PMLR: Proceedings of Machine Learning Research, vol. 62, 1-12, 2017 ISIPTA '17

Differences of Opinion

Dionissi Aliprantis DIONISSI.ALIPRANTIS@CLEV.FRB.ORG
Federal Reserve Bank of Cleveland (US)

Abstract

This paper considers the resolution of ambiguity according to the scientific ideal of direct observation when there is a practical necessity for social learning. An agent faces ambiguity when she directly observes low-quality data yielding set-identified signals. I suppose the agent's objective is to choose the single belief replicating what would occur with high-quality data yielding point-identified signals. I allow the agent to solve this missing data problem using signals 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; social learning; partial identification; causal inference; DeGroot learning rule; bounded confidence.

1. Introduction

We all hold beliefs based on limited personal experience. This is often due to logistical, and not philosophical, limitations. The scientific ideal of "seeing for one's self" is subject to time and resource constraints that make it infeasible to personally verify all claims. How do we form beliefs based on evidence beyond our personal experience?

This paper studies scenarios of partial identification in which personal experience offers no guidance for belief formation beyond a range of possibilities. Consider the example of a high school guidance counselor advising minority students on whether to attend a selective or non-selective college. What is the probability that an advisee will graduate from the selective college? The counselor would face partial identification if the high school had not tracked the experiences of recent graduates, or had sent few students to selective colleges.

When facing partial identification, the counselor could provide his students with a range of probabilities. Alternatively, the counselor could provide a single probability based on information beyond his directly-observed data. The choice of a single probability would use the counselor's judgment to combine his own experience; his discussions with others like counselors or teachers; and the conflicting estimates in the literature (Arcidiacono and Lovenheim, 2016; Alon and Tienda, 2005). This paper models the counselor's choice of a single probability.

The general setting begins with an agent who must form beliefs about a set of propositions. The agent can use a model to translate data into signals about the truth of each proposition. Under frequentist inference she may form beliefs as the mean of her signals observed over discrete time.[1]

There are many situations in which the available data might only allow the agent to partially identify a signal. An obvious scenario pertains to causal propositions when one cannot easily ob-

1. For independent and identically distributed (iid) signals, the Law of Large Numbers ensures such beliefs will converge to the mean of the signal distribution.

serve the Data Generating Process (DGP) under controlled intervention. This situation is widespread in economics, with many important counterfactual outcomes waiting to be definitively quantified.[2] Beliefs of an agent observing iid partially-identified signals will converge to a set when formed by averaging signals observed over discrete time (Artstein and Vitale, 1975).

When the agent has a set of possible beliefs, or faces ambiguity, prominent decision rules instruct her to choose the single belief generating an extreme utility (Gilboa and Marinacci, 2013). 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).

This paper separates belief formation from preferences: When choosing one belief, the objective is to accurately represent the DGP. While the scientific ideal is to attain this objective based on direct observation, no single belief cannot satisfy this ideal when directly-observed data are only capable of partial identification. However, a single belief can approximate the scientific ideal if data yielding point-identification can be inferred from second-hand observations.

I specify the agent's problem as an attempt to replicate the beliefs she would have formed had she directly observed data yielding point-identified signals. The agent's problem can be viewed as a missing data problem to be solved with signals observed through her social network. I assume that communication is imperfect, so that socially-observed signals are communicated alone, without the data or model used in their construction.

I first show that if the agent uses linear opinion pooling of signals, a common method for combining forecasts and estimates, she will follow the canonical DeGroot (1974) learning rule under a special case of observed data. I then show that such DeGroot updating solves the agent's problem under additional assumptions on the homogeneity of data and models in the agent's network.

Two issues argue for pushing beyond the assumptions necessary for DeGroot updating to solve the agent's problem. The first is that the assumptions justifying DeGroot updating are strong. For example, individuals can be justified in using different models to interpret the same data (Al-Najjar, 2009), and the agent might observe new data over time (Jadbabaie et al., 2012).

