Differences of Opinion

Dionissi Aliprantis

First Version: January 2016. This Version: July 4, 2017

Abstract: The relationship between data and beliefs can be difficult to explain. This paper studies how problems identifying causal effects can lead to ambiguity, or a set of possible beliefs, and how an agent might use social learning to resolve that ambiguity. Data from an ideal experiment for identifying causal effects yield a point-valued signal about a binary state of the world. I assume an agent must form beliefs when observing imperfect data yielding set-valued signals. I specify the agent's objective as choosing the belief that she would have formed with data yielding a point-valued signals. I allow the agent to impute missing data using information observed through her network in combination with a model of social learning. In some cases, the agent's belief formation reduces to DeGroot updating, and beliefs in a network reach a consensus. In other cases, the agent's updating can generate polarization and sustain clustered disagreement, even on a connected network where everyone observes the same data and processes that data with the same model.

Keywords: Belief Formation, Subjective Probability, Imprecise Probability, Social Learning, Partial Identification, Causal Inference, Network, DeGroot Learning Rule, Bounded Confidence

---

Contact: Federal Reserve Bank of Cleveland, +1(216)579-3021, *dionissi.aliprantis@clev.frb.org*.

Acknowledgements: Many of the ideas in this paper originated from discussions with Alon Bergman and Gregorio Caetano. I also thank Ben Craig, Michalis Haliassos, Charles Manski, Jan-Peter Siedlarek, Alireza Tahbaz-Salehi, seminar participants at the Cleveland Fed and Goethe University Frankfurt, and ISIPTA '17 conference participants for helpful comments.

1 Introduction

Transforming data into beliefs is not always straightforward. The data arriving between 2006 and 2012 from the United States' National Oceanic and Atmospheric Administration (NOAA) show a clear continuation of the upward trend in average global temperature over previous decades. Yet many more Americans believed there was an upward trend in 2006 than in 2012 (Pew (2012)). Is it possible to rationalize this type of belief updating?

Many cases of inference generate disagreement that can be viewed through the lens of a coarse, binary model. Citing just three examples based on their relevance for policy, there is disagreement on whether: extending unemployment benefits increases unemployment (Hagedorn et al. (2013), Farber et al. (2015)); government spending stimulates the economy (Ramey (2011), Serrato and Wingender (2016)); the US labor market was at full employment in June 2017.

The need for data with specific features can make this type of disagreement especially common for causal inference. The example of neighborhood effects on employment illustrates: We do not have access to data from an ideal experiment randomly allocating households across neighborhoods. The data we do have can be interpreted as evidence of the absence (Ludwig et al. (2013)) or presence (Aliprantis and Richter (2016)) of neighborhood effects.

This paper studies how the obstacles to identifying causal effects can lead to ambiguity, or a set of possible beliefs, and how an agent might use social learning to resolve that ambiguity. To infer missing data from another person, an agent can use their disagreement over multiple issues. In some cases this inductive rule can lead the agent to over-interpret differences of opinion, creating polarization in situations that might be unexpected.

To frame the issues, I first consider the case of inference when there are no problems of identification. An agent observes a sequence of data, where the sampling procedure is identical over time. At each point in time the agent uses her model to translate the observed data into a point-valued signal $\sigma_t \in [0, 1]$ about a binary state of the world $s \in \{0, 1\}$. Beliefs will converge if subjective beliefs $\lambda_t \in [0, 1]$ are formed by taking the average of signals over time.

In the general case when causal effects are not so cleanly identified, we might suppose that an agent observes iid data yielding coarse, or set-valued, signals. This circumstance is widespread in the social sciences, where we rarely observe data from an ideal experiment for identifying causal effects. Beliefs formed by averaging imprecise probabilities over discrete time will converge to a set (Artstein and Vitale (1975)), a scenario of partial identification.

How might a decision maker choose one belief from a set of possible beliefs? When an agent faces ambiguity, prominent decision rules instruct her to choose the single belief generating an extreme utility (Gilboa and Marinacci (2013)).[1] Alternatively, the agent might form beliefs based on data inferred from others' beliefs. This approach to resolving ambiguity separates belief formation from

[1]For example, the maxmin expected utility decision rule maximizes expected utility after choosing the belief that would be set by a malevolent nature minimizing the agent's utility for any decision (Gilboa and Schmeidler (1989)). The minimax regret decision rule maximizes expected utility after choosing the belief maximizing the agent's lost utility from not knowing the true state of the world (Manski (2011)).

preferences, which is natural if the agent's objective is to accurately represent the Data Generating Process. Moreover, looking for information in one's social network is often done in settings of information poverty.[2]

I specify the agent's problem as choosing the belief she would have formed with access to data yielding point-valued signals. Such an agent is committed to the scientific ideal of direct observation, but faces time, resource, or ethical constraints making it infeasible to personally verify the claims in question.

I allow the agent to solve her missing data problem with information observed through her social network. I assume that communication is restricted to signals and beliefs. In the face of such imperfect communication, the agent needs a model of social learning to infer the data she does not observe from the signals she does observe.

I first show that DeGroot (1974) updating, the benchmark model of non-Bayesian social learning, solves a special case of the agent's problem. If the agent addresses her problem of inference with linear opinion pooling of signals, a common method for combining forecasts and estimates, she will follow a DeGroot learning rule under a special case of observed data. Strong assumptions on the data and models in the agent's network are required, however, for DeGroot updating to solve the agent's problem.[3]

The major contribution of this paper is to show that by searching for a general solution to the agent's problem, one can find generalizations of DeGroot updating capable of generating polarization. Such a learning rule is a desideratum of the literature on social learning for its ability to reconcile theory and evidence (Golub and Sadler (2016)). An empirical analogue of a connected network – individuals exposed to sources of information contradicting their beliefs – is often observed together with persistent disagreement.[4] However, DeGroot learning and many of its generalizations converge to a degenerate distribution for connected networks (Jackson (2008)).[5]

I show that while linear opinion pooling can solve the agent's problem, in contrast to standard DeGroot updating, a general solution also requires a first stage in which signals are properly transformed. I present the selection of a model that properly interprets signals as a statistical learning problem, and show that this problem is not well-posed. That is, frictions from communication generate a fundamental problem of inference, in that signals do not convey the same information as directly observed data, and the agent cannot know whether she is properly interpreting signals without this information.

I study how the agent might extract information under the "reasonable" heuristic that differences in interpreting data are consistent across propositions, and that such differences can be in-

[2]For evidence related to the example of neighborhood and school choice, see the ethnography in de Souza Briggs et al. (2008). More empirical examples are given in Section 3.

[3]Individuals can be justified in using different models to interpret the same data (Al-Najjar (2009)), the agent might observe new data over time (Jadbabaie et al. (2012)), and individuals might directly observe different data.

[4]There is persistent disagreement over propositions like Iraq had an active WMD program, President Obama was born in the US, vaccines cause autism, and global warming is occurring despite public debate. This disagreement persists despite exposure to opposing views (Gentzkow and Shapiro (2011)).

[5]Time to consensus, though, is not invariant across all connected network structures (Golub and Jackson (2012)).

ferred from beliefs. I allow the agent to assess the credibility of applying this heuristic to each sender by using the relative entropy of disagreement over all propositions. This gives weight to a sender based not on agreement, but based on understanding how someone interprets data (Sethi and Yildiz (2016)).

Although the agent's updating rule tends to reach a consensus, I show that the rule is also capable of generating polarization and can sustain clustered disagreement, even on a connected network where everyone directly observes the same data and processes that data with the same model. Polarization is possible because in contrast to updating in DeGroot or bounded confidence models, the agent can update her beliefs away from a signal if it comes from a sender with whom she tends to disagree. In other words, the agent's updating rule need not lead to constricting belief updating (Mueller-Frank (2015)). Two keys for generating polarization are low-quality data and perceptions about the distribution of models for interpreting directly observed data.

The paper proceeds as follows: Section 2 sets the stage for the agent's problem, describing how she could arrive at a set of beliefs when directly observing data. Section 3 explores one way the agent might try to resolve the ambiguity she faces, using the signals she receives from individuals in her social network to form her beliefs. Section 3.4 shows why finding a model of social learning to solve the agent's problem is an ill-posed problem and describes the implications of a heuristic the agent might use to specify a model of social learning. Section 4 investigates the implications of this heuristic in greater detail, studying belief dynamics under one specification of the updating rule for several parameterizations under various network and proposition structures. Section 5 concludes, and Appendix A lists notation to help the reader.

2 Belief Formation via Directly-Observed Data

Suppose there is a finite set of propositions $\{p^1, p^2, \dots, p^K\} = \mathcal{K}$, none of which can be written as a compound proposition using other propositions in the set.[6] An agent must determine the truth value of the statements, $T(p^k) \in \{0, 1\}$, and agent i's beliefs at time t are denoted by $\lambda_{it}^k = \Pr(T(p^k) = 1)$.[7] The agent directly observes data W_{it} sampled under identical procedures at each point in time.

Consider a classical (frequentist) setting. With high-quality data W_{it}^*, the agent would be able to use her model φ_i^k to translate her data into an independent and identically distributed (iid) sequence of signals $\{\sigma_{it}^{k*}\}_{t=1}^T$, where

$$\sigma_{it}^{k*} = \varphi_i^k(W_{it}^*) \in [0, 1].$$

The law of large numbers ensures convergence to the mean of the signal distribution, which I will

[6] This greatly simplifies the analysis. See Paris and Vencovská (1990) and Wilmers (2010) for implications of propositional calculus when considering propositions formed as compound propositions.

[7] A proposition is a statement that is either true ($T(p^k) = 1$) or false ($T(p^k) = 0$).

denote by μ_i^{k*}, for beliefs formed as

$$\begin{aligned}\lambda_{it+1}^{k*} &= \frac{1}{t}\sum_{n=1}^{t}\sigma_{in}^{k*} \\ &= \beta_t\sigma_{it}^{k*} + (1-\beta_t)\lambda_{it}^{k*} \quad \text{where} \quad \beta_t = 1/t. \end{aligned} \tag{1}$$

Now consider a setting in which the agent's directly-observed data W_{it} only allows her to set identify the true iid signal σ_{it}^{k*}. Inspired by the literature on partial identification (Manski (2007), Tamer (2010)), suppose the agent's model and data allow her to determine a signal σ_{it}^{k} and its quality θ_{it}^{k},

$$(\sigma_{it}^{k}, \theta_{it}^{k}) = \varphi_i^k(W_{it}) \in [0,1]^2,$$

where the true signal is related to the observed signal by

$$\sigma_{it}^{k*} \in [\ \max\{0, \sigma_{it}^{k} - (1-\theta_{it}^{k})\}\ ,\ \min\{\sigma_{it}^{k} + (1-\theta_{it}^{k}), 1\}\ ]\ \equiv\ [\ \underline{\sigma}_{it}^{k*}\ ,\ \overline{\sigma}_{it}^{k*}\ ]. \tag{2}$$

The agent then knows from her signals of imperfect quality that the average

$$\lambda_{it+1}^{k*} = \frac{1}{t}\sum_{n=1}^{t}\sigma_{in}^{k*} \quad \in \quad \Lambda_{it+1}^{k*} = \left[\ \frac{1}{t}\sum_{n=1}^{t}\underline{\sigma}_{in}^{k*}\ ,\ \frac{1}{t}\sum_{n=1}^{t}\overline{\sigma}_{in}^{k*}\ \right],$$

where the sets $[\underline{\sigma}_{it}^{k*}, \overline{\sigma}_{it}^{k*}]$ and Λ_{it+1}^{k*} are often referred to as "imprecise probabilities" (Coolen et al. (2011)). The set $[\underline{\sigma}_{it}^{k*}, \overline{\sigma}_{it}^{k*}]$ is what can be learned about p^k from the directly-observed data under the most credible assumptions. While the agent can also determine a point estimate σ_{it}^{k}, doing so requires less credible assumptions, so the agent cannot be sure that $\mathbb{E}[\sigma_{it}^{k}] = \mu_i^{k*}$ unless $\theta_{it}^{k} = 1$. Manski (2016) discusses how the length of an interval estimate can be reduced, even as far as a point estimate, by adopting assumptions of decreasing credibility.

