False Information and Disagreement in Social Networks*

Evan Sadler†

June 1, 2018

Abstract

Disagreement, including on matters of fact, is a pervasive phenomenon, yet this is incompatible with existing work on social learning. I propose a model of information processing with two key features: (i) the agent encounters false information, and (ii) the agent cannot distinguish true propositions from false ones. I study two families of axioms for update rules, finding that "willingness-to-learn" axioms are incompatible with "non-manipulability" axioms. I also provide an axiomatic characterization of several update rules. In a simple social learning model, disagreement is not just possible, but generic. I characterize the influence of each agent on steady-state beliefs and apply the framework to study echo chambers and belief manipulation.

1 Introduction

People often disagree about facts. There is widespread belief in the United States that climate change is a hoax. A significant number of people believe that vaccines cause autism. Studys of fake news during the 2016 elections reveal that people did believe many false stories propagating on social media (Allcott and Gentzkow, 2017). This phenomenon presents a puzzle because economic models of social learning do not permit long-run disagreement. Indeed, from Bayesian approaches to ones with heuristic updating, long-run consensus is perhaps the most robust prediction across these models (Golub and Sadler, 2016). That this fails to match our casual observations of the world should give us pause: to productively comment on what information and network structures best promote information aggregation and consensus, we need a model that generates realistic disagreement.

Behind theoretical results on consensus lies an unspoken assumption about the way people process information. Implicit in essentially all economic models is that every piece of information an agent encounters is true—the agent observes a partition element of a state

*I am especially grateful to David Hirshleifer and Ben Golub, and to participants at the 2017 Retreat on Information, Networks, and Social Economics, for invaluable feedback on this project. I also thank Daron Acemoglu, Yeon-Koo Che, Barton Lipman, Doron Ravid, and Yuval Salant for helpful conversations and suggestions.

†Harvard University – esadler@fas.harvard.edu

space, and this partition element contains the true state. In reality—whether in the form of fake news stories, rumors, or sincere but mistaken statements—we are constantly bombarded with falsehoods. We cannot always tell what is true or not, and we have to decide what to believe. In this paper, I propose a theory of information processing in which agents can and do hear false information. This theory feeds into a social learning model in which disagreement is a generic outcome. By taking seriously agents' epistemological challenges, we obtain a natural framework to study echo chambers and belief manipulation.

An agent sequentially encounters propositions. I represent a proposition as a set of states—what we typically call an event—but a proposition need not contain the true state. The agent holds in memory a set of propositions she believes are true at any given moment—these are her beliefs, meaning she thinks the true state lies in the intersection of these propositions. On encountering a new proposition, the agent applies an update rule to arrive at new beliefs. The update rule must allow the agent to process *any* proposition, whether true or not. I first ask: is there a "good" update rule in this setting?

I consider two classes of axioms we might wish to satisfy. "Willingness-to-learn" axioms capture the idea that a good update rule should allow the agent to learn about the world, and hopefully arrive at the truth. In contrast, "non-manipulability" axioms capture the idea that a good update rule should yield the same beliefs given the same information, no matter how it is presented. An impossibility theorem shows that no rule can simultaneously satisfy both classes: the ability to learn necessarily opens the agent to manipulation.

I subsequently provide an axiomatic characterization of update rules, a representation theorem. These rules capture common psychological biases, such as confirmation bias and motivated reasoning,[1] as well as skepticism. Every rule encompasses a particular constellation of willingness-to-learn and non-manipulability axioms. In light of the impossibility result, this reveals how each update rule embodies distinctive tradeoffs. I make a normative argument in favor of two particular update rules—a stubborn rule under which agents always retain their current beliefs and a variant under which agents accept a proposition only if it resolves all uncertainty—that inform our later study of social learning.

We should view this framework as complementary to the theory of subjective probability, rather than in competition with it. The present paper says nothing about how to reason over possible states—an agent may hold a subjective prior and make decisions to maximize expected utility. Likewise, subjective probability theory says nothing about how to update following probability zero events. After observing the impossible, an agent must still decide what states to consider going forward. I draw an analogy to Kuhn's theory of scientific paradigms. Savage's subjective probability theory corresponds to "normal science," while the framework of this paper details the shifts between paradigms.

In principle, we can accommodate false statements within the standard paradigm. However, taking this approach to its natural conclusion precludes learning in most situations. We rarely have direct observations of payoff relevant signals. Instead, we hear others' reports

[1]See Nickerson (1998) for a survey of evidence on confirmation bias. Within the economics literature, Rabin and Schrag (1999) provide a canonical analysis of confirmation bias. Kahan (2013) and Tappin et al. (2017) offer recent evidence on motivated reasoning.

through conversations, newspapers, television, or the internet. Since these reports are logically independent of the actual signal—that John said P is logically independent of whether P is true—each necessitates an expansion of our state space to account for a potentially false report. If the veracity of reports is always in question, we obtain no actual information. A sequence of reports yields no "approach to certainty," only overwhelming possibilities: the agent's prior, not the data, remains the dominant influence on any decision.

In practice, we escape the issue through what philosopher Paul Grice termed the "cooperative principle." Grice's insight was that listeners make certain presumptions of a speaker in normal conversation. We presume a speaker is being truthful (maxim of quality), is providing all needed information, and no extra (maxim of quantity), is being relevant to the topic at hand (maxim of relation), and is trying to be clear (maxim of manner). Speakers in turn exploit these presumptions to convey meaning beyond their literal words. The maxim of quality is most clearly relevant to this paper. If we presume a speaker avoids saying that which she knows is false, and that for which she has insufficient evidence, we can treat her report as if it were a direct observation. If our presumption is ever incorrect, or people make honest mistakes, we open the door for falsehoods to propagate. This necessitates a way to resolve conflicting statements that are each made in good faith.

Drawing on these ideas, I introduce a stylized model of social learning in a network. Some agents accept new propositions if they resolve all uncertainty. Others are stubborn and never give up their beliefs. We might view the latter as "experts" with direct observations, who are therefore unwilling to defer to others. All agents are endowed with beliefs that are more likely correct than not and have random opportunities to communicate with neighbors. Belief evolution here is strikingly different from that in other learning models. Generically, agents fail to reach consensus—if two stubborn agents ever disagree, they always do so—and even when they do, the probability of a correct consensus is independent of both the population size and the network structure: there is no meaningful aggregation of information.

Given the network structure and the identities of the stubborn agents, we can characterize the long-run probability that each agent holds correct beliefs as well as the distinct influence that each stubborn agent exerts on each other agent. In general, influential agents are those with a large number of relatively short paths to others in the network. In expectation, the probability that any agent holds correct beliefs in the long run is equal to the probability that any one stubborn agent begins with correct beliefs.

Echo chambers are a natural phenomenon in this framework. Cohesive groups can sustain distinctive beliefs, even if most agents are open to being persuaded, because certain propositions are frequently repeated. This suggests questions about how to shift societal beliefs. In examples, I explore the effects of adding stubborn agents, sowing doubt, changing links, and fact-checking—the model offers a straightforward way to evaluate each intervention.

There are long-standing and prominent disagreements in politics, science, health, education, and many other contexts. Policies that leverage social learning can help promote better collective decision making, but to be effective we need a deeper understanding of the underlying influence mechanisms. Our typical view of uncertainty fails to grapple with a fundamental question: how do we know if something is true? Uncertainty about a data generating process imposes limitations on rational choice. To work within "small worlds" as Savage implored, we must take some things for granted. We can make a fuzzy distinction between things of which we are absolutely sure and things that in essence we take on faith. Most information that guides our choices falls closer to the latter end of this spectrum, which makes it crucial to understand learning dynamics in this environment.

1.1 Related Work

In a seminal contribution, Savage (1954) gave foundations for expected utility maximization with respect to a subjective probability measure, a framework that underlies much of modern economics. Savage discusses limitations at length, contrasting the proverbs "one can cross that bridge when one comes to it" and "look before you leap" to represent two extreme approaches to decision making. Recognizing the absurdity of optimizing over every possibile contingency over the course of one's life (i.e. "look before you leap"), he argues that:

> *[T]o cross one's bridges when one comes to them means to attack relatively simple problems of decision by artificially confining attention to so small a world that the "Look before you leap" principle can be applied there. (pp. 16)*

However, Savage never says how to select these "small worlds," nor how to deal with the challenges that inevitably arise after choosing the "wrong" one. The present paper partially addresses this unfinished work.

My framework vaguely resembles theories of belief revision in propositional logic (e.g. Alchourrón et al., 1985). These theories represent beliefs as the logical closure of a set of propositions. A set of axioms characterizes a rule in which one always accepts a new proposition, adding it to the existing knowledge base if it is consistent and otherwise discarding the old knowledge base. Representing uncertainty through a state space lends a different structure to the present paper, and our axioms do not have direct analogs in earlier work.

Rubinstein and Salant (2006) provide a more recent inspiration. They study choice among alternatives that are presented as an ordered list, and natural axioms lead to choice rules that are order dependent. I adopt a similar approach to belief updating. An agent processes information as an ordered list of propositions, and the list order often affects beliefs. However, belief updating is different from choice among alternatives in important ways—an agent may believe multiple propositions simultaneously, and there are consistency constraints on which ones. My results complement the earlier paper, enhancing our understanding of how procedural aspects of choice and belief formation affect decision making.

This paper fits into a broader context of research that studies decision making when agents face cognitive limitations or have imperfect models of the environment. For instance, Spiegler (2016) considers agents with possibly incorrect causal models of the world, which distorts expectations and can lead to different action choices. Similarly, Jehiel (2005) and Mullainathan et al. (2008) explore the consequences of categorical thinking, which can create opportunities for persuasion, while Gennaioli and Schleifer (2010) and Wilson (2014) show how imperfect recall accounts for many cognitive biases. This research has given us new and useful explanations for behavior that deviates from standard models, and we should view the present paper as furthering this agenda.

