Opinion dynamics beyond social influence

Benedikt V Meylahn[1] ![ORCID] and Christa Searle[2,3] ![ORCID]

[1]Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, The Netherlands, [2]Edinburgh Business School, Heriot-Watt University, Edinburgh, UK, and [3]Stellenbosch Unit for Operations Research in Engineering, Department of Industrial Engineering, Stellenbosch University, Stellenbosch, South Africa
Corresponding author: Benedikt V. Meylahn; Email: b.v.meylahn@uva.nl

Abstract
We present an opinion dynamics model framework discarding two common assumptions in the literature: (a) that there is direct influence between beliefs of neighboring agents, and (b) that agent belief is static in the absence of social influence. Agents in our framework learn from random experiences which possibly reinforce their belief. Agents determine whether they switch opinions by comparing their belief to a threshold. Subsequently, influence of an alter on an ego is not direct incorporation of the alter's belief into the ego's but by adjusting the ego's decision-making criteria. We provide an instance from the framework in which social influence between agents generalizes majority rules updating. We conduct a sensitivity analysis as well as a pair of experiments concerning heterogeneous population parameters. We conclude that the framework is capable of producing consensus, polarization and fragmentation with only assimilative forces between agents which typically, in other models, lead exclusively to consensus.

1. Introduction

The opinions held by an agent may be of crucial importance to their expressed behavior. Models that consider opinion formation tend to focus exclusively on social influence mechanisms of opinion change. We suggest a framework which includes a calculating, rational component in terms of how the agent incorporates information resulting from life's experience, as well as an affective component in terms of the effect of the opinions held by alters on that of the agent. This gives an explicit formulation of an agent's internal thought process which we believe should not be governed exclusively by social influence.

Our research question is: What elements constitute a model of opinion dynamics which goes beyond social influence while still including it? In particular our aim is to present a framework to support believable models of opinions allowing for two common phenomena:

1. agents who change their opinion based on personal experience (possibly in the absence of social influence), and
2. agents who retain their opinion in spite of (possibly strong) social influence urging them to change it.

There is a common idea driving both of these phenomena. Few people when asked why they hold the opinion they do will answer: "Because my neighbors hold this opinion." We believe it more likely they provide reasoning, substantiation and possibly evidence from their own experience. This points to a cognitive element which is often ignored in the literature pertaining to

opinion dynamics. In the model we propose agents learn (gather experience possibly reinforcing their convictions) about the opinion they hold, while also exposed to social influence from their neighbors. While the idea of modeling the evolution of opinions under social influence is not new, the inclusion of learning about an opinion through personal experience has not received much attention.

1.1 Relation to the literature

The literature in the field of opinion dynamics is expansive which is attested to by the abundance of review papers aiming to capture a moment of the state of the art of the field (see e.g., Castellano et al. (2009), Flache et al. (2017), Proskurnikov and Tempo (2017, 2018), Noorazar et al. (2020), Zha et al. (2021) and Bernardo et al. (2024)). As such, an exhaustive review of the literature is beyond the scope of this paper. The discussion that follows focuses on the commonalities between the models in the field and where these may be expanded upon. Furthermore, we restrict ourselves to literature pertinent to this paper in particular.

There is a stream of literature in which the agents incorporate a (possibly) weighted average of their neighbors' beliefs into their own (see e.g., the seminal works of French (1956), Harary (1959), and DeGroot (1974), and more recently Altafini (2013), Proskurnikov, et al. (2016), Liu et al. (2017) and Chan et al. (2024)). A second stream of literature follows the voter model (Clifford and Sudbury, 1973; Holley and Ligget, 1975) in which agents copy the opinion held by someone in their neighborhood. Castellano et al. (2009) made a significant improvement by means of the q-voter model in which instead of copying a random neighbor, agents copy the opinion held by at least q of their neighbors. For an overview the interested reader may consult Redner (2019).

The models within these two streams can further be categorized according to modeling decisions:

- Opinion representation being continuous or discrete;
- Opinion updating happening simultaneously or asynchronously;
- Forces between neighboring agents consisting only of attractive forces or including repulsive forces.

Despite these differences, the common thread is that agents are initialized with an opinion, the evolution of which is governed only by inter-agent communication. Another similarity of these models is that there is no distinction between an agent's opinion and their belief (which we define as the strength of conviction in that opinion), or the implication that an agent adjusts their belief (rather than their opinion) as a direct result of observing another agent's belief.

Remark 1. *Note that we distinguish between opinions and beliefs, terms that are often used interchangeably in the literature. We do this so that we can refer to the opinion an agent holds and the conviction in that opinion independently. As such we define*

- *opinion: a discrete opinion from the set of opinions which indicates support for the point of view represented by the opinion, and*
- *belief: the agent's strength of conviction in the opinion they hold.*

Giardini, et al. (2015) present a model in which an agent's opinion is a combination of: subjective truth value, a level of confidence therein and a perceived sharedness. Subsequently these three variables may change as result of interactions between agents.[1] Empirical studies by Johnson and Grayson (2005) and Ozdemir et al. (2020) show support for social influence effects which do not act directly on an agent's cognitive belief but rather on their emotive connection to the topic.

More recently Baccelli et al. (2017) challenge the assumption that agents change their beliefs[2] only as consequence of their network. They present a model in which noisy signals between agents

represent the possible endogenous evolution of belief within an agent. Though the agents have the possibility (by means of noise) to change their belief without social influence, the influence of another agents' belief is still directly on their own belief.

Flache et al. (2017) highlight the need for closer inspection of the assumptions underlying social influence and the modeling decisions that take place as a result of these assumptions. Flache et al. (2017) acknowledge that Giardini, et al. (2015) make such an effort and call upon researchers in the field to follow suit.

Noorazar et al. (2020) explicitly mention the need for more models like those of Baccelli et al. (2017) which question the assumption that beliefs should evolve exclusively as a result of social interaction. This evolution of beliefs outside the confines of social interaction is characteristic of sophisticated agents who have internal thought processes beyond copying their neighbors or behaving as an average of their social connections.

The gap in the existing literature is evident: There is a clear need for a model of opinion dynamics in which simultaneously (a) social influence between agents is not acting directly on the belief of agents (yet still directly on their decision making) and (b) agents have an internal process by which belief change may occur beyond the effects of social influence. Such a model would complement the existing literature which focuses on direct influence between beliefs. It would also provide a useful tool for the modeling of complex systems which are concerning more than the evolution of opinions.

1.2 Contribution

In an effort to address this gap and our research question, our core contribution is a model of opinion dynamics in which the social influence between agents is present but not all-dominating. Agents can both change and hold onto their opinion despite influence from their neighbors urging them to do the opposite. The mechanism by which this is achieved follows an idea from social psychology which has long been neglected in the modeling of opinion dynamics—that opinions (or attitudes) are shaped by experience.

In particular, to address the identified gap, we present a framework in which an opinion is modelled as a lens through which experiences are interpreted. This constitutes a random process by which an opinion sometimes successfully aligns with an experience had by an agent and sometimes fails to do so. This random process is a means to the end of modeling agents who can learn through experience about the opinions they hold. In doing so we align our model with the theory of attitude formation of Fazio et al. (2004). The agent's opinion then is a choice of lens, hoping for alignment between their opinion and experiences which creates cognitive harmony. In our model this is a decision-making process by which each agent asks themself if the opinion they hold aligns with a sufficient portion of life's experiences. This lends some sophistication to the agents capable of reasoning about the opinion they hold. We model the influence from one agent to another by means of adjusted decision-making criteria rather than a direct incorporation of a neighbors belief into one's own. That is, an agent is inclined to require a lower reliability from an opinion they share with a large portion of their neighborhood in order to maintain that opinion. By including this learning process as well as a social influence process we aim to align our framework with Gerard and Orive (1987) suggesting that we (humans) make our choice of opinion using both social and non-social information about it.

The result is a lightweight framework which may easily be implemented on top of other agent-based simulation models. We showcase the framework by means of a model instance. The model instance (and therefore the framework) has desirable properties which we confirm by a sensitivity analysis as well as a pair of experiments concerning heterogeneous population parameters: The framework instance enables polarization, consensus and fragmentation as steady state outcomes, all in the context of exclusively assimilative forces between agents. The framework features agents with sophistication in their view of the world yet does not incur a large computational load.

Table 1. Key concepts used in the model

Symbol	Name	Interpretation
$\mathcal{A}$	Opinion set	Discrete options for which opinions may be held by an agent
$a \in \mathcal{A}$	Opinion	Opinion from the set of possible opinions
$\theta_a \in (0, 1)$	Reliability	Likelihood that an experience positively reinforces opinion a
$b(x)$	Belief density	Believed probability density function of opinion reliability
$B(x)$	Belief distribution	Believed probability distribution function of opinion reliability
$\widehat{\theta_a}$	Estimated θ_a	Point estimate of the opinion reliability based on $B(x)$
$X \in \{1, 0\}$	Experience	Reinforcement ($X = 1$) or contradiction ($X = 0$) of held opinion
$\kappa \in (0, \infty)$	Stubbornness	Resistance of an agent toward social pressure

1.3 Nomenclature

There are key terms which we disambiguate to facilitate the exposition of the rest of the paper. We use the word *opinion* to refer to a choice a from a discrete set of opinions $\mathcal{A}$. Each opinion a from the set has a true *reliability* $\theta_a \in (0, 1)$ which relates to the likelihood that an *experience* $X \in \{1, 0\}$ reinforces that opinion ($X = 1$). The agents hold a *belief density* function b and a corresponding *belief distribution* function B pertaining to the reliability of their opinion in the Bayesian sense. These can be used to create a *point estimate* $\widehat{\theta}$ of the reliability of the opinion they hold (how reliable the agent believes their opinion to be). We use *stubbornness* $\kappa \in (0, \infty)$ as it relates to resistance toward social influence, not resistance to changing opinions in general. These key terms are summarized in Table 1.

1.4 Organization of paper

The remainder of the paper is presented in two parts. The first part details the framework: In Section 2 we describe an opinion dynamics model framework for a single agent which we extend to many agents in Section 3. The second part of the paper entails a model instance of the framework with a sensitivity analysis and a set of experiments. Specifically, Section 4 deals with the model instance from the framework. In Section 5 we describe the process and results of a sensitivity analysis of the model. In Section 6 we discuss experiments conducted on the model. We close with a discussion of our work and possible avenues of future research in Section 7.

2. Solo agent opinion dynamics model

For ease of exposition we first present a solo agent opinion dynamics model. We believe that agents should be able to adjust their opinion also in the absence of social influence. This model grew out of, and therefore closely follows, the model of Meylahn, et al. (2024) that investigates the problem of trusting institutions as a learning problem. Specifically we present a generalized framework which covers the single agent model of Meylahn, et al. (2024) as a special case. We posit that holding an opinion is akin to trusting that this opinion provides a good enough lens through which to interpret experiences and therefore may be modeled to have a reliability.

We note that modeling the evolution of an opinion using both social and non-social information aligns with the ideas from social psychology (cf. Gerard and Orive, 1987). We model the non-social information by means of experiences which drive belief formation as suggested by Fazio, et al. (2004).

