

# How can beliefs alter opinions? Joint opinion and belief evolution

Hiro Kataoka<sup>1</sup>, Jérôme Euzenat<sup>2</sup> and Koji Hasebe<sup>1</sup>

<sup>1</sup>*Department of Computer Science, University of Tsukuba, Tsukuba, Japan*

<sup>2</sup>*Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, F-38000 Grenoble, France*  
*kataoka@mas.cs.tsukuba.ac.jp, Jerome.Euzenat@inria.fr, hasebe@cs.tsukuba.ac.jp*

**Keywords:** Social networks, Multi-agent simulation, Opinion dynamics, Belief merge, Belief propagation

**Abstract:** Multi-agent opinion dynamics and belief propagation have been studied independently. However, people beliefs may influence their opinions and beliefs should not be too dissonant with opinions. This paper considers how to combine opinions and beliefs so that they preserve the cognitive coherence of agents. For that purpose, it integrates social opinion and belief propagation using classical procedures and two new specific operations guided by values held by agents. In order to assess the effect of this model, we experiment with such agents, varying the processing workflow, graph topology, initial beliefs and opinions, and values. Results show that, by maintaining the coherence between beliefs and opinions of individual agents, the social beliefs and opinions resulting from their propagation are indeed affected. In particular, contrary to what happens with the classical opinion dynamics and belief propagation procedures, connecting opinions and beliefs makes them not necessarily converge and not even stabilize. This means that the model supports agents having polarized opinions and beliefs. The outcome of the propagation depends on the workflow, topics, values, and initial graph topology, beliefs and opinions.

## 1 INTRODUCTION

The words ‘beliefs’ and ‘opinions’ are often used interchangeably (Friedkin et al., 2016; Tang et al., 2021; Wen et al., 2025). This does not raise issues when talking about beliefs only or opinions only. This is to the point that (Schwind et al., 2016) has ‘opinion’ in the title and no occurrence in the text, though (Defuant et al., 2000) has ‘belief’ in the title and a single occurrence, in an example, in the first paragraph of the text. The former is about beliefs as considered here and the latter about opinions. However, beliefs and opinions are different and not making the distinction becomes confusing when both are taken into account. An agent’s opinions reflect its preference towards hypothetical statements, though beliefs represent what it holds as true. These do not necessarily have to coincide: an agent’s beliefs may be related to the current state of the world, e.g., that no ‘tram line is built’, and its opinion may be positive towards ‘building a tram line’. Beliefs can be seen as a logical theory, though opinions are often expressed as the scoring of statements on a scale. Statements in logical theories are interdependent and may be requested to be non contradictory, though opinions are often considered inde-

pendently.

Both opinions and beliefs can spread within a society of agents.

The dynamics of agents’ opinion propagation is studied in the area of opinion dynamics from the standpoint of social influence (Sîrbu et al., 2017), i.e., how the opinions of others influence an agent’s opinions. Agents’ opinions are represented by a numeric value representing the agreement or disagreement of the agent with a statement.

There has also been work on modeling how agents exchange knowledge and beliefs and update theirs with those that they receive (Schwind et al., 2015). While opinion dynamics mostly rely on social influence among neighbors, knowledge exchange takes into account the logical aspect of knowledge and, in particular, the willingness of agents to ensure the consistency of their beliefs. In consequence, this work is often based on belief revision or belief merge (Fermé and Hansson, 2018).

However, in spite of their differences, beliefs and opinions should not be regarded as independent. Opinions are at least partially grounded on beliefs and beliefs may be affected by opinions (Oktar and Lombrozo, 2025). Moreover, people may care about

maintaining some coherence between their beliefs and opinions.

We aim at exploring the mutual influence of opinions and beliefs. For that purpose, we provide a model in which agents maintain both opinions and beliefs and update one from the other through belief alignment and opinion formation operations relying on cultural values inducing preferences between the possible states of the world.

The questions that this approach poses concern the actual effect of connecting beliefs and opinions. In particular, to which extent this connection can alter the standard behavior of opinion dynamics and belief revision. We consider this from the standpoint of the stabilization and convergence of the processes. In particular, given a model that makes agent opinion converge, does connecting it to beliefs can make it non converging (polarized)? Conversely, given a model in which beliefs do not converge, can connecting them to opinions make them converge? The goal of opinion-belief connection is to keep these coherent but the consequence of this coherence on social propagation remains unclear.

To answer these questions, we run simulations varying the parameters affecting these procedures: the network topology, the agent initial opinions and beliefs, and the values that agents use for connecting beliefs and opinions.