The social learning model I study is closely related to "voter models" from statistical physics (Liggett, 1997; Mossel and Tamuz, 2014). In a voter model, each agent has one of two possible opinions at any moment in time, and neighbors stochastically influence each other to switch their opinions. Here, agents have stochastic opportunities to communicate propositions to one another, and applying an update rule leads to discrete shifts in beliefs. Yildiz et al. (2013) is closest to this paper, studying a voter model in which some agents are stubborn and do not change their opinions—this is formally equivalent to the model in this paper if we assume a binary state space. I make two innovations on this literature. The first is conceptual. Voter models are typically used to represent *opinion* dynamics rather than beliefs about a factual state of the world. The axiomatization of update rules justifies a broader interpretation and suggests alternative ways to manipulate beliefs. Second, when we consider learning over a larger state space, this framework permits a richer range of behavior.

The closest analogs to voter models in the economics literature are the learning models based on DeGroot updating. Golub and Jackson (2010, 2012) provide the canonical analysis, giving conditions under which a large population converges on the truth and assessing the speed of convergence. Mueller-Frank (2018) considers a class of related update rules and studies belief manipulation. The key analytical tools come from the study of finite state Markov chains, and this paper leverages the same tools, characterizing steady-state beliefs in terms of the eigenvectors and eigenvalues of a transition matrix.

As in this paper, Molavi et al. (2017) axiomatize a non-Bayesian learning model. They assume agents cannot remember the entire history (imperfect recall) and provide a natural set of axioms that lead to a log-linear update rule. The authors prove a long-run learning result that applies to a large class of rules, showing that updating under imperfect recall generically leads to long-run consensus. This paper focuses instead on the role of false information and differs in that *disagreement* is the generic outcome.

A few recent papers exhibit mechanisms that prevent agents from reaching consensus. Li and Tan (2017) study agents who know only the local network structure and repeatedly share posterior beliefs with their neighbors. Assuming a naïve update rule, under which agents attribute unexpected changes to new signals, the authors show that in some networks beliefs oscillate in perpetuity. Similarly, Dasaratha and He (2017) explore sequential social learning with agents who fail to account for information redundancy, finding that imperfectly segregated groups can herd on different actions. There are two important differences in the way the present paper generates disagreement. First, I provide a microfoundation for an alternative belief updating procedure. Instead of assuming that agents make mistakes of a particular kind, the agents here face a more fundamental limitation in how they deal with false information. Second, earlier work requires carefully constructed examples to produce disagreement, while disagreement is a generic outcome in this paper.

Acemoglu et al. (2016) study a statistical learning problem in which two agents receive the same sequence of signals about an underlying state. If the agents disagree slightly on the signal distribution (i.e. they have different priors on the joint distribution of the state and the signals), then asymptotic beliefs can diverge. Disagreement arises from a lack of identification. The present paper offers a complementary perspective as a lack of identification is not always a compelling explanation. Moreover, the fragility of asymptotic agreement is sensitive to how one measures the "closeness" of the agents' priors.

Among models of disagreement, Zhao (2017) is closest in spirit to this paper, axiomatizing an alternative update rule and briefly considering social learning. Zhao studies updating in response to information of the form "event A is more likely than event B." He derives a unique rule that minimizes the amount that agents must change their beliefs. Agents following this rule display recency bias and may be vulnerable to persuasion. The author presents an example of three agents communicating in a line network that leads to oscillating beliefs. However, the lack of consensus is fragile, depending on the timing of communcation.

1.2 Savage's Example

Before introducing a formal model, I borrow an example from Savage to highlight the difficulty of studying falsehoods within the standard paradigm. Suppose you are making an omelet and have broken five of six eggs into a bowl. You are uncertain whether the remaining egg is fresh or rotten, and you therefore face the following choice. You can break the sixth egg into the bowl with the others, which risks ruining the omelet if it is rotten, or you can break the sixth egg into a separate saucer for inspection, which forces you to wash the saucer. This example neatly illustrates the primitives of Savage's model: states (rotten or fresh), acts (break into bowl or saucer), and consequences (ruin the omelet or not, dirty saucer or not). Washing the saucer is the cost of observing the underlying state.

We implicitly assume that breaking the egg into the saucer allows you to know whether it is rotten. Consider a twist on the example. Your sense of smell is not good, so you cannot verify the egg's freshness yourself. Should you break it into the saucer, you must ask your partner to check it for you and report its condition. However, you and your partner recently had a fight, and you suspect there is lingering bitterness—a bitter partner may wish to ruin your omelet. There are now four states to consider—egg rotten or not, partner bitter or not—and your prior over the four states will affect how you interpret your partner's report. How then are you to learn the state of the egg?

You might solicit further reports, say from a neighbor, but this compounds the problem. Perhaps your neighbor in on your partner's side. If we take seriously the possibility that any report could be false, then a new report only grows the state space. On a fundamental level, we learn nothing about the egg, and it becomes impossible to confine our attention to a "small world" of possible states. The way we resolve this difficulty in practice is through the cooperative principle: we do not take seriously that any report may be false.

2 States, Propositions, and Beliefs

We start with a finite collection of **states** Ω an agent has in mind. A state is a description of the world, including details relevant to decisions our agent might make.[2] Write ω^* for the true state. Over the course of her life, our agent learns **propositions**. A proposition is a subset of states $P \subseteq \Omega$. If $\omega^* \in P$, then P is true; otherwise it is false. Two propositions P_1 and P_2 **contradict** one another if $P_1 \cap P_2 = \emptyset$. We explore how an agent might process the propositions that she learns, knowing some are false.

At any time, we represent the agent's **beliefs** B as a collection of propositions $\{P_i\}_{i \in I}$. The agent makes choices as if these propositions are true. Beliefs are **consistent**, meaning $\cap_{i \in I} P_i \neq \emptyset$, and we write $\mathcal{B}$ for the set of consistent beliefs. The agent encounters propositions sequentially. On learning a new proposition $\hat{P}$, the agent applies an **update rule** $U(B, \hat{P}) : \mathcal{B} \times 2^S \to \mathcal{B}$ to arrive at new beliefs. We assume $U(B, \hat{P}) \subseteq B \cup \{\hat{P}\}$.

Analysis centers on the agent's beliefs after encountering different information. We represent information as an ordered list of propositions $L = (P_1, P_2, ..., P_k)$. We write $U(B, L)$ for the agent's beliefs after applying U to each element of L in sequence, starting from the beliefs B. For convenience, I write $U(L)$ for $U(\emptyset, L)$, and I abuse notation by writing P for a list containing the single proposition P (e.g. write $U(P)$ instead of $U(\emptyset, (P))$). Given beliefs $B = \{P_i\}_{i \in I}$, define

$$P(B) = \bigcap_{i \in I} P_i,$$

the most informative proposition B implies—if the propositions in list L are mutually consistent, define $P(L)$ analogously. Beliefs B and B' are **equivalent**, written $B \cong B'$, if $P(B) = P(B')$. Equivalent beliefs permit the agent to entertain the same set of possible states. If B and B' are equivalent, an agent should make the same choices under either.

2.1 Remarks

We should think of propositions as statements about the world that we can express in natural language. For instance, "Ben has a dog" or "the temperature outside is 75 degrees Fahrenheit." Viewed this way, it makes sense to remember particular propositions as opposed to their logical implications. If I later learn that Ben has a cocker spaniel, I do not forget the proposition "Ben has a dog." If I talk to someone about Ben, I may mention that he has a dog without specifying what type. Similarly, if two propositions imply another, I may not infer this without additional prompting. If I know that a lawyer has passed the bar, and that Ben is a lawyer, it may not immediately occur to me that Ben has passed the bar.

The structure of update rules embeds at least two implicit assumptions. First, the agent does not recall the order in which she learns propositions—or at least the order does not affect updating. Second, the agent forgets any propositions that she rejects as false. Once the agent accepts a proposition, there is nothing special about it relative to others that she

[2] This need not be a complete description. We should interpret a state as including the maximum amount of detail that the agent is capable of processing.

believes, and once the agent rejects a proposition, she does not retain or communicate it to others. These assumptions exclude any middle ground where the agent can tentatively accept a proposition or entertain one she thinks is most likely false.

The question what propositions our agent believes is orthogonal to whether the agent forms a subjective prior and draws inferences according to Bayes' rule. When faced with a decision, the agent can update a prior over Ω to its conditional distribution given $P(B)$, choosing an action to maximize expected utility with respect to the posterior. Our framework says nothing about what weights to assign different states; it asks which states to consider at all. Conversely, Bayes' rule offers no guidance on how to reconcile contradictory propositions. While the Savage axioms tell us how to work within an accepted paradigm, the present paper looks at how an agent might move between paradigms.

2.2 Learning Axioms

To find a good update rule, we first need to define what "good" means. At a basic level, to capture a meaningful notion of belief, an update rule should be stable in the following sense: if I hear a proposition that I already believe, I should not change my beliefs. Throughout the paper, I assume update rules satisfy the stability axiom.

Definition 1 (Stability Axiom)**.** *The update rule U satisfies **stability** if $U(B, P) = B$ whenever $P \in B$.*

In a similar spirit, own-frame independence requires the agent to update in the same way from beliefs B and B' whenever $B \cong B'$. This makes belief equivalence significant for updating as well as for decisions. When I impose this axiom, I say so explicitly.

Definition 2 (Own-Frame Independence)**.** *The update rule U satisfies **own-frame independence** if for any beliefs B and B' with $B \cong B'$, we have*

$$U(B, P) \cong U(B', P).$$

I divide the remaining axioms into two intuitive categories:

- Willingness-to-learn: I am willing to learn new things and can converge on the truth
- Non-manipulability: changing how information is presented does not change my beliefs

Each category captures a desirable feature of belief updating: we want the ability to learn, and we want to avoid manipulation.