2.1 Definition and interpretation of opinions

The dynamics evolve over rounds indexed $t = 1, 2, \ldots$. At the start of each round our agent holds an opinion a from the set of possible opinions $\mathcal{A}$. We refer to the opinion held by the agent in round $t \in \mathbb{N}$ as $a_t \in \mathcal{A}$. The agent is subsequently exposed to an experience. We call the outcome of an experience in round t holding opinion $a \in \mathcal{A}$: $X_a^t \in \{0, 1\}$. When X_a^t takes the value one, the agent's opinion a aligns with (is reinforced by) the experience in round t. Conversely, when X_a^t takes the value zero, the agent's opinion a does not align with the experience in round t. Specifically X_a^t for all $a \in \mathcal{A}$ and $t \in \mathbb{N}$ are random variables:

$$X_a^t = \begin{cases} 1, & \text{with probability } \theta_a \\ 0, & \text{with probability } 1 - \theta_a, \end{cases} \quad \text{with } \theta_a \in (0, 1), \forall a \in \mathcal{A}. \quad (1)$$

We call the probability that opinion a aligns with an experience, opinion a's *reliability*, $\theta_a \in (0, 1)$ for all $a \in \mathcal{A}$. Experiences $\{X_a^t : t \in \mathbb{N}\}$ are i.i.d. for each opinion $a \in \mathcal{A}$. We note the conscious decision to not fix $X_a^t = 1 - X_b^t$ or $\theta_a = 1 - \theta_b$ for $a \neq b$ and $a, b \in \mathcal{A}$. While opinions $a \neq b$ from $\mathcal{A}$ are indeed concerning the same subject, we allow for overlap in the sets of experiences which may align with different opinions. Consider, for example, that dissatisfaction with one political party does not necessarily imply satisfaction with an alternative. The agent only interprets experiences using the opinion they hold. This means they do not observe X_b^t for $b \neq a_t$. We assume this for simplicity. If the agent did observe X_b^t for $b \neq a_t$ would imply that agents interpret all of their experiences with each possible opinion. Furthermore, they do not know the respective opinion's reliabilities θ_a for $a \in \mathcal{A}$. We suppose that the agent receives utility $p \in \mathbb{N}$ when an experience aligns with their opinion ($X_a^t = 1$) and loses utility $l \in \mathbb{N}$ when it not($X_a^t = 0$).

The analogy of viewing an opinion as a lens through which to interpret experiences may be elaborated into a set of views which comprise something like a *narrative* about a particular subject. The narrative we follow may provide: explanations for experiences, a guideline on how to act, heuristic answers to questions, and an ideal to strive for. We present an example of our interpretation from the folklore of Robin Hood.

Example 1 (Robin Hood). *Two possible narratives, or opinions, provide a means for the townsfolk to make sense of what is happening in Nottingham. The two opinions support the Sheriff and Robin, respectively:*

- *Supporters of the Sheriff of Nottingham believe that the people of Nottingham are indebted to him and should work hard and pay taxes to him.*
- *Others, who support Robin, believe that the people of Nottingham have been treated unfairly and should be helped, rather than taxed and punished.*

Consider the position of one of the Sheriff's friends. They accept that some people of the town have nothing, because they believe this to be the result of laziness. Their opinion is affirmed when they observe a lazy person lose their fortune i.e., an experience which aligns with their opinion. Alternatively, if they meet someone whose livelihood was unjustly taken by the Sheriff, their views might be challenged.

Similarly one of Robin's supporters might be affirmed in their views when witnessing the widespread poverty in the town. Yet, their views too may be challenged if they see the Sheriff protecting townsfolk from injustices that may befall them. This experience disagrees with their opinion by providing an example of how the taxes are being used to the benefit of the town.

The binary expression of the opinions in this example is support for Robin or the Sheriff. The opinions on the other hand are complex objects. In the case of political preferences our "narrative" interpretation aligns with the description of an ideology summarized by Weber (2019). We use

"opinion" rather than "narrative," in order to avoid (direct) connection to how narratives within media might influence opinions.

2.2 Agent belief

The agent holding opinion $a \in \mathcal{A}$ has a belief which measures the strength of their conviction of their opinion. This takes the form of a belief density (function) $b(x)$ and a corresponding belief distribution (function) $B(x)$. These are related in the usual way:

$$B(x) = \int_0^x b(s)\mathrm{d}s. \tag{2}$$

In particular the belief distribution denotes the subjective probability (as in the Bayesian interpretation of probability) they place on the true reliability of the opinion they hold being below a given value,

$$B(x) := \mathbb{P}(\theta_a \le x) \quad x \in [0, 1]. \tag{3}$$

The agent uses their belief distribution to attain an estimate for the true value of θ_a. In the absence of other evidence, the agent uses their prior belief. After any number of experiences, the agent adjusts their belief accordingly.

2.2.1 Prior belief distribution

We model the agent to start with a prior belief distribution which is what they believe the distribution function of θ_a to be for all opinions $a \in \mathcal{A}$ that they have no other information about. The agent has prior belief density $b_0(x)$ and distribution $B_0(x)$ meaning that they initially believe that the probability relating to the reliability of opinion a_0 is such that:

$$\mathbb{P}(\theta_{a_0} \le x) = B_0(x) = \int_0^x b_0(s)\mathrm{d}s. \tag{4}$$

If the agent switches their opinion at some time t_0 to opinion $a_{t_0} \in \mathcal{A}$ then they revert to their prior belief $b_0(x)$ regardless of which opinion they are switching to. This models the agent's forgetting of experiences with an opinion they may have held in the past. The formulation (4) generalizes the formulation of belief in Meylahn et al. (2024) by allowing general belief distributions instead of only the Beta-distribution.

In agent-based models with more richness, agent features may play an important role in determining the prior an agent has for each of the opinions. Examples of characteristics that may play a role are age, education attained and current employment. We assume that their prior beliefs are all the same for simplicity.

2.2.2 Belief distribution update

We call the consecutive rounds in which the agent held opinion a, a *run* with opinion a. We refer to the most recent switching time as S_t, which identifies the round in which the current run started:

$$S_t := \min\{n : a_m = a_t, \forall m \in [n, t]\}. \tag{5}$$

We model the agent to "forget" previous runs with an opinion. In constructing their belief distribution of an opinion, they only use their most recent run's history which started at time S_t. This means that they only use the experiences gained since they most recently switched to the opinion they are currently holding to formulate their belief distribution of that opinion. We define the

agent's current experience history until the end of round $t \in \mathbb{N}$ as H_t. That is the set of experiences observed up until (and including) round t during their current run:

$$H_t = \{X_{a_n}^n : n \in [S_t, t]\} \quad \text{for } t \in \mathbb{N}. \tag{6}$$

This information along with their prior belief is used by the agent to formulate a belief distribution at time $t \in \mathbb{N}$:

$$B_0 \times H_t \to B_t(x). \tag{7}$$

With this incorporation of an agent's ability to "forget" it may be noted that more complicated mechanisms are possible, and we do not insist that ours is the best, however, we do suggest there is value in having at least some form of forgetting in a model of opinion dynamics which includes learning.

2.2.3 Point estimate of opinion reliability

We model the agent to use this belief to attain a point estimate $\widehat{\theta}_t \in [0, 1]$ of the reliability, θ_{a_t}. A common example is the mean value of the belief density,

$$\widehat{\theta}_t = \int_0^1 x b_t(x) \mathrm{d}x, \quad \in [0, 1]. \tag{8}$$

Alternatives to the mean value are the upper or lower confidence intervals and many more. The way in which the history is incorporated is part of the modeller's choice. A logical choice for a Beta-distributed prior belief is Bayesian updating. Alternatively if the prior is a point mass on the estimate, simple exponential smoothing might better serve the task.

Note that each agent has only one belief distribution at any time which is their belief in the opinion they currently hold. We assume, as simplification, that the agents do not continuously compare the alternative interpretations (resulting from different opinions) of an experience. This is what would be required to track the reliability of each opinion in order to explicitly compare these. Should the number of possible opinions be large, agents which compare between all opinions often would be faced with a heavy computational effort.

2.3 Threshold decision making (choosing an opinion)

The agent is faced with deciding whether to place their trust in the opinion which they held in the previous round. They do so by a satisficing procedure (cf. Simon 1956; Artinger et al. 2022); checking if the current opinion is good enough. The agent decision-making process we use has mechanical similarities with the "state system" in the "consumats" model proposed by Jager and co-authors (see e.g., Jager et al. 1995, 1999; Janssen and Jager, 1999). In this consumer behavior model, agents who are satisfied with their consumption do not spend cognitive resources to find alternatives as long as they deem their current options to be good enough. It is only the dissatisfied consumer who investigates other avenues of consumption as motivated by their dissatisfaction.

In choosing an opinion to hold the agent asks themselves whether they expect positive utility from the opinion they are currently holding. In other words they check the truth of the inequality:

$$p\widehat{\theta}_t - l(1 - \widehat{\theta}_t) \geq 0. \tag{9}$$

This inequality may be rearranged and so equivalently the agent asks themselves whether:

$$\widehat{\theta}_t \geq \frac{l}{p + l} =: \theta_{\mathrm{crit}} \in (0, 1). \tag{10}$$

Here we have defined θ_{crit}, the minimum reliability point estimate the agent requires an opinion to have in order to continue holding that opinion. If the agent chooses to switch opinions (because

they are sufficiently dissatisfied), they choose a new one from the set of opinions excluding the opinion they are switching from. The choice of the agent at time $t \in \mathbb{N}$ can now be defined:

$$a_t = \begin{cases} a_{t-1}, & \text{if } \widehat{\theta}_t \geq \theta_{\text{crit}}, \\ b \in \mathcal{A} \setminus a_{t-1}, & \text{otherwise.} \end{cases} \tag{11}$$

This generalizes the decision making in the single agent model presented by Meylahn et al. (2024) from trusting or not trusting to holding one of the opinions in $\mathcal{A}$. The protocol used to choose which of the alternative opinions the agent chooses is up to the modeller. For a set $\mathcal{A}$ of only two opinions the choice is straightforward; simply the other opinion. If there are three or more opinions the choice could make use of a direct comparison between the remaining options to decide which opinion to switch to. These mechanisms of switching in the solo agent model serve as a baseline. In the multiple agent model there are more sophisticated and believable mechanisms by which an agent might choose which opinion to switch to.

This is a simple satisficing model in which the agent is not concerned with continuous comparisons between opinions. Instead the agent has a desired level of reliability and retains the opinion they are holding if they believe it to satisfy this level.

3. Many agent opinion dynamics framework

In this section we extend the solo agent framework by placing N solo agents into a network. The reason for delaying this exposition is clarity. There is interdependence of the actions taken by agents and the actions taken by their neighbors. By first introducing a solo agent model which contains the basic elements of the many agent model, all that remains is to describe how these are influenced by the interplay between agents. The agents in the network communicate their opinion with their neighbors which is how we induce social influence between them. The crucial difference between this framework and other opinion dynamics models is that the effect of the inter-agent communication is not on the agents' belief distributions but rather on their decision-making threshold θ_{crit}.

We start by recapitulating those elements of the single agent model which are specified per agent in the many agent model. As such §3.1 will have much overlap with §2 with only small differences. Thereafter, in §3.2 we describe the new elements which create the social influence between agents.

3.1 A network of solo agents

The framework does not prescribe a population structure but assumes that one is given. That is, either the modeller uses a network empirically sourced or makes use of an appropriate random graph model to generate a network. The interested reader may consult the work of Robins, et al. (2001) relating to models of network generation. We suppose that there exists a population of agents of finite size $N \in \mathbb{N}$. The population is embedded in a network $G = (V, E)$, with $|V| = N$ vertices representing agents and a set of social ties represented by the edges, E. The edges are directional in nature; an edge $(u, j) \in E$ indicates that agent u exerts social influence on agent j.

We define the opinion held by agent $j \in V$ in round t as $a_t(j) \in \mathcal{A}$. In each round every agent (a) holds an opinion, (b) has an experience which corroborates or contradicts their opinion and (c) observes the opinions held by their neighbors. The outcome of the experience had by agent $j \in V$, holding opinion $a \in \mathcal{A}$ at time $t \in \mathbb{N}$ is the random variable:

$$X_a^t(j) = \begin{cases} 1 & \text{with probability } \theta_a, \\ 0 & \text{with probability } 1 - \theta_a, \end{cases} \quad \text{with } \theta_a \in (0, 1), \forall a \in \mathcal{A}. \tag{12}$$

Similarly to the solo agent model, agent $j \in V$ receives utility $p_j \in \mathbb{N}$ when an experience agrees with their opinion and loses utility $l_j \in \mathbb{N}$ when an experience disagrees with their opinion. We think it is a reasonable starting point to assume that θ_a for $a \in \mathcal{A}$ is the same for all agents because it is a property of the opinion rather than the agents.