In order to give some interpretation to the above notation, suppose that two researchers had access to the same sequence of data generated by many states randomly increasing and decreasing their state budgets. The first researcher is interested in learning about the proposition p^1= "Increasing state spending stimulates the state economy." Suppose that this proposition were true ($T(p^1) = 1$). In the case of high-quality data where $\theta_{it}^{1} = 1$ for all t, $\sigma_{it}^{1} = \sigma_{it}^{1*}$, and so λ_{it+1}^{1*} can be calculated from Equation (1) as the mean signal. Supposing that the true mean of the signal distribution were 0.9, the Law of Large Numbers ensures that the first researcher's belief will converge to 0.9 as $t \to \infty$.

Now suppose that the second researcher is interested in using the same data to learn the truth of proposition p^2= "Increasing federal spending stimulates the national economy." Given the differences between the state and national economies, as well as differences in state and federal government purchases, the researcher judges her state-level data to be of low-quality, mapping into signals represented by $\theta_{it}^{2} = 0.2$ for all t. A stylized example would be that the signal distribution took on a discrete support, with probability 0.25 that $\sigma_{it}^{2} = 0$, probability 0.25 that $\sigma_{it}^{2} = 0.5$, and probability 0.5 that $\sigma^2_{it} = 1$.

In this case, the second researcher will be subject to ambiguity in addition to risk.[8] If the observed signal is $\sigma^2_{it} = 0$, 0.5, or 1, then based on Equation 2 the researcher can bound the true signal to be within, respectively, $[0, 0.8]$, $[0.1, 0.9]$, or $[0.2, 1]$ (See Figure 1.). Thus, as $t \to \infty$, the second researcher will infer that the mean of the true signals is $\mu^{2*}_i \in \Lambda^{2*}_i = [0.125, 0.925]$.[9]

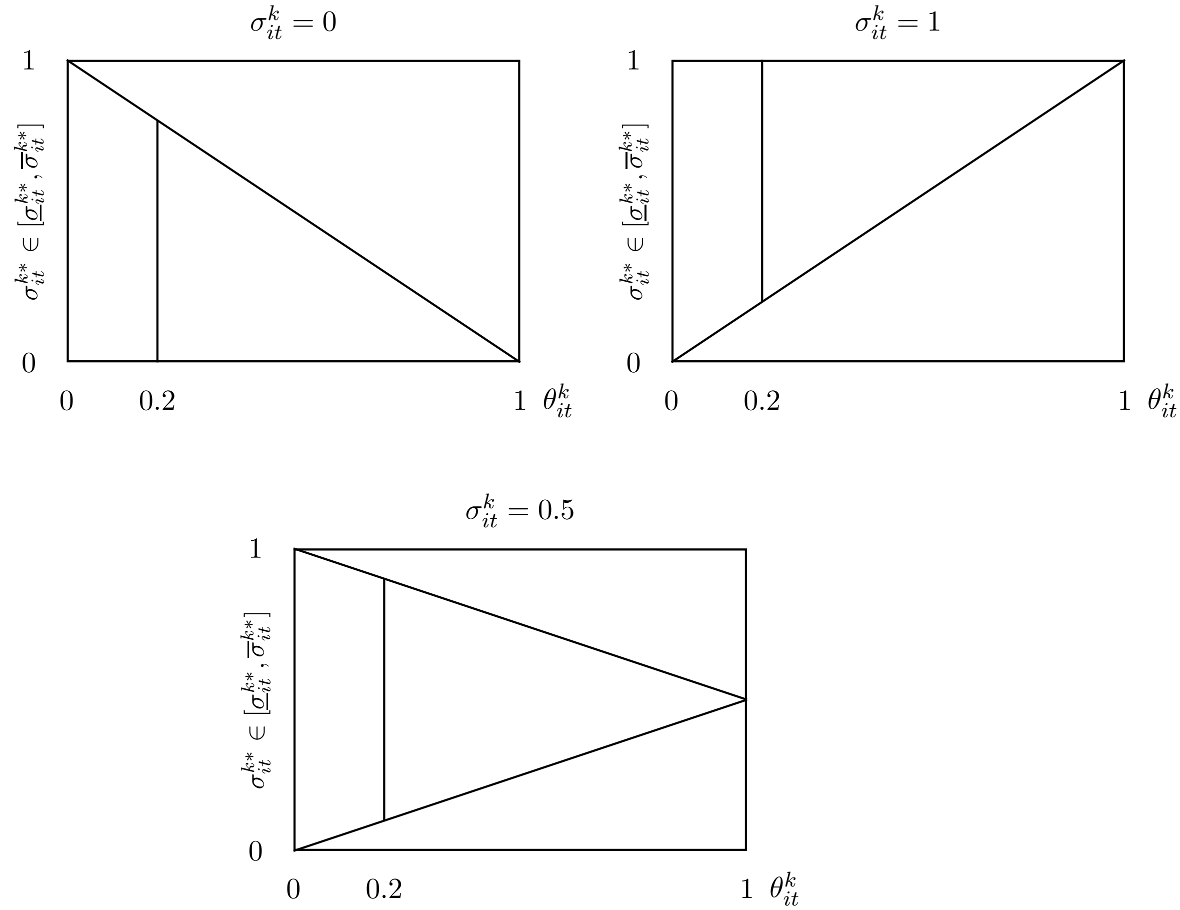


Figure 1: $\varphi^k_i(W_{it}) = (\sigma^k_{it}, \theta^k_{it}) \quad \Rightarrow \quad [\ \underline{\sigma}^{k*}_{it}\ ,\ \overline{\sigma}^{k*}_{it}\ ]$

Non-causal propositions can also have low-quality signals for reasons like survey non-response (Manski (2015)). The above model can apply to any proposition for which the agent faces a missing data problem.

In addition to describing signals, throughout the analysis I will use "high-quality" (relative to the agent's model) to describe data yielding point-identified signals ($\theta^k_{it} = 1$), and "low-quality" to describe data yielding set-identified signals ($\theta^k_{it} < 1$). Another interpretation of an extremely low-quality signal, $\theta^k_{it} = 0$, is that the agent does not directly observe any data for a given proposition

[8]In this context a point-valued belief $\lambda^k_{it} \in (0, 1)$ represents risk, while a set-valued belief $\Lambda^k_{it} \subseteq [0, 1]$ represents Knightian uncertainty or ambiguity.

[9]Confidence intervals for the identified set Λ^{k*}_i are studied in Imbens and Manski (2004) and Stoye (2009), more generally as confidence regions in Chernozhukov et al. (2007) and Romano and Shaikh (2010), and using Bayesian methods in Moon and Schorfheide (2012) and Bollinger and van Hasselt (2017).

p^k, so that $\varphi_i^k(\emptyset) = (\sigma_{it}^k, 0) \Rightarrow \sigma_{it}^{k*} \in [0,1]$. It could also be the case that the agent's model φ_i^k is not capable of extracting information from data. For example, an agent ignorant of genetics and molecular biology would likely have a model incapable of interpreting data on the human genome. In such cases, one could assign $\varphi_i^k(W_{it}^*) = \varphi_i^k(W_{it}) = (\sigma_{it}^k, 0) \Rightarrow \sigma_{it}^{k*} \in [0,1]$ for any data set. For this analysis I will assume that the agent's model produces a point-identified signal given a high-quality data set.

3 Belief Formation via Social Learning

A criticism of Bayesian decision theory is that in some circumstances, it might not be possible for the agent to express her beliefs using a distribution over the set Λ_{it}^{k*}. Bayesian decision theory is difficult to apply to these circumstances, since an imprecise probability cannot be used to make decisions according to the standard Savage axioms (Gilboa and Marinacci (2013)).

When holding beliefs represented by an imprecise probability Λ_{it}^{k*}, several approaches to decision making can be interpreted as picking one belief from the set Λ_{it}^{k*}, and then using this probability as a subjective belief with which to make decisions following the Savage axioms. The chosen probability is typically pessimistic, assuming the worst case in some sense of utility. For example, the Γ-maxmin utility decision rule maximizes expected utility after choosing the belief that would be set by a malevolent nature minimizing the agent's utility for any decision (Gilboa and Schmeidler (1989)). Similarly, the Γ-minimax regret decision rule chooses the single belief that maximizes the loss from making decisions with the chosen belief rather than the true probability when the agent makes decisions to minimize this loss (Manski (2011)).

When a decision maker does not observe all of the relevant data to form beliefs, there is empirical evidence that such a decision maker will often try to infer the relevant information from social observations. Examples are as diverse as neighborhood and school choice (de Souza Briggs et al. (2008)); adoption of a new technology (Conley and Udry (2010), Foster and Rosenzweig (1995)); consumption of a new good (Moretti (2011)); investment decisions (Bursztyn et al. (2014)); retirement savings decisions (Beshears et al. (2015)); and choice of health insurance plans (Sorensen (2006)).[10] The subsequent model explores belief formation when the agent chooses one belief from Λ_{it}^{k*} using information from her social network.

3.1 The Agent's Problem

Suppose the agent is a member of a network of $J+1$ individuals from which she might gather information. The agent directly-observes the information

$$\mathcal{I}_{it} \equiv \left\{ (\lambda_{it}^1, \sigma_{it}^1, \theta_{it}^1) \ , \ \ldots \ , \ (\lambda_{it}^K, \sigma_{it}^K, \theta_{it}^K) \right\}.$$

[10] This type of social learning is not restricted to causal propositions. As an example, the Federal Reserve looks at the beliefs of others when forming beliefs about whether the economy is in a recession or is at full employment.

To initialize the process we might let $\lambda_{i1}^k = \sigma_{i1}^k$; assume that the agent observes point identified signals from $t = -T$ until $t = 1$ and then set identified signals for $t > 1$; or else assume that the agent has just randomly reset $t = 1$ (as a random mutation in an evolutionary algorithm). The agent also observes information in her social network about the truth of propositions. We denote the set of others in the agent's network as $\mathcal{J}$. However, the agent does not directly observe the data individuals in her network ($j \in \mathcal{J}$) directly observe. Instead, the agent observes individuals' beliefs and their interpreted data in the form of their signals. Thus, the socially-observed information available to the agent is

$$\mathcal{I}_{Jt} \equiv \left\{ \ \{\lambda_{jt}^1, \sigma_{jt}^1\}_{j\in\mathcal{J}^1} \ , \ \ldots \ , \ \{\lambda_{jt}^K, \sigma_{jt}^K\}_{j\in\mathcal{J}^K} \ \right\},$$

where the agent receives information about proposition p^k from individuals in $\mathcal{J}^k \subseteq \mathcal{J}$.

