, 2, ...$ The state ω is drawn by nature at time zero. Agents are indexed by $i \in N$ and agent i enters period t with prior beliefs $\pi\left(\omega \mid s_i^{t-1}\right) \equiv \pi_{i,t-1}$, where s_i^{t-1} is the history of signals observed by agent i. We denote agent i's beliefs in period zero, before he has observed any signals, as $\pi_{i,0}(\omega)$.

4.2. **The dynamic channel choice.** In each period, agent i chooses the precision of the channel $S_{i,t}$, defined by the error probabilities $\epsilon_{0i,t} \equiv p(s_t = 1 \mid \omega = 0)$ and $\epsilon_{1i,t} \equiv p(s_t = 0 \mid \omega = 1)$. Agent i chooses his channel to maximize the sum of the within-period objective function Φ and the discounted expected continuation value. The optimization problem can thus be expressed as a Bellman equation with the prior $\pi_{i,t}$ as the state variable

$$\mathbb{V}\left(\pi_{i,t}\right) = \max\left\{U\left(\epsilon_{0i,t}, \epsilon_{1i,t}, \pi_{i,t-1}\right) - \theta\Gamma(\epsilon_{0i,t}, \epsilon_{1i,t}, \pi_{i,t-1}) + \delta E_{i,t}\left[\mathbb{V}\left(\pi_{i,t+1}\right)\right]\right\}. \tag{4.1}$$

The parameter $\delta \in (0, 1)$ is the rate at which an agent discount future utility. $U\left(\epsilon_{0i,t}, \epsilon_{1i,t}, \pi_{i,t-1}\right)$ denotes the period t utility of the channel $S_{i,t}(\epsilon_{0i,t}, \epsilon_{1i,t})$ conditional on the prior $\pi_{i,t-1}$. The function $\theta\Gamma(\epsilon_{0i,t}, \epsilon_{1i,t}, \pi_{i,t-1})$ denotes the associated period t cost.

Proposition 4. *(The precision of channels) For a given prior $\widehat{\pi}$, the optimal channel in the dynamic model is at least as precise as the optimal channel in the static model.*

Proof. First, note that a more precise signal today translates into a more precise expected prior and a higher expected period utility in the next period so that

$$\frac{\partial E_{i,t}\left[U\left(\epsilon_{0i,t}, \epsilon_{1i,t}, \pi_{i,t-1}\right)\right]}{\partial\left|\pi_{i,t} - \frac{1}{2}\right|} > 0. \tag{4.2}$$

Since the expected cost of future channels is independent of the prior we have that

$$\frac{\partial E_{i,t}\left[\mathbb{V}\left(\pi_{i,t+1}\right)\right]}{\partial\left|\pi_{i,t} - \frac{1}{2}\right|} > 0 \tag{4.3}$$

implying that

$$\epsilon_{0i,t}^*\left(\widehat{\pi}\right) \leq \epsilon_0^*\left(\widehat{\pi}\right), \epsilon_{1i,t}^*\left(\widehat{\pi}\right) \leq \epsilon_1^*\left(\widehat{\pi}\right)$$

□

Proposition 5. *(Dynamic information shutdown). For a sufficiently precise prior $\pi\left(\omega \mid s_i^{t-1}\right)$ an agent will choose a completely uninformative channel in period t.*

Proof. In the Appendix. □

The intuition for the information shutdown in the dynamic model is the same as in the static set up. With a sufficiently precise prior coming into the period, the marginal benefit of any informative channel is smaller than its marginal cost.

Proposition 6. *(Instant information satisfaction) For a sufficiently imprecise prior, an agent will choose a precise enough channel to reach the information shutdown point within the same period. That is, there exists a prior $\pi\left(\omega \mid s^{t-1}\right)$ such that $\pi\left(\omega \mid s^t\right) > \overline{\pi}_t$ or $\pi\left(\omega \mid s^t\right) < \underline{\pi}_t$ implying that $C\left(\Omega, S_{t+1}^*\right) = 0$.*