Now each agent $j \in V$ is equipped with a prior belief density $b_0^j(x)$ and distribution $B_0^j(x)$ relating to the reliability of opinion $a_0(j)$ as before:

$$B_0^j(x) := \mathbb{P}(\theta_{a_t(j)} \leq x) = \int_0^x b_0^j(s)\mathrm{d}s, \quad x \in [0, 1]. \tag{13}$$

Agents may switch between opinions and so we refer to agent j's most recent switching time:

$$S_t(j) := \min\{n : a_m(j) = a_t(j), \forall m \in [n, t]\}, \quad \text{for } j \in V \text{ and } t \in \mathbb{N}. \tag{14}$$

Subsequently, agent j's current experience history until the end of round $t \in \mathbb{N}$ is $H_t(j)$. That is the set of experiences observed up until (and including) round t during their current run:

$$H_t(j) = \{X_{a_n(j)}^n(j) : n \in [S_t(j), t]\}, \quad \text{for } j \in V, \text{and } t \in \mathbb{N}. \tag{15}$$

This information along with their prior belief is used by the agent to formulate a belief distribution for the opinion they are holding at time $t \in \mathbb{N}$:

$$B_0^j \times H_t(j) \to B_t^j(x). \tag{16}$$

This belief, in turn gives an updated estimate $\widehat{\theta}_{t,j}$ (assuming use of the mean value),

$$\widehat{\theta_t}(j) = \int_0^1 x b_t^j(x)\mathrm{d}x, \tag{17}$$

representing the strength of their conviction in the opinion they are holding.

3.2 Social influence

The process of social influence we present aligns with the *internalization* process of opinion change presented by Kelman (1961). That is, the agents are using the information provided by their network to facilitate their decision making, rather than simply trying to comply or identify with one another. An agent $u \in V$ influences agent $j \in V$ if there is an edge $(u, j) \in E$. Each agent $j \in V$ has a set of social ties which we call the neighborhood of agent j:

$$N(j) := \left\{u : (u, j) \in E\right\}, \tag{18}$$

that is the set of agents who are said to influence agent j. For ease of notation we assume that the edges between agents are unweighted. Agent $j \in V$ observes the opinions held by the agents in their neighborhood $N(j)$. We thus model agents to communicate only which opinion they hold to their neighbors and not their belief, *i.e.* the strength of their conviction in that opinion. This assumption is based on the idea that it is much easier to convey a discrete choice of opinion to a social connection than it is to elucidate the nuances involved in the procedure by which such a choice was made. This provides agent $j \in V$ with their network influence set for time $t \in \mathbb{N}$:

$$I_t(j) := \{a_t(k) : k \in N(j)\}. \tag{19}$$

In formulating the influence of $N(j)$ on agent $j \in V$ we aim to follow empirical literature which shows that peer-to-peer influence has its effect not in cognitive elements of belief but rather affective elements influencing decision making (Johnson and Grayson, 2005; Ozdemir et al., 2020). In doing so, we posit that the role of peer-to-peer influence is similar in the context of brand loyalty and the expression of opinions. Indeed the methods of marketing are commonly used in politics (Newman. 2002) which is one of the main arenas of opinions dynamics. The affective elements may of course still influence decision making, but the channel they follow does not effect

the rational calculating elements associated with decision making. Instead the agent's threshold is adjusted based on the information gained from their network. Consider an agent's thoughts: 'If it is good enough for my neighbors, why should it not be good enough for me?'

Define $\theta^*_{\mathrm{crit},t}(j)$ as the *network adjusted critical reliability* of agent $j \in V$ at time $t \in \mathbb{N}$:

$$\theta^*_{\mathrm{crit},t}(j) : p_j \times l_j \times I_t(j) \mapsto [0, 1]. \tag{20}$$

The elements this function may use are thus contained in p_j, l_j and $I_t(j)$. We define the number of agents in j's neighborhood expressing the same opinion as agent $j \in V$ at time $t \in \mathbb{N}$ as:

$$m_t(j) := \left| \left\{ b : \left(b \in I_t(j) \right) \wedge \left(b = a_t(j) \right) \right\} \right|, \tag{21}$$

i.e. the number of agents in agent j's neighborhood *matching* agent j's opinion. Similarly we define the number of agents in agent j's neighborhood *not matching* agent j's opinion as:

$$n_t(j) := |N(j)| - m_t(j). \tag{22}$$

Note that if it is desirable to have weighted connections between agents, the above may be replaced with total weight within agent j's neighbohood in agreement with agent j's opinion and the total weight remaining, respectively. The functions $\theta^*_{\mathrm{crit},t}(i)$ can be chosen in various ways. We suggest the following properties of the network influence function:

- If there is equal support and opposition ($m_t(j) = n_t(j)$), the effect is null, $\theta^*_{\mathrm{crit},t}(j) = \theta_{\mathrm{crit}}(j)$.
- If there is more support than opposition ($m_t(j) > n_t(j)$), then the threshold is lowered, $\theta^*_{\mathrm{crit},t}(j) < \theta_{\mathrm{crit}}(j)$.
- If there is more opposition than support ($m_t(j) < n_t(j)$), then the threshold is increased, $\theta^*_{\mathrm{crit},t}(j) > \theta_{\mathrm{crit}}(j)$.

The belief updating of the agents thus has not changed, yet their decision making is affected by the communication of their neighbors. This means that $\widehat{\theta}_t(j)$ is still constructed as in the solo agent model. Instead of comparing this believed point estimate to a constant threshold $\theta_{\mathrm{crit}}(j)$, agent $j \in V$ uses the truth of the inequality $\widehat{\theta}_t(j) > \theta^*_{\mathrm{crit},t}(j)$ to determine whether they keep holding their opinion. Note that there are two possible triggers for an agent to switch their opinion. Their estimate may drop below their critical reliability as a result of one too many experiences which did not align with their opinion. Alternatively, a change in the opinions held by their neighbors may increase their critical reliability above their current estimate. Once an agent has decided to switch their opinion, they might make use of one of a number of switching rules to figure out which opinion to switch to:

- Agents can make a direct comparison of the remaining opinions based on their rational beliefs of those opinions, choosing the 'best' one. Such reasoned behavior has been studied extensively (Fishbein and Ajzen, 1977; Ajzen. 1991), for a recent description and discussion of developments consult (Ajzen, 2020).
- The agent might imitate a neighbor. Jager and various co-authors (Jager et al., 1995, 1999 Janssen and Jager, 1999) suggest that imitation is most likely to occur when agents are faced with high uncertainty about their options and are not willing to spend a lot of cognitive resources.
- Or the agent can adopt the majority opinion within their neighbohood. This corresponds to the expected choice under the random imitation described in the previous option.

The choice of switching rule is up to the modeller using the framework and may well be heavily influenced by the topic of the opinions. For a more detailed discussion on agent decision making the interested reader may consult the survey of Balke and Gilbert (2014).
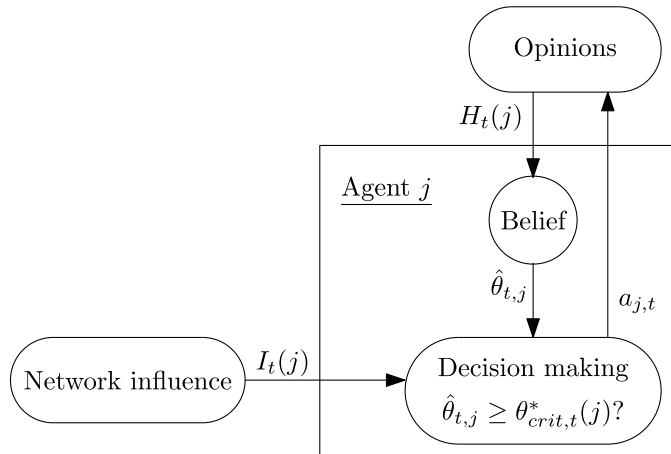
Figure 1. A graphical illustration of the opinion dynamics framework proposed.

   This formulation of social influence is in stark contrast with the mechanism of social influence in the dual agent model presented by Meylahn et al. (2024). In their study, agents use the observation of their neighbors' action rationally to adjust their belief, our agents simply use the heuristic of changing the threshold which they use for the decision making. This saves a lot of computation making the model tractable for more than two agents.

3.3 Summary of framework

Agents choose an opinion to hold in each round $a_t(j) \in \mathcal{A}$. Subsequently they experience agreement or disagreement. The effect of this experience is an updated belief distribution B_t^j based on $H_t(j)$. In order to choose whether to switch opinion in the following round, they compare a point estimate from their belief $\widehat{\theta}_t(j)$ with a critical reliability $\theta^*_{crit,t}(j)$. This critical reliability is adjusted according to the opinions expressed by the agent's neighbors captured in $I_t(j)$. Their own opinion expression is also how the agent influences their neighbors. Figure 1 serves to illustrate how the elements of the framework fit together. In the single agent model, the agent's opinion boils down to a 'strong enough' belief. In the multi-agent model, an agent's opinion is the result of a process combining cognitive processing of experiences in the agent belief and the affective forces of social influence.

Example 2 (Political preferences). *The opinion set $\mathcal{A}$ may consist of political parties, while the opinion held by an agent represents their current voting inclination. The experiences had by the agents correspond to daily experiences which may enhance or diminish an agent's support for their political party. For instance, an agent may observe a system implemented by their political party failing (or succeeding) thereby weakening (or strengthening) their support for the party. It need not be the case that an agent grows in support for an alternative party simply because they are dissatisfied with their own. Our framework assumes that agents only consider other parties once they are sufficiently dissatisfied with their previous one, as opposed to performing a constant comparison.*

4. Framework instance

We consider an opinion set of two opinions $\mathcal{A} = \{0, 1\}$. The opinions 0 and 1 have reliability $\theta_0, \theta_1 \in (0, 1)$, respectively. The agents in our framework instance are embedded in a Watts and Strogatz (1998) generated network. The workings of the Watts-Strogatz model are illustrated in

the Appendix. Note that for this model instance we assume that if agent u influences agent v, then also agent u is influenced by agent v. That is if $(u, v) \in E$ then also $(v, u) \in E$. Studying the effect of asymmetrical influence is an interesting avenue for future work which falls outside the scope of this paper.