The agent might try Bayesian updating, or Bayesian social learning, according to Bayes' rule:

$$Pr(T(p^k) = 1|\sigma_{it}^k, \{\sigma_{jt}^k\}_{j\in\mathcal{J}^k}) = \frac{Pr(\sigma_{it}^k, \{\sigma_{jt}^k\}_{j\in\mathcal{J}^k}|T(p^k) = 1)Pr(T(p^k) = 1)}{Pr(\sigma_{it}^k, \{\sigma_{jt}^k\}_{j\in\mathcal{J}^k})}$$

Using beliefs λ_{it}^k as the agent's prior, this would imply updating as

$$\lambda_{it+1}^k = \frac{f(\sigma_{it}^k, \{\sigma_{jt}^k\}_{j\in\mathcal{J}^k}|T(p^k) = 1)\lambda_{it}^k}{f(\sigma_{it}^k, \{\sigma_{jt}^k\}_{j\in\mathcal{J}^k}|T(p^k) = 1)\lambda_{it}^k + f(\sigma_{it}^k, \{\sigma_{jt}^k\}_{j\in\mathcal{J}^k}|T(p^k) = 0)(1 - \lambda_{it}^k)}.$$

Acemoglu et al. (2016) show in a related setting that strong restrictions would be required on the conditional pdfs $f(\cdot|T(p^k))$ for there to be asymptotic agreement across agents. More fundamentally, correctly specifying the likelihood function $f(\sigma_{it}^k, \{\sigma_{jt}^k\}_{j\in\mathcal{J}^k}|T(p^k))$ can require unrealistic assumptions about the information and computation available to the agent (Acemoglu and Ozdaglar (2011)).[11] Weakening these assumptions is a key motivation of the literature on non-Bayesian social learning (Molavi et al. (2015)).

Correctly specifying the likelihood function is the same as specifying $f(\varphi_{it}^k(W_{it}^*), \{\varphi_j^k(W_{jt})\}_{j\in\mathcal{J}^k}|T(p^k))$, which would require not only that the agent know the sampling processes for W_{it}^* and W_{jt}^* conditional on $T(p^k)$, but also the models $\{\varphi_j^k\}_{j\in\mathcal{J}^k}$. I rule out Bayesian social learning by restricting social information to beliefs and signals, assuming that the agent does not observe the additional information required to specify the likelihood function:

(A1) Imperfect Communication: Agent i can only observe point estimates λ_{jt}^k and σ_{jt}^k. She cannot observe measures of the sender's ambiguity $\Lambda_{jt}^{k*}, \theta_{jt}^k$ or their model $\varphi_j^k \ \forall \ j, t, k$

The issue captured by A1 is that data must be transformed into information using a model, and it is difficult for individuals to communicate this process. Therefore, valuable details are lost relative

[11] Benoît and Dubra (2015) and Andreoni and Mylovanov (2012) study polarization under private learning when agents disagree about $f(\sigma_{it}^{k*}|T(p^k))$. Alternatively, in this context their analyses could be interpreted as agents having different models for private learning φ_i^k, each proposition p^k being a conjunction of simple propositions $p^k = p^{k'} \wedge p^{k''}$, and W_{it}^* being revealed at different subperiods of t for $p^{k'}$ and $p^{k''}$.

to directly observing the data when information is obtained socially.[12] Among other reasons, this assumption is descriptively appealing because there is a well-documented tendency for researchers and statistical agencies to focus on communicating their point estimates σ_{it}^k without communicating about their models φ_i^k or measures of uncertainty θ_{it}^k (Manski (2007), Manski (2015)).

With A1 ruling out Bayesian social learning, I assume that the agent uses signals in an effort to replicate classical inference. Given a loss function $\mathcal{L}$, the agent's problem is to choose functions f^k from some set $\mathcal{F}$ to solve the problem

$$
\begin{aligned}
&\min_{f^1,\ldots,f^K \in \mathcal{F}} \quad \sum_{k=1}^{K} \mathcal{L}\left(\mathbb{E}\left[\mu_i^{k*} - \lim_{t\to\infty} \lambda_{it+1}^k\right]\right) \qquad (3)\\
s.t.\ \ &(\mathcal{I}_{it}, \mathcal{I}_{Jt})\\
&\widehat{\sigma}_{it}^k = f^k(\mathcal{I}_{it}, \mathcal{I}_{Jt}) \qquad \text{for } \ k = 1, \ldots, K\\
&\lambda_{it+1}^k = \beta_t \widehat{\sigma}_{it}^k + (1-\beta_t)\lambda_{it}^k \qquad \text{for } \ k = 1, \ldots, K
\end{aligned}
$$

I will refer to the agent's construction of her unobserved, high-quality signals $\widehat{\sigma}_{it}^k$ as her inferred signals. A natural restriction on $\mathcal{F}$ is to make inferred signals a weighted average of directly- and socially-observed signals. In this case, f^k can be written as

$$
\widehat{\sigma}_{it}^k = \underbrace{\theta_i^k}_{\substack{\text{share of signal}\\ \text{directly-observed}}} \sigma_{it}^k + \underbrace{(1-\theta_{it}^k)}_{\substack{\text{share of signal}\\ \text{socially-observed}}} \sigma_{Jt}^k.
$$

This restriction reframes the choice of f^k as the choice of σ_{Jt}^k.[13] Posing the inferred signals as weighted averages also gives an interpretation to θ_{it}^k as the agent's subjective judgment about the credibility of her modeling assumptions and/or a measure of the quality of her data.

3.2 Solutions to Special Cases of the Agent's Problem

When faced with problems like the agent's problem, a popular set $\mathcal{F}$ is linear opinion pooling (Ranjan and Gneiting (2010)). It turns out that using repeated linear opinion pooling to solve the agent's problem results in DeGroot updating if data are only observed in the first period, and signals continue to be sent in later periods.

Proposition 1 (DeGroot)**.** *If data are only observed once at $t = 1$, the agent sets $\lambda_{i1}^k = \sigma_{i1}^k$, $\theta_{it}^k = \theta_{i1}^k$ for all $t > 1$, and subsequent signals are interchangeable with beliefs ($\sigma_{it}^k = \lambda_{it}^k$ and $\sigma_{jt}^k = \lambda_{jt}^k$ for $j \geq 2$), then linear opinion pooling where the agent constructs her inferred signals for*

[12] A1 imposes a version of word-of-mouth learning (Ellison and Fudenberg (1995), Banerjee and Fudenberg (2004)).

[13] Assuming that $\{W_{it}\}_{t=1}^{\infty}$ and $\{\varphi_i^k\}_{k=1}^{K}$ are exogenous, both $\{\sigma_{it}^k\}_{t=1}^{\infty}$ and $\{\theta_{it}^k\}_{k=1,t=1}^{K,\infty}$ are given. Thus, in an abuse of notation, I will refer to f^k both as the function determining $\widehat{\sigma}_{it}^k$ and as the function determining σ_{Jt}^k.

$t \geq 2$ as

$$\widehat{\sigma}_{it}^k = \theta_i^k \sigma_{it}^k + (1-\theta_i^k)\sigma_{Jt}^k \qquad \textit{where} \tag{4}$$

$$\sigma_{Jt}^k = \sum_{j \in \mathcal{J}^k} \underbrace{w_j^k}_{\substack{\textit{share of social signal} \\ \textit{from individual } j}} \sigma_j^k \quad \textit{with } \; w_j^k \geq 0 \;\; \forall \;\; j \in \mathcal{J}^k, \;\; \sum_{j \in \mathcal{J}^k} w_j^k = 1 \tag{5}$$

is equivalent to DeGroot updating where $\boldsymbol{\lambda}_{t+1}^k = \Omega_t^k \boldsymbol{\lambda}_t^k$ and the entries of Ω_t^k are

$$\omega_{iit}^k = \beta_t \theta_i^k + (1-\beta_t)$$
$$\omega_{ijt}^k = \beta_t (1-\theta_i^k) w_j^k.$$

Proof. As hypothesized, set $\lambda_{i1}^k = \sigma_{i1}^k$. For $t \geq 2$, the equality of beliefs and signals, together with the updating equation in the agent's problem (3) imply that

$$\begin{aligned} \sigma_{it+1}^k &= \beta_t \widehat{\sigma}_{it}^k + (1-\beta_t)\sigma_{it}^k \\ &= \beta_t \theta_i^k \sigma_{it}^k + (1-\beta_t)\sigma_{it}^k + \beta_t(1-\theta_i^k) \sum_{j \in \mathcal{J}^k} w_j^k \sigma_{jt}^k. \end{aligned}$$

□

Furthermore, when the data observed in $t = 1$ generate unbiased point-estimates of signals, repeated linear opinion pooling/DeGroot updating solves the agent's problem.

Proposition 2 (Unbiased Signals)**.** *Assume again, as we did in the case of private learning, that*

(A2) *Averaging Signals: $\beta_t = 1/t$, so that $\beta_t \widehat{\sigma}_{it}^k + (1-\beta_t)\lambda_{it}^k = \frac{1}{t}\sum_{n=1}^t \widehat{\sigma}_{in}^k$*

If the observed data yield unbiased signals

(A3) *Private signals are iid with $\mathbb{E}[\sigma_{it}^{k*}] \equiv \mu_i^{k*} = \mu_i^k \equiv \mathbb{E}[\sigma_i^k]$, and*

(A4a) *Social signals are iid for each $j \in \mathcal{J}^k$ with $\mathbb{E}[\sigma_{it}^{k*}] \equiv \mu_i^{k*} = \mu_j^k \equiv \mathbb{E}[\sigma_{jt}^k] \;\; \forall j \in \mathcal{J}^k$,*

then repeated linear opinion pooling/DeGroot updating following Equations 4 and 5 solves the agent's problem.

Proof. Proposition 6 in Golub and Sadler (2016) states that as long as Ω^k is strongly connected and primitive, then

$$\lim_{t \to \infty} \sigma_{it+1}^k = \sum_{n=1}^{J+1} \pi_n^k \sigma_{n1}^k$$

where π_n^k is n's left-hand eigenvector centrality in Ω^k.[14] Additionally assuming that each weight $w_{i\cdot}$ is strictly positive and that $\theta^k \in (0,1)$ to ensure Ω^k is primitive, then since $\sum_{n=1}^{J+1} \pi_n^k = 1$ and

[14] A network is strongly connected if any agent i has a directed path in the network to any agent j. A strongly connected network is primitive if each agent attaches a non-zero weight to each agent.

$\mathbb{E}[\sigma^k_{n1}] = \mu^{k*}_i$ for all n, we know that

$$\mathbb{E}[\mu^{k*}_i - \lim_{t\to\infty} \lambda^k_{it+1}] = \mathbb{E}[\mu^{k*}_i - \sum_{n=1}^{J+1} \pi_n \sigma^k_{n1}] = \mu^{k*}_i - \mu^{k*}_i = 0.$$

□

3.3 A General Solution to the Agent's Problem

A4a is a very restrictive assumption that need not hold in the general case of the agent's problem. That is, while the agent observes data and signals in each period, this information is potentially biased. In this case, the agent can still solve her problem if she has a model capable of accurately interpreting the social signals she receives.