Proof. In the Appendix. □

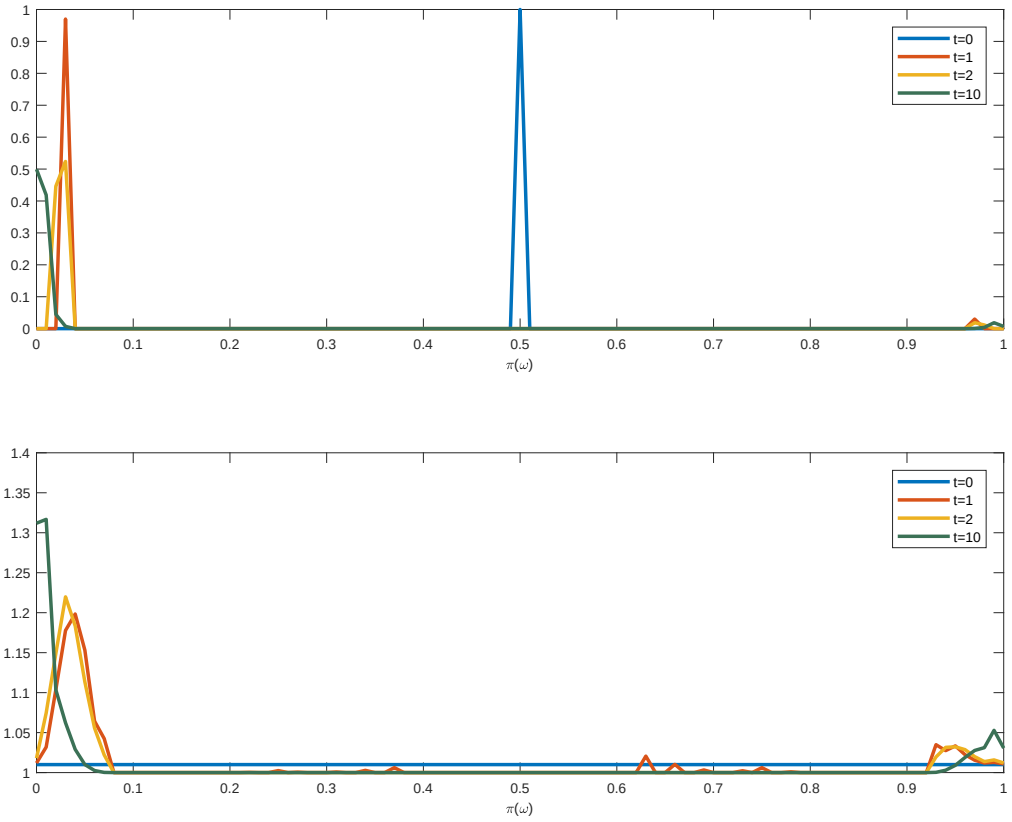


FIGURE 2. Belief dynamics with $\omega = 0$. Blue lines describe initial distribution of beliefs, red lines illustrate distribution of posterior beliefs in period 1, yellow lines illustrate posterior beliefs in period 2 and green lines illustrate posterior beliefs in period 10. In the top panel, the simulation is initialized from degenerate distribution of priors, i.e. $\pi_{i,0}(\omega) = 1/2 \forall i$. Bottom panel is initialized from a uniform distribution of priors, i.e. $\pi_{i,0}(\omega) \sim U(0,1)$.

4.3. **Population dynamics of beliefs.** We can illustrate the dynamics of agent's beliefs and the endogenous permanence of different beliefs by simulating the model. In Figure 2 below, we have plotted the initial distribution of beliefs across the population together with how the distribution of beliefs evolve over time under the assumption that $\omega = 0$. In the top panel, we set the initial distribution of beliefs to be degenerate at $\pi(w) = 1/2$ for all agents. All agents then choose the same error probabilities. A fraction $\left(1 - \epsilon^*_{0i,t}\right)$ then observes the signals $s = 0$ and the fraction $\epsilon^*_{0i,t}$ observes the signal $s = 1$. The fact that agents with prior $\pi(w) = 1/2$ choose relatively precise channels has two consequences. First, it implies that only a small fraction of agents get the wrong signal. Second, all agents have relatively precise posteriors. Together this means that a smaller fraction of the population will believe that the wrong stet is more likely and that the posterior beliefs are quite certain, even when they are "wrong". The latter implies that in the next period all agents will choose completely

uninformative channels. The disagreement that is solely due to that a fraction $\epsilon^*_{0i,t}$ observed a different signal from the rest then become permanent.

In the bottom panel of Figure 2, we set the initial distribution of beliefs to be uniform on the interval $(0, 1)$ and we assumed that $\omega = 0$. After one period, we can see that the mass of beliefs have moved away from the midpoint toward more certain (but not necessarily more correct) beliefs. After 10 periods, the distribution has virtually converged to the limiting distribution. While the mass of beliefs that are close to the correct state is larger than the mass close to the incorrect state, both are positive.