We observe that, connecting beliefs and opinions drastically impact the results. In particular, opinions which always converge with some models (DeGroot, 1974) may be polarized when the same model is connected to beliefs. Although beliefs could diverge, they diverge more when connected to opinions, i.e., in a situation in which beliefs and opinions converge, connecting them may lead to both diverging. Beside process convergence and stabilization, the connection affects the agents' opinions and beliefs in a variety of ways more complex to analyze than their independent propagation.

The rest of this paper is organized as follows. After introducing some preliminary notation (Section 2), the paper briefly presents previous work on opinion dynamics and belief propagation as well as the problem of their connection (Section 3). It then introduces a model to organize this connection through opinion formation and belief alignment (Section 4). This model is used to perform experiments whose setup is presented (Section 5) and whose results are discussed (Section 6) before concluding (Section 7).

## 2 PRELIMINARIES

In order to unify the notations, we define those which are used throughout the paper.

### 2.1 Agents and network

We consider a set  $A$  of agents, such that each agent  $a \in A$  has a set  $N_a$  of neighbors (or acquaintances).  $N_a$  may also be specified as a weighting function  $N_a : A \rightarrow [0, 1]$  denoting the influence of neighbors on  $a$ . In this case,  $a' \in N_a$  is interpreted as  $N_a(a') > 0$ . This defines a directed graph  $G = \langle A, N \rangle$  such that  $N = \{ \langle a, a' \rangle ; a \in A, a' \in N_a \}$ . Consider three colleagues Alice, Bob and Carol, so that Alice and Bob and Alice and Carol trust each other ( $N_{\text{Alice}}(x) = 1$  and  $N_x(\text{Alice}) = 1$ ) but Carol does not trust Bob and Bob is defiant of Carol ( $N_{\text{Carol}}(\text{Bob}) = 0$  and  $N_{\text{Bob}}(\text{Carol}) = 0.5$ ).

Each agent  $a \in A$  has a cognitive state defined as  $\langle O'_a, B'_a \rangle$ , i.e., the pair of its opinions  $O'_a$  and beliefs  $B'_a$  at time  $t$  ( $t$  is omitted if it is clear from the context).

### 2.2 Beliefs

Agent  $a$ 's beliefs consist in a set of formulas held true by the agent. Here we consider agents' beliefs as theories in a logic language  $\mathcal{L}$ , usually with a model-theoretic semantics ( $\mathcal{M}(B)$  denotes the set of models of beliefs  $B \subseteq \mathcal{L}$ ). We express them here in propositional logic in the language  $\mathcal{L}_P$ , built from a set  $P$  of propositions as signature. Such propositions may be that 'a tram be built' (noted as  $p$ ), 'traffic jams are low' ( $q$ ) or 'taxes are low' ( $r$ ). For instance, if Alice believes that trams reduce traffic jams and increase taxes, her beliefs may be  $p \rightarrow (q \wedge \neg r)$ . More precisely, this set of beliefs is expressed, without loss of generality, as a single formula  $\phi$  in full disjunctive normal form (FDNF), i.e., each literal appearing exactly once in each disjunct. Hence, Alice's beliefs are expressed as  $(\neg p \wedge \neg q \wedge r) \vee (\neg p \wedge q \wedge r) \vee (\neg p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge q \wedge \neg r) \vee (p \wedge q \wedge \neg r)$ . In such a case, there is a one-to-one correspondence between the disjuncts of the formula and the interpretations it has as models  $\mathcal{M}(\phi)$ . The symbol  $\top$  (top) denotes the formula satisfied by all interpretations (or the empty theory), thus  $\mathcal{M}(\top)$  is the set of all possible interpretations. Conversely, to a set of models  $M$  corresponds a single FDNF formula  $[M]$ .

### 2.3 Opinions

Given a set  $T$  of topics also expressed as propositional formulas ( $T \subseteq \mathcal{L}_P$ ), agent  $a$  holds opinions  $O'_a$

over these topics. An opinion toward a topic  $\phi \in T$  scores, on a numeric scale, how an agent is for or against  $\phi$ . For instance, Bob and Alice may be in favor that ‘a tram be built’, though Carol would prefer not and that ‘Roads be larger’. An agent’s opinions map topics to real numbers ( $O_a^t : T \mapsto \mathbb{R}$ ). This map will be equivalently expressed as its graph on  $T \times \mathbb{R}$ :  $\{\langle \phi, n \rangle; O_a^t(\phi) = n\}$ .

### 3 STATE OF THE ART

We briefly review how the social propagation of opinions (Section 3.1) and beliefs (Section 3.2) has been considered independently, before discussing attempts at combining them (Section 3.3).

#### 3.1 Opinion dynamics

Opinion dynamics studies how agents’ opinions evolve through interactions between agents (Sîrbu et al., 2017; Peralta et al., 2022).