Proposition 3 (Biased Social Signals)**.** *Now suppose that the agent receives biased signals in the sense that* $\mathbb{E}[\sigma^k_{jt}] \neq \mu^{k*}_{it}$, *but that the agent has successfully engaged in statistical learning in the following sense:*

(A4b) *The agent has a model of social learning* g^k *that interprets social signals as* $s^k_{jt} = g^k(\mathcal{I}_{it}, \mathcal{I}_{Jt})$. *The* s^k_{jt} *are iid for each* $j \in \mathcal{J}^k$ *with* $\mathbb{E}[\sigma^{k*}_{it}] \equiv \mu^{k*}_i = \mathbb{E}[s^k_{jt}] \ \forall \ j \in \mathcal{J}^k$.

Then linear opinion pooling where the agent constructs unobserved high-quality signals with her model as

$$\widehat{\sigma}^k_{it} = \theta^k_{it}\sigma^k_{it} + (1-\theta^k_{it})\sigma^k_{Jt} \quad \textit{where} \tag{6}$$

$$\sigma^k_{Jt} = \sum_{j\in\mathcal{J}^k} w^k_{jt} s^k_{jt} \quad \textit{with} \ w^k_{jt} \geq 0 \ \forall \ j \in \mathcal{J}^k, \ \sum_{j\in\mathcal{J}^k} w^k_{jt} = 1 \tag{7}$$

$$s^k_{jt} = g^k(\mathcal{I}_{it}, \mathcal{I}_{Jt}) \tag{8}$$

solves the agent's problem.

Proof. By A2 we know that $\lim_{t\to\infty} \lambda^k_{it+1} = \lim_{t\to\infty} \frac{1}{t}\sum_{n=1}^{t} \widehat{\sigma}^k_{in}$. If the signals are iid, then by the law of large numbers we know that $\lim_{t\to\infty} \lambda^k_{it+1} = \mathbb{E}[\widehat{\sigma}^k_{it}]$. After repeatedly applying the linearity of the expectations operator, A3 and A4a imply that

$$\begin{aligned}
\lim_{t\to\infty} \lambda^k_{it+1} = \mathbb{E}[\widehat{\sigma}^k_{it}] &= \mathbb{E}[\overline{\theta}^k_i \sigma^k_{it} + (1-\overline{\theta}^k_i)\sigma^k_{Jt}] = \overline{\theta}^k_i \mathbb{E}[\sigma^k_{it}] + (1-\overline{\theta}^k_i)\mathbb{E}[\sigma^k_{Jt}] \\
&= \overline{\theta}^k_i \mathbb{E}[\sigma^k_{it}] + (1-\overline{\theta}^k_i)\mathbb{E}[\sum_{j\in\mathcal{J}^k} w^k_{jt}\sigma^k_{jt}] = \overline{\theta}^k_i \mathbb{E}[\sigma^k_{it}] + (1-\overline{\theta}^k_i)\sum_{j\in\mathcal{J}^k} w^k_{jt}\mathbb{E}[\sigma^k_{jt}] \\
&= \overline{\theta}^k_i \mu^k_{it} + (1-\overline{\theta}^k_i)\sum_{j\in\mathcal{J}^k} w^k_{jt}\mu^k_{jt} \\
&= \mu^{k*}_i .
\end{aligned} \tag{9}$$

□

Determining the correct model for interpreting signals could be viewed as a statistical learning problem. To put the problem in the notation of statistical learning, here we adapt some notation from Chapter 1 of Schölkopf and Smola (2002). Suppose that at time t the agent observes outcomes of high-quality data for herself and everyone in her network,

$$(x_i^1, y_i^1), \ldots, (x_i^K, y_i^K) \in \mathcal{X} \times [0,1]$$

where $x_i^k = x_i$ for $k = 1, \ldots, K$ and

$$\begin{aligned} (x_i^k, y_i^k) &= \left([\mathcal{I}_i, \mathcal{I}_j], \sigma_i^{k*} \right) \\ &\equiv \left(\begin{bmatrix} \lambda_{it}^1 & \sigma_{it}^1 & \theta_{it}^1 & \lambda_{1t}^1 & \cdots & \lambda_{Jt}^1 & \sigma_{1t}^1 & \cdots & \sigma_{Jt}^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{it}^K & \sigma_{it}^K & \theta_{it}^K & \lambda_{1t}^K & \cdots & \lambda_{Jt}^K & \sigma_{1t}^K & \cdots & \sigma_{Jt}^K \end{bmatrix}, \ \sigma_{it}^{k*} \right). \end{aligned} \tag{10}$$

The agent constructs her high-quality signal as:

$$y_{it}^k = f^k(x_{it}^k)$$

where the function $f^k \in \mathcal{F}$ is chosen to minimize the expectation of an empirical risk function like mean squared error

$$R[f^k] = \mathbb{E}_t[c(x_{it}^k, y_{it}^k, f^k(x_{it}^k))] = \mathbb{E}_t[(y_{it}^k - f^k(x_{it}^k))^2],$$

and the set of functions $\mathcal{F}$ is chosen with the help of Vapnik-Chervonenkis (VC) theory to avoid overfitting.

A fundamental problem, however, is that the agent never observes $y_{it}^k = \sigma_{it}^{k*}$. Thus, the agent cannot construct the risk function, so choosing f^k according to this criterion is not a well-posed problem. Facing this ill-posed problem, the agent could stop here and forego the information in the social signals she receives. She could choose a belief based on decision rules equipped to deal with ambiguity, like the maxmin utility or minimax regret decision rules.

3.4 A Heuristic for the Agent's Problem

Using socially-observed information to reconstruct the unobserved quantity of interest from Equation 10, $y_{it}^k = \sigma_{it}^{k*}$, is similar to the approach taken in many applied problems. One example is causal inference.[15] When facing the fundamental problem of evaluation, which is that all required counterfactuals (potential outcomes) are not observed (Holland (1986)), researchers typically combine a model with observations to infer the unobservable quantities of interest.[16]

[15] Another example is speech recognition (Raj et al. (2004)). This approach can also be seen as a forecasting problem.

[16] Although this practice has been criticized for utilizing terms not subject to empirical scrutiny (Dawid (2000)), it is how both the Rubin Causal Model constructs past outcomes under ideal interventions to treatment

When people are asked how they interpret information from different sources, they report forming beliefs as if they were trying to impute σ_{it}^{k*}, often by interpreting socially-observed information according to a function like f^k. For example, American voters were interviewed (BBC (2016)) about how they use different sources of information when forming beliefs, and one respondent discussed the desire to find a news source reporting his σ_{it}^{k*}:

> "I don't really know that there is an honest, genuine, unbiased news source available to any of us. Ideologically, you go to the news source that's going to report one incident the way you want to hear it."

Another respondent reported the use of a function like f^k for imputing σ_{it}^{k*} from "untrusted" news sources:

> "I don't trust the national media. Whatever they say, if you really want to be sure, you just take a position 180 degrees opposite, and you'll be closer to the truth."

More broadly, the beginning of this Section refers to studies showing that people attempt to address the agent's problem by engaging in social learning. Studies of animal behavior also indicate that animals resolve uncertainty of the type facing the agent by combining directly-observed and socially-observed information (Kendal et al. (2005)), sometimes even differentiating between sources of information (van Bergen et al. (2004)).

Based on this empirical evidence, it might be reasonable to investigate heuristics for inference that are motivated by the agent's problem. The agent's problem is inferring the value of $\pi_{ijt}^k \equiv \sigma_{it}^{k*} - \sigma_{jt}^k$ so that she can properly adjust each socially observed signal σ_{jt}^k. To think about the agent's problem in terms of inference under uncertainty about π_{ijt}^k, first denote the sampling process for data W_{it}^* by Γ_i^*, which results in the sampling distribution $\sigma_{it}^{k*} \sim G_i^{k*}$, with mean μ_i^{k*}. Denote the sampling process for data W_{it} by Γ_i, where the sampling distribution $\sigma_{it}^k \sim G_i^k$ has mean μ_i^k. Define analogous sampling processes and sampling distributions for the signals from each agent j.

Then there are five factors that drive the distribution of π_{ijt}^{k*}:

I (Random Sampling Error): $\Gamma_i^* = \Gamma_j$

II (Biased Sampling Process): $\Gamma_i^* \neq \Gamma_j$

III (Different Models): $\varphi_i^k \neq \varphi_j^k$

IV (Social Influence): j's model of social learning and/or her network

V (Strategic Reporting): Part of what determines the function φ_j^k is strategic reporting

(Imbens and Rubin (2015)), Structural Causal Models construct past outcomes under additional interventions to the DGP (Pearl (2009)), and both types of models construct future outcomes under interventions (Aliprantis (2015)).

Assuming that only factors (I)-(III) drive the distribution of $\sigma_{it}^{k*} - \sigma_{jt}^{k}$, we have

$$\mathbb{E}\left[\lambda_{it}^{k*} - \lambda_{jt}^{k}\right] = \mathbb{E}\left[\sigma_{it}^{k*} - \sigma_{jt}^{k}\right] = \mu_i^{k*} - \mu_j^{k},$$

so that

$$s_{jt}^{k} = \sigma_{jt}^{k} + \left(\lambda_{it}^{k*} - \lambda_{jt}^{k}\right) \tag{11}$$

solves the agent's problem.[17] Not observing the right hand side of Equation 11, the agent might replace the unobserved quantity λ_{it}^{k*} with the observed quantity λ_{it}^{k} to follow the heuristic:

$$\textbf{(H1):} \quad s_{jt}^{k} = \sigma_{jt}^{k} + \left(\lambda_{it}^{k} - \lambda_{jt}^{k}\right).$$

The agent knows that (H1) does not necessarily solve her problem if factors (IV) and (V) help to drive the distribution of π_{ijt}^{k}. Thus, she would like to lean on (H1) as little as possible, or to use signals from the senders for whom (H1) is the most credible.

Define Δ_{ijt} as the credibility that the agent i deems to H1 as applied to sender j's signal, and $\Delta_{it}^{k} = f\left(\{\Delta_{ijt}\}_{j\in\mathcal{J}^k}\right) \in [0,1]$ as the total credibility agent i deems to H1 as applied to the socially-available information on proposition k. Then if we define $s_{jt}^{k} = g^k(\mathcal{I}_{it}, \mathcal{I}_{Jt})$ by (H1) and denote the weights attached to each of these interpreted signals as the sender's relative credibility $w_{jt}^{k} = \frac{\Delta_{ijt}}{\sum_{j=1}^{J^k}\Delta_{ijt}}$, we can define

$$s_{Jt}^{k} = \sum_{j=1}^{J^k} w_{jt}^{k} s_{jt}^{k}$$

(assuming $\Delta_{ijt} \neq 0$ for some $j \in \mathcal{J}^k$), and the agent could infer signals using $\widehat{\sigma}_{it}^{k} = f^k(\mathcal{I}_{it}, \mathcal{I}_{Jt})$ defined as

$$\widehat{\sigma}_{it}^{k} = \theta_{it}^{k}\sigma_{it}^{k} + (1-\theta_{it}^{k})\left[\Delta_{it}^{k} s_{Jt}^{k} + (1-\Delta_{it}^{k})\sigma_{it}^{k}\right]. \tag{f^k}$$

Note that H1 allows for moving beliefs away from a signal, which violates the Monotonicity assumption in Molavi et al. (2015), and characterizes the difference between this model's heterogeneous confidence learning rule and bounded confidence models like those developed in Hegselmann and Krause (2002) or Sotiropoulos et al. (2015).