5. Extensions

5.1. **Irreducible noise.** In some circumstances, it may be reasonable to think that the channels available to agents contain a noise component that is irreducible. To model such situations, we use the modified channel $\widetilde{S}_{i,t} \in (0,1) \times (0,1)$ defined by the probabilities $(1 - \epsilon_{0i,t}) \times (1 - \widetilde{\epsilon}_0) \equiv p(s = 0 \mid \omega = 0)$ and $(1 - \epsilon_{1i,t}) \times (1 - \widetilde{\epsilon}_1) \equiv p(s = 1 \mid \omega = 1)$ where $\widetilde{\epsilon}_0, \widetilde{\epsilon}_1 \in (0, 1)$. The parameters $\widetilde{\epsilon}_0$ and $\widetilde{\epsilon}_1$ thus determine an exogenous upper bound on the precision of the channel $\widetilde{S}$.

The main consequence of introducing the irreducible noise is that it takes longer for agents to reach the information shutdown point in the dynamics setting.

[**TBC**]

5.2. **Asymmetric preferences across states.** [**TBC**]

6. Conclusions

In this paper we have studied optimal information choice in a setting where agents can choose how precise their signals are in different state of the world. We showed that regardless of whether the cost of a channel is proportional to an agent's subjective entropy reduction, or to the channel's channel capacity, channel choices will display what we call the confirmation effect, by which agents choose channels that are more precise in the state that they believe to be a priori more likely. Regardless of cost function, agents will also choose channels that are less precise, the more precise their priors are. We call this effect the complacency effect. The complacency effect is stronger when channel cost is proportional to channel capacity. This is so because cost is then independent of an agent's prior, while the cost of a channel is decreasing in the precision of a prior when cost is proportional to entropy reduction. We showed that the complacency effect can be strong enough so that agents stop updating their beliefs entirely when their priors are sufficiently precise. This is not a limit result: Information shutdown can occur even when an agent still attaches a substantial probability to both states. It may also occur for agents that attached a larger probability to the incorrect state than to the correct state. That agents may stop updating their beliefs in this way that make it possible for agents to disagree permanently. permanently disagree about which state of the world is more likely.

In the model, the true state of the world does not affect agents immediately and their utility only depends on how uncertain their beliefs are. The proposed model is thus suitable to study disagreement about facts that agents cannot evaluate the truth of immediately. Consider for instance the question of whether climate change caused by human activity is

real. For most people, there is nothing in their everyday experience that would directly indicate whether the scientific consensus or the climate change sceptics are right. However, the optimal voting and consumption behavior may depend on who is right. Individuals who initially attach a higher probability to climate change being real, will seek out information from sources that are likely to be right if in fact, climate change is real. However, individuals who initially believe that it is Chinese hoax perpetrated in conspiracy with liberal elites and special interest at universities, will seek out information from sources that are more likely to be accurate if these initial beliefs are correct. The mechanism can thus explain why a population may self-sort into different media bubbles.

In this paper we used two different functions to model the cost of acquiring information. Both of them build on Shannon's (1948) concept of entropy and both imply that attention is a scarce resource that should be allocated rationally. However, we saw that letting the information cost be proportional to the subjective reduction in entropy lead to different implications compared to when the cost is proportional to a channel's channel capacity. The main difference arise from the fact that observing a given signal leads to small reduction in entropy for a well-informed agent while the channel capacity of a given signal structure is independent of an agents prior. In this paper, we argued that modeling cost as proportional to channel capacity captures the time cost of reading articles or observing or listening to broadcasts for which the agent does not have complete control over the content. Modelling attention using channel capacity has the potential to also explain other phenomena that rely on agents not updating their beliefs.

References

Baliga, S., E. Hanany, and P. Klibanoff (2013). Polarization and ambiguity. *..., Kellogg CMS-EMS ...*.

Baumeister, R. F. and L. S. Newman (1994, February). Self-Regulation of Cognitive Inference and Decision Processes. *Personality and Social Psychology Bulletin 20*(1), 3–19.

Berk, R. H. (1966, February). Limiting Behavior of Posterior Distributions when the Model is Incorrect. *The Annals of Mathematical Statistics 37*(1), 51–58.

Besley, T. and A. Prat (2006). Handcuffs for the grabbing hand? media capture and government accountability. *The American Economic Review 96*(3), 720–736.

Blackwell, D. and L. Dubins (1962, September). Merging of Opinions with Increasing Information. *The Annals of Mathematical Statistics 33*(3), 882–886.

Cover, T. M. and J. A. Thomas (2006). Elements of information theory 2nd edition.

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

Freedman, D. A. (1963, December). On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case. *The Annals of Mathematical Statistics 34*(4), 1386–1403.

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

Gentzkow, M. and J. M. Shapiro (2006). Media bias and reputation. *Journal of political Economy 114*(2), 280–316.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American economic review 70*(3), 393–408.

Hong, H. and M. Kacperczyk (2010). Competition and bias. *The Quarterly Journal of Economics 125*(4), 1683–1725.