In opinion dynamics, agents exchange their opinions with that of their neighbors and compute a new opinion by aggregating them. For instance, this may be performed by computing:

$$O_a^{t+1}(\phi) = \frac{\sum_{a' \in A} N_a(a') \times O_{a'}^t(\phi)}{\sum_{a' \in A} N_a(a')}. \quad (1)$$

If the weights of  $N_a$  are required to sum up to 1, then the adjacency matrix of  $G$  is stochastic and this equation corresponds to the DeGroot model (DeGroot, 1974). Agents may update their opinions asynchronously, i.e., one agent after the other (Berenbrink et al., 2022), or synchronously, i.e., all agents at the same time, as in the DeGroot model. In this model, when there exists an agent  $a$  that is reachable by all other agents in  $G$ , opinions towards a topic converge to a single value.

Well-known classical opinion dynamics models are based on weighted average (DeGroot, 1974), bounded confidence (Deffuant et al., 2000; Hegselmann and Krause, 2002) and votes (Clifford and Sudbury, 1973; Holley and Liggett, 1975), but more models are continuously proposed (Dong et al., 2018; Noorazar, 2020; Hassani et al., 2022; Bernardo et al., 2024).

#### 3.2 Belief propagation

Other work has made agents learn beliefs from the environment (Tang et al., 2021; Bourahla et al., 2021). This anchors their beliefs. Then, they exchange their beliefs either explicitly (Tang et al., 2021) or through

performing common tasks (Bourahla et al., 2021). Agents then adapt their beliefs, in function of the other beliefs and experience based on direct measures of accuracy or less direct measures of success.

Agents may also propagate beliefs socially by exchanging their beliefs with their neighbors and update their own beliefs accordingly. Like in opinion dynamics, this process can be synchronous. For instance, Belief Revision Games (BRG) (Schwind et al., 2015) consider a set of agents holding propositional beliefs. At each turn, each agent updates its beliefs with that of its neighbors ( $N_a$ ) with a belief merge operator ( $R$ ):

$$B_a^{t+1} = R(\langle B_{a'}^t \rangle_{a' \in N_a}).$$

In principle, this can be performed with different revision and/or merging operators (Fermé and Hansson, 2018). As a distance-based additive merging operator, we will use the operation  $R$  defined as <sup>1</sup>:

$$R(\mathbf{B}) = \left[ \operatorname{argmin}_{m \in \mathcal{M}(T)} \sum_{B \in \mathbf{B}} \min_{m' \in \mathcal{M}(B)} d^H(m, m') \right]. \quad (2)$$

such that  $\mathbf{B}$  is a set of beliefs to be merged. (Schwind et al., 2015) shows what properties (AGM-style postulates) are needed, for each specific operator, to achieve consistency preservation, unanimity preservation, agreement preservation, monotony, convergence, stability, etc. The  $R$  operator above reaches stability when the network is acyclic (what is called convergence in (Schwind et al., 2015) corresponds to stability in this paper).

Belief Flow Networks (BFN) (Schwind et al., 2024) have recently been introduced to perform iterative belief revision. This means that agents update an epistemic state comprising beliefs and change conditions. In BFN, agents interact asynchronously, i.e., one of them communicates its beliefs to another which updates its own beliefs using an improvement operator, instead of merging the beliefs of all neighbors as in BRG. In this framework, if the network they form is strongly connected, then agents reach a consensus (Schwind et al., 2024).

#### 3.3 Connection between opinions and beliefs

In spite of socio-psychological evidence of their relation (Oktar and Lombrozo, 2025; Wen et al., 2025), there has been few models explicitly connecting changing opinions and beliefs as defined in opinion dynamics and belief propagation. Some models constrain Boolean public opinions with common

<sup>1</sup> $d^H$  is the Hamming distance between two interpretations, i.e.,  $d^H(m, m') = |\{p \in P; (m \models p \wedge m' \not\models p) \vee (m \not\models p \wedge m' \models p)\}|$ .

logically expressed integrity constraints (Botan et al., 2019). Other work uses an influence matrix between opinions in order to enforce their dependencies (Friedkin et al., 2016; Nedić et al., 2019). However, these constraints do not have the status of beliefs and do not change.

In (Tang et al., 2021), the distinction is made between beliefs and experience. Experience is modeled as data incompletely sampling the environment. From samples, agents learn beliefs, under the form of a classifier based on feature correlation. Then they explicitly exchange both data and beliefs with their neighbors according to different modalities: they exchange one piece of sample that they add to their own sample, they exchange their classifier and if it is more accurate than their own classifier on their own sample, then they adopt it. This last procedure is called ‘epistemic rationality’, meaning that agents try to maintain as much coherence as possible between their beliefs and their experience. It is shown theoretically that agents converge towards the same beliefs but that this can be suboptimal in terms of accuracy. Although, this work does not deal with opinions, only belief and data, it provides the notion of internal coherence of agents’ beliefs and experience.