4 Opinion Dynamics

4.1 Empirical Implementation

To empirically implement the agent's model of social learning, we must first define the credibility she gives to (H1) as a means of adjusting a signal from sender j, Δ_{ijt}. The agent might use the

[17] DeGroot (1974) studies social learning under (I). Manski (2004) studies social learning under a version of (II) focusing more attention on the construction of the interval $\Lambda_i^{k*} \equiv \mathbb{E}[\Lambda_{it}^{k*}]$. Manski (2004) models communication differently than (A1): assuming that outcomes are directly observed rules out (III) and makes the agent's problem more specifically about how to shrink the interval Λ_i^{k*} after observing potentially non-iid data. This paper adopts a more stylized version of (II) in order to study social learning under (I)+(II)+(III).

uncertainty of $\delta_{ijt}^k \equiv \lambda_{it}^k - \lambda_{jt}^k$ over the proposition space as a means of assigning credibility to H1 applied to sender j. Entropy is one measure of uncertainty developed in information theory.[18]

Figure 2 helps to illustrate the idea of (relative) entropy using the notion of Shannon entropy from information theory (Shannon (1948)). The agent would be most informed about agent j if the distribution of δ_{ijt}^k were a degenerate distribution, and would be least informed were δ_{ijt}^k to follow a uniform distribution. In the Figure, $f(\delta_{i\max t}^k) = U[-1,1]$ has the maximum entropy (representing the least informative sender), senders $j = 1, 2, 3$ have high entropy (representing low information), senders $j = 4, 5, 6$ have medium entropy (representing moderate information), and senders $j = 7, 8, 9$ have low entropy (representing high information).

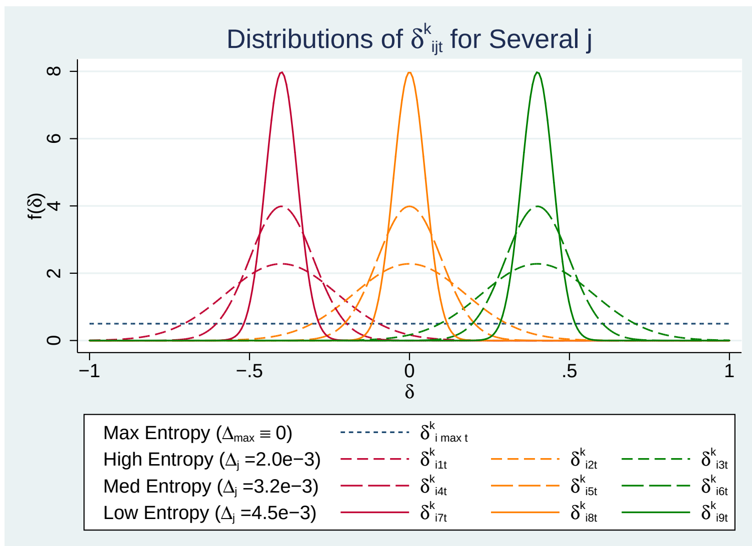


Figure 2: Entropy of Sender j

Although senders $j = 2$ and $j = 5$ have the same average disagreement across propositions with the agent, the agent will give more weight to interpreted signals from sender $j = 5$ because their disagreement has lower entropy than that of sender $j = 2$. That is, the agent is more certain about how she will disagree with sender $j = 5$. On the other hand, note that while the agent expects to disagree differently with senders $j = 1$, $j = 2$, and $j = 3$, the agent deems interpreted signals from these senders to be equally credible because they all have the same entropy. This helps to illustrate that what matters to the agent's judgment about credibility is not her average disagreement with a sender, but how uncertain she is about her disagreement with a sender.

Assuming that the distribution of δ_{ijt}^k across propositions $k \in \{1, \ldots, K\}$ has a probability density function Q_{ijt}, the informational content of sender j's signal (or certainty about δ_{ijt}^k) can be defined as the Kullback-Leibler divergence from the uniform distribution over $[-1, 1]$:

$$D_{KL}(Q_{ijt} : U) = \int_{-1}^{1} Q_{ijt}(\delta) \log \left(\frac{Q_{ijt}(\delta)}{1/2} \right) d\delta,$$

[18]We can think of other alternatives. For example, the agent might also measure credibility via the Root Mean Square Error of using δ_{ijt}^k for a given proposition k as a predictor for $\delta_{ijt}^{k'}$ when $k \neq k'$.

which is a measure of the difference in entropy of Q_{ijt} relative to the maximum entropy (uncertainty) distribution.[19] I assume that the credibility the agent assigns to interpreted signals from sender j is $\Delta_{ijt} \equiv \rho(D_{KL}(Q_{ijt} : U))$.

4.2 Simulations

I now show simulations illustrating some of the belief dynamics that f^k can generate. I first show that beliefs can polarize and clusters can be sustained in a relatively simply setting. I then show that this result is not an artifact of the simple setting by replicating it in more nuanced settings.

In both numerical experiments, I consider a network of $J + 1 = 300$ individuals learning about $K = 30$ propositions, with $\theta_{it}^k = 0.1$ for all k, i, and t. Each agent assesses the credibility of (H1) applied to sender j's signal according to

$$\Delta_{ijt} \equiv \rho(D_{KL}(Q_{ijt} : U)) = [\gamma_1 D_{KL}(Q_{ijt} : U)]^{\gamma_2}$$

where $(\gamma_1, \gamma_2) = (100, 8) \in [0, \infty) \times [0, \infty)$ can be thought of as distrust parameters, and the credibility given to the sum of social information in the network $\Delta_{it}^k = 1$ for all k, i, and t. I consider directly-observed data in the network most likely to result in convergence to a degenerate distribution. I assume all individuals directly-observe data generating the same signals at all times for all propositions:

$$\sigma_{it}^k = 0.5 \quad \forall \quad t = 2, 3, 4, \ldots; \quad k = 1, \ldots, K; \quad \text{and} \quad i = 1, \ldots, J + 1.$$

What varies across Experiments 1 and 2 is the initial distribution of beliefs at $t = 1$. In Experiment 1 I assume there are two clusters of individuals, $\mathcal{C}_1$ and $\mathcal{C}_2$, with $card(\mathcal{C}_1) = 100$ and $card(\mathcal{C}_2) = 200$. Initial beliefs λ_{i1}^k are generated as follows:

$$\overline{\lambda}_i^k \sim \begin{cases} \mathcal{N}(0.8, 0.1) & \text{if } i \in \mathcal{C}_1 \;\; \forall k = 1, \ldots, K; \\ \mathcal{N}(0.2, 0.01) & \text{if } i \in \mathcal{C}_2 \;\; \forall k = 1, \ldots, K. \end{cases}$$

with

$$\lambda_{i1}^k = \min\left\{\max\left\{\overline{\lambda}_i^k, 0\right\}, 1\right\} \quad \forall k = 1, \ldots, K.$$

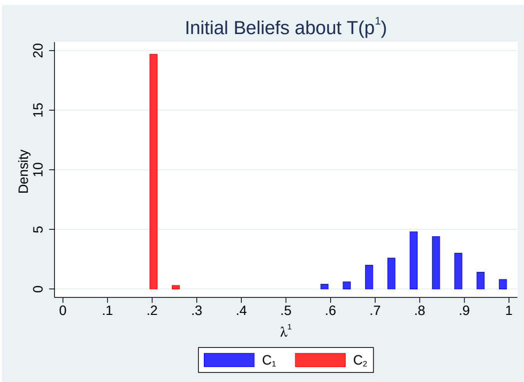
Figure 3a shows the initial distribution of beliefs by clusters for proposition p^1.

Figure 3b shows that as individuals in the network update using f^k, the updating maintains the clustering, with $\mathcal{C}_1$ and $\mathcal{C}_2$ still clearly distinguishable from one another. The basic idea is that if a given agent tends to agree with those in a widely-distributed cluster (unbiased but imprecise), but tends to disagree with those in a tightly-distributed cluster (biased but precise), that agent will rely

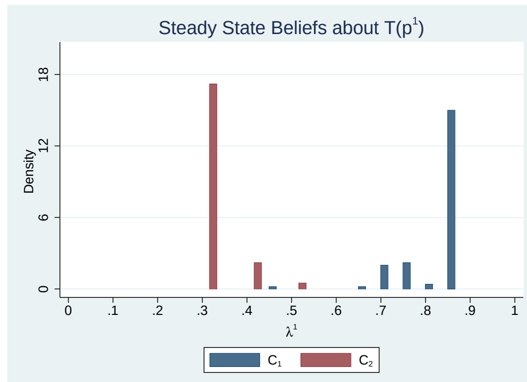
[19] In another setting, Zanardo (2017) shows that Kullback-Leibler relative entropy satisfies desirable axioms as a measure of disagreement.

more on interpreted signals from the disagreeing cluster, and this can cause her to overcompensate when they provide her with unbiased signals.

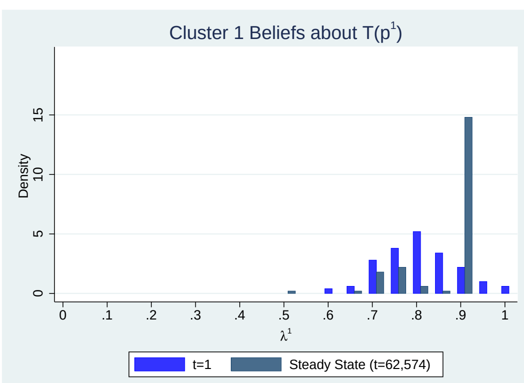
This is what happens to agents in cluster $\mathcal{C}_1$ in Experiment 1 (Figure 3c). Given their disagreement, for an individual in $\mathcal{C}_1$ the interpretation is that any signal from someone in $\mathcal{C}_2$ must, on average, be an understatement of σ_{it}^{1*}, so the interpreted signal s_{jt}^1 adjusts the received signal upwards. Since the disagreement for individuals within $\mathcal{C}_1$ is more uncertain than the disagreement across individuals in $\mathcal{C}_1$ and $\mathcal{C}_2$, individuals in $\mathcal{C}_1$ give more credibility to the interpreted signals from those in $\mathcal{C}_2$. As a result, individuals in $\mathcal{C}_1$ overcompensate and move their beliefs away from 0.5.



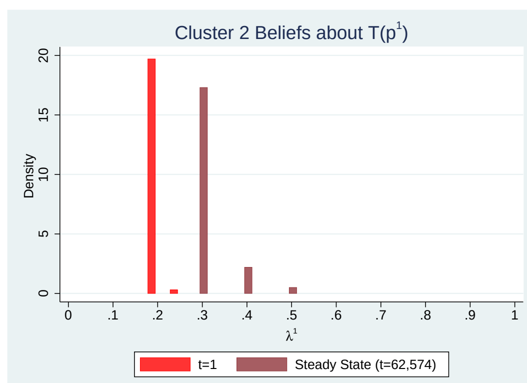
(a) Initial Distribution of Beliefs about $T(p^1)$



(b) Steady State Distribution of Beliefs about $T(p^1)$



(c) Distributions of Beliefs about $T(p^1)$ for $\mathcal{C}_1$



(d) Distribution of Beliefs about $T(p^1)$ for $\mathcal{C}_2$

Figure 3: Beliefs in Experiment 1
Steady state belief distributions are those for which $\max_{k,i} ||\lambda_{it}^k - \lambda_{it+1}^k|| <$ 1e-6.