In the willingness-to-learn category, I include three separate axioms. Credulity is an expression of Grice's cooperative principle: if everyone communicates in good faith, we should always retain a maximal set of propositions. Openness captures the idea that we should be able to arrive at truth if we encounter the right information. Finally, a weaker version of credulity requires only that we have some non-empty belief after processing any list of propositions.

Definition 3 (Willingness-to-Learn Axioms). *The rule U satisfies **credulity** if for any B and P, there is no consistent $B' \subseteq B \cup \{P\}$ such that $U(B,P)$ is a strict subset of B'.*

*The rule U satisfies **openness** if for any set of beliefs B and any state ω, there exists a list of propositions such that*

$$\bigcap_{P \in U(B,L)} P = \{\omega\}.$$

*The rule U satisfies **weak credulity** if $U(L) \neq \emptyset$ for any non-empty list L.*

An agent with an open update rule is always willing to change her mind. This axiom leaves room to choose what particular sequence of propositions can change the agent's beliefs. Credulity is in some sense more restrictive. If a new proposition P is consistent with B, credulity dictates $U(B,P) = B \cup \{P\}$. We have more flexibility if P is inconsistent with B. One option is to reject P. If we accept P, we may have a choice which propositions we retain from B. For instance, suppose $\Omega = \{1,2,3\}$, and B contains $\{1,2\}$ and $\{2,3\}$. If we accept $P = \{1,3\}$, then credulity implies we retain exactly one of the other two propositions.

The non-manipulability category includes axioms of label neutrality, order independence, and frame independence. Label neutrality means that permuting or relabeling the states does not change how the update rule processes information. Likewise, order independence means that our final beliefs are insensitive to the order in which we learn propositions. Frame independence is a bit more subtle. Given a list of consistent propositions L, frame independence says that we can replace L with any logically equivalent list and arrive at equivalent beliefs. For instance, if $\Omega = \{1,2,3\}$, then learning $\{1,2\}$ followed by $\{2,3\}$ should be equivalent to learning the single proposition $\{2\}$. Note that frame independence implies own-frame independence, but the converse of this is not true.

If π is a permutation of the states, and $P = \{\omega_1, \omega_2, ..., \omega_k\}$, write P_π for the proposition $\{\pi(\omega_1), \pi(\omega_2), ..., \pi(\omega_k)\}$. Similarly, if $L = (P_1, P_2, ..., P_k)$, write L_π for the list $(P_{1\pi}, P_{2\pi}, ..., P_{k\pi})$. Define B_π analogously for a collection of propositions B. If π is a permutation on a collection of propositions, we write $\pi(L)$ for $(\pi(P_1), \pi(P_2), ..., \pi(P_k))$. Finally, given two lists L and L', let $L + L'$ denote concatenation. I occasionally abuse notation by writing $L + P$ $(P + L)$ for the list L with proposition P appended to the end (beginning) of the list L.

Definition 4 (Non-Manipulability Axioms). *The rule U satisfies **label neutrality** (LN) if for any permutation π of Ω, any beliefs B, and any proposition P, we have*

$$U(B,P)_\pi = U(B_\pi, P_\pi).$$

*The rule U satisfies **order independence** (OI) if for any permutation π of L, we have*

$$U(L) = U(\pi(L)).$$

*The rule U satisfies **frame independence** (FI) if for any consistent list L, we have*

$$U(L_1 + L + L_2) \cong U(L_1 + P(L) + L_2).$$

The non-manipulability axioms only have significance in the presence of false information. If every proposition the agent hears is true, then the update rule that accepts all new propositions trivially satisfies all three axioms. If propositions can be false, each axiom imposes meaningful restrictions on our update rule.

2.3 Examples

I describe a few examples of update rules, illustrating how we can satisfy some axioms but not others.

Skeptical Updating

Some people are intensely skeptical of anything they are told. The **skeptic rule** U rejects all propositions, maintaining empty beliefs. This rule trivially satisfies all of the non-manipulability axioms and none of the willingness-to-learn axioms.

Wishful Thinking

Wishful thinking occurs when people believe something because they want it to be true. We can capture this idea through an update rule characterized via an ordering $\prec$ on the set of states. Let $\omega(B)$ denote the highest ranked state contained in $P(B)$, and let $\omega(P)$ denote the highest ranked state in P. A **wishful thinking rule** updates as follows:

- For any proposition P, we have $U(\emptyset, P) = P$
- If $\omega(P) = \omega(B)$, then $U(B, P) = B \cup \{P\}$
- If $\omega(P) \prec \omega(B)$, then $U(B, P) = B$
- If $\omega(P) \succ \omega(B)$, then $U(B, P) = \{P\}$

Under a wishful thinking rule, the agent has an underlying preference relation over states. The agent keeps track of the most preferred state in her beliefs and rejects any proposition that contains only less preferred states. If the agent encounters a proposition containing a more preferred state, she accepts it and discards her prior beliefs—none of the older propositions allow her to entertain the new preferred state. A wishful thinking rule is both order independent and frame independent, though it is obviously not label neutral. Of the willingness-to-learn axioms, it satisfies only weak credulity.

Stubborn Updating

People often retain their existing beliefs when faced with contradictory evidence. The **stubborn update rule** U accepts propositions that are consistent with current beliefs and rejects propositions that conflict with those beliefs. Formally, we have $U(B, P) = B \cup \{P\}$ if $\{P\}$ is consistent with B, and $U(B, P) = B$ otherwise. This can lead to belief polarization.

For instance, if one agent initially believes $\{2,3\}$ and another initially believes $\{1,2\}$, then giving both the list $(\{1,3\},\{1,2\})$, identifying the true state 1, eliminates their overlap: the first agent becomes sure of state 3, while the second learns the truth.

The stubborn rule is not open, but it does satisfy credulity and label neutrality. The order in which information is received clearly matters. Suppose $\Omega = \{1,2,3\}$, and the agent encounters $\{1,2\}$, $\{2,3\}$, and $\{1,3\}$. Depending on the order in which information arrives, we can convince the agent of any of the three states. For intstance, the list $L = (\{1,2\},\{2,3\},\{1,3\})$ results in beliefs $\{\{1,2\},\{2,3\}\}$, which are equivalent to $\{2\}$. Even if the agent subsequently receives the more specific proposition $\{1\}$, which is consistent with one of the other two, this information is discarded in favor of current beliefs. The rule also fails to satisfy frame independence—consider $L = (\{1,2\},\{2,3\},\{1,3\})$ and $L' = (\{1,2\},\{3\})$—but it does satisfy own-frame independence: we only add propositions consistent with $P(B)$.

With a slight adjustment, we can satisfy openness. We say that U is **stubborn with a preference for certainty (SPC)** if $U(B,P)$ follows the stubborn rule whenever $|P| \geq 2$, but it always accepts P if $|P| = 1$. The agent following this rule rejects new information that conflicts with her beliefs, unless that information completely pins down the state. In this case, the agent accepts the new proposition. The SPC rule is clearly open, credulous, and label neutral. It also satisfies own-frame independence. To see this, we just need to show that $U(B,P) \cong U(B',P)$ whenever $B \cong B'$. This is obvious if P is consistent with B and B'. If P is inconsistent with B and B' and $|P| \geq 2$, then $U(B,P) = B \cong B' = U(B',P)$. If instead $|P| = 1$, then both updates yield beliefs equivalent to P.

3 (Im)possibility Theorems

This section proves two results on update rules. The first is an impossibility theorem, showing that willingness-to-learn axioms and non-manipulability axioms are generically incompatible. The second is a representation theorem that characterizes the update rules of section 2.3. Giving primacy to the credulity axiom—and hence to the cooperative principle—I make a normative case for the two stubborn rules.

Theorem 1 (Impossibility)**.** *Suppose Ω has at least 5 states.*

(a) There is no update rule satisfying both openness and order independence.

(b) There is no update rule satisfying both credulity and order independence.

(c) There is no update rule satisfying both openness and frame independence.

(d) There is no update rule satisfying both credulity and frame independence.

Proof. For part (a), suppose there exists a rule U satisfying openness and order independence. Suppose $a, b \in \Omega$ are distinct states, and let $\overline{L}$ denote a list containing every conceivable proposition—such a list exists because Ω is finite. By openness, there exists a list L_a such that $U(\overline{L} + L_a) \cong \{a\}$ and a list L_b such that $U(\overline{L} + L_b) \cong \{b\}$. By order independence, we can reorder $\overline{L} + L_a$ so that each proposition in L_a appears right after its counterpart in $\overline{L}$. By stability, this implies $U(\overline{L}) \cong \{a\}$. Similarly, we arrive at $U(\overline{L}) \cong \{b\}$, a contradiction.

For part (b), suppose there exists a rule U satisfying credulity and order independence. Suppose $a, b, c, d, e \in \Omega$ are distinct states, and define the propositions $A = \{a, b\}$, $B = \{b, c\}$, $C = \{c, d\}$, $D = \{d, e\}$ and $E = \{a, e\}$. Without loss of generality, suppose $U(A, C) = U(C, A) = \{A\}$. We then necessarily have $U(A, D) = U(D, A) = \{A\}$. If not then $U(A, D, C) = U(D, C) = \{C, D\}$ and $U(A, C, D) = U(A, D) = \{D\}$, violating order independence. Similarly, we have $U(B, D) = U(D, B) = \{B\}$; if not, then $U(D, B, A) = U(D, A) = \{A\}$ and $U(D, A, B) = U(A, B) = \{A, B\}$, again violating order independence. Continuing in this manner, we find that $U(B, E) = U(E, B) = \{B\}$ because otherwise we get $U(E, B, D) = U(E, D) = \{D, E\}$ and $U(D, B, E) = U(B, E) = \{E\}$. However, we must also have $U(E, C) = U(C, E) = \{E\}$; the alternative leads to $U(E, C, A) = U(C, A) = \{A\}$ and $U(A, C, E) = U(A, E) = \{A, E\}$. This then implies $U(B, E, C) = U(B, C) = \{B, C\}$ and $U(C, E, B) = U(E, B) = \{B\}$, contradicting order independence.