4.1 Belief in the opinion

The agents in our model have a prior belief density in the form of a Beta-distribution with shape parameters $\alpha, \beta \in \mathbb{N}$,

$$b_0(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\mathrm{d}y}. \tag{23}$$

At time $t \in \mathbb{N}$, each agent $j \in V$ is only aware of the most recent history $H_t(j)$, which pertains to the opinion they currently hold. As such the agent keeps track of the number of *confirming* experiences during their most recent history $H_t(j)$ up until time $t \in \mathbb{N}$ using:

$$c_t(j) = \sum_{n=S_t(j)}^{t} X_n. \tag{24}$$

Similarly they use

$$r_t(j) = \sum_{n=S_t(j)}^{t} (1 - X_n), \tag{25}$$

to keep track of the number of experiences during their most recent history $H_t(j)$ up until time $t \in \mathbb{N}$ which *refute* their current opinion. Furthermore, the agents make use of Bayesian belief updating to keep:

$$B_t^j(x) = \mathbb{P}\left(\theta_{a_t(j)} \leq x \mid H_t(j)\right), \quad \text{for } t \in \mathbb{N}. \tag{26}$$

With the convention that if $H_t(j) = \emptyset$, then $\mathbb{P}(\theta_{a_t(j)} \leq x) = B_0^j(x)$. The agent j thus holds a belief density $b_t^j(\theta)$ at time $t \in \mathbb{N}$

$$b_t^j(x) = \frac{x^{c_t(j)}(1-x)^{r_t(j)}b_0^j(x)}{\int_0^1 y^{c_t(j)}(1-y)^{r_t(j)}b_0^j(y)\mathrm{d}y}, \quad x \in [0, 1]. \tag{27}$$

As point estimate the agents use the mean of their belief distribution at time $t \in \mathbb{N}$. Conveniently, for the combination of Beta-distributed prior belief and Bayesian updating the mean of the belief distribution is given by

$$\widehat{\theta}_t(j) = \frac{\alpha + c_t(j)}{\alpha + \beta + c_t(j) + r_t(j)}, \quad \forall t \in \mathbb{N}, \text{ and all } j \in V. \tag{28}$$

4.2 Threshold and network influence

In equations (21) and (22) in §3.2 we have defined the number of agents in agent j's neighborhood in agreement with agent $j \in V$ at time $t \in \mathbb{N}$ as $m_t(j)$ and the number of agents in disagreement as $n_t(j)$, respectively. We define the network influence function $\theta^*_{\text{crit},t}(j)$ representing the influence of agent $j \in V$'s neighborhood on their decision making as a mapping $\theta^*_{\text{crit},t}(j) : p_j \times l_j \times I_t(j) \mapsto [0, 1]$ by

$$\theta^*_{\text{crit},t}(j) = \frac{p_j + f_j(I_t(j))}{p_j + l_j}, \quad \text{for } t \in \mathbb{N}, \tag{29}$$
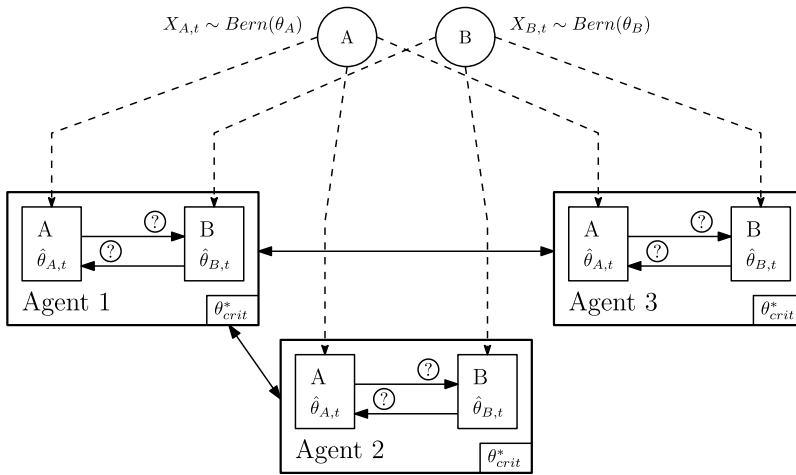
Figure 2. A graphical illustration of an example network agent model.

with the crucial element $f_j(I_t(j))$ defined:

$$f_j(I_t(j)) := \frac{n_t(j) - m_t(j)}{\kappa_j |N(j)|}, \quad \text{for } t \in \mathbb{N}, j \in V, \tag{30}$$

where $\kappa_j \in (0, \infty)$ is agent j's *stubbornness* parameter. A large κ represents agents who are not very influenced by their neighborhood. In fact if κ is large enough for an agent, their opinion switching happens independently from their neighborhood.[3] It should be noted that even the most stubborn agents in our model may change their opinion based on their individual belief updating. For the most stubborn agent possible ($\kappa_j = \infty$) the network influence becomes negligible ($f_j \to 0$) and so their criteria remains unchanged $\theta^*_{\text{crit},t} = \theta_{\text{crit}}$. It is possible that this agent's estimate of their opinion's reliability drops below this threshold and so they can still switch their opinion. This highlights that the stubbornness in our model is not a stubbornness of a particular opinion but rather a stubbornness with respect to influence from others.

It is worth mentioning that if κ is small enough for all agents then the model reduces to one in which agents adopt the opinion being held by the majority of their neighborhood.[4]

We use a natural starting point for utility parameters. By setting $l_j = p_j = 1$ for all $j \in V$ we have that every agent's starting threshold is $\theta_{\text{crit}} = 1/2$. This means that an agent in the absence of a network influence will continue to hold their current opinion if their belief satisfies: $\widehat{\theta}_t(j) > 0.5$. The resulting decision making is made by checking the inequality of $\widehat{\theta}_t(j) > \theta^*_{\text{crit},t}(j)$:

$$\frac{\alpha + c_t(j)}{\alpha + \beta + c_t(j) + r_t(j)} \geq \left(1 + \frac{n_t(j) - m_t(j)}{\kappa_j |N(j)|}\right) / 2. \tag{31}$$

A graphical representation of this model is presented in Figure 2. This example has three agents $V = \{1, 2, 3\}$ with connections $E = \{(1, 2), (2, 1), (1, 3), (3, 1)\}$. The question mark icons represent a check of the inequality (31). A transition from trusting one opinion to trusting the other should take place whenever this inequality is not true.

4.3 Observable outcomes of a simulation

In the model we have presented, agents are able to converge onto an opinion. For an agent $j \in V$ holding opinion $a \in \mathcal{A}$, with enough time (and thus sampling of experiences) the agent's estimated reliability may tend towards the true reliability, $\widehat{\theta}_t(j) \to \theta_a$ and if $\theta_a > \theta^*_{\text{crit},t}(j)$ it is possible that the agent holds this opinion indefinitely. If at some time $t_0 \in \mathbb{N}$ this is true for all the agents in the

network, the process has reached a steady state, as anymore switching of opinions becomes more unlikely. For the purposes of the simulation, we use a proxy for steady state: 100 simulated rounds in which no agent switches their opinion. We are interested in whether there is consensus in such a steady state or if there is some level of discordance. We define the probability of consensus as the probability of each agent holding the same opinion at a steady state time $t_0 \in \mathbb{N}$:

$$C := \mathbb{P}\left(a_{t_0}(j) = a_{t_0}(1), \forall j \in V\right). \tag{32}$$

The choice of reference to the first agent's opinion $a_{t_0}(1)$ is arbitrary as all agents are required to be in agreement.

We define the proportion of discordance as the number of discordant edges divided by the total number of edges. A discordant edge (u, v) is one in which the opinions of the agents are different: $a_u \neq a_v$. We label the set containing the discordant edges at time $t \in \mathbb{N}$, $E_D(t)$:

$$E_D(t) := \{(u, v) \mid a_t(u) \neq a_t(v), \text{with } (u, v) \in E\}. \tag{33}$$

As such we can define the asymptotic proportion of discordance as:

$$D := \mathbb{E}\left[\liminf_{t \to \infty} \frac{|E_D(t)|}{|E|}\right]. \tag{34}$$

Considering that we are performing a simulation study we can only approximate the quantities of interest with empirical measures at the end of simulation runs. To this end we consider the empirical value:

$$\widehat{C} = \frac{z_c}{Z_{\text{sim}}}, \tag{35}$$

where z_c is the number of simulation runs in which $a_{t_0}(j) = a_{t_0}(1), \forall j \in V$ at termination time t_0, and Z_{sim} is the total number of simulations that were run. Similarly, for the proportion of discordance:

$$\widehat{D} = \frac{\sum_{n=1}^{Z_{\text{sim}}} |E_D(t_0, n)|}{Z_{\text{sim}} \times |E|}, \tag{36}$$

where $E_D(t_0, n)$ is the set of discordant edges at the time when the n-th simulation iteration terminates $t = t_0$. In other words, we sum the number of discordant edges at the termination time of each simulation run and divide this by the total number of edges. This gives the average proportion of edges that were discordant at termination time across all simulations.
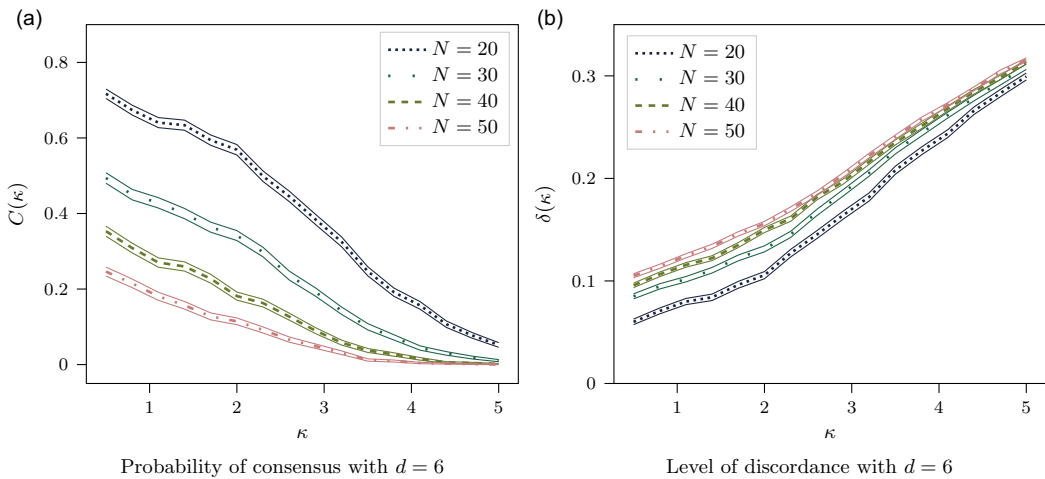
5. Sensitivity analysis

We are interested in the effect of the model parameters on the probability of consensus and the level of discordance. To study this we conduct a sensitivity analysis of which we now describe the setup and the results. The base network consists of $N = 20$ agents in the Watts-Strogatz model Watts and Strogatz (1998) with $d = 4$ nearest neighbors and a rewiring probability of $w = 0.20$. Note that we generate a new network for each simulation run. All the agents in the population hold a uniform prior belief distribution which is a Beta-distribution with shape parameters $\alpha = 4$, and $\beta = 2$. We reiterate that whenever an agent switches their opinion they restart their learning process from their prior belief. The value of the stubbornness is varied along with the parameters being tested for the sensitivity though is kept homogeneous between agents, $\kappa_j = \kappa, \forall j \in V$. We take $\kappa = 0.5$ to $\kappa = 5$ in increments of 0.3. In general the trend we observe is that as κ increases the probability of consensus decreases and the level of discordance increases (Figure 3).