Recent work uses an altered confidence-bounded opinion propagation models to mimic online social network behavior (Sasahara et al., 2020) and belief revision games. It shows that the connection between the two processes leads to increase the number of echo chambers (Kataoka et al., 2025). In this model, the echo chambers are formed by the agents actively modifying their network based on a notion of tolerance.

Here we consider a fixed network allowing us to reuse independently proposed opinion dynamics and belief propagation mechanisms. We aim at assessing that connecting both may affect the dynamic properties of belief and opinion propagation, and, specifically their stabilization and convergence. For that purpose, we introduce the following model.

## 4 A BELIEF-OPINION SYNCHRONIZATION AND PROPAGATION MODEL

We introduce below a model for connecting opinions and beliefs in the spirit of (Kataoka et al., 2025). It is based on a set of operations affecting agent states (Section 4.1) which are split between social operations dealing with relations with neighbors (Section 4.2) and cognitive operations dealing with coher-

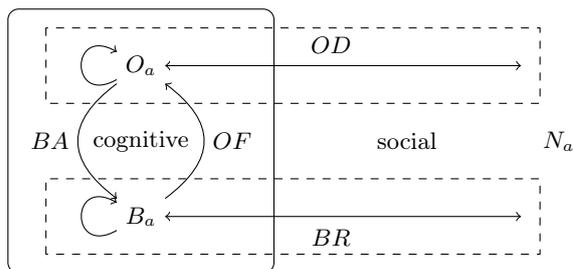


Figure 1: The four beliefs-opinions update operations.

ence across agents’ beliefs and opinions (Section 4.3).

### 4.1 Operations affecting beliefs and opinions

Agents’ beliefs and opinions being defined as in Section 2, we identify processes affecting them in two groups:

- Externally, the role of social processes, such as opinion dynamics and belief propagation, is that agents maintain a form of social coherence across their beliefs and opinions with respect to those of their neighbors.
- Internally, the role of cognitive processes to be introduced below is that agents maintain a form of cognitive coherence, or reduce the cognitive dissonance (Festinger, 1957), between their own beliefs and opinions.

Following the usual opinion dynamic and belief revision game protocols, these processes are performed in a synchronous way (all agents perform them at the same time). In this paper, processes are expressed as functions which modify one component of agents’ states.

Two social processes, opinion dynamics (*OD*) and belief revision (*BR*) modify respectively opinions and beliefs (Section 4.2). In principle, these two operations are independent and can be very dissimilar, e.g., numerical vs. logical. Connecting them requires the ability to align beliefs with opinions and opinions to beliefs. For that purpose, we introduce two cognitive operations (Section 4.3):

- Opinion Formation (*OF*) which generates opinions corresponding to beliefs, and
- Belief Alignment (*BA*) which generates beliefs supporting opinions.

Figure 1 presents the four operations and how they affect the opinions and beliefs of an agent.

The four procedures  $OD$ ,  $BR$ ,  $OF$ , and  $BA$  may be scheduled in various ways. In this paper, they are expressed as workflows, using a sequence operator (expressed with function composition  $\circ$ ). A typical workflow is  $BA \circ OD \circ OF \circ BR$  (called  $\text{bfirst}$  below), which performs belief propagation, then opinion formation from the resulting beliefs, before opinion dynamics propagates opinions among agents and the resulting opinions are used to update the beliefs. This workflow is performed repetitively.

## 4.2 Social processes

The first two processes  $OD$  and  $BR$  correspond to existing opinion dynamics models and belief revision frameworks since they update agents' opinions and beliefs based on their neighbors', respectively. Hence, we simply use existing models to define them.

For  $OD$ , we use the DeGroot model as expressed in Equation 1. For  $BR$ , we use the belief merge operator described in (Schwind et al., 2015) and more specifically  $R$  (Equation 2). In both cases, the agents own opinions may be taken into account when aggregating neighbors opinions and beliefs according to the network. However, these have no specific status with respect to the other neighbors.

In summary, the two already available operations are:

$$BR(\langle O_a, B_a \rangle) = \langle O_a, R(\langle B_{a'} \rangle_{a' \in N_a}) \rangle, \quad (\text{BR})$$

$$OD(\langle O_a, B_a \rangle) = \left\langle \left\{ \left\langle \phi, \frac{\sum_{a' \in N_a} N_a(a') \times O_{a'}(\phi)}{\sum_{a' \in N_a} N_a(a')} \right\rangle \right\}_{\phi \in T}, B_a \right\rangle. \quad (\text{OD})$$

## 4.3 Cognitive processes

It remains to explain how to connect beliefs and opinions. In order to implement opinion formation and belief alignment, agents rely on cultural values expressing their preferences between interpretations.