For part (c), suppose there exists a rule U satisfying openness and frame independence. First, we note that $U(\emptyset, P) = \{P\}$ for any proposition P. Let a denote any state in P, and let b denote a state not in P. By openness, there exists L_a such that $U(L_a) \cong a$. By frame independence, we have $U(L_a) \cong U(\{a\}) \cong U(P + \{a, b\})$, which is impossible if U rejects P starting from empty beliefs. Similarly, the rule U must accept any proposition P' that contradicts P. For $b \notin P$, there exists L_b such that $U(P + L_b) \cong \{b\} \cong U(P + \{b\})$, and we can perform the same construction with P' a proposition containing $\{b\}$. Now, suppose $a, b, c, d \in \Omega$ are distinct states, and consider the two lists $L = (\{b\}, \{c, d\})$ and $L' = (\{a, b\}, \{c\})$. We then have $U(L) = \{c, d\}$ and $U(L') = \{c\}$, which are not equivalent. However, taking $L^* = (\{a, b\}, \{b, c\}, \{c, d\})$, frame independence requires $U(L) \cong U(L^*) \cong U(L')$, a contradiction.

A similar construction establishes part (d). Taking L and L' as above, credulity implies $U(L)$ is either $\{b\}$ or $\{c, d\}$, and $U(L')$ is either $\{a, b\}$ or $\{c\}$. We can never have $U(L) \cong U(L')$, but frame independence requires it via the intermediate list L^*.

□

Intuitively, order independent update rules should resemble the wishful thinking rule: if U satisfies at least weak credulity, then $U(\{a\}, \{b\}) = U(\{b\}, \{a\})$ defines an ordering on states, which is incompatible with openness. Note that part (a) is the only part of Theorem 1 that uses the stability axiom. The proof of part (b), as the only part requiring 5 states, is more intricate. We essentially construct an ordering on propositions that contradict one another, writing $P_1 \succ P_2$ if $U(P_1, P_2) = U(P_2, P_1) = P_1$. The proof rests on the observation that if proposition Q is disjoint from both P_1 and P_2, then order independence implies we have either $Q \succ P_1$ and $Q \succ P_2$ or $Q \prec P_1$ and $Q \prec P_2$. Applying this rule within a cycle of overlapping propositions yields a contradiction.

Frame independence is also highly restrictive. The counterexample highlights that it matters how we group propositions together. A proposition consistent with two others can narrow the conclusions we draw from each. If those two propositions contradict one another, then we can only narrow the conclusion from one of them. Own-frame independence is easier to satisfy because it commits to one way of grouping propositions—it groups propositions in the agent's current beliefs. Collectively, these results show that we cannot avoid manipulation if our update rule allows learning: beliefs depend on the order and framing of information.

The examples in section 2.3 show that we can satisfy some combinations of axioms. A representation theorem shows that these update rules are special in this regard.

Theorem 2 (Possibility)**.** *We have the following axiomatic characterizations of update rules.*

(a) If U satisfies label neutrality, order independence, and frame independence, then U is the skeptic rule.

(b) If U satisfies order independence, frame independence, and weak credulity, then U is a wishful thinking rule for some preference ordering on states.

(c) If U satisfies credulity, own-frame independence, and label neutrality, then U is either the stubborn rule or the SPC rule. If U also satisfies openness, then U is the SPC rule.

Proof. For part (a), suppose there exists a proposition P such that $U(\emptyset, P) = \{P\}$. By label neutrality, we have $U(\emptyset, P') = \{P'\}$ for any P' such that $|P'| = |P|$. Let L denote a list containing all such P'. Label neutrality implies that for any permutation π of S, we have $U(L)_\pi = U(L_\pi)$. Order independence implies that $U(L) \cong U(L_\pi)$, so $U(L) \cong U(L)_\pi$, which can only be true for any π if $U(L) = \emptyset$. However, this means that $U(L + P) = P$ and $U(L + P') = P'$. Since P and P' both appear in L, we can reorder the lists $L + P$ and $L + P'$ so that duplicate propositions appear consequtively. Stability then implies that $U(L + P) \cong U(L + P')$, a contradiciton. Hence, the skeptic rule is the only update rule satisfying all three non-manipulability axioms.

For part (b), consider a list L^* containing every singleton proposition. By weak credulity, we have $U(L^*) = \{\omega\}$ for some $\omega \in \Omega$. By order independence the update rule must yield beliefs $\{\omega\}$ from any list of singletons that contains $\{\omega\}$. First, I claim that whenever the agent holds beliefs inconsistent with $\{\omega\}$, the agent accepts any proposition that contains ω. Suppose there exists a list $L = (P_1, P_2, ..., P_k)$ in which no proposition contains ω and a proposition $P \notin L^*$ such that $P \notin U(L^* + P)$, and $\omega \in P$. We can construct a proposition $P' = \{\omega, \omega'\}$ such that $P \cap P' = \{\omega\}$. We can similarly construct non-singleton propositions P_i' for each $i \in 1, 2, ..., k$ such that $P_i \cap P_i' = \{\omega_i\} \neq \{\omega\}$. As long as P_i does not contain every state except ω, we can do this so that P_i' does not contain ω as well. If P_i does contain all other states, take $P_i' = P'$. Let $L' = (P_1', P_2', ..., P_k')$. By successive applications of order and frame independence, we have

$$\begin{aligned} U(L + P + P' + L') &\cong U(L + \{\omega\} + L') \\ &\cong U(\{\omega\} + L + L') \\ &\cong U(\{\omega\} + \{\omega_1\} + \{\omega_2\} + ... + \{\omega_k\}) \\ &= \{\omega\}. \end{aligned}$$

If L' has no propositions containing ω, then in order to have $U(L + P + P' + L') \cong \{\omega\}$, we must retain the proposition P. If L' contains copies of P', we can remove them using order independence and stability, so again, we must retain the proposition P. Note that by order independence, this also means we can never discard a proposition containing ω after learning one that does not.

Next, I show that after accepting a proposition that contains ω, the agent cannot accept a proposition that does not contain ω—the beliefs $U(L + \hat{P})$ cannot contain $\hat{P}$ with $\omega \notin \hat{P}$ if ω is contained in one of the propositions of L. By frame independence, it is without loss to take L containing a single proposition P. If $P = \{\omega\}$, the result is clear from our work above—we cannot discard $\{\omega\}$, so we cannot accept a proposition that does not contain ω. If $P \neq \{\omega\}$, let $\hat{P}$ be a proposition such that $\omega \notin \hat{P}$. If $P \cap \hat{P} = \emptyset$, then again we are done because of the previous paragraph. If the intersection is non-empty, let $P' = \{\omega, \omega'\}$ for some $\omega' \notin \hat{P}$, and define $\underline{P} = P \cap \hat{P}$. If the update rule accepts $\hat{P}$, we have

$$U(P + \hat{P} + P') \cong U(\underline{P} + P'),$$

which cannot be equivalent to $\{\omega\}$ since $\omega \notin \underline{P}$. However, our earlier work implies

$$U(P + \hat{P} + P') \cong U(P + P' + \hat{P}) \cong U(\{\omega\} + \hat{P}) = \{\omega\},$$

a contradiction. We conclude that the update rule rejects $\hat{P}$. Also, note this implies (by order independence) that after accepting a proposition containing ω, we must discard all propositions that do not contain ω. Finally, note that frame independence implies the update rule must accept any proposition that both contains ω and shrinks the intersection of propositions in the agent's beliefs.

To complete the proof for part (b), we iterate the above argument. We now start from a list containing all singletons *except* $\{\omega\}$, and we consider only propositions not containing ω. This establishes the properties of the wishful thinking rule for the second highest ranked state, and we can repeat the argument for the rest of the ordering.

For part (c), note that credulity implies we accept any proposition consistent with existing beliefs. We now show that U must reject any inconsistent proposition containing at least two states. Suppose there exists $\hat{P}$ with at least two states and beliefs B inconsistent with $\hat{P}$ such that $\hat{P} \in U(B, \hat{P})$. Consider two cases. First, suppose $P(U(B, \hat{P})) = \hat{P}$, and define $P^* = P(B) \cup \{\omega\}$ for some $\omega \in \hat{P}$. Using credulity, we have either $U(B, P^* + \hat{P}) \cong P(B)$ or $U(B, P^* + \hat{P}) \cong \{\omega\}$. Neither option is equivalent to $\hat{P}$ since $\hat{P}$ contains more than just ω. Alternatively, if $P(U(B, \hat{P}))$ is a strict subset of $\hat{P}$, then $U\left(P(B), \hat{P}\right)$ is equivalent to either $P(B)$ or $\hat{P}$, and this contradicts own-frame independence. Therefore, we must reject $\hat{P}$. Label neutrality implies we must treat all inconsistent singleton propositions the same: either reject all of them (the stubborn rule) or accept all of the (the SPC rule). $\square$

Theorem 2 illuminates fundamental tradeoffs between the axioms. One cannot learn without permitting some form of manipulation—the only way to avoid all manipulation is to remain skeptical of any proposition. Wishful thinking appears as a natural compromise that leans towards non-manipulability. The last part establishes a close connection between the cooperative principle, as embodied in the credulity axiom, and stubborn belief updating.