Second, while DeGroot updating is the benchmark for non-Bayesian learning on social networks, a combination of theory and evidence motivates the desideratum of an alternative capable of generating polarization (Golub and Sadler, 2016). DeGroot learning and many of its generalizations converge to a degenerate distribution for connected networks (Jackson, 2008; Dandekar et al., 2013).[3] However, an empirical analogue of a connected network - individuals exposed to sources of information contradicting their beliefs - is often observed together with persistent disagreement. Examples include scientific opinions when journals publish opposing research and public opinion when individuals are exposed to diverse news sources (Gentzkow and Shapiro, 2011).[4] The emergence of "fake news" highlights this limitation of DeGroot updating.

2. In microeconomics alone it has proven difficult to ascertain outcomes under controlled interventions to neighborhood characteristics (Ludwig et al., 2008; Aliprantis, 2017), teacher characteristics (Rothstein, 2010; Kinsler, 2012), educational attainment (Angrist and Krueger, 1991; Aliprantis, 2012), minimum wages (Card and Krueger, 1994; Neumark and Wascher, 2000), unemployment benefits (Hagedorn et al., 2013; Farber and Valletta, 2015), income taxes (Manski, 2014), and right-to-carry laws (Manski and Pepper, 2015; Durlauf et al., 2016).

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

4. For example, 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.

I show that linear opinion pooling can still solve the agent's problem after weakening the assumptions justifying DeGroot updating. In contrast to DeGroot updating, though, this solution 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.

The agent might nevertheless choose a model for interpreting signals, just as methods for causal inference attempt to overcome the fundamental problem of evaluation. I study how the agent might use the model implied by a "reasonable" heuristic. The agent first interprets signals according to the model. The agent then uses the relative entropy of disagreement over all propositions to assess the credibility of applying the heuristic to each sender. The agent then combines interpreted signals, giving more weight to the interpreted signals from senders deemed most credible.

Although the 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. A key mechanism is generated by the use of relative entropy to assess the credibility of interpreted signals. 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 more on interpreted signals from the disagreeing cluster, and this can cause her to overcompensate when they provide her with unbiased signals.

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. In the full paper I also show why finding a model of social learning to solve the agent's problem is an ill-posed problem, and describe the implications of a heuristic the agent might use to specify a model of social learning. I further investigate the implications of this heuristic in greater detail, studying belief dynamics under one specification of the updating rule for several paramaterizations under various network and proposition structures. Section 4 concludes.

2. Belief Formation via Directly-Observed Data

Suppose there is a finite set of propositions $\{p^1, p^2, \ldots, p^K\} = \mathcal{K}$, none of which can be written as a compound proposition using other propositions in the set.[5] 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 $\lambda_{it}^k = \Pr(T(p^k) = 1)$.[6] The agent directly observes data W_{it}.

5. 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.

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

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

$$\sigma^{k*}_{it} = \varphi^k_i(W^*_{it}) \in [0,1].$$

The law of large numbers ensures convergence to the mean of the signal distribution, which I will denote by μ^{k*}_i, for beliefs formed as

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

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^{k*}_{it}. 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^k_{it} and its quality θ^k_{it},

$$(\sigma^k_{it}, \theta^k_{it}) = \varphi^k_i(W_{it}) \in [0,1]^2,$$

where the true signal is related to the observed signal by

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

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

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

where the sets $[\underline{\sigma}^{k*}_{it}, \overline{\sigma}^{k*}_{it}]$ and Λ^{k*}_{it+1} are often referred to as "imprecise probabilities" (Coolen et al., 2011). The set $[\underline{\sigma}^{k*}_{it}, \overline{\sigma}^{k*}_{it}]$ is what can be learned about p^k from the directly-observed data under the most credible assumptions. While the agent can also determine σ^k_{it}, doing so requires less credible assumptions, so the agent cannot be sure that $\mathbb{E}[\sigma^k_{it}] = \mu^{k*}_i$ unless $\theta^k_{it} = 1$.