The opinions in the sensitivity analysis have identical reliability $\theta_0 = \theta_1 = 0.6$. Keeping these the same allows us to focus on the agent dynamics rather than questions of convergence to a 'better' opinion. The simulation starts with a warm-up phase $t_s = 5$ rounds. In these rounds the
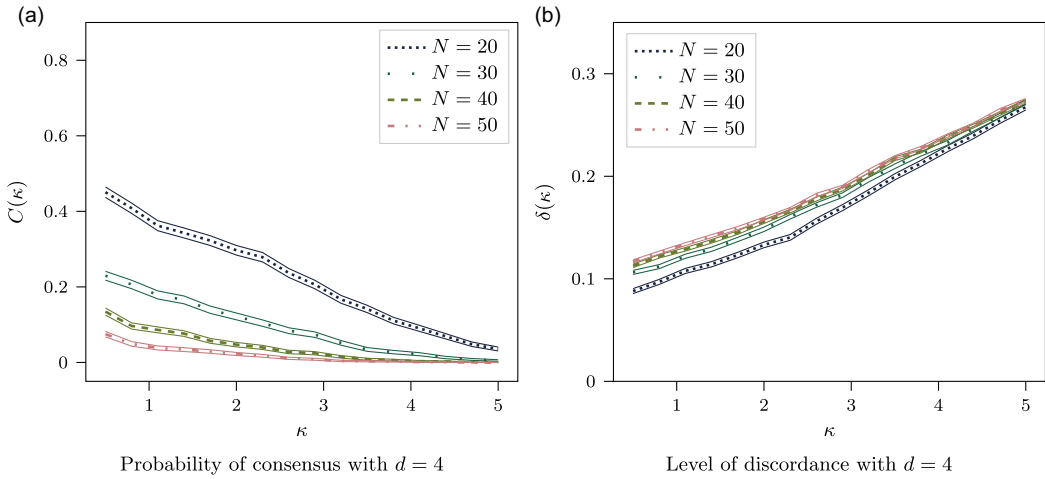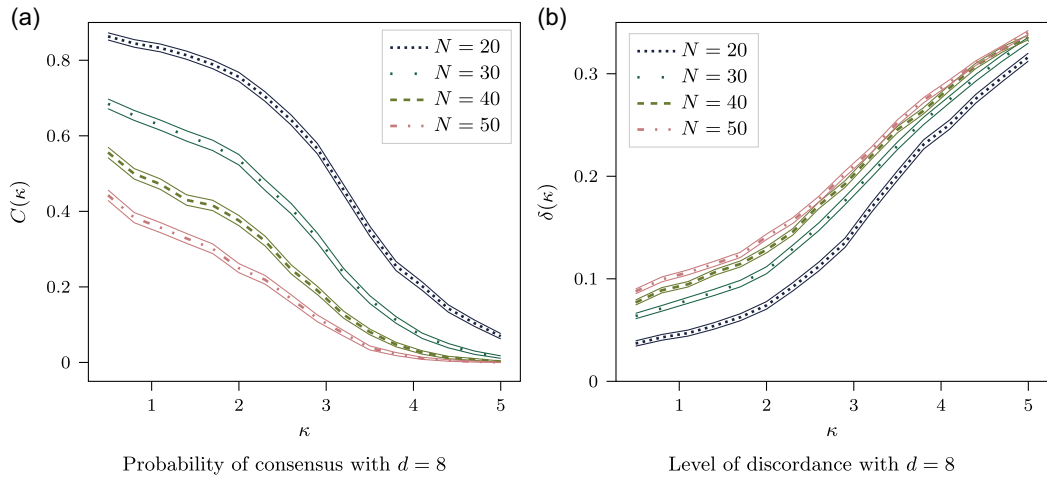
Figure 3. Results of the sensitivity analysis inspecting the number of agents in the system with 4 nearest neighbors. Parameters: $d = 4$, $w = 0.2$, $t_s = 5$, $\alpha = 4$, $\beta = 2$, and $\theta_0 = \theta_1 = 0.6$.

Figure 4. Results of the sensitivity analysis inspecting the number of agents in the system with 6 nearest neighbours. Parameters: $d = 6$, $w = 0.2$, $t_s = 5$, $\alpha = 4$, $\beta = 2$, and $\theta_0 = \theta_1 = 0.6$.

agents follow the solo agent model. Thereafter they start communicating with neighbors and take this communication into consideration changing their $\theta^*_{\text{crit},t}$ accordingly. We run the simulation model 5 000 times under each of the parameter settings in order to obtain 95% confidence intervals for the quantities of interest (Figure 4), (Figure 5).

5.1 Number of agents

We vary the total number of agents $N \in \{20, 30, 40, 50\}$. Additionally, for each of these values of N, we vary the nearest number of neighbors taking values $l \in \{4, 6, 8\}$. We present the results grouped according to the number of nearest neighbors. Within Figures 3a, 4a, and 5a, depicting the probability of consensus, we observe that as the population grows, the probability of consensus

Figure 5. Results of the sensitivity analysis inspecting the number of agents in the system with 8 nearest neighbors. Parameters: $d = 8$, $w = 0.2$, $t_s = 5$, $\alpha = 4$, $\beta = 2$, and $\theta_0 = \theta_1 = 0.6$.

decreases. By comparing between these figures we see that as the number of neighbors increases the probability of consensus increases. Both of these results are conceivable. A larger population (keeping the number of neighbors constant) is likely to make it difficult for an opinion to spread throughout the entire network. Similarly, the more neighbors the agents have (keeping the population size constant), the easier it should be for an opinion to spread throughout the population. Quite logically, we observe the inverse effect in Figures 3b, 4b, and 5b, on the level of discordance. We would like to draw the reader's attention to the fact that both of these effects (comparing between population size) becomes smaller as the stubbornness κ increases to 5. We believe this is conceivable because a greater κ means that the agents in the population become increasingly independent of one another and thus are less effected by the network they form a part of.

5.2 Probability of rewiring

In order to see how the probability of rewiring affects the outcome of the model, we take the probability of rewiring each edge from the set $w \in \{0, 0.05, 0.10, 0.15, 0.20, 0.25\}$. We present the probability of consensus for these probabilities of rewiring in Figure 6a and the level of discordance in Figure 6b. Again we can see the effect of the network decreasing as κ increases in Figure 6a by the fact that the probabilities of consensus seem to merge at roughly $\kappa = 4$. Sensibly, we see that the probability of consensus is higher for networks with more rewiring. We posit that this is because of the greater number of cross population connections which make the formation of two (or more) polarized groups more difficult.

We also observe an interesting phenomenon in Figure 6b. Namely, that level of convergence for different probabilities of rewiring all seem to intersect just after $\kappa = 2$. Furthermore, these have an inverse ordering before and after this crossing point. Our explanation for this effect follows: Before the crossing point, a higher probability of rewiring makes polarization harder, and so the population tends toward consensus which has the lowest level of discordance. The setting with low level of rewiring has much more structure which allows more easily for polarization. After the crossing point, a high enough κ creates more room not only for polarization but also fragmentation. As the agents become more independent, the cross population connections at greater w create a higher level of discordance. More structured populations with low w have more of a balance between polarization (with relatively low discordance) and some fragmentation.
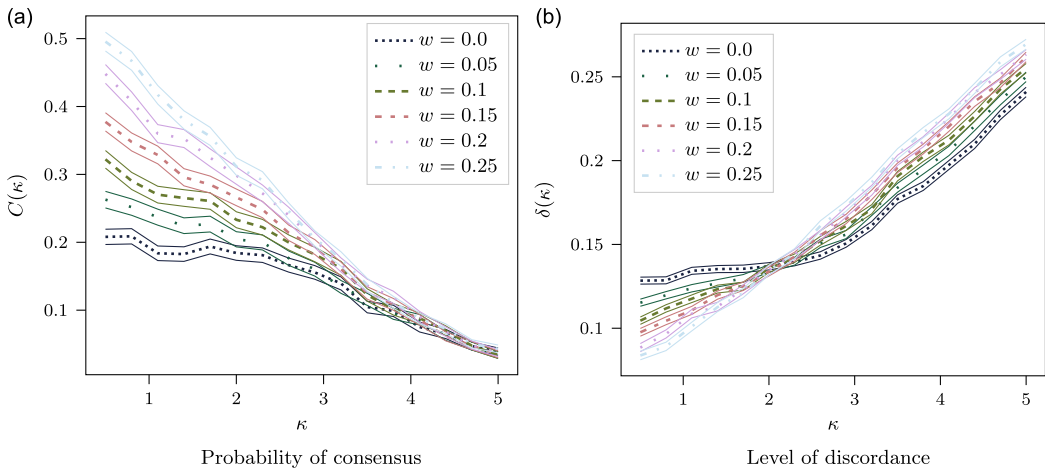
Figure 6. Results of the sensitivity analysis inspecting the probability of rewiring. Parameters: $N = 20$, $d = 4$, $t_s = 5$, $\alpha = 4$, $\beta = 2$, and $\theta_0 = \theta_1 = 0.6$.
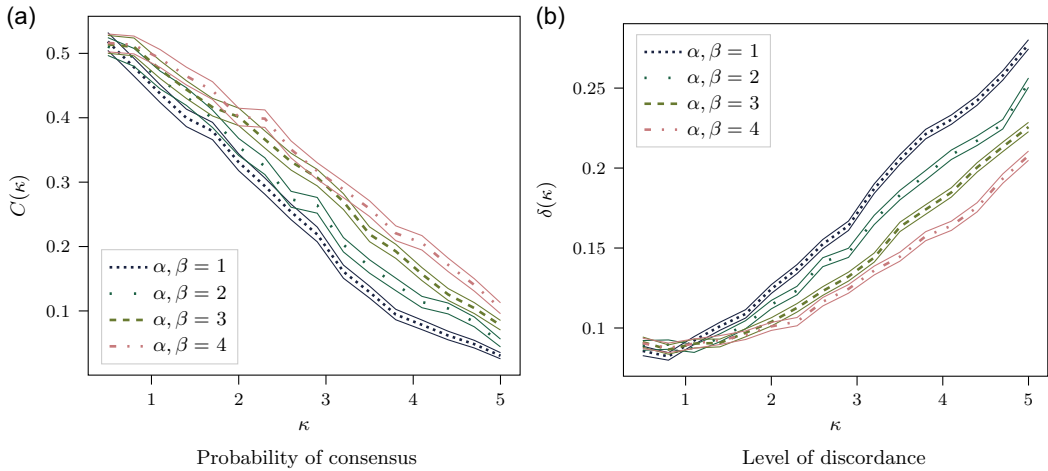
5.3 Prior belief

The parameters α and β of the beta-distributed belief are a measure of optimism in the agents in which a larger ratio α/β indicating greater optimism. To identify the effect of this prior belief distribution, we choose values $(\alpha, \beta) \in \{(4, 4), (4, 2), (3, 3), (3, 2), (2, 2), (2, 1), (1, 1)\}$. This takes into consideration that $\alpha \geq \beta$ must hold. This is required to ensure that $\widehat{\theta_0} > \theta_{\text{crit}} = 0.5$ and the agents do not switch immediately away from an opinion they have just switched to. First, we consider those prior beliefs in which $\alpha = \beta$. Thereafter, we consider the prior belief combinations in which $\alpha > \beta$.

5.3.1 Priors with $\alpha = \beta$

In these cases, switching early is quite likely as the initial estimate is on the cusp of the critical level. The general trend we observe in Figure 7a is that the lower α and β lead to lower probability of consensus than higher vales of α and β. This is explained by the fact that greater values of α and β mean that the change in their estimate from one round to the next is comparatively small at the start of a run with an opinion. As an illustration, consider an agent with 5 affirming experiences during the warm-up. If this agent has the $\alpha = \beta = 1$ prior, their estimate at the end of the warm-up is $\widehat{\theta} = 0.86$, which may allow them to retain this opinion in the face of a disagreeing neighborhood. If instead this agent had the $\alpha = \beta = 4$ prior, their estimate at the end of the warm-up would be $\widehat{\theta} = 0.69$. This lower estimate can withstand less disagreement and so it makes sense that greater initial values of $\alpha = \beta$ lead to more consensus. We see the opposite effect on the level of discordance in Figure 7b.