### 4.3.1 Cultural values

In order to maintain the coherence (or 'epistemic rationality' (Tang et al., 2021)) in agent behavior, it seems preferable that belief alignment and opinion formation follow the same principles. This enforces the coherence between opinions and beliefs. Such principles should be of higher order than opinions and beliefs.

A good candidate for driving these processes are values (or cultural values). Values are deep beliefs that help agents to recognize what is 'good' or 'bad' (Schwartz, 1999; Vanhée, 2015; Mercur

Table 1: Valuations of the interpretations induced by three different cultural value sets.

interpretations			$V_1$	$V_2$	$V_3$
$p$	$q$	$r$	$\text{rqp}$	$\text{p}\bar{q}\bar{r}$	$\text{pqr}$
t	t	t	7/7	1/7	7/7
t	t	f	6/7	5/7	3/7
t	f	t	5/7	3/7	5/7
t	f	f	4/7	7/7	1/7
f	t	t	3/7	0/7	6/7
f	t	f	2/7	4/7	2/7
f	f	t	1/7	2/7	4/7
f	f	f	0/7	6/7	0/7

et al., 2019). They may be shared among a population, but they never change (or at least they change more slowly than opinions, beliefs and knowledge). In particular, agents use their cultural values to determine their preferences between alternative states of the world.

Values are rendered as a map  $V$  from interpretations in  $\mathcal{M}(T)$  to real numbers between 0 and 1. Intuitively, the cultural values toward an interpretation  $I$  is close to 0 and 1, if the agent considers  $I$  as bad and good, respectively. This entails a relation  $\prec$  between interpretations defined from  $V$  as  $I \prec I'$  if  $V(I) < V(I')$ . For example, Table 1 represents values  $V_1$  and  $V_3$  which consider as good, or preferable, worlds consistent with  $p$  ('tram is built'),  $q$  ('traffic jams are low'), and  $r$  ('taxes are low'). The difference between  $V_1$  and  $V_3$  is that if one of the three propositions has to not be satisfied, then  $V_1$  prefers that it be  $r$  and  $V_3$  prefers that it be  $p$ .  $V_2$  in turn considers traffic jam ( $q$ ) and tax ( $r$ ) reduction as bad (or not desirable) but tram building ( $p$ ) as good.

Here, the function  $V$  between interpretations is directly derived from a strict preference on the importance of propositions in  $P$  (denoted by the order of position of propositions in  $\text{p}\bar{q}\bar{r}$ ) and their strict preferred truth values  $\text{p}$  or  $\bar{\text{p}}$ . A function  $V$  defined in this way is thus always injective.

In the following, the same values are shared among all agents and do not change over time. These values are involved in opinion formation and belief alignments as discussed below.

### 4.3.2 Opinion formation

In order to form their opinions, i.e., to decide which opinion to assign to a topic, agent  $a$  considers:

- Its beliefs ( $B_a$ ) which determine the truth of formulas;
- Its current opinions ( $O_a$ ) which it may want to

maintain or change minimally;

- Its (cultural) values ( $V$ ) which evaluate the desirability that the statement be true with respect to these values.

For instance, consider that Bob

- believes that ‘if tram lines are built, then there will be less traffic jams’, his beliefs are  $p \rightarrow q$  represented as  $B_{\text{Bob}} (= (p \wedge q \wedge r) \vee (p \wedge q \wedge \neg r) \vee (\neg p \wedge q \wedge r) \vee (\neg p \wedge \neg q \wedge r) \vee (\neg p \wedge q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge \neg r))$ ;
- has neutral opinions toward building tram lines (i.e., the topic  $\phi \equiv p$ ):  $O_{\text{Bob}}(\phi) = \frac{1}{2}$ ;
- holds the values  $V_1$  in Table 1.

If an agent  $a$  is asked its opinion about a statement  $\phi$ , it will answer how preferable  $\phi$  would be to itself if  $\phi$  happens in the worlds it believes that it is in, i.e.,  $\mathcal{M}(B_a)$ . Thus,  $a$ 's opinion about  $\phi$  is regarded as the overall evaluation of the preference of each model in  $\mathcal{M}(B_a)$  such that  $\phi$  can happen, i.e.,  $\mathcal{M}(B_a \wedge \phi)$ . However,  $\phi$  may be inconsistent with  $B_a$  and yet be desirable, i.e., the agent wishes it would be in another situation. For that reason we use the revision of  $B_a$  by  $\phi$ . The value for  $a$  of statement  $\phi$  with respect to beliefs  $B$  (denoted by  $v(B, \phi)$ ) is defined as:

$$v(B, \phi) = \frac{\sum_{m \in \mathcal{M}(R(\langle B, \phi \rangle))} V(m)}{|\mathcal{M}(R(\langle B, \phi \rangle))|} \quad (3)$$

For coherence's sake, the same merging operator  $R$  as in  $BR$  is used. Note that the divisor is always positive because  $R$  always returns consistent beliefs.