The next section explores the implications of update rules in a model of social learning. Here, I give primacy to the cooperative principle, assuming that people generally trust one another, and update rules satisfy credulity. Own-frame independence and label neutrality have a particularly strong normative appeal, demanding that our update rule is insensitive to how we describe our own beliefs and how we label the underlying states. I therefore focus on the stubborn rule and the SPC rule. While openness holds appeal as well, pointing us to the SPC rule, I do not insist on this axiom going forward. The core premise of this paper is that much of what people believe is taken on faith, but people who have learned through their own careful research are unlikely to defer to hearsay. If social learning primarily entails the exchange of unverified propositions, it seems inappropriate to rule out stubborn updating.

4 Social Learning and Disagreement

The literature on social learning addresses three main questions. Do agents reach a consensus in the long-run? Who most influences the consensus belief? Do agents effectively aggregate dispersed information? Broadly the answers are, almost always, the most central agents in the network (for an appropriate centrality measure), and yes as long as no one is too influential. I ask how the model in this paper changes our answers to these core questions.

A population of n agents interacts in a network, which we represent as a directed graph G. I assume the network is strongly connected and write G_i for the set of neighbors of i. The state space is $\Omega = \{1, 2, ..., z\}$, and we assume $n \geq z$. Without loss of generality, suppose $\{1\}$ is true. At time zero, each agent i is endowed with an initial set of beliefs $B_i(0)$ drawn independently according to a distribution $\mu \in \Delta(\mathcal{B})$. We write $\mu(B)$ for the probability of initial belief B, and $\mu(\mathcal{B}')$ for the probability of realizing some belief $B \in \mathcal{B}' \subseteq \mathcal{B}$. For convenience, define $\mathcal{B}_1 = \{B \in \mathcal{B} \,|\, 1 \in P(B)\}$, the set of beliefs consistent with the true state. Define also $\mathcal{S}$ as the set of beliefs B that contain a singleton proposition, and $\mathcal{S}^*$ as the set of beliefs B that contain only a singleton proposition. I refer to the belief endowment as the agent's *signal*, and I assume signals are both *informative* and *conditionally symmetric*.

Definition 5. *A signal distribution $\mu \in \Delta(\mathcal{B})$ is* ***informative*** *if $\mu(\emptyset) < 1$, and for any permutation π of the states, we have*

$$\mu(B) > \mu(B_\pi)$$

whenever $\mu(B_\pi) > 0$ and $1 \in P(B)$ but $1 \notin P(B_\pi)$.

The signal distribution μ is ***conditionally symmetric*** *if $\mu(B) = \mu(B_\pi)$ whenever both $1 \in B$ and $1 \in B_\pi$ or both $1 \notin B$ and $1 \notin B_\pi$.*

An informative signal means that $P(B)$ tends to include the true state—a statistician who observes the belief endowments could identify the true state. This is important for how we interpret results on information aggregation. Conditional symmetry is not a substantive assumption, but it simplifies statements.

Each agent i has an update rule U_i. Unless otherwise stated, I assume U_i satisfies credulity, own-frame independence, and label neutrality. In each discrete period, exactly one edge ij is chosen from G uniformly at random. Agent i communicates a single proposition $\hat{P}$ in her beliefs, drawn uniformly at random, to agent j. Writing B_j for agent j's current belief, agent j then updates to the new belief $U_j(B_j, \hat{P})$. The population achieves **consensus** if at some time, there exists a proposition P^* such that

$$P(B_i) = P^*$$

for all agents i.

4.1 Lack of Consensus

Our first theorem shows that lack of consensus is a generic outcome in this model.

Theorem 3. *The network achieves consensus with probability one if and only if at least one of the following statements is true:*

(a) We have $\mu(\mathcal{B}_1) = 1$

(b) At least $n - 1$ *agents follow the SPC rule, and* $\mu(\mathcal{S}) = 1$

Proof. If $\mu(\mathcal{B}_1) = 1$, then the true state is contained in all agents' belief endowments, and in particular, the true state is contained in any proposition that is ever shared. Agents trivially accept every proposition they hear, and since the network is connected they reach a consensus on the intersection of the initial beliefs.

Alternatively, suppose $n - 1$ agents follow the SPC rule and $\mu(\mathcal{S}) = 1$. Beliefs evolve according to an irreducible Markov chain for which a state is absorbing if and only if it entails consensus—hence we reach consensus with probability one. If the remaining agent is stubborn, then the only possible consensus is consensus on the stubborn agent's belief. If all agents follow the SPC rule, any consensus is possible.

If neither (a) nor (b) hold, then the following event occurs with positive probability: no agent's initial belief contains a singleton proposition, and for each $\omega \in \Omega$ there exists an agent i such that $\omega \notin P(B_i)$.[3] Since no agent's initial belief contains a singleton proposition, all agents follow the stubborn update rule. The intersection of the initial beliefs is empty, so consensus is impossible. □

Theorem 3 presents a stark constrast with standard results. We need strong conditions to ensure consensus. Condition (a) requires that belief endowments only include true propositions—this is the case in typical learning models. Condition (b) requires both that almost all agents follow the SPC rule and that all belief endowments include a singleton. Once we exclude these possibilities—allowing false initial beliefs and multiple stubborn agents—disagreement is possible. Under slightly stronger assumptions, disagreement becomes almost certain in large networks—it gets harder to obtain consensus as the network grows.

[3]Note this claim relies on informativeness and conditional symmetry.

Let $\{G_n\}_{n\in\mathbb{N}}$ denote a sequence of networks, and let s_n denote the number of stubborn agents in G_n. For any signal distribution μ and network G, let $C_\mu(G)$ denote the probability of reaching consensus. The next result shows that generically in a large network, long-run disagreement is not just possible but inevitable.

Proposition 1. *Suppose $\mu(\mathcal{B}_1) < 1$. If either $\lim_{n\to\infty} s_n = \infty$ or $\mu(\mathcal{S}) = 0$, then*

$$\lim_{n\to\infty} C_\mu(G_n) = 0.$$

Proof. Note that if $\mu(\mathcal{S}) = 0$, we can treat all agents as stubborn agents. For each $\omega \in \Omega$, there is some positive probability that a stubborn agent's belief endowment excludes the state ω. If the number of stubborn agents grows without bound, then with probability approaching one, for each state $\omega \in \Omega$, there exists a stubborn agent whose belief endowment excludes ω. In this event, consensus is impossible. □

4.2 Influence and Steady-State Beliefs

Lack of consensus complicates the concept of influence. In standard settings, everyone converges on the same belief, and we can ask how that consensus depends on each individual's initial belief. Here, we need to address who influences whose beliefs. Further complicating matters, the long-run outcome is typically path-dependent: with stubborn updating, early chance events determine key features of the long-run steady state.

We say that beliefs *harden* in the population if after some period t, no agent changes her beliefs. Given a network G and a signal distribution μ, we write $H_\mu(G)$ for the probability that beliefs harden—the first result of this subsection characterizes when this occurs. As in the statement of Proposition 1 in the last subsection, we let $\{G_n\}_{n\in\mathbb{N}}$ denote a sequence of networks and s_n the number of stubborn agents in G_n.

Proposition 2. *Beliefs harden with probability one if and only if one of the following conditions holds:*

(a) We have $\mu(\mathcal{S}) = 0$

(b) The only singleton proposition in the support of μ is $\{1\}$

(c) At least $n-1$ agents follow the SPC rule

Suppose μ puts positive probability on beliefs containing two distinct singleton propositions. If $\lim_{n\to\infty} s_n = \infty$, then

$$\lim_{n\to\infty} H_\mu(G_n) = 0.$$

Proof. I first note that conditions (a) - (c) each imply that beliefs harden. If $\mu(\mathcal{S}) = 0$, this is obvious because SPC agents must stop updating if there are no singleton propositions. If there is only one singleton proposition in the belief endowments, informativeness implies it must be $\{1\}$. Any SPC agent who gets exposed to this proposition ceases to update, and it is clear that agents must cease updating eventually. If all agents follow the SPC rule and at least one singleton proposition is in the belief endowments, then all absorbing states feature consensus. Suppose there are $n-1$ agents following the SPC rule. If the stubborn agent ever learns a singleton proposition, consensus on this belief becomes a unique absorbing state. Otherwise, the SPC agents reach consensus on one of the singleton propositions as they would without a stubborn agent present.

Conversely, if there exist two or or more stubborn agents and two or more singleton propositions in the signal support, then with positive probability, two stubborn agents have belief endowments featuring different singleton propositions. SPC agents with paths to both must oscillate between these two beliefs in perpetuity. The last part follows from noting that as s_n grows, the probability that there exist two stubborn agents who believe distinct singleton propositions approaches one. □

Belief hardening requires either the absence of multiple stubborn agents or the absence of multiple singleton propositions in the belief endowments. Under mild assumptions, the probability $H_\mu(G_n)$ converges to zero for large n. We generically reach a steady state in which some stubborn agents disagree with one another, and those following the SPC rule continually change their minds.

In the generic case, stubborn agents who refuse to update exert influence on others. Although beliefs continually change, and there is never consensus, we reach a steady-state distribution that depends on both the update rules and the network structure. The belief evolution process is a complex object, and we cannot in general specify what this steady-state distribution is. However, under simplifying assumptions we can give a precise characterization, one that offers intuition for the general case.

Given a network G, write A for the set of stubborn agents, and recall $\mathcal{S}^*$ is the set of belief endowments that contain only singleton propositions. Going forward, we assume $\mu(\mathcal{S}^*) = 1$. Let $d_i = |G_i|$ denote the degree of agent i, and define $m = \sum_{i=1}^n d_i$ as the number of directed edges in G. The *direct influence network* $\tilde{G}$ is a weighted graph that we define as follows:

- For each $i \in A$, we have $\tilde{g}_{ii} = 1$ and $\tilde{g}_{ij} = 0$ for all $j \neq i$
- For each $i \notin A$, we have $\tilde{g}_{ii} = \frac{m-d_i}{m}$ and $\tilde{g}_{ij} = \frac{1}{m}$ if $j \in G_i$, with $\tilde{g}_{ij} = 0$ otherwise.