5.3.2 Priors with $\alpha > \beta$

The case with $\alpha > \beta$ exemplifies a greater optimism of an agent in their opinion at the start of a run. The greater level of optimism is represented by a ratio of 2 to 1 in the settings of $(\alpha, \beta) \in \{(2, 1), (4, 2)\}$. The remaining settings $(\alpha, \beta) \in \{(3, 2), (4, 3)\}$ also exhibit optimism but to a lesser extent. In Figures 8a and 8b, respectively, we see that the more optimistic settings lead to less consensus and more discordance than the somewhat less optimistic settings. Furthermore, we see that the more optimistic settings (both having a ratio of 2:1) do not differ greatly from one another. The slight difference we do see is that there is less consensus and more discordance when $(\alpha, \beta) = (4, 2)$ than $(2, 1)$. This is because the greater values of α and β mean that the belief

Figure 7. Results of the sensitivity analysis inspecting the prior belief distribution of the agents with $\alpha = \beta$. Parameters: $N = 20, d = 4, w = 0.2, t_s = 5$, and $\theta_0 = \theta_1 = 0.6$.
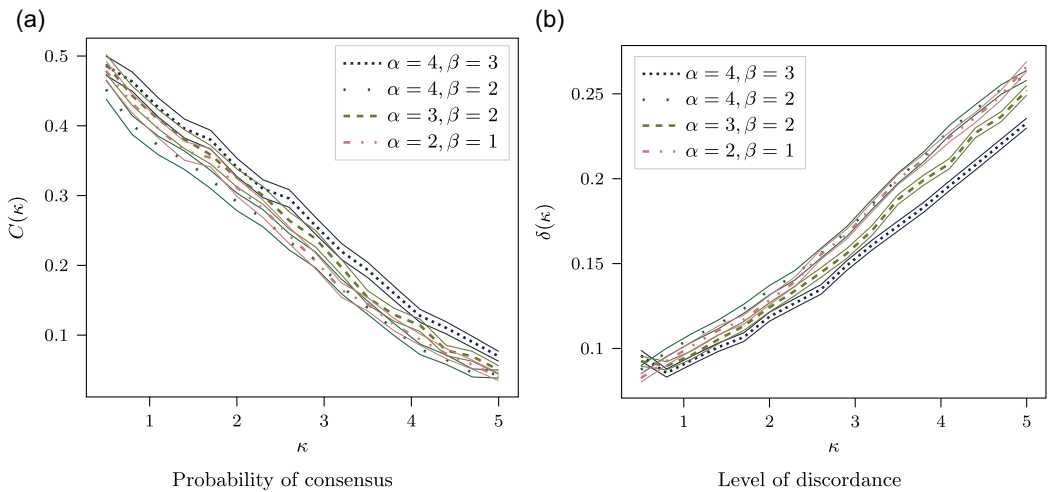
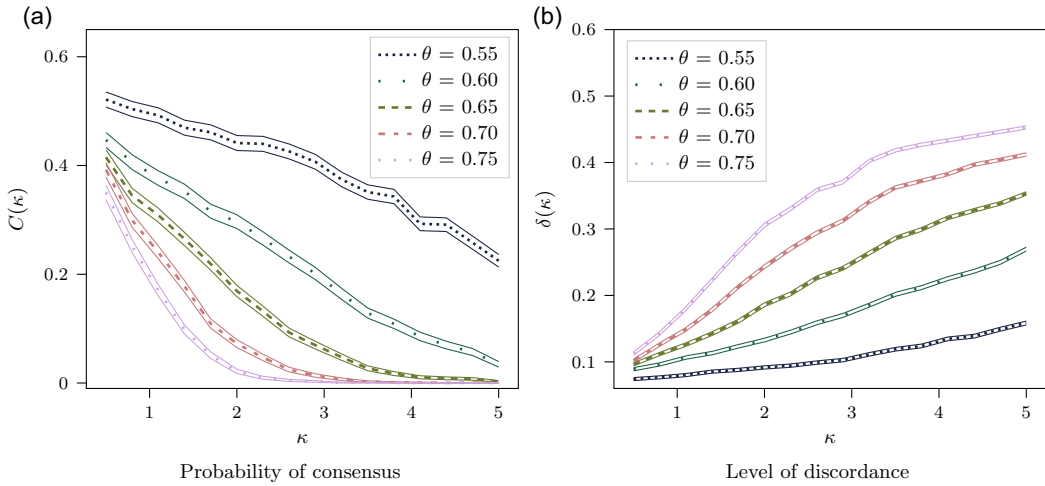**Figure 8.** Results of the sensitivity analysis inspecting the prior belief distribution of the agents with $\alpha > \beta$. Parameters: $N = 20, d = 4, w = 0.2, t_s = 5$, and $\theta_0 = \theta_1 = 0.6$.

estimate changes less in the first couple of rounds, and so there is a greater chance of staying with the starting opinion. Within the less optimistic settings, we note that the slightly more optimistic of the two ($\alpha = 3, \beta = 2$) leads to less consensus and more discordance than ($\alpha = 4, \beta = 3$). In summary, as optimism increases, the probability of consensus decreases and the proportion of discordance increases.

5.4 Opinion reliability

To elucidate the effect of the level of reliability of the opinions, we choose $\theta_0 = \theta_1 \in \{0.55, 0.60, 0.65, 0.70, 0.75\}$. We plot the probability of consensus for these settings in Figure 9a and the level of discordance in Figure 9b. We see that a greater reliability of both opinions lead to

Figure 9. Results of the sensitivity analysis inspecting the effect of the reliability of the opinions $\theta_0 = \theta_1$. Parameters: $N = 20$, $d = 4$, $w = 0.2$, $t_s = 5$, $\alpha = 4$, and $\beta = 2$.
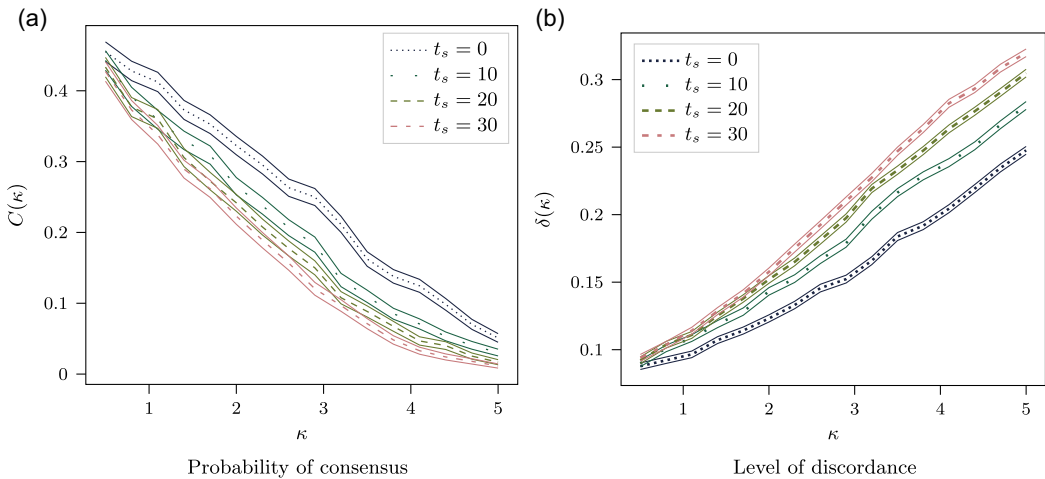
Figure 10. Results of the sensitivity analysis inspecting the effect of the warm-up period. Parameters: $N = 20$, $d = 4$, $w = 0.2$, $\alpha = 4$, $\beta = 2$, and $\theta_0 = \theta_1 = 0.6$.

less consensus and more discordance. This is conceivable as when the reliability is greater, convergence of their point estimate of the reliability to this true parameter allows for neighborhoods with more disagreement (and thus greater $\theta^*_{\text{crit},t}$). As the reliability increases, agents are less dependent on their network to agree with them (thus decreasing $\theta^*_{\text{crit},t}$) in order for their point estimate of an opinion's reliability to converge.

5.5 Warm-up length

We vary the warm-up length t_s taking values $t_s \in \{0, 10, 20, 30\}$. In Figure 10a, which plots the probability of consensus for differing warm-up lengths, we see that a shorter warm-up period leads to more consensus. Similarly in Figure 10b, we see that longer warm-up periods lead to more discordance. This result is conceivable as in the early stages of an interaction with an opinion, the

estimated reliability is changing a lot more than toward the end. In other words, having a longer warm-up period allows for an agent's belief distribution to "settle" before having to compete with a network adjusted threshold $\theta^*_{\text{crit},t}$.

5.6 Discussion

We highlight the fact that in the model, the agents are exposed only to attractive forces between each other. That is we follow the assumption of assimilative social influence between agents and yet we have rich results illustrating a range of possible outcomes from polarization to consensus. We believe that especially for agent-based simulation modelers who would like to include an opinion dynamic within a greater context, this model may prove to be useful because of the diversity of its outcomes which interact in a conceivable way with the parameter settings. We acknowledge that stubbornness does facilitate the reaching of polarization and fragmentation states. We highlight that a high stubbornness is not a requirement for avoiding consensus which is clear from the simulation results showing that even when $\kappa = 0.5$, the greatest probability of consensus observed in Figure 5a is at $C(\kappa) \approx 0.85$. We thus hypothesise the origin of the disagreement between neighboring agents in steady states to be rooted in the differences in agent experiences (and interpretations thereof) and not in the incorporation of a stubbornness parameter.

6. Experiments on framework instance

The sensitivity analysis conducted in §5 showcases that the model is capable of a variety of end states and that these interact with the model parameters in a logical way. A strength of agent-based modeling is describing micro-behavior rules of agents resulting from the combination of their characteristics and their environment (possibly interactions between agents) and subsequently observing the resulting macro behavior of the population. The strength of our model then is the possibility of modeling agents with different parameters; prior belief distributions, stubbornness parameters, or interpretation of agreement and disagreement with an opinion (p_j and l_j for $j \in V$). This section is devoted to the results of two experiments in which we introduce heterogeneity to the model. We do this by drawing the stubbornness parameter for different agents from a distribution in one experiment and defining the reliability of one opinion to be greater than the other in another experiment.

As a result of the sensitivity analysis, a suitable set of parameters were chosen as the baseline for the experiments presented in §6. The definition of suitable parameters we use are those which present a richness in the type of results that may be obtained in the steady state. We chose a population of $N = 30$ agents, connected to their $d = 6$ nearest neighbors with a rewiring probability of $w = 0.2$. As prior belief parameters, we chose $\alpha = 4$ and $\beta = 2$. This was done in order to keep frivolous switching back and forth between opinions to a minimum, which is more likely when the agent's prior estimate is closer to typical threshold values. The opinion's reliability is kept at $\theta = 0.6$ as this setting provides a relatively high probability of consensus at low κ and reaches a minimum toward the end of our chosen range at $\kappa = 5$. The warm-up period is set to 10 rounds, which we believe to strike a good balance between allowing the agents' belief to settle without blocking dynamics completely.

6.1 Heterogeneous agent stubbornness

The value of κ plays an important role in determining how sensitive the agents are to the opinions held by their neighbors. We conducted simulation runs in which the stubbornness of each agent was drawn from a Gaussian distribution centered on $\mu \in \{1.5, 2.5, 3.5, 4.5\}$. The standard deviation of this distribution was varied taking values $\sigma \in \{0.5, 1.0, 1.5\}$. In fact, we use the truncated
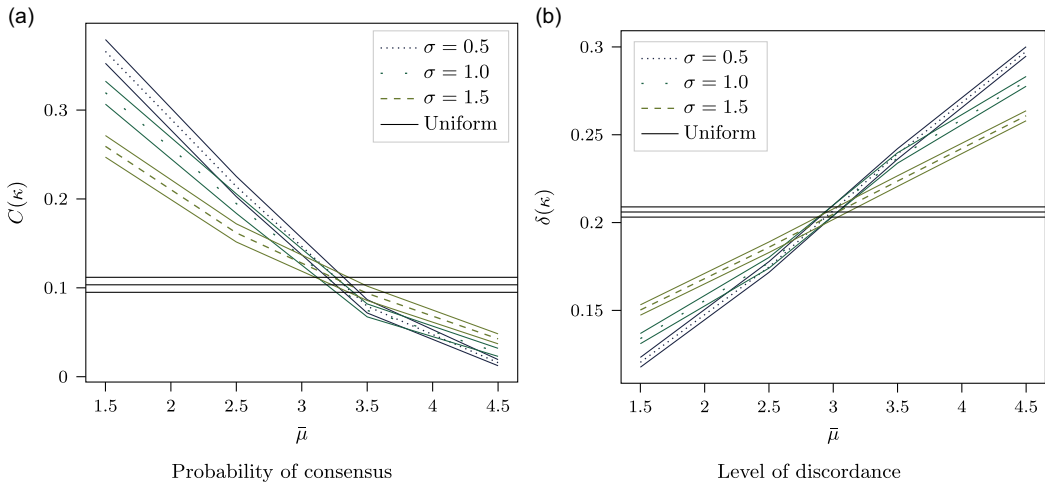
Figure 11. Results of experiment with agent stubbornness drawn from a Gaussian distribution with mean depicted on the horizontal axis and standard deviation shown in the legend. Parameters: $N = 30$, $d = 6$, $w = 0.2$, $t_s = 10$, $\alpha = 4$, $\beta = 2$, and $\theta_0 = \theta_1 = 0.6$.