Appendix B shows two further experiments illustrating that this type of clustered polarization is not an artifact of having two clusters, and that there are possibilities for interesting questions about which clusters polarize and which ones converge.

5 Conclusion

This paper presented a positive theory of belief formation. I proposed one way that an agent might choose a single subjective probability from a set of possible probabilities. When the agent faces ambiguity because her directly-observed data only allow her to partially identify a signal about the truth of a proposition, she might seek to learn from individuals in her social network. Assuming that communication is imperfect, so that individuals can only communicate a point estimate of their signals and beliefs, the agent must determine how to combine the signals she observes.

I showed that when signals are unbiased, linear opinion pooling of signals generates DeGroot updating, and is able to replicate classical inference with high-quality data yielding point-identified signals. When individuals in the agent's network have different models or access to different quality data, then their signals will not be unbiased. In this case, the agent might still form beliefs by linear opinion pooling on interpreted signals. I considered a heuristic that the agent might use to select a social learning model for interpreting signals, and how this model can lead to a non-constricting rule of thumb capable of generating polarization and clustered, permanent disagreement on a connected network where everyone observes the same data and processes that data with the same model.

Topics for future investigation include understanding when a network is wise under various definitions, whether beliefs must necessarily become unidimensional as in DeMarzo et al. (2003), and how one might endogenize the agent's network along the lines in Sethi and Yildiz (2012) so as to generate a model of rational inattention (Sims (2003)). Finally, we might be interested in how the heuristic in this paper relates to the ambiguity aversion literature, and whether it can address some of the empirical and theoretical concerns from Al-Najjar and Weinstein (2009).

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz (2016). Fragility of asymptotic agreement under Bayesian learning. *Theoretical Economics 11*, 187–225.

Acemoglu, D. and A. Ozdaglar (2011). Opinion dynamics and learning in social networks. *Dynamic Games and Applications 1*(1), 3–49.

Al-Najjar, N. I. (2009). Decision makers as statisticians: Diversity, ambiguity, and learning. *Econometrica 77*(5), 1371–1401.

Al-Najjar, N. I. and J. Weinstein (2009). The ambiguity aversion literature: a critical assessment. *Economics and Philosophy 25*(03), 249–284.

Aliprantis, D. (2015). Covariates and causal effects: The problem of context. *Federal Reserve Bank of Cleveland Working Paper 13-10R.*

Aliprantis, D. and F. G.-C. Richter (2016). Evidence of neighborhood effects from MTO: LATEs of neighborhood quality. *Federal Reserve Bank of Cleveland WP 12-08R.*

Andreoni, J. and T. Mylovanov (2012). Diverging opinions. *American Economic Journal: Microeconomics 4*(1), 209–232.

Artstein, Z. and R. A. Vitale (1975). A strong law of large numbers for random compact sets. *The Annals of Probability 3*(5), 879–882.

Banerjee, A. and D. Fudenberg (2004). Word-of-mouth learning. *Games and Economic Behavior 46*, 1–22.

BBC (2016, December 5). 'No truth in journalism': Trump voters cast wide net for news. *BBC News*. Retrieved from http://www.bbc.com/news/av/world-us-canada-38191313/no-truth-in-journalism-trump-voters-cast-wide-net-for-news.

Benoît, J.-P. and J. Dubra (2015). A theory of rational attitude polarization. *Mimeo., London Business School*.

Beshears, J., J. J. Choi, D. Laibson, B. C. Madrian, and K. L. Milkman (2015). The effect of providing peer information on retirement savings decisions. *The Journal of Finance 70*(3), 1161–1201.

Bollinger, C. R. and M. van Hasselt (2017). A Bayesian analysis of binary misclassification. *Economics Letters 156*, 68 – 73.

Bursztyn, L., F. Ederer, B. Ferman, and N. Yuchtman (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica 82*(4), 1273–1301.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*(5), 1243–1284.

Conley, T. G. and C. R. Udry (2010). Learning about a new technology: Pineapple in Ghana. *The American Economic Review 100*(1), 35–69.

Coolen, F. P., M. C. Troffaes, and T. Augustin (2011). Imprecise probability. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science*. Springer.

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association 95*(450), 407–424.

de Souza Briggs, X., K. S. Ferryman, S. J. Popkin, and M. Rendon (2008). Why did the Moving to Opportunity experiment not get young people into better schools? *Housing Policy Debate 19*(1), 53–91.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association 69*(345), 118–121.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel (2003). Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics 118*(3), 909–968.

Ellison, G. and D. Fudenberg (1995). Word-of-mouth communication and social learning. *The Quarterly Journal of Economics 110*(1), 93–125.

Farber, H. S., J. Rothstein, and R. G. Valletta (2015). The effect of extended unemployment insurance benefits: Evidence from the 2012-2013 phase-out. *American Economic Review 105*(5), 171–176.

Foster, A. D. and M. R. Rosenzweig (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy 103*(6), 1176–1209.

Gentzkow, M. and J. M. Shapiro (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics 126*(4), 1799–1839.

Gilboa, I. and M. Marinacci (2013). Ambiguity and the Bayesian paradigm. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics: Tenth World Congress: Economic Theory*, Volume I, Chapter 7, pp. 179–242. Cambridge University Press.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Golub, B. and M. O. Jackson (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics 127*(3), 1287–1338.

Golub, B. and E. Sadler (2016). Learning in social networks. In Y. Bramoullé, A. Galeotti, and B. Rogers (Eds.), *The Oxford Handbook of the Economics of Networks*. Oxford University Press.

Hagedorn, M., F. Karahan, I. Manovskii, and K. Mitman (2013). Unemployment benefits and unemployment in the Great Recession: The role of macro effects. *NBER Working Paper 19499*.

Hegselmann, R. and U. Krause (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation 5*(3).

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

Imbens, G. W. and C. F. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica 72*(6), 1845–1857.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton: Princeton University Press.

Jadbabaie, A., P. Molavi, A. Sandroni, and A. Tahbaz-Salehi (2012). Non-Bayesian social learning. *Games and Economic Behavior 76*(1), 210 – 225.

Kendal, R. L., I. Coolen, Y. van Bergen, and K. N. Laland (2005). Trade-offs in the adaptive use of social and asocial learning. Volume 35 of *Advances in the Study of Behavior*, pp. 333 – 379. Academic Press.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2013). Long-term neighborhood effects on low-income families: Evidence from Moving to Opportunity. *American Economic Review 103*(3), 226–231.

Manski, C. F. (2004). Social learning from private experiences: The dynamics of the selection problem. *The Review of Economic Studies 71*(2), 443–458.

Manski, C. F. (2007). *Identification for Prediction and Decision.* Harvard University Press.

Manski, C. F. (2011). Choosing treatment policies under ambiguity. *Annual Review of Economics 3*, 25–49.

Manski, C. F. (2015). Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern. *Journal of Economic Literature 53*(3), 631–653.

Manski, C. F. (2016). Credible interval estimates for official statistics with survey nonresponse. *Journal of Econometrics 191*(2), 293–301. Innovations in Measurement in Economics and Econometrics.

Molavi, P., A. Tahbaz-Salehi, and A. Jadbabaie (2015). Foundations of non-Bayesian social learning. *Mimeo., Columbia University.*

Moon, H. R. and F. Schorfheide (2012). Bayesian and frequentist inference in partially identified models. *Econometrica 80*(2), 755–782.

Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies 78*(1), 356–393.

Mueller-Frank, M. (2015). Reaching consensus in social networks. *IESE Working Paper 1116-E.*

Paris, J. B. and A. Vencovská (1990). A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning 4*(3), 183–223.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.

Pew (2012, October 15). More say there is solid evidence of global warming. *The Pew Research Center*. Retrieved from http://www.people-press.org/2012/10/15/more-say-there-is-solid-evidence-of-global-warming/.

Raj, B., M. L. Seltzer, and R. M. Stern (2004). Reconstruction of missing features for robust speech recognition. *Speech Communication 43*(4), 275–296.

Ramey, V. A. (2011). Can government purchases stimulate the economy? *Journal of Economic Literature 49*(3), 673–685.

Ranjan, R. and T. Gneiting (2010). Combining probability forecasts. *Journal of the Royal Statistical Society. Series B 72*(1), 71–91.

Romano, J. P. and A. M. Shaikh (2010). Inference for the identified set in partially identified econometric models. *Econometrica 78*(1), 169–211.

Schölkopf, B. and A. J. Smola (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.

Serrato, J. C. S. and P. Wingender (2016). Estimating local fiscal multipliers. *NBER WP 22425*.

Sethi, R. and M. Yildiz (2012). Public disagreement. *American Economic Journal: Microeconomics 4*(3), 57–95.

Sethi, R. and M. Yildiz (2016). Communication with unknown perspectives. *Econometrica 84*(6), 2029–2069.

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

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

Sorensen, A. T. (2006). Social learning and health plan choice. *The Rand Journal of Economics 37*(4), 929–945.

Sotiropoulos, D. N., C. Bilanakos, and G. M. Giaglis (2015). A time-variant and non-linear model of opinion formation in social networks. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 872–879.

Stoye, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica 77*(4), 1299–1315.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics 2*(1), 167–195.

van Bergen, Y., I. Coolen, and K. N. Laland (2004). Nine-spined sticklebacks exploit the most reliable source when public and private information conflict. *Proceedings of the Royal Society of London B: Biological Sciences 271*(1542), 957–962.

Wilmers, G. (2010). The social entropy process: Axiomatising the aggregation of probabilistic beliefs. *Mimeo., Manchester Institute for Mathematical Sciences*.

Zanardo, E. (2017). How to measure disagreement? *Mimeo., Columbia University*.