For example, Bob performs opinion formation by considering  $R(\langle B_{\text{Bob}}, \phi \rangle) = p \wedge q$ , corresponding to the FDNF formula or  $(p \wedge q \wedge r) \vee (p \wedge q \wedge \neg r)$ , and how good (or bad) it is. Then,  $v(B_{\text{Bob}}, \phi) = \frac{1}{2} \times (\frac{7}{7} + \frac{6}{7}) = \frac{13}{14}$ .

Thus  $V$  allows an agent  $a$  to determine which opinion to hold with respect to its beliefs through  $v(B_a, \phi)$  (Equation 3). However,  $a$  may show inertia in changing its opinions. Hence,  $v$  is balanced with its current opinions  $O_a$  through a parameter  $\alpha$ .

The opinion formation operation is defined as:

$$OF(\langle O_a, B_a \rangle) = \langle \{ \langle \phi, \alpha \times O_a(\phi) + (1 - \alpha) \times v(B_a, \phi) \rangle \}_{\phi \in T}, B_a \rangle \quad (\text{OF})$$

with  $\alpha \in [0, 1]$  expressing the inertia of the agent, or its unwillingness to change opinions from fully ignoring neighbor beliefs ( $\alpha = 1$ ) to not take its previous opinions into account ( $\alpha = 0$ ). The former corresponds to having no opinion formation ( $O_a(\phi)$  does not change). The latter maximizes adherence to beliefs ( $O_a(\phi) = v(B_a, \phi)$ ).

For instance, assuming  $\alpha = 0.5$ , Bob's aligned opinion toward  $\phi$  is  $\frac{1}{2} \times \frac{1}{2} + \frac{13}{14} \times \frac{1}{2} = \frac{5}{7}$ . This means that his beliefs drive Bob to consider the topic of building a tram line ( $\phi$ ) as more desirable than before. However, if Bob had the values  $V_2$  in the table,  $v(B_{\text{Bob}}, \phi) = \frac{1}{2} \times (\frac{1}{7} + \frac{5}{7}) = \frac{3}{7}$ , hence his aligned opinion toward  $\phi$  is  $\frac{1}{2} \times \frac{1}{2} + \frac{3}{7} \times \frac{1}{2} = \frac{13}{28}$ . This means that he retains his neutral opinion toward the topic.

### 4.3.3 Belief alignment

The opinions that an agent holds due to its social interactions may have an influence on this agent's beliefs. This may be due to coherence constraints which minimize the difference between the opinions that the beliefs support and the opinions actually held.

Belief alignment consists in adjusting their beliefs to their opinions. It is more difficult than opinion formation because it goes from valuations to theories.

The preference order  $\prec$  can be extended into a strict total order  $\ll$  between beliefs in full disjunctive normal form by considering the interpretations along that order (preferred first) and stating that a belief is preferred to another if it satisfies the first interpretation for which their satisfaction differs:

$$B \ll B' \text{ iff } \exists m \in \mathcal{M}(T); \begin{cases} m(B') = t \text{ and} \\ m(B) = f \text{ and} \\ \forall m' \succ m, m'(B) = m'(B') \end{cases}$$

such that  $m(B) = t$  if  $m \models B$  and  $m(B) = f$  otherwise. If there do not exist two equivalent formulas, like in FDNF, then  $\ll$  is a strict total order.

To express the quantity that belief alignment has to reduce, we introduce a distance  $\Delta_a(B, T)$  between the current opinions and how beliefs support them:

$$\Delta_a(B, T) = \sum_{\phi \in T} |v(B, \phi) - O_a(\phi)|$$

Belief alignment selects the preferred interpretation among those that minimize  $\Delta_a$ :

$$BA(\langle O_a, B_a \rangle) = \langle O_a, \max_{\ll} \text{argmin}_{B \in \text{FDNF}^*(\mathcal{L}_P)} \Delta_a(B, T) \rangle \quad (\text{BA})$$

such that  $\text{FDNF}^*(\mathcal{L}_P)$  is the set of consistent FDNF formulas (excluding the empty disjunction). This yields a single formula, in case there are several beliefs at the minimal distance, because  $\ll$  is total.