Gaussian distribution on the support $\{0.5, 5.5\}$. We present the results from a set of simulation runs in which the stubbornness parameter for each agent was taken from the uniform distribution $\mathcal{U}_{[0.5,5.5]}$. The resulting probability of consensus is depicted in Figure 11a, and the proportion of discordance is depicted in Figure 11b.

In Figure 11a and Figure 11b, we see that simulations with greater σ behave more like the simulation in which the stubbornness is uniformly distributed than the other simulation runs. This means that when μ is relatively low, there is more discordance and less consensus for these runs compared to runs with lower σ. The opposite holds for greater μ. In words: A population which has, in general, a lower stubbornness, greater variability between agents helps diversify opinions. Conversely, in population with, in general, a greater stubbornness (stronger sense of individuality), greater variability between agents hinders diversity of opinions. This showcases a subtle (though possibly expected) paradox: In populations that are in general individualistic (greater μ), a greater diversity of agent characteristics (greater σ) results in lower diversity of agent opinion. Reason for this is that with a greater σ there are more agents who also have a lower stubbornness. The agents with a lower stubbornness drive the system to more agreement. Thus, an increase in diversity of agent *stubbornness* causes a decrease of diversity in agent *opinions*.

6.2 Opinions with different reliability

This experiment bears the flavor of models of learning in populations. The agents in question are given a homogeneous κ, but the true reliability of the opinions is set unequal: $\theta_0 > \theta_1$. Specifically, we simulated the pairs $(\theta_0, \theta_1) \in \{(0.65, 0.60), (0.70, 0.60), (0.75, 0.60)\}$, which showcase a growing difference between the more reliable opinion and the other. We also experiment on the effect of a constant difference in the reliability between two opinion's reliabilities by the pairs $(\theta_0, \theta_1) \in \{(0.65, 0.60), (0.70, 0.65), (0.75, 0.70)\}$, which highlight the effect of a greater general reliability while keeping the nominal difference between the two opinion's reliability constant.

6.2.1 Growing difference

We depict the probability of consensus in the growing difference experiment in Figure 12a. The corresponding portion of discordance is depicted in Figure 12b. In Figure 12a, we see that a greater
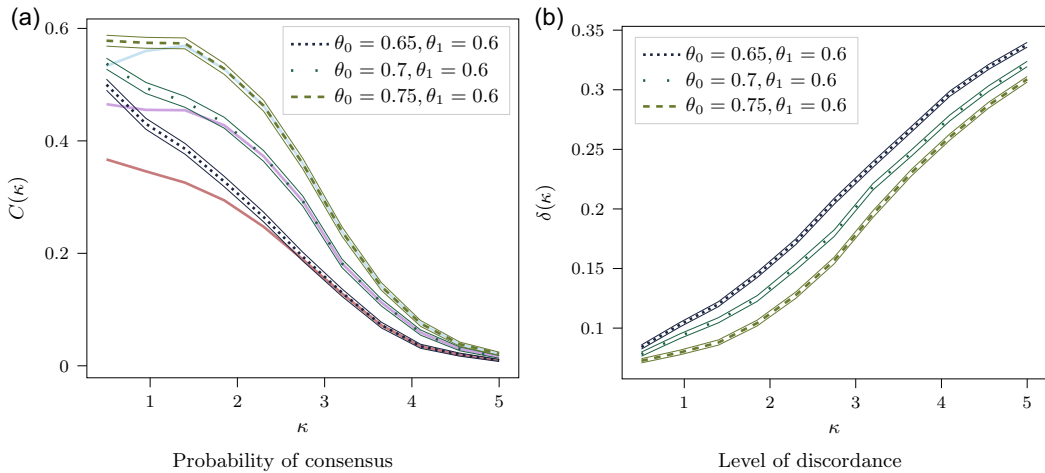
Figure 12. Results of the experiment in which the difference between θ_0 and θ_1 is growing. Additionally plotted in solid lines is the probability of consensus on opinion 0 keeping in mind that $\theta_0 > \theta_1$. As κ increases the solid lines join their simulation's counterpart: $\theta_0 = 0.65$, $\theta_1 = 0.6$ in blue, $\theta_0 = 0.7$, $\theta_1 = 0.6$ in purple, and $\theta_0 = 0.75$, $\theta_1 = 0.6$ in red. Parameters: $N = 30$, $d = 6$, $w = 0.2$, $t_s = 10$, $\alpha = 4$, and $\beta = 2$.

difference in the opinion's reliability fosters a greater probability of consensus. We also note that as the stubbornness of the population grows, the less likely it becomes that the population reaches consensus on the "inferior" opinion. This becomes so extreme that if κ is great enough, if there is consensus in the population, this is on the opinion with the greater reliability. Similarly in Figure 12b, we see that a greater difference in reliability implies a lower portion of discordance in expectation.
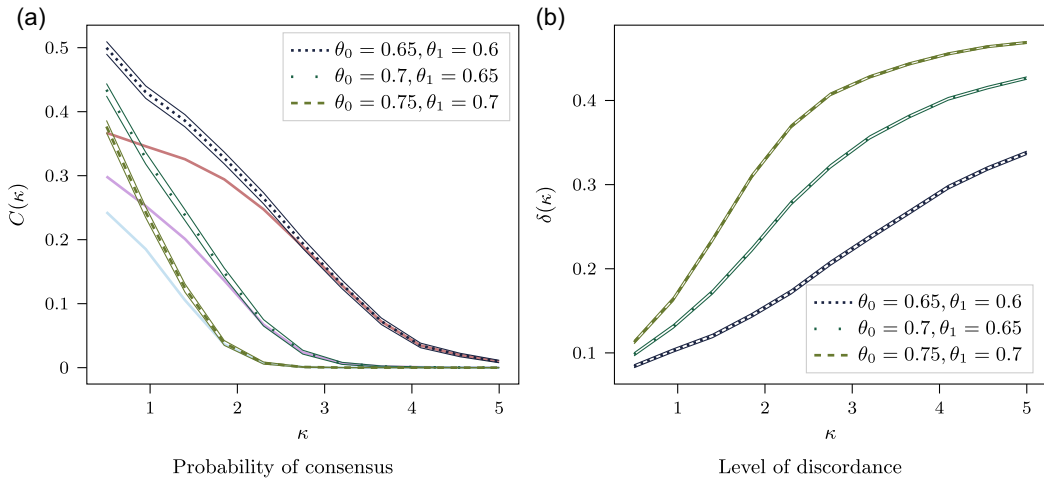
6.2.2 Constant difference
We plot the probability of consensus in the subexperiment with constant difference between θ_0 and θ_1 in Figure 13a. We plot the corresponding portion of discordance in Figure 13b. We see a similar trend in Figure 13a to the one present in the sensitivity analysis: Lower reliability leads to more consensus of opinion. As in the experiment with a growing difference between opinion reliability, we see that greater stubbornness in the population leads to a greater chance of agreeing on the "better" opinion. In Figure 13b, we see again (as in the sensitivity analysis) that a greater reliability leads to more discordance.

7. Discussion

In this section, we discuss the results of the experiments conducted. In doing so, we also reflect on the merits of the model when interpreted as a heuristic for communication interpretation between agents. Subsequently, we discuss differences between our framework and relevant literature.

7.1 Interpretation of the experiments
As a result of the experiment with different κ per agent, we see that an increase in heterogeneity (greater standard deviation of the distribution from which we sample the stubbornness κ) decreases the differences which arise from shifting the mean μ of the distribution. From a modeler's perspective this may be intuitive, as a greater spread in the distribution should decrease the

Figure 13. Results of the experiment in which the difference between θ_0 and θ_1 is constant. Additionally plotted in solid lines is the probability of consensus on opinion 0 keeping in mind that $\theta_0 > \theta_1$. As κ increases the solid lines join their simulation's counterpart: $\theta_0 = 0.65, \theta_1 = 0.6$ in red, $\theta_0 = 0.7, \theta_1 = 0.65$ in purple, and $\theta_0 = 0.75, \theta_1 = 0.7$ in blue. Parameters: $N = 30$, $d = 6$, $w = 0.2$, $t_s = 10$, $\alpha = 4$, and $\beta = 2$.

effect of shifting its mean. It does, however, also hint to the important difference between individualism and diversity. In this context: Populations with lower individualism tend toward more consensus. Furthermore, populations with diversity in the extent of its agent's individualism may increase or decrease the probability of consensus depending on the mean value of the population's individualism.

The experiment with opinions of different reliability show us that populations with greater stubbornness may be more sure that if they reach consensus, it is upon the better alternative. Furthermore, the greater the difference between two opinions, the less discordance one expects in the population. The more clear-cut the difference between two opinions, the easier it should be for the population to learn this and subsequently reach consensus on the better of the two. It should be said here that it is also true that enough stubbornness leads to general disagreement in the population. Thus, balance may be important to the goal of reaching consensus on the better of two opinions. In general, it seems that the agents in the model make good use of the information provided by their network: Consensus on the better opinion is more likely than on the worse opinion, and this is increasingly the case the greater the difference between the two opinions. Though agents are not modeled 'rationally' this outcome suggests that the heuristic method by which agents incorporate their neighbors opinions does aid the agents in making good decisions. We note that the model captures situations in which because agents are in agreement with one another, they do not critically assess their decision. This is underlined by the fact that when stubbornness is low enough, consensus may be reached upon the 'inferior' opinion. We believe this results from the modeling decision for agents to be influenced by the opinion expressed rather than the belief held by their neighbors.

7.2 Contributions and future work

The framework we present in Sections 2 and 3 addresses the current lack of models in the opinion dynamics literature which have sophisticated agents who may a) adjust their opinion in absence of network influence beyond the introduction of noise and b) retain their opinion despite network influence. This is achieved by following social psychological theory that attitudes are driven by

experience (Gerard and Orive, 1987; Fazio et al., 2004), which was also eluded to by Giardini, et al. (2015). Furthermore, it entails models from the same basic assumptions of assimilative forces between agents yet with a novel aspect: Opinions of alters do not affect an agent's strength of conviction (belief) of an opinion directly but rather the decision- making process by which an agent chooses their opinion. This is an attempt towards modeling interacting agents as opposed to what Giardini et al. (2015) call interacting opinions. Furthermore, the framework for social influence we present generalizes the celebrated majority rule dynamics (Galam, 2002). Our model thus also presents a possible explanation as to what internal process may result in dynamics that resemble majority rules. The framework also is computationally light despite the relatively high level of detail of the agents. This enables the modeling of agents with reasonable sophistication in general models (not focused exclusively on opinion dynamics) using this framework.

The results of the framework instance and experiments we present in Sections 4, and 6 highlight that models from the framework have desirable and reasonable characteristics: An array of outcomes is possible entailing consensus, polarization as well as fragmentation, all without the need for repulsive forces between agents of different opinions. The current framework shows that fragmentation is possible without explicitly modeling *selection* dynamics (see for instance Kempe et al. (2016)), which describe the tendency of individuals to interact with others who are similar to them. An interesting line of future work is thus to incorporate evolution of the network structure alongside the evolution of the opinions using the principle of selection.