When beliefs contain only singletons, the stubborn agents never update, and agents following the SPC rule simply copy the beliefs of the last person with whom they speak.

The entry $\tilde{g}_{ij}$ represents the probability in any period that agent i replaces her belief with that of agent j. The powers of $\tilde{G}$ therefore describe the probabilities with which different agents adopt others' initial beliefs after some time: the ijth entry of $\tilde{G}^t$ is the probability that agent i adopts agent j's initial belief at time t. The matrix $\tilde{G}$ is a stochastic matrix—we recognize it as the transition matrix for an n state Markov chain in which each $i \in A$ corresponds to an absorbing state—so there exists a limit $M \equiv \lim_{t\to\infty} \tilde{G}^t$ that describes the steady-state distribution of beliefs. The following theorem restates a standard result about the Markov chain, which we can reinterpret in light of the social learning model.

Theorem 4. *Suppose $\mu(\mathcal{S}^*) = 1$. There exists a matrix $M = \lim_{t\to\infty} \tilde{G}^t$ satisfying*

$$M = \sum_{i \in A} \boldsymbol{\nu}^{(i)} \mathbf{e}_i',$$

where $\mathbf{e}_i$ denotes a vector of zeros with a one in the ith position, and the $\boldsymbol{\nu}^{(i)}$ are linearly independent right eigenvectors of $\tilde{G}$ with eigenvalue 1 such that $\boldsymbol{\nu}_i^{(i)} = 1$ and $\boldsymbol{\nu}_j^{(i)} = 0$ for each $j \in A$ with $j \neq i$.

Proof. See a standard text on random processes, such as Gallager (2013). □

At steady state, the value m_{ji} is the probability that agent j adopts the initial belief of agent i—note this can be positive only if $i \in A$. The ith column of M is the eigenvector $\boldsymbol{\nu}^{(i)}$ for $i \in A$, and this vector represents the influence of agent i's initial belief on the population steady state. Unlike in a DeGroot model, we cannot reduce an agent's influence to a single number since there is no consensus. The entries of $\boldsymbol{\nu}^{(i)}$ capture the distinct influence of i on each other agent j. The L_1 norm of $\boldsymbol{\nu}^{(i)}$ measures i's overall influence. Similarly, the jth row of M tells us the sources that influence agent j and how much. Agents following the SPC rule exert no influence on steady-state beliefs. Theorem 4 makes it straightforward to compute steady-state beliefs once we specify the initial beliefs of the stubborn agents.

Another important difference relative to the DeGroot model is that an agent's influence depends both on the network G and on who else is stubborn. As one example, consider the diamond network of Figure 1. If agents 1 and 4 are stubborn, then symmetry implies they each exert the same influence $\frac{1}{2}$ on the other two agents. If agent 2, rather than agent 4, is stubborn, agent 1 suffers a loss of influence. In this case, agent 1 exerts influence $\frac{2}{5}$ on agent 3 and influence of only $\frac{1}{5}$ on agent 4. This drop occurs because agent 2 has more direct contact with agent 4, and agent 4's contact with agent 3 reinforces the influence of 2 on 3.

We can gain additional intuition about influential agents based on techniques used to study voter models (e.g. Yildiz et al., 2013). The learning process is closely related to an unbiased random walk on the graph G. The influence $\nu_j^{(i)}$ of the stubborn agent i on agent j is equal to the probability that a simple random walk initiated at j hits agent i before hitting any other stubborn agent. With this connection to random walks, it becomes clear that influential agents are those with many short paths to other agents—provided these paths are not blocked by other stubborn agents.

In the general case where $\mu(\mathcal{S}^*) < 1$, difficulty arises because SPC agents need not accept every proposition they hear. The steady state depends not just on the network and who the stubborn agents are, but also on what propositions stubborn agents come to believe. Intuitively, having many short paths to others still correlates with influence, but believing non-singleton propositions reduces this influence. How we frame information matters, and this has implications for belief manipulation strategies that I highlight in the next section.

4.3 Lack of Aggregation

As a corollary to Theorem 4, we see that generically there is no information aggregation. Without some way for agents to differentiate between true and false, there is no reliable way

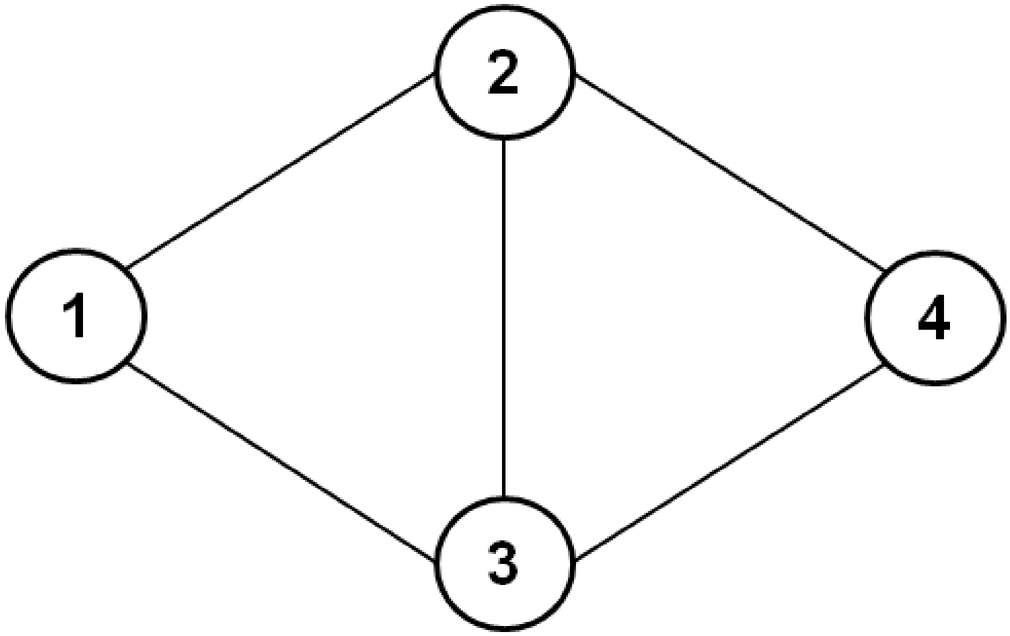

Figure 1: The diamond network.

to spread the truth, even if most agents start with correct beliefs.

Corollary 1. *Suppose* $\mu(\mathcal{S}^*) = 1$*. We have*

$$\mathbb{P}\left(B_i(t) = \{1\}\right) = \mu(\{1\})$$

for all i *and all* t*.*

Proof. In each period, the probability that an agent believes the truth is a linear weighted average of the initial probabilities. The result follows since the initial beliefs are i.i.d. draws from μ. □

4.4 Discussion

This section began with three questions about social learning. Do agents reach consensus? Who is influential? Can we aggregate dispersed information? Compared with earlier work, our model yields fundamentally different answers. Disagreement, not consensus, is generic. Influence is multifaceted. Aggregation is unachievable. These differences arise because our agents confront falsehoods that are indistinguishable from the truth.

Beyond these basic questions, our results highlight the role of language in social learning. Our update rules privilege propositions that eliminate all uncertainty—we can view this as an expression of what psychologists call the need for cognitive closure, that people dislike uncertainty and ambiguity. One consequence is that less precise communication is less persuasive: the possibility of communicating non-singleton propositions can reduce an agent's influence. Even if the way information is framed has no affect on any individual's choices, it nonetheless affects how that information spreads through a society.

I have prioritized the credulity axiom in the analysis of this section, but one might ask how different learning rules affect outcomes. The skeptic rule is an easy case: a skeptic agent may as well not exist. Since skeptics never accept any proposition, they never communicate any proposition, and they exert no influence. As I discuss in the next section, inducing skepticism can be an effective way to manipulate steady-state beliefs. Wishful thinkers on the other hand behave similarly to stubborn agents. After arriving at the preferred belief, a wishful thinker ceases to update. While Corollary 1 highlights that learning with stubborn agents never affects the probability of holding correct beliefs, wishful thinking changes this result. If we add wishful thinkers to the network, their implicit preference orderings over states, and not their signals, will influence the steady state. This suggests another normative argument in favor of the stubborn rule and the SPC rule: these rules at least maintain the overall quality of information in a society, while wishful thinking can worsen it.

5 Echo Chambers and Belief Manipulation

I highlight two applications of the learning model in this section. First, I study the phenomenon of echo chambers, illustrating how repetition of certain proposition can drive beliefs within insular groups. Second, I study strategies to manipulate beliefs. The framework helps us assess a wide range of interventions, including changes to the network, sowing doubt, fact-checking, and obfuscation.

5.1 Echo Chambers

A large empirical literature documents communities that share and reinforce particular beliefs.[4] There are several questions to ask from a theoretical perspective. Why do echo chambers form? Are they driven more by network structure or the idiosyncratic views of individuals? What types of communities are prone to form echo chambers? How can we avoid developing echo chambers? An example sheds light on these questions.

Consider a network containing two cliques—a clique is a complete subgraph—one of size n and the other of size m. Suppose the state space is $\Omega = \{1, 2\}$, and each of the two cliques contains one stubborn agent who never updates. Assume the stubborn agent believes $\{1\}$ in the first clique and $\{2\}$ in the second, and all other agents follow the SPC rule. There is exactly one link between the two cliques. Figure 2 depicts the case with $n = 5$ and $m = 4$.