A Appendix: Notation

The Agent's Problem

p^k	Proposition $k \in \{1, \ldots, K\}$
$T(p^k)$	Truth-value of p^k, taking values in $\{0, 1\}$
λ_{it}^k	Agent i's beliefs at time t about $T(p^k)$, taking values in $[0, 1]$
φ_i^k	Agent i's model for transforming data into signals about $T(p^k)$
W_{it}^*	Data directly-observed by agent i yielding point-identified signals about $T(p^k)$ via φ_i^k
σ_{it}^{k*}	A point-identified signal
μ_i^{k*}	Agent i's ideal belief when $\mu_i^{k*} = \mathbb{E}_t[\sigma_{it}^{k*}]$
W_{it}	Data yielding set-identified signals about $T(p^k)$ via φ_i^k
Λ_{it}^{k*}	Set of possible beliefs (imprecise probability)
θ_{it}^k	Parameter that, along with σ_{it}^k, defines the correspondence $\varphi_i^k : W_{it} \rightrightarrows [\underline{\sigma}_{it}^{k*}, \overline{\sigma}_{it}^{k*}]$

The Agent's Model of Social Learning

$\mathcal{J}$	Network of J individuals sending signals to agent i
$\mathcal{J}^k \subseteq \mathcal{J}$	Set of J^k individuals sending signals to agent i about proposition p^k
$\mathcal{I}_{it}$	Agent i's directly-observed information set
$\mathcal{I}_{Jt}$	Agent i's socially-observed information set
φ_j^k	Individual j's model for transforming data into signals about $T(p^k)$
W_{jt}	Individual j's directly-observed data
σ_{jt}^k	Individual j's signal (point/single probability)
f^k	Agent i's model of social learning, or for interpreting signals from senders in her network
s_{jt}^k	Agent i's interpreted signal from sender j regarding proposition p^k
δ_{ijt}^k	Disagreement between agent i and sender j about proposition k , or $\lambda_{it}^k - \lambda_{jt}^k$
Δ_{ijt}	Credibility agent i assesses to (H1) when applied to sender j's signals
Δ_{it}^k	Total credibility of interpreted signals on proposition p^k
w_{jt}	Weight given to interpreted signals from sender j
s_{Jt}^k	Inferred signal from all socially-observed signals
$\widehat{\sigma}_{it}^k$	Inferred signal from combining directly-observed data and socially-observed signals

B Appendix: Experiments 2 and 3

In Experiment 2, I assume there are four clusters of individuals, $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, and $\mathcal{C}_4$ with $card(\mathcal{C}_1) = 50$, $card(\mathcal{C}_2) = 100$, $card(\mathcal{C}_3) = 50$, and $card(\mathcal{C}_4) = 100$. Initial beliefs λ_{i1}^k are generated as follows:

$$\overline{\lambda}_i^k \sim \begin{cases} \mathcal{N}(0.8, 0.1) & \text{if } i \in \mathcal{C}_1 \;\; \forall \;\; k = 1, \ldots, K; \\ \mathcal{N}(0.2, 0.01) & \text{if } i \in \mathcal{C}_2 \;\; \forall \;\; k = 1, \ldots, K; \\ \mathcal{N}(0.8, 0.1) & \text{if } i \in \mathcal{C}_3 \quad \text{and } k \text{ even}; \\ \mathcal{N}(0.5, 0.1) & \text{if } i \in \mathcal{C}_3 \quad \text{and } k \text{ odd}; \\ \mathcal{N}(0.2, 0.01) & \text{if } i \in \mathcal{C}_4 \quad \text{and } k \text{ even}; \\ \mathcal{N}(0.5, 0.01) & \text{if } i \in \mathcal{C}_4 \quad \text{and } k \text{ odd}; \end{cases}$$

with

$$\lambda_{i1}^k = \min\left\{\max\left\{\overline{\lambda}_i^k, 0\right\}, 1\right\} \quad \forall k = 1, \ldots, K.$$

Figures 4a and 5a shows the initial distribution of beliefs by clusters for propositions p^1 and p^2, respectively. Figures 4 and 5 show how beliefs evolve by cluster.

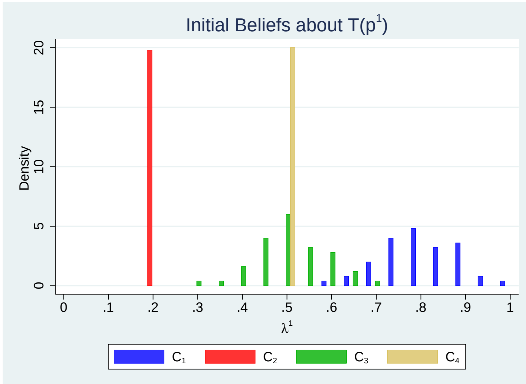
In Experiment 3, I also assume there are four clusters of individuals, $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, and $\mathcal{C}_4$ with $card(\mathcal{C}_1) = 50$, $card(\mathcal{C}_2) = 100$, $card(\mathcal{C}_3) = 50$, and $card(\mathcal{C}_4) = 100$. Initial beliefs λ_{i1}^k are generated as follows:

$$\overline{\lambda}_i^k \sim \begin{cases} \mathcal{N}(0.8, 0.1) & \text{if } i \in \mathcal{C}_1 \;\; \forall \;\; k = 1, \ldots, K; \\ \mathcal{N}(0.2, 0.01) & \text{if } i \in \mathcal{C}_2 \;\; \forall \;\; k = 1, \ldots, K; \\ \mathcal{N}(0.9, 0.1) & \text{if } i \in \mathcal{C}_3 \;\; \forall \;\; k = 1, \ldots, K; \\ \mathcal{N}(0, 0) & \text{if } i \in \mathcal{C}_4 \;\; \forall \;\; k = 1, \ldots, K; \end{cases}$$

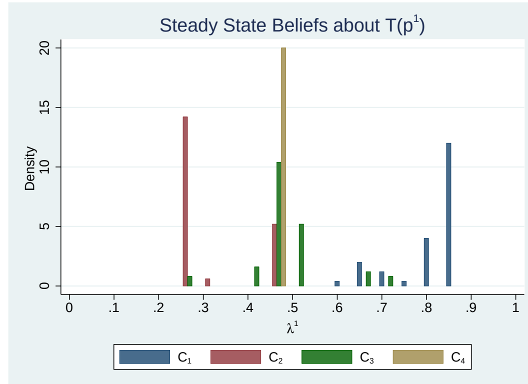
with

$$\lambda_{i1}^k = \min\left\{\max\left\{\overline{\lambda}_i^k, 0\right\}, 1\right\} \quad \forall k = 1, \ldots, K.$$

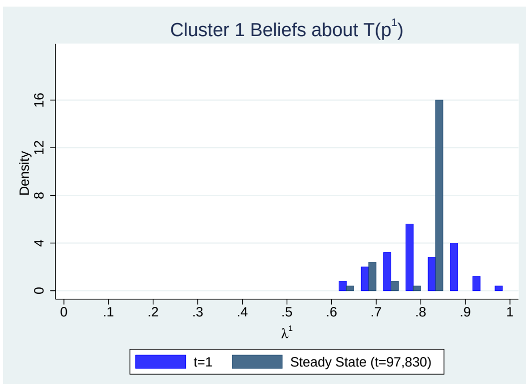
Figures 6a and 7a shows the initial distribution of beliefs by clusters for propositions p^1 and p^2, respectively. Figures 6 and 7 show how beliefs evolve by cluster.



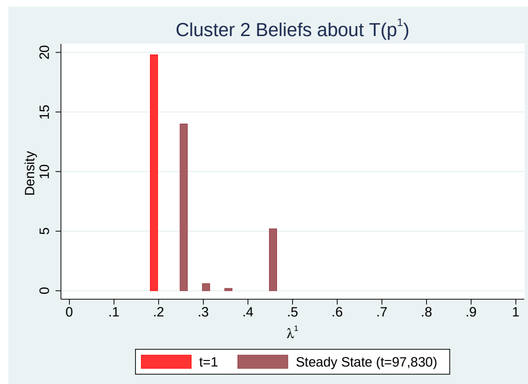
(a) Initial Distribution of Beliefs about $T(p^1)$



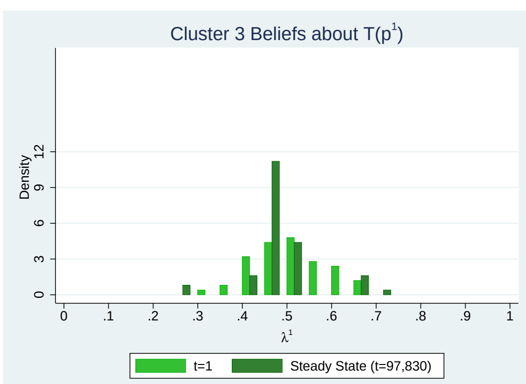
(b) Steady State Distribution of Beliefs about $T(p^1)$



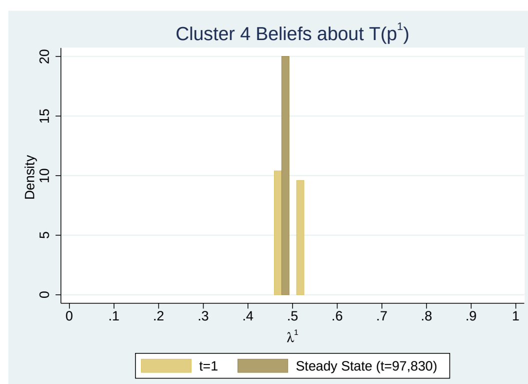
(c) Distributions of Beliefs about $T(p^1)$ for $\mathcal{C}_1$



(d) Distribution of Beliefs about $T(p^1)$ for $\mathcal{C}_2$



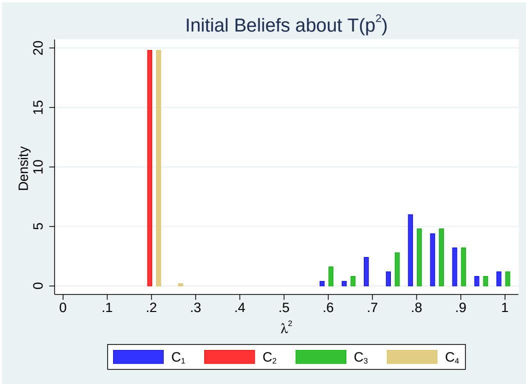
(e) Distributions of Beliefs about $T(p^1)$ for $\mathcal{C}_3$



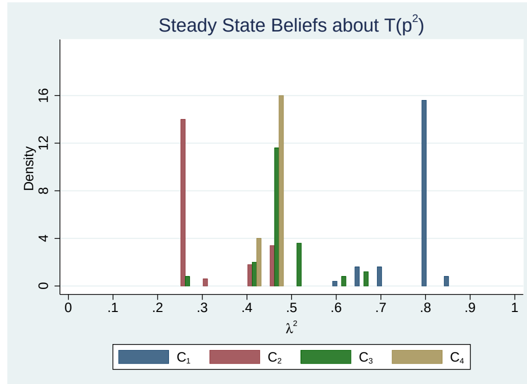
(f) Distribution of Beliefs about $T(p^1)$ for $\mathcal{C}_4$

Figure 4: Beliefs in Experiment 2

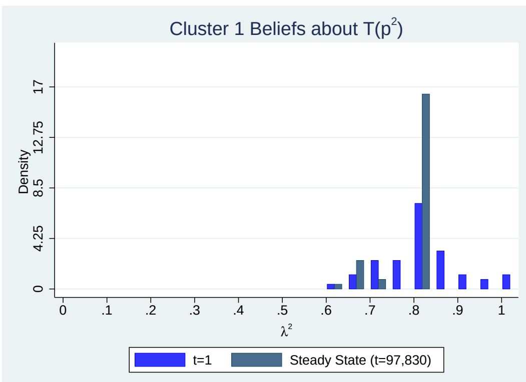
Steady state belief distributions are those for which $\max_{k,i} ||\lambda_{it}^k - \lambda_{it+1}^k|| < 1\text{e-}6$.