The definition of the opinions used in our framework is broad and allows for interesting future work in which agent behavior may be coupled back to the reliability of an opinion. For example consider a population of agents who are faced with the choice of a means of transportation. The agent's belief on the reliability of the options available is likely to play a role in their decision making. Closing the feedback loop: The agent decision making (the number of people using each type) is likely to influence the reliability of the options available. The fact that our model is lightweight means that it may be straightforwardly implemented in agent-based models which investigate more than opinions but rather the interface between opinion dynamics and their effect on agent behavior.

With careful adjustment, the framework we present can also be applied to the diffusion of innovation. Martins, et al. (2009) draw the connection between the diffusion of innovation and opinion dynamics by modeling agents who learn the quality of a new product through their social interactions. Fu and Riche (2021) present a theoretical model for the adoption of a new technology within a market. Their model includes learning of the quality of the new technology, yet does not include the effect of network ties. Valente (1996) and Iyengar et al. (2011) empirically investigate the adoption of innovation and show that social influence plays an important role. In particular, Valente (1996) does so using a (social network) threshold based decision-making model. We could interpret such adoption behavior through our framework: two technologies (opinions) of differing quality (reliability) in the context of a network in which the older technology is dominant at initialization. Due to a small group of early adopters using the new technology as well as dissatisfaction with the older technology, agents start adopting (and learning about) the newer technology. An extension to the current framework is to incorporate the effects of marketing and other media on the agents which, as pointed out by Van den Bulte and Lillien (2003), are important in this context.

Competing interests. None.

Data availability. Data and code available on request to corresponding author.

Notes

1 While they do distinguish between opinions and beliefs, they use direct influence between the elements which constitute the belief.

2 Baccelli et al. (2017) do not distinguish between beliefs and opinions and use the word "opinion." We use "belief" in discussing their paper because this aligns with our definitions.

3 Yildiz et al. (2013) studied an extension of the voter model including stubborn agents who were unable to adjust their opinion. We consider our stubbornness parameter a generalization as an agent might be anywhere between the two extremes: completely unaffected by their neighborhood or completely susceptible to the majority opinion in their neighborhood.

4 Majority rules models have received attention in their own right by Mossel et al. (2014); Tamuz and Tessler (2015); Benjamini et al. (2016) and more recently Nguyen et al. (2020).

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. *Human Behavior and Emerging Technologies*, *2*(4), 314–324.

Altafini, C. (2013). Consensus problems on networks with antagonistic interactions. *IEEE Transactions on Automatic Control*, *58*(4), 935–946.

Artinger, F. M., Gigerenzer, G., & Jacobs, P. (2022). Satisficing: integrating two traditions. *Journal of Economic Literature*, *60*(2), 598–635.

Baccelli, F., Chatterjee, A., & Vishwanath, S. (2017). Pairwise stochastic bounded confidence opinion dynamics: Heavy tails and stability. *IEEE Transactions on Automatic Control*, *62*(11), 5678–5693.

Balke, T., & Gilbert, N. (2014). How do agents make decisions? A survey. *Journal of Artificial Societies and Social Simulation*, *14*(4).

Benjamini, I., Chan, S.-O., O'Donnell, R., Tamuz, O., & Tan, L.-Y. (2016). Convergence, unanimity and disagreement in majority dynamics on unimodular graphs and random graphs. *Stochastic Processes and their Applications*, *126*(9), 2719–27333.

Bernardo, C., Altafini, C., Proskurnikov, A., & Vasca, F. (2024). Bounded confidence opinion dynamics: a survey. *Automatica*, *159*(11), 111302.

Castellano, C., Muñoz, M. A., & Pastor-Satorras, R. (2009). Nonlinear *q*-voter model. *Physical Review E*, *80*, article no. 041129.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Review of Modern Physics*, *81*(2), 591–646.

Chan, K. M. D., Duivenvoorden, R., Flache, A., & Mandjes, M. (2024). A relative approach to opinion formation. *The Journal of Mathematical Sociology*, *48*(1), 1–41.

Clifford, P., & Sudbury, A. (1973). A model for spatial conflict. *Biometrika*, *60*(3), 581–588.

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

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology*, *87*(3), 293–311.

Fishbein, M., & Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research. *Philosophy and Rhetoric*, *10*(2), 130–132.

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S. , & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, *20*(4).

French, J. R. (1956). A formal theory of social power. *Psychological Review*, *63*(3), 181–194.

Fu, W., & Riche, A. L. (2021). Endogenous growth model with Bayesian learning and technology selection. *Mathematical Social Sciences*, *114*, 58–71.

Galam, S. (2002). Minority opinion spreading in random geometry. *The European Physical Journal B*, *25*(4), 403–406.

Gerard, A. B., & Orive, R. (1987). The dynamics of opinion formation. *Advances in Experimental Social Psychology*, *20*, 171–202.

Giardini, F., Vilone, D., & Conte, R. (2015). Consensus emerging from the bottom-up: The role of cognitive variables in opinion dynamics. *Frontiers in Physics*, *3*.

Harary, F. (1959). A criterion for unanimity in French's theory of social power. In: *Studies in social power* (Cartwright, D., ed.)168–182. Ann Arbor (MI): Institute for Social Research.

Holley, R. A., & Ligget, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability*, *3*(4), 643–663.

Iyengar, R., Van den Bulte, C., & Valente, T. W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science*, *30*(2), 195–212.

Jager, W., van Asselt, M. B. A., Boodt, B. C., Rotmans, J., & Vlek, C. A. J. (1995). *Consumer behaviour, a modelling perspective in the context of integrated assessment of global change*. In *Rijksinstituut voor Volksgezondheid en Milieu RIVM*

Jager, W., Janssen, M. A., & Vlek, C. A. J. (1999). Consumats in a commons dilemma - testing the behavioural rules of simulated consumers, Rijksuniversiteit Groningen, Tech. Rep. COV 99-01.

Janssen, M. A., & Jager, W. (1999). An integrated approach to simulating behavioural processes: A case study of the lock-in of consumption patterns. *Journal of Artificial Societies and Social Simulation*, *2*(2).

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationship. *Journal of Business Research*, *58*(4), 500–507.

Kelman, H. C. (1961). Processes of opinion change. *Public Opinion Quarterly*, *25*(1), 57–78.

Kempe, D., Kleinberg, J., Oren, S., & Slivkins, A. (2016). Selection and influence in cultural dynamics. *Network Science*, *4*(1), 1–27.

Liu, J., Chen, X., Başar, T., & Belabbas, M. A. (2017). Exponential convergence of the discrete- and continuous-time Altafini models. *IEEE Transactions on Automatic Control*, *62*(12), 6168–6182.

Martins, A. C., de B. Pereira, C., & Vicente, R. (2009). An opinion dynamics model for the diffusion of innovations. *Physica A: Statistical Mechanics and its Applications*, *388*(15), 3225–3232.

Meylahn, B. V., den Boer, A. V., & Mandjes, M. R. H. (2024). Trusting: Alone and together. *The Journal of Mathematical Sociology*, *48*(4), 424–478.

Mossel, A., Neeman, J., & Tamuz, O. (2014). Majority dynamics and aggregation of information in social networks. *Autonomous Agents and Multi-Agent Systems*, *28*(3), 408–429.

Newman, B. I. (2002). The role of marketing in politics. *Journal of Political Marketing*, *1*(1), 1–5.

Nguyen, V. X., Xiao, G., Xu, X.-J., Wu, Q., & Xia, C.-Y. (2020). Dynamics of opinion formation under majority rules on complex social networks. *Scientific Reports*, *10*(1), 456.

Noorazar, H., Vixie, K. R., Talebanpour, A., & Hu, Y. (2020). From classical to modern opinion dynamics. *International Journal of Modern Physics C*, *31*(7), 2050101.

Ozdemir, S., Zhang, S., Gupta, S., & Bebek, G. (2020). The effects of trust and peer influence on corporate brand–consumer relationship and consumer loyalty. *Journal of Business Research*, *117*, 791–805.

Proskurnikov, A. V., & Tempo, R. (2017). A tutorial on modeling and analysis of dynamic social networks. Part I. *Annual Reviews in Control*, *43*, 65–79.

Proskurnikov, A. V., & Tempo, R. (2018). A tutorial on modeling and analysis of dynamic social networks. Part II. *Annual Reviews in Control*, *45*, 166–190.

Proskurnikov, A. V., Matveev, A. S., & Cao, M. (2016). Opinion dynamics in social networks with hostile camps: Consensus vs. polarization. *IEEE Transactions on Automatic Control*, *61*(6), 1524–1536.

Redner, S. (2019). Reality-inspired voter models: A mini-review. *Comptes Rendus. Physique*, *20*(4), 275–292.

Robins, G., Pattison, P., & Elliott, P. (2001). Network models for social influence processes. *Psychometrika*, *66*(2), 161–189.

Simon, H. A. (1956). Rational choice and the strucuture of the environment. *Psychological Review*, *63*(2), 129–138.

Tamuz, O., & Tessler, R. J. (2015). Majority dynamics and the retention of information. *Israel Journal of Mathematics*, *206*(1), 483–507.

Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, *18*(1), 69–89.

Van den Bulte, C., & Lillien, G. L. (2003). Medical innocation revisited: Social contagion versus marketing effort. *Americal Journal of Sociology*, *106*(5), 1409–1435.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, *393*(6684), 440–442.

Weber, C. (2019). Ideology and values in political decision making. In *Oxford Research Encyclopedia of Politics*. Oxford University Press.

Yildiz, E., Ozdaglar, A., Acemoglu, D., Saberi, A., & Scaglione, A. (2013). Binary opinion dynamics with stubborn agents. *ACM Transactions on Economics and Computation*, *1*(4), 30–30.

Zha, Q., Kou, G., Zhang, H., Liang, H., Chen, X., Li, C.-C., & Dong, Y. (2021). Opinion dynamics in finance and business: a literature review and research opportunities. *Financial Innovation*, *6*(1), 1–22.

Appendix A. Watts–Strogatz network

The agents in our example are embedded within a Watts–Strogatz random graph model Watts and Strogatz ([1998](#)). The creation of a Watts–Strogatz random graph is illustrated in three steps. This uses, $N \in \mathbb{N}$ the number of agents in the population, $d \in 2\mathbb{N}$ the initial number of nearest neighbors to each agent, and $w \in (0, 1)$ the rewiring probability.

1. First, we arrange the population of N agents, on a cycle graph and connected each agent to their d-nearest neighbors.

2. Second, for each edge in the circulant created, we flip a coin that lands heads with probability w and if it lands heads, we "cut" the edge off of one of its vertices.

3. Finally, each of the edges cut in this way is rewired to another vertex uniformly at random.
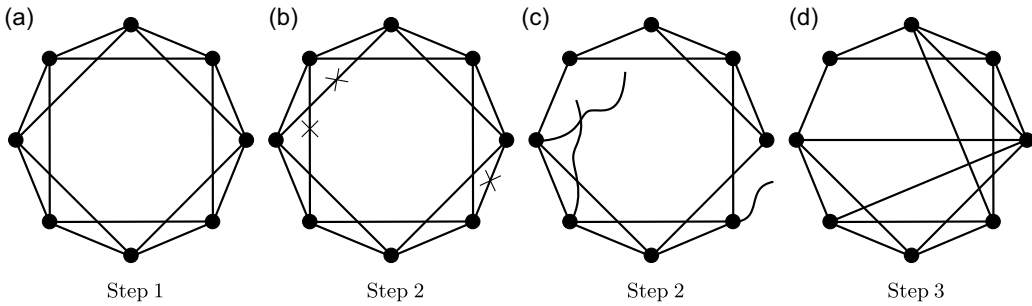
Figure 14. The steps to create a Watts–Strogatz random graph on $N = 8$ agents with $d = 4$ nearest neighbors.

This network structure has the property that average path lengths between vertices are short, yet there is still high clustering of vertices. We illustrate this process in Figure 14 for a network on $N = 8$ agents, with $d = 4$ nearest neighbors.