Taking advantage of symmetries, we can compute the steady-state beliefs in this network—we can represent a belief at steady state as the probability that an agent believes the proposition $\{1\}$. Let x_1 and x_2 denote the steady-state beliefs of SPC agents inside the two cliques,

[4] See Barberá et al. (2015), Jasny et al. (2015), and Del Vicario et al. (2016) for recent examples.

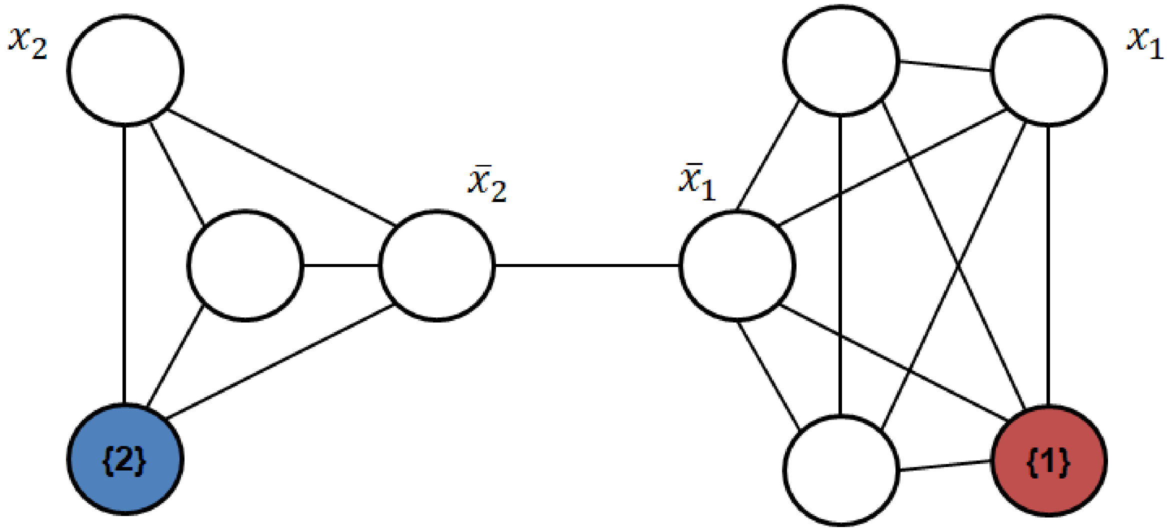

Figure 2: Two cliques with a connecting link.

and let $\overline{x}_1$ and $\overline{x}_2$ denote the steady-state beliefs for the two agents who are linked across the cliques. We have

$$x_1 = \frac{1}{n-1} + \frac{n-3}{n-1}x_1 + \frac{1}{n-1}\overline{x}_1, \qquad \overline{x}_1 = \frac{1}{n} + \frac{n-2}{n}x_1 + \frac{1}{n}\overline{x}_2.$$

$$x_2 = \frac{m-3}{m-1}x_2 + \frac{1}{m-1}\overline{x}_2, \qquad \overline{x}_2 = \frac{m-2}{m}x_2 + \frac{1}{m}\overline{x}_1,$$

Solving these equations yields

$$x_1 = \frac{nm+m+2n}{nm+2m+2n}, \qquad x_2 = \frac{n}{nm+2m+2n},$$

with $\overline{x}_1 = 2x_1 - 1$ and $\overline{x}_2 = 2x_2$.

Increasing both n and m together leads to increased polarization between the groups—the belief x_1 converges to one, and x_2 converges to zero. The reason is exactly the echo chamber effect. Even though each clique only contains one stubborn agent, the frequent repetition of this belief by other members of the clique reinforces the influence of this agent. The narrow bottleneck connecting the two cliques cannot counterbalance the effect. Notice that increasing n while leaving m fixed does increase the steady-state belief of the second group x_2, but there is a limit: we always have $x_2 < \frac{1}{m+2}$. Small close-knit groups can largely sustain their views, even if the other group is far larger.

This example broadly suggests that echo chambers emerge because of highly clustered networks. If individuals cannot or do not directly verify the veracity of statements themselves, instead relying on what they hear from friends and neighbors, then the mere repetition of statements can influence beliefs. Network structure is more important for this effect than having many stubborn agents in a group—in our example, adding stubborn agents to the cliques, particularly when the cliques are large, has minimal impact on steady-state beliefs.

5.2 Belief Manipulation

There are many potential ways to manipulate beliefs. Most obviously, we can broadcast information to the agents. To represent this in the model, we might add stubborn agents to the network or convert SPC agents to stubborn ones with particular beliefs. Alternatively, sowing doubt is a time-honored manipulation strategy.[5] To capture this, we might induce agents to adopt the skeptic update rule, which effectively removes them from the network. In some cases, altering the link structure is feasible, changing who effectively communicates with whom. For instance, Facebook can change whose posts appear on an individual's news feed. We might also urge people to check information with reliable sources, or inundate them with redundant information. Our model allows us to assess the effects of each of these interventions—under the simplifying assumptions of Theorem 4, we can exactly compute the effects; otherwise, we can simulate the model.

Adding Stubborn Agents

Changing an open agent into a stubborn one clearly affects steady-state beliefs, and Theorem 4 offers a straightforward way to assess the impact. As in the previous subsection, suppose $\Omega = \{1, 2\}$. Consider the network in figure 3 with two connected 3-cliques, each containing a single stubborn agent. Steady-state beliefs are

$$x_1 = \frac{6}{7}, \quad \overline{x}_1 = \frac{5}{7}, \quad \overline{x}_2 = \frac{2}{7}, \quad x_2 = \frac{1}{7}.$$

Suppose we want to convince more agents of the true state 1.

What happens if we convert the upper right agent into a stubborn agent with beliefs $\{1\}$? The steady-state equations become

$$\overline{x}_1 = \frac{2 + \overline{x}_2}{3}, \quad \overline{x}_2 = \frac{x_2 + \overline{x}_1}{3}, \quad x_2 = \frac{\overline{x}_2}{2}.$$

Solving yields

$$\overline{x}_1 = \frac{10}{13}, \quad \overline{x}_2 = \frac{4}{13}, \quad x_2 = \frac{2}{13}.$$

The average belief increases modestly from $\frac{1}{2}$ to $\frac{7}{13}$. Suppose instead that we convert the boundary agent with belief $\overline{x}_2$ in the figure. We then get $x_2 = \frac{1}{2}$, and $x_1 = \overline{x}_1 = 1$, raising the average belief from $\frac{1}{2}$ to $\frac{2}{3}$, a far larger effect.

Targeting clearly matters. In the first intervention, we reinforce the existing bias in $\overline{x}_1$, which then weakly influences the other agents. In the second intervention, we more directly

[5] For instance, it was a key part of how tobacco companies responded to negative health studies and how oil companies responded to climate science.

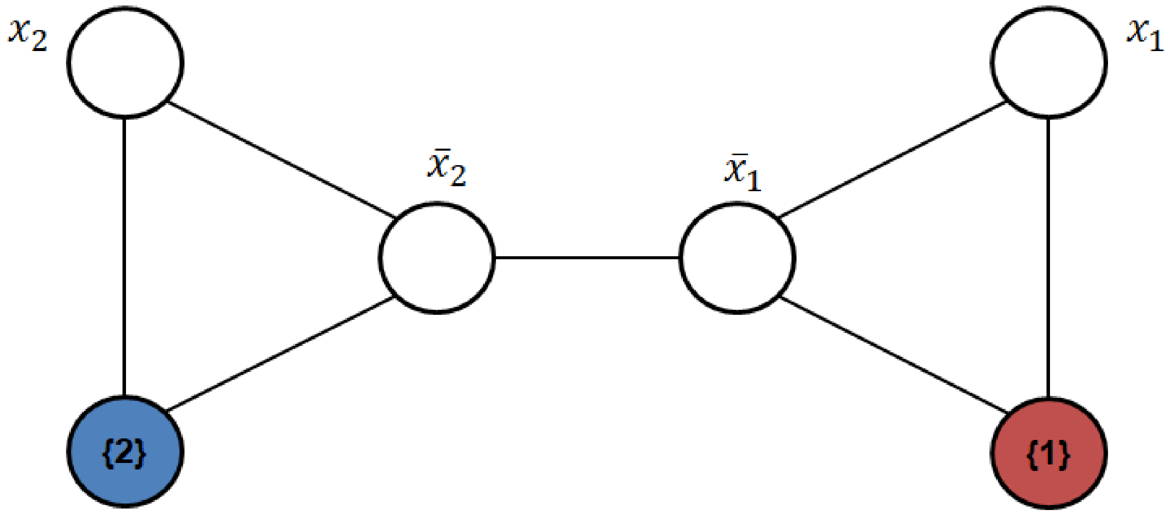

Figure 3: Connected 3-Cliques

influence x_2 and we cut off the right clique from the other stubborn agent's influence. In practice, converting some agents may be easier than others, and this should play a role in any targeting strategy. As agents do not distinguish the true state in their update rules, the analysis applies equally well if we interpret the new stubborn agent as an expert or as an adversary.

Sowing Doubt

A skeptic neither accepts propositions nor communicates them, so sowing doubt is formally equivalent to removing agents from the network. Returning to the example in Figure 3, suppose we eliminate the upper left agent with belief x_2. The steady-state equations become

$$x_1 = \frac{1 + \overline{x}_1}{2}, \quad \overline{x}_1 = \frac{1 + x_1 + \overline{x}_2}{3}, \quad \overline{x}_2 = \frac{\overline{x}_1}{2}.$$

Solving gives

$$x_1 = \frac{7}{8}, \quad \overline{x}_1 = \frac{3}{4}, \quad \overline{x}_2 = \frac{3}{8}.$$

If we interpret the skeptical agent as having average belief $\frac{1}{2}$, then we have raised the average belief in the population from $\frac{1}{2}$ to $\frac{7}{12}$. Notice the change is larger than when we made the upper right agent stubborn. This highlights that sowing doubt among the opposition can prove more effective than reinforcing existing beliefs within one's own group. By reducing repetition of the proposition $\{2\}$, we increase the relative influence of the group on the right.