Again, let us consider that Bob has the same beliefs  $B_{\text{Bob}} \equiv p \rightarrow q$  and the same values  $V_1$  as before and has opinion  $O_{\text{Bob}}(p) = 0.45$ . We obtain

$$\text{argmin}_{B \in \text{FDNF}^*(\mathcal{L}_P)} \Delta_a(B, \{p\}) = \{ \neg p \wedge \neg q \wedge r, \neg p \wedge \neg r, (\neg p \wedge q \wedge \neg r) \vee (\neg p \wedge \neg q) \}. \quad (4)$$

Among these beliefs, the last one is the maximal one with respect to  $\ll$ . Hence, the aligned beliefs are  $(\neg p \wedge q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge r) \vee (\neg p \wedge \neg q \wedge \neg r)$  in FDNF. On the other hand, if Bob hold values  $V_2$  from Table 1, we obtain

$$\operatorname{argmin}_{B \in \text{FDNF}^*(\mathcal{L}_p)} \Delta_a(B, \{p\}) = \{(\neg p \wedge q \wedge r) \vee (\neg p \wedge \neg q)\},$$

hence the aligned beliefs are  $(\neg p \wedge q \wedge r) \vee (\neg p \wedge \neg q \wedge r) \vee (\neg p \wedge \neg q \wedge \neg r)$  in FDNF. In both cases, the slightly negative opinion over building tram lines, has driven Bob to adopt beliefs that deny the building of tram lines.

## 5 EXPERIMENTAL SETTING

In order to measure the effect that the provided connection has on the resulting opinions and beliefs of agents, we run experiments with the following setting.

### 5.1 Experiment plan

In the reported experiments, the following parameters are frozen:

**number of agents:**  $|A| = 100$ .

**number of edges:**  $|N| = 400$ .

**signature:**  $P = \{p, q, r\}$ .

**number of iterations:** 1000.

There are five controlled parameters:

**topics** Two topics are used to take into account interdependency:  $\text{nooverlap} = \{p, \neg p\}$  and  $\text{overlap} = \{p, q\}$ . The former topics have no common model, hence should be separated; the latter on the contrary shows two compatible topics.

**inertia** considers intermediate values  $\alpha \in \{0.25, 0.5, 0.75\}$  to span its range without considering the extremes.

**values**  $V_1, V_2$  and  $V_3$  as presented in Table 1.

**workflows** compare  $\text{indep} = OD \circ BR$ ,  $\text{ofirst} = OF \circ BR \circ BA \circ OD$ , and  $\text{bfirst} = BA \circ OD \circ OF \circ BR$ . The former does not trigger the beliefs-opinions connection; the remainder are the same process starting with different operations.

**initial graph, beliefs and opinions** are randomly determined by 20 different seeds.

This makes  $\underbrace{2}_{\text{topics}} \times \underbrace{3}_{\alpha} \times \underbrace{3}_{\text{values}} \times \underbrace{3}_{\text{workflows}} \times \underbrace{20}_{\text{seeds}} = 1080$  runs.

The network is initialized as a random graph with  $|A|$  nodes (agents) and  $|N|$  edges. For each agent  $a$ , the weights of  $N_a$  sum to 1. and are evenly distributed among connected agents. It is also checked that the graphs have the DeGroot property that there is at least one node reachable by all others. Each agent's initial opinions and beliefs are generated randomly.

Before running the experiment, beliefs and opinions are synchronized by performing  $BA \circ OF$  until the distance between opinions before the operation and after it is less than  $\epsilon$  and  $\delta$  (see below), respectively. There are two motivations for performing this: (a) experiments thus start with internally coherent states, and (b) we expect to neutralize the effect of workflows. After all,  $\text{ofirst}$  and  $\text{bfirst}$  are the same workflow except for the initial social operation.

### 5.2 Measures

While running this experiment, agents' beliefs and opinions are collected at each tick. This allows for the computation of the following measures which will be considered as dependent variables:

*Oc/Bc* The number of agents having changed their opinions/beliefs at the last tick.

*Od/Bd* The maximal distance between agents' opinions/beliefs.

*Bu* The number of unique beliefs.

We measure the distance between opinions by the Manhattan distance:

$$d_O(O, O') = \sum_{\phi \in T} |O(\phi) - O'(\phi)|.$$

Moreover, we measure the distance between beliefs by the Hamming distance between their models:

$$d_B(B, B') = |(\mathcal{M}(B) \setminus \mathcal{M}(B')) \cup (\mathcal{M}(B') \setminus \mathcal{M}(B))|.$$

From these measures it is possible to decide if the process is stabilized (when  $Oc = Bc = 0$  during the last  $\eta$  interactions) or has converged (when  $Bd \leq \delta$  and  $Od \leq \epsilon$ ). Hereafter, we always use  $\eta = 20$ ,  $\epsilon = 10^{-5}$  and  $\delta = 0$  which provide enough confidence.