The canonical example of the proposition p^1= "A given coin will land Heads." helps to illustrate the difference between these settings. Suppose that high-quality data maps into signals generated by iid draws from a binomial distribution with probability 0.5 where $\sigma = 1$ if the coin lands Heads and $\sigma = 0$ if the coin lands Tails. In the case of high-quality data where $\theta^1_{it} = 1$ for all t, $\sigma^k_{it} = \sigma^{k*}_{it}$, and so λ^{k*}_{it+1} can be calculated from (1) as the relative frequency of Heads, and will converge to 0.5 as $t \to \infty$.

In contrast, an agent with low-quality data mapping into signals represented by $\theta^1_{it} = 0.2$ for all t will be subject to ambiguity in addition to risk.[7] If the observed signal is Heads, then the agent can bound the true signal to be within $[0.2, 1]$. If the observed signal is Tails, then the agent bounds the true signal to be in $[0, 0.8]$. Thus as $t \to \infty$, the agent will infer that the mean of the true signals is $\mu^{k*}_i \in \Lambda^{k*}_i = [0.1, 0.9]$.[8]

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

8. 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 (2008).

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 ($\theta_{it}^k = 1$), and "low-quality" to describe data yielding set-identified signals ($\theta_{it}^k < 1$). For causal propositions, the difficulty of achieving identification is an obvious interpretation of signals having low quality. Examples abound of counterfactual outcomes that are difficult to quantify in microeconomics, macroeconomics, and finance because one cannot easily observe the Data Generating Process (DGP) under controlled intervention.[9]

Non-causal propositions can also have low-quality signals for reasons like survey non-response (Manski, 2015). Another interpretation of an extremely low-quality signal, $\theta_{it}^k = 0$, is that the agent does not directly observe any data for a given proposition 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).

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

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

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

9. See Footnote 2 for some examples from microeconomics.

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

$$\mathcal{I}_{Jt} \equiv \left\{ \; \{\lambda^1_{jt}, \sigma^1_{jt}\}_{j\in\mathcal{J}^1} \;\; , \;\; \ldots \;\; , \;\; \{\lambda^K_{jt}, \sigma^K_{jt}\}_{j\in\mathcal{J}^K} \; \right\},$$

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:

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

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

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

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 $f(\sigma^k_{it}, \{\sigma^k_{jt}\}_{j\in\mathcal{J}^k}|T(p^k))$ can require unrealistic assumptions about the information and computation available to the agent (Acemoglu and Ozdaglar, 2011).[10] 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

$$f(\varphi^k_{it}(W^*_{it}), \{\varphi^k_j(W_{jt})\}_{j\in\mathcal{J}^k}|T(p^k)),$$

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 $\{\varphi^k_j\}_{j\in\mathcal{J}^k}$. 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^k_{jt} and σ^k_{jt}. She cannot observe measures of the sender's ambiguity $\Lambda^{k*}_{jt}, \theta^k_{jt}$ or their model $\varphi^k_j \;\; \forall \;\; j, t, k$

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 to directly observing the data when information is obtained socially. Among other reasons, this assumption is positively appealing because there is a well-documented tendency for researchers and statistical agencies to focus on communicating their point estimates σ^k_{it} without communicating about their models φ^k_i or measures of uncertainty θ^k_{it} (Manski, 2007, 2015).

10. Benoît and Dubra (2015) and Andreoni and Mylovanov (2012) study polarization under private learning when agents disagree about $f(\sigma^{k*}_{it}|T(p^k))$. Alternatively, in this context their analyses could be interpreted as agents having different models for private learning φ^k_i, 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 $p^{k''}$.