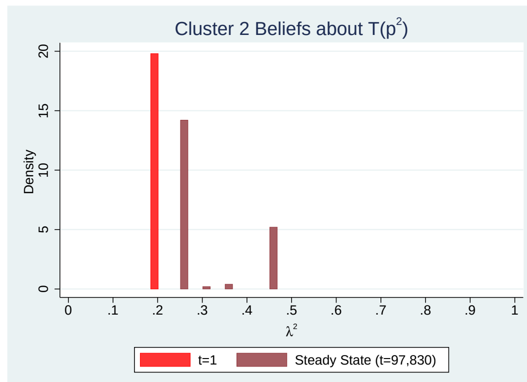
(a) Initial Distribution of Beliefs about $T(p^1)$



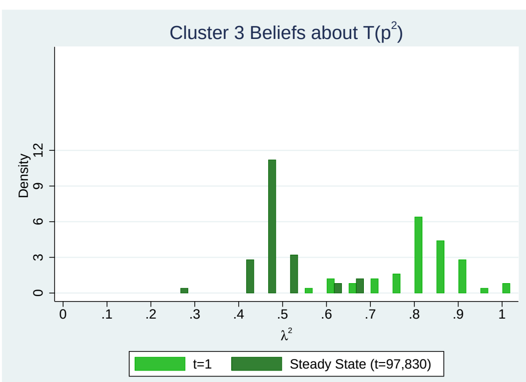
(b) Steady State Distribution of Beliefs about $T(p^1)$



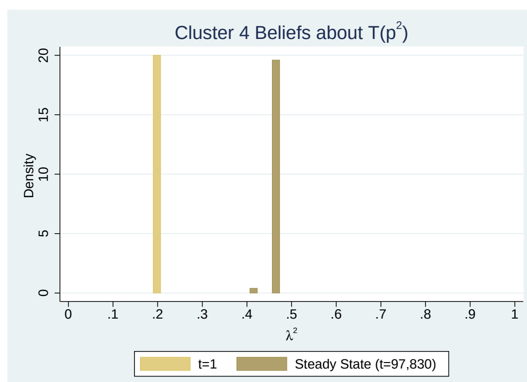
(c) Distributions of Beliefs about $T(p^1)$ for $\mathcal{C}_1$



(d) Distribution of Beliefs about $T(p^1)$ for $\mathcal{C}_2$

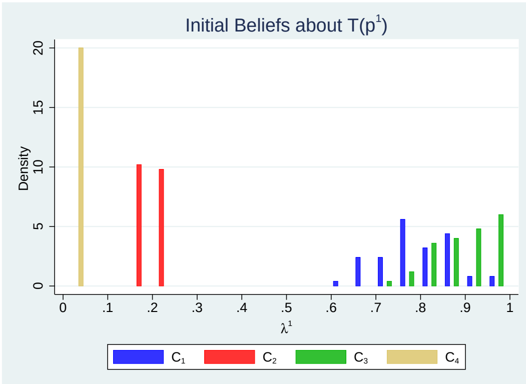


(e) Distributions of Beliefs about $T(p^1)$ for $\mathcal{C}_3$

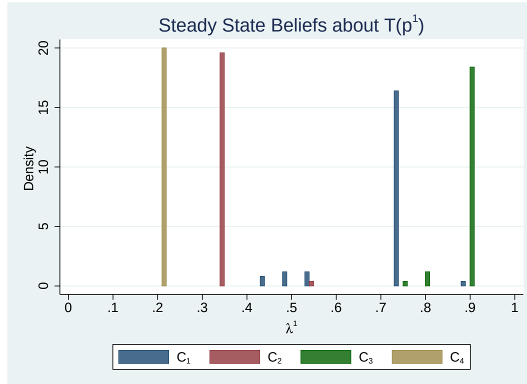


(f) Distribution of Beliefs about $T(p^1)$ for $\mathcal{C}_4$

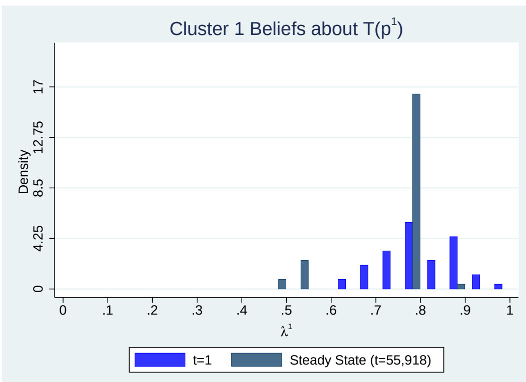
Figure 5: Beliefs in Experiment 2
Steady state belief distributions are those for which $\max_{k,i} ||\lambda_{it}^k - \lambda_{it+1}^k|| < 1\text{e-}6$.



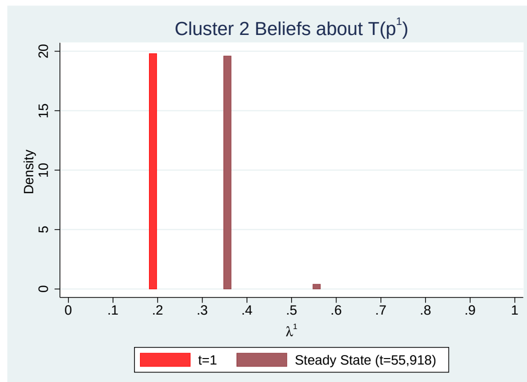
(a) Initial Distribution of Beliefs about $T(p^1)$



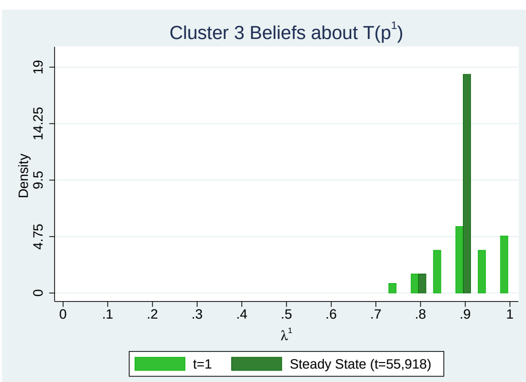
(b) Steady State Distribution of Beliefs about $T(p^1)$



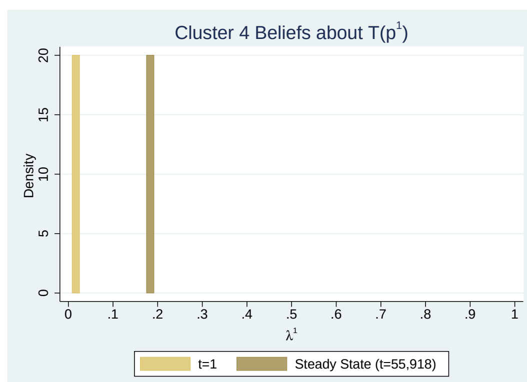
(c) Distributions of Beliefs about $T(p^1)$ for $\mathcal{C}_1$



(d) Distribution of Beliefs about $T(p^1)$ for $\mathcal{C}_2$



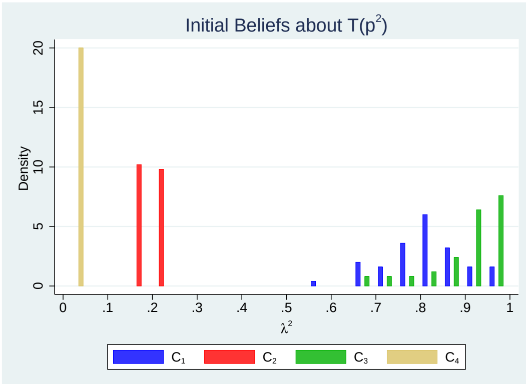
(e) Distributions of Beliefs about $T(p^1)$ for $\mathcal{C}_3$



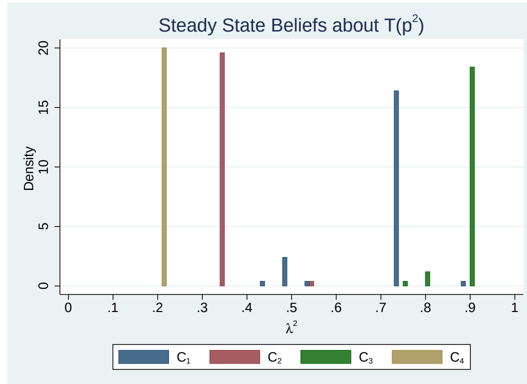
(f) Distribution of Beliefs about $T(p^1)$ for $\mathcal{C}_4$

Figure 6: Beliefs in Experiment 3

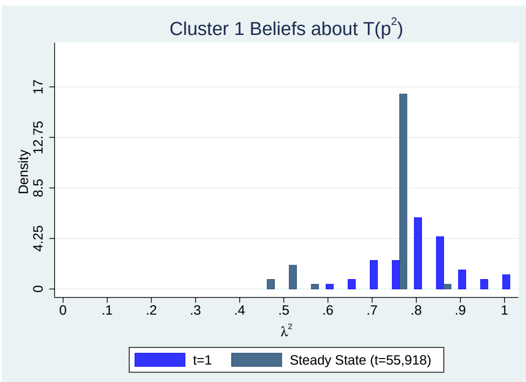
Steady state belief distributions are those for which $\max_{k,i} ||\lambda^k_{it} - \lambda^k_{it+1}|| < 1\text{e-}6$.



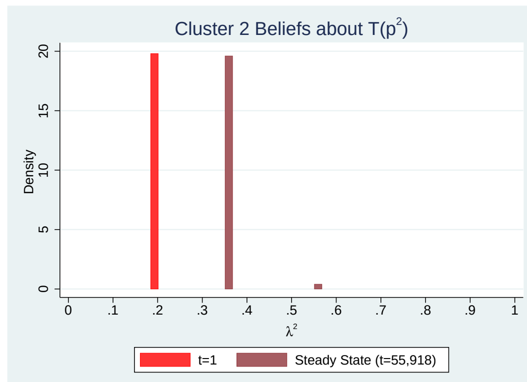
(a) Initial Distribution of Beliefs about $T(p^1)$



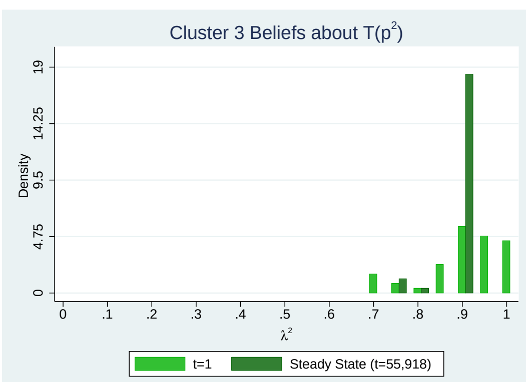
(b) Steady State Distribution of Beliefs about $T(p^1)$



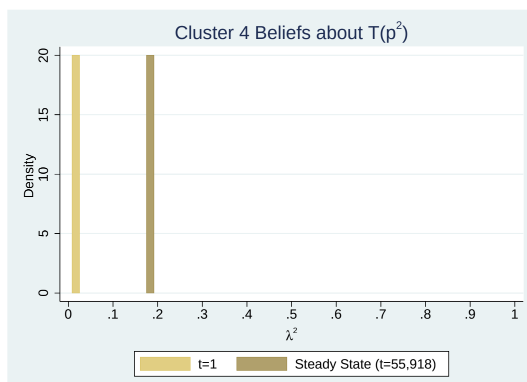
(c) Distributions of Beliefs about $T(p^1)$ for $\mathcal{C}_1$



(d) Distribution of Beliefs about $T(p^1)$ for $\mathcal{C}_2$



(e) Distributions of Beliefs about $T(p^1)$ for $\mathcal{C}_3$



(f) Distribution of Beliefs about $T(p^1)$ for $\mathcal{C}_4$

Figure 7: Beliefs in Experiment 3
Steady state belief distributions are those for which $\max_{k,i} ||\lambda_{it}^k - \lambda_{it+1}^k|| < 1\text{e-}6$.