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

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

Maćkowiak, B. and M. Wiederholt (2009). Optimal sticky prices under rational inattention. *The American Economic Review 99*(3), 769–803.

Matějka, F. (2015). Rigid pricing and rationally inattentive consumer. *Journal of Economic Theory 158*, 656–678.

Matejka, F. and A. McKay (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review 105*(1), 272–98.

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

Mullainathan, S. and A. Shleifer (2005). lthe market for newsm. *American Economic Review 95*(4), 1005–1030.

Nickerson, R. 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* (February).

Savage, L. J. (1954). *The foundations of statistics*. Dover Books.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal 27*(3), 379–423.

Sims, C. A. (1998). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, Volume 49, pp. 317–356. Elsevier.

Sims, C. a. (2003, April). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Suen, W. (2004). The Self Perpetuation of Biased Beliefs. *The Economic Journal 114*, 377–396.

Woodford, M. (2009). Information-constrained state-dependent pricing. *Journal of Monetary Economics 56*, S100–S124.

Woodford, M. (2012). Inattentive Valuation and Reference-Dependent Choice. *Working Paper*.

Appendix A. Proof of Lemma 4

[**TBC**]

Appendix B. Proof of Proposition 1

[**TBC**]

Appendix C. Proof of Proposition 2

[**TBC**]

Appendix D. Proof of Proposition 3

Proof. We want to prove that under Γ_C, there exists sufficiently precise priors $\overline{\pi} \in [\frac{1}{2}, 1)$ and $\underline{\pi} \in (0, \frac{1}{2}]$ so that an agent will choose a completely uninformative channel. By symmetry, we only need to prove the statement for $\overline{\pi} \in [\frac{1}{2}, 1)$. We first show that as $\pi \to 0$ the marginal utility of more precise signals tends to zero.

Using the variance of a binary variable $\omega \in \{0, 1\}$ is given by $p(\omega = 1)(1 - p(\omega = 1))$ we can write the utility function as

$$U(S, \pi) = -\pi + \frac{(1 - \epsilon_1)^2 \pi^2}{(1 - \epsilon_1)\pi + \epsilon_0 (1 - \pi)} + \frac{\epsilon_1^2 \pi^2}{\epsilon_1 \pi + (1 - \epsilon_0)(1 - \pi)} \tag{D.1}$$

The marginal utility of ϵ_0 and ϵ_1 are then given by

$$\frac{\partial U}{\partial \epsilon_0} = \frac{(1 - \pi)\epsilon_1^2 \pi^2}{\left[\epsilon_1 \pi + (1 - \epsilon_0)(1 - \pi)\right]^2} - \frac{(1 - \pi)(1 - \epsilon_1)^2 \pi^2}{\left[(1 - \epsilon_1)\pi + \epsilon_0 (1 - \pi)\right]^2} \tag{D.2}$$

$$\frac{\partial U}{\partial \epsilon_1} = \frac{\pi \epsilon_0^2 (1 - \pi)^2}{\left[\epsilon_0 (1 - \pi) + (1 - \epsilon_1)\pi\right]^2} - \frac{\pi (1 - \epsilon_0)^2 (1 - \pi)^2}{\left[\epsilon_0 (1 - \pi) + (1 - \epsilon_1)\pi\right]^2} \tag{D.3}$$

so that

$$\lim_{\pi \to 1} \frac{\partial U}{\partial \epsilon_0} = \lim_{\pi \to 1} \frac{\partial U}{\partial \epsilon_1} = 0 \tag{D.4}$$

Because Γ_C is non-negative and convex in the precision of signals and independent of the prior π, the marginal cost of more precise signals is strictly positive. Taking the limit of the first order conditions thus gives

$$\lim_{\pi\to 1}\left(\frac{\partial U\left(\epsilon_0,\epsilon_1,\pi\right)}{\partial\epsilon_0}-\theta\frac{\partial\Gamma\left(\epsilon_0,\epsilon_1,\pi\right)}{\partial\epsilon_0}\right) \;<\; 0 \tag{D.5}$$

$$\lim_{\pi\to 1}\left(\frac{\partial U\left(\epsilon_0,\epsilon_1,\pi\right)}{\partial\epsilon_1}-\theta\frac{\partial\Gamma\left(\epsilon_0,\epsilon_1,\pi\right)}{\partial\epsilon_1}\right) \;<\; 0 \tag{D.6}$$

for any $\epsilon_0+\epsilon_1\leq 1$. There is thus no filter that satisfies the first order conditions as $\pi\to\infty$ and the optimal choice is then to choose an uninformative filter with cost zero. Because the inequalities in (D.5) and (D.6) are strict, this holds for some $\overline{\pi}$ strictly smaller than 1, proving the proposition. □