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

$$\min_{f^1,\ldots,f^K\in\mathcal{F}} \quad \sum_{k=1}^{K}\mathcal{L}\left(\mathbb{E}\left[\mu_i^{k*}-\lim_{t\to\infty}\lambda_{it+1}^k\right]\right) \tag{3}$$

$$\begin{aligned} s.t.\quad & (\mathcal{I}_{it},\mathcal{I}_{Jt}) \\ & \widehat{\sigma}_{it}^k = f^k(\mathcal{I}_{it},\mathcal{I}_{Jt}) && \text{for}\quad k=1,\ldots,K \\ & \lambda_{it+1}^k = \beta_t\widehat{\sigma}_{it}^k + (1-\beta_t)\lambda_{it}^k && \text{for}\quad k=1,\ldots,K \end{aligned}$$

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.[11] 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 Some Solutions to 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 $t\geq 2$ as*

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

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

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

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

11. 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.

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

$$\begin{aligned}\sigma^k_{it+1} &= \beta_t \widehat{\sigma}^k_{it} + (1-\beta_t)\sigma^k_{it} \\ &= \beta_t \theta^k_i \sigma^k_{it} + (1-\beta_t)\sigma^k_{it} + \beta_t(1-\theta^k_i)\sum_{j\in\mathcal{J}^k} w^k_j \sigma^k_{jt}.\end{aligned}$$

■

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 $\beta_t \widehat{\sigma}^k_{it} + (1-\beta_t)\lambda^k_{it} = \frac{1}{t}\sum_{n=1}^{t} \widehat{\sigma}^k_{in}$*

If the observed data yield unbiased signals

(A3) *Private signals are iid with $\mathbb{E}[\sigma^{k*}_{it}] \equiv \mu^{k*}_i = \mu^k_i \equiv \mathbb{E}[\sigma^k_i]$, and*

(A4a) *Social signals are iid for each $j \in \mathcal{J}^k$ with $\mathbb{E}[\sigma^{k*}_{it}] \equiv \mu^{k*}_i = \mu^k_j \equiv \mathbb{E}[\sigma^k_{jt}]$ $\forall j \in \mathcal{J}^k$,*

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

$$\lim_{t\to\infty} \sigma^k_{it+1} = \sum_{n=1}^{J+1} \pi^k_n \sigma^k_{n1}$$

where π^k_n is n's left-hand eigenvector centrality in Ω^k. Since $\sum_{n=1}^{J+1} \pi^k_n = 1$ and $\mathbb{E}[\sigma^k_{n1}] = \mu^{k*}_i$ for all n, we know that

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

■

We can imagine scenarios in which the agent observes data and signals in each period, but this additional 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

$$\widehat{\sigma}_{it}^k = \theta_{it}^k \sigma_{it}^k + (1-\theta_{it}^k)\sigma_{Jt}^k \quad \textit{where} \tag{6}$$

$$\sigma_{Jt}^k = \sum_{j \in \mathcal{J}^k} w_{jt}^k s_{jt}^k \qquad \textit{with} \;\; w_{jt}^k \geq 0 \;\; \forall \;\; j \in \mathcal{J}^k, \;\; \sum_{j \in \mathcal{J}^k} w_{jt}^k = 1 \tag{7}$$

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

solves the agent's problem.

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

$$\begin{aligned}
\lim_{t\to\infty} \lambda_{it+1}^k = \mathbb{E}[\widehat{\sigma}_{it}^k] \;\; &= \mathbb{E}[\overline{\theta}_i^k \sigma_{it}^k + (1-\overline{\theta}_i^k)\sigma_{Jt}^k] \;= \overline{\theta}_i^k \mathbb{E}[\sigma_{it}^k] + (1-\overline{\theta}_i^k)\mathbb{E}[\sigma_{Jt}^k] \\
&= \overline{\theta}_i^k \mathbb{E}[\sigma_{it}^k] + (1-\overline{\theta}_i^k)\mathbb{E}[\sum_{j\in\mathcal{J}^k} w_{jt}^k \sigma_{jt}^k] = \overline{\theta}_i^k \mathbb{E}[\sigma_{it}^k] + (1-\overline{\theta}_i^k)\sum_{j\in\mathcal{J}^k} w_{jt}^k \mathbb{E}[\sigma_{jt}^k] \\
&= \overline{\theta}_i^k \mu_{it}^k + (1-\overline{\theta}_i^k)\sum_{j\in\mathcal{J}^k} w_{jt}^k \mu_{jt}^k \qquad (9) \\
&= \mu_i^{k*}.
\end{aligned}$$

■

4. 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.