Changing the Network

Introducing people to one another, or blocking contact, also affects belief evolution. Suppose we add a link in Figure 3 between the upper right agent and the agent with belief $\overline{x}_2$. The steady-state equations become

$$\overline{x}_1 = x_1 = \frac{1 + x_1 + \overline{x}_2}{3}, \quad x_2 = \frac{\overline{x}_2}{2}, \quad \overline{x}_2 = \frac{x_2 + \overline{x}_1 + x_1}{4}.$$

Solving gives

$$\overline{x}_1 = x_1 = \frac{7}{10}, \quad x_2 = \frac{1}{5}, \quad \overline{x}_2 = \frac{2}{5}.$$

Relative to the original network, we have increased beliefs in the left clique and slightly decreased them in the right—integrating the two groups brings beliefs closer together.

Disrupting links in the network has an effect similar to sowing doubt. Suppose instead that we delete the link between the upper left agent and the stubborn agent with belief $\{2\}$. The resulting steady-state beliefs are

$$\overline{x}_1 = \frac{3}{4}, \quad x_1 = \frac{7}{8}, \quad x_2 = \overline{x}_2 = \frac{3}{8}.$$

Removing the direct link to the stubborn agent reduces the echo chamber effect. The upper left agent no longer reinforces the belief of the group on the left.

Fact-Checking

Fact-checking was a prominent part of efforts to counter misinformation during the 2016 United States national elections. News organizations published numerous articles debunking common misconceptions, and independent websites emerged to keep track of false stories and encourage people to check what they hear with a reliable source. Our framework can help evaluate the efficacy of these interventions.

Suppose that when a SPC agent hears a singleton proposition from a friend, with probability $\gamma \in (0, 1)$ she checks a reliable source before updating. Assume the reliable source is always correct, and conditional on checking, the agent updates to the correct belief. True and false beliefs can still coexist in steady state, but this biases updating towards the truth.

To study the evolution of beliefs under the assumptions of Theorem 4, we need to adjust the transition matrix $\tilde{G}$. As before, a stubborn agent $i \in A$ never revises her beliefs. If $i \notin A$, then with probability $\frac{m-d_i}{m}$ she does not hear anything and does not update. For each neighbor $j \in G_i$, with probability $\frac{1-\gamma}{m}$, agent i hears from j, fails to fact-check, and adopts j's belief. Finally, with probability $\frac{\gamma d_i}{m}$, agent i hears from a neighbor, decides to fact-check, and adopts the correct belief.

Hence, we can represent transitions using an $(n+1) \times (n+1)$ row-stochastic matrix $\tilde{G}_\gamma$. We add a dummy agent $n+1$ who holds correct beliefs and never updates. The upper left $n \times n$ submatrix looks like $\tilde{G}$ from before. For agents $i \notin A$ we take $\tilde{g}_{ij} = \frac{1-\gamma}{m}$ for $j \in G_i$, and $\tilde{g}_{i,n+1} = \frac{\gamma d_i}{m}$. Theorem 4 applies, and we can interpret the eigenvector $\boldsymbol{\nu}^{(n+1)}$ as the expected impact of fact-checking.

Obfuscation

Let $\Omega = \{1, 2, 3\}$, and suppose there are three agents in a complete graph. Two agents are stubborn, starting with beliefs $\{1\}$ and $\{2\}$ respectively. Assume the third agent follows the SPC rule: she maintains her current beliefs unless the proposition she hears is a singleton, in which case she accepts it. At the steady state, the third agent is equally likely to believe $\{1\}$ or $\{2\}$.

Now suppose the stubborn agent who believes $\{1\}$ learns the additional proposition $\{1, 3\}$. Now, in half of her interactions with the third agent, she communicates $\{1, 3\}$ instead of $\{1\}$. If the third agent believes $\{2\}$ when this happens, she rejects the new proposition. At the steady state, the third agent now believes $\{2\}$ with probability $\frac{2}{3}$. Adding another proposition with multiple states crowds out the proposition that the third agent is willing to accept. Staying "on message" is valuable.

6 Final Remarks

Once we allow that some statements we hear are false, it becomes far less clear how an individual should update her beliefs. By axiomatizing update rules in this setting, we clarify the inherent tradeoffs in how we process information obtained from others. Being receptive to new information entails being vulnerable to belief manipulation. The only way to avoid this is to abandon the cooperative principle and refuse to believe what others tell us.

On bringing this framework into a model of social learning, persistent disagreement is no longer such a puzzle. If some individuals are stubborn—whether because they have convincing direct observations or because they have a separate agenda—then communication of beliefs between agents generically fails to bring consensus. Relative to standard approaches, we also get different answers to questions about influence and information aggregation. Both network position and the update rules that others use are important determinants of influence. We not only are unable to eliminate false beliefs, but we cannot achieve any meaningful aggregation of information. Moreover, the model naturally captures echo chambers and provides a way to assess the efficacy of many interventions.

An important missing feature is any distinction between information sources. We typically do not place the same amount of trust in everyone with whom we speak—our trust that others act in accordance with the cooperative principle may depend on attributes like status and group identity. One way to incorporate this in future work is to have agents apply different update rules depending on the identity of the neighbor with whom they communicate. Such a model would allow us to explore how varying levels of trust within and between communities affects social learning and examine the special influence of those who span multiple communities.

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz (2016), "Fragility of Asymptotic Agreement Under Bayesian Learning." *Theoretical Economics*, 11, 187–225.

Alchourrón, Carlos, Peter Gärdenfors, and David Makinson (1985), "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions." *The Journal of Symbolic Logic*, 50, 510–530.

Allcott, Hunt and Mathew Gentzkow (2017), "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, 31, 211–236.

Barberá, Pablo, John Jost, Jonathan Nagler, Joshua Tucker, and Richard Bonneau (2015), "Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?" *Psychological Science*, 26, 1531–1542.

Dasaratha, Krishna and Kevin He (2017), "Network Structure and Naïve Sequential Learning." Working Paper.

Del Vicario, Michela, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Qalter Quattrociocchi (2016), "Echo Chambers: Emotional Contagion and Group Polarization on Facebook." *Scientific Reports*.

Gallager, Robert (2013), *Stochastic Processes: Theory for Applications*. Cambridge University Press.

Gennaioli, Nicola and Andrei Schleifer (2010), "What Comes to Mind." *Quarterly Journal of Economics*, 125, 1399–1433.

Golub, Ben and Matthew Jackson (2010), "Naïve Learning in Social Networks and the Wisdom of Crowds." *American Economic Journal: Microeconomics*, 2, 112–149.

Golub, Ben and Matthew Jackson (2012), "How Homophily Affects the Speed of Learning and Best-Response Dynamics." *The Quarterly Journal of Economics*, 127, 1287–1338.

Golub, Ben and Evan Sadler (2016), "Learning in Social Networks." In *Oxford Handbook on the Economics of Networks* (Yann Bramoullé, Andrea Galeotti, and Brian Rogers, eds.), Oxford University Press.

Jasny, Lorien, Joseph Waggle, and Dana Fisher (2015), "An Empirical Examination of Echo Chambers in US Climate Policy Networks." *Nature Climate Change*, 782–786.

Jehiel, Philippe (2005), "Analogy-Based Expectation Equilibrium." *Journal of Economic Theory*, 123, 81–104.

Kahan, Dan (2013), "Ideology, Motivated Reasoning, and Cognitive Reflection." *Judgment and Decision Making*, 8, 407–424.

Li, Wei and Xu Tan (2017), "Learning in Local Networks." Working Paper.

Liggett, Thomas (1997), "Stochastic Models of Interacting Systems." *Annals of Probability*, 25, 1–29.

Molavi, Pooya, Alireza Tahbaz-Salehi, and Ali Jadbabaie (2017), "Foundations of Non-Bayesian Social Learning." Working Paper.

Mossel, Elchanan and Omer Tamuz (2014), "Opinion Exchange Dynamics." *Probability Surveys*, 14, 155–204.

Mueller-Frank, Manuel (2018), "Subtle Manipulation of Opinions in Social Networks." Working Paper.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer (2008), "Coarse Thinking and Persuasion." *Quarterly Journal of Economics*, 123, 577–619.

Nickerson, Raymond (1998), "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology*, 2, 175–220.

Rabin, Matthew and Joel Schrag (1999), "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics*, 114, 37–82.

Rubinstein, Ariel and Yuval Salant (2006), "A Model of Choice from Lists." *Theoretical Economics*, 1, 3–17.

Savage, Leonard (1954), *The Foundations of Statistics*, second edition. Dover Publications Inc. 2nd Ed. published 1972.

Spiegler, Ran (2016), "Bayesian Networks and Boundedly Rational Expectations." *Quarterly Journal of Economics*, 131, 1243–1290.

Tappin, Ben, Leslie van der Leer, and Ryan McKay (2017), "The Heart Trumps the Head: Desirability Bias in Political Belief Revision." *Journal of Experimental Psychology: General*, 146, 1143–1149.

Wilson, Andrea (2014), "Bounded Memory and Biases in Information Processing." *Econometrica*, 82, 2257–2294.

Yildiz, Ercan, Daron Acemoglu, Asuman Ozdaglar, Amin Saberi, and Anna Scaglione (2013), "Discrete Opinion Dynamics with Stubborn Agents." *ACM Transactions on Economics and Computation*, 1.

Zhao, Chen (2017), "Pseudo-Bayesian Updating." Working Paper.