Acknowledgment

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

The opinions expressed are those of the author and do not represent views of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System.

References

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

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

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

D. Aliprantis. Redshirting, compulsory schooling laws, and educational attainment. *Journal of Educational and Behavioral Statistics*, 37(2):316–338, 2012.

D. Aliprantis. Assessing the evidence on neighborhood effects from Moving to Opportunity. *Empirical Economics*, 2017. doi:10.1007/s00181-016-1186-1.

S. Alon and M. Tienda. Assessing the "mismatch" hypothesis: Differences in college graduation rates by institutional selectivity. *Sociology of Education*, 78(4):294–315, 2005.

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

J. D. Angrist and A. B. Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.

P. Arcidiacono and M. Lovenheim. Affirmative action and the quality-fit trade-off. *Journal of Economic Literature*, 54(1):3–51, 2016.

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

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

C. R. Bollinger and M. van Hasselt. A Bayesian analysis of binary misclassification: Inference in partially identified models. *Mimeo., University of Kentucky*, 2008.

D. Card and A. B. Krueger. Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4):772–793, 1994.

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

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

P. Dandekar, A. Goel, and D. T. Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, 2013. doi:10.1073/pnas.1217220110.

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

S. N. Durlauf, S. Navarro, and D. A. Rivers. Model uncertainty and the effect of shall-issue right-to-carry laws on crime. *European Economic Review*, 81:32–67, 2016.

H. S. Farber and R. G. Valletta. Do extended unemployment benefits lengthen unemployment spells?: Evidence from recent cycles in the U.S. labor market. *Journal of Human Resources*, 50 (4):873–909, 2015.

M. Gentzkow and J. M. Shapiro. Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839, 2011. doi:10.1093/qje/qjr044.

I. Gilboa and M. Marinacci. Ambiguity and the Bayesian paradigm. In D. Acemoglu, M. Arellano, and E. Dekel, editors, *Advances in Economics and Econometrics: Tenth World Congress: Economic Theory*, volume I, chapter 7, pages 179–242. Cambridge University Press, 2013.

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

B. Golub and M. O. Jackson. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338, 2012. doi:10.1093/qje/qjs021.

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

M. Hagedorn, F. Karahan, I. Manovskii, and K. Mitman. Unemployment benefits and unemployment in the great recession: The role of macro effects. Technical report, 2013.

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

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

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

J. Kinsler. Assessing Rothstein's critique of teacher value-added models. *Quantitative Economics*, 3(2):333–362, 2012.

J. Ludwig, J. B. Liebman, J. R. Kling, G. J. Duncan, L. F. Katz, R. C. Kessler, and L. Sanbonmatsu. What can we learn about neighborhood effects from the Moving to Opportunity experiment? *American Journal of Sociology*, 114(1):144–188, 2008.

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

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

C. F. Manski. Identification of income-leisure preferences and evaluation of income tax policy. *Quantitative Economics*, 5(1):145–174, 2014.

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

C. F. Manski and J. V. Pepper. How do right-to-carry laws affect crime rates? coping with ambiguity using bounded-variation assumptions. *NBER Working Paper 21701*, 2015.

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

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

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

D. Neumark and W. Wascher. Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania: Comment. *The American Economic Review*, 90(5): 1362–1396, 2000.

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

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

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

J. Rothstein. Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1):175–214, 2010.

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

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

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