## 6 RESULTS

We discuss the experimental results obtained from the standpoint of the influence of connecting beliefs and opinions on the stability and convergence of the process (Section 6.1), the kind of effects this connection may have in the resulting opinions and beliefs (Section 6.2), and the failure of the synchronization procedure to obtain results independent from whether

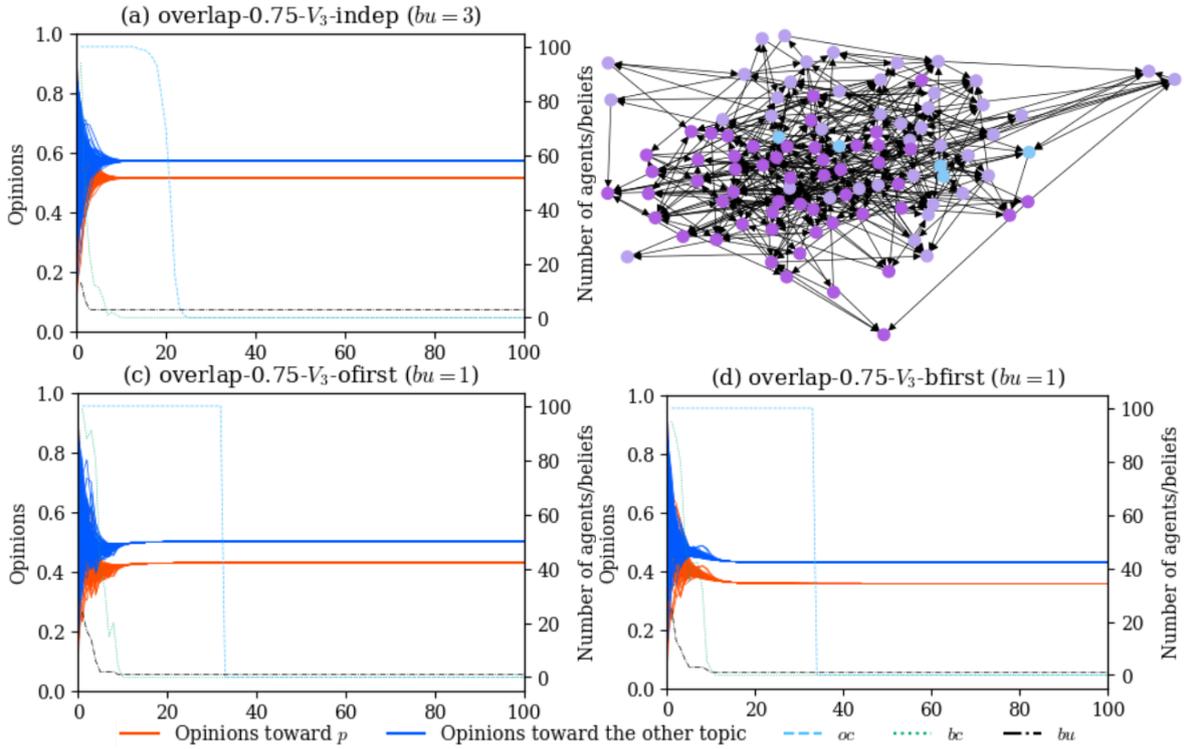


Figure 2: Three plots (seed: 94821) with (a) a non converging run with respect to beliefs with indep, (b) the beliefs of its network (different colors = different belief sets) at iteration 1000 and two fully converging runs with (c) ofirst and (d) bfirst.

agents start propagating opinions ofirst or beliefs bfirst (Section 6.3).

All the compared plots below are from runs with the same topics, the same inertia and the same initial situation. Only workflows and values differ.

## 6.1 Influence of the connection on stability and convergence

Table 2 displays the distribution of runs as converging, stabilizing or none of these with respect to the workflow used. It shows that connecting beliefs and opinions, which only occurs with ofirst and bfirst affects the convergence and stability of the process.

Concerning the indep workflow results are split among whether the property affects opinions or beliefs; for the two other workflows, the behavior of beliefs and opinions is the same in each run.

All of the runs with the workflow indep converge on opinions as predicted by the properties of the DeGroot model (DeGroot, 1974). This is different for beliefs as this property does not hold in general (Schwind et al., 2015). A very few cases (3 out of 360) do not converge however.

Table 2: Distribution of runs converging, stabilizing or none after 1000 ticks depending on the used workflow. In each column, the conditions of the leftward columns are not satisfied.

workflow	converges		stabilizes		else
	opinions $Od \leq \epsilon$	beliefs $Bd \leq \delta$	opinions $Oc = 0$	beliefs $Bc = 0$	
indep	360 (100%)	357 (99.17%)	0	3 (0.83%)	0
ofirst	349 (96.94%)		11 (3.06%)		0
bfirst	351 (97.50%)		8 (2.22%)		1 (0.28%)

If the workflow is changed from indep to either ofirst or bfirst, opinions and beliefs do not necessarily converge. This means that the connection of beliefs and opinions can alter the behavior of DeGroot opinion propagation model by introducing polarization.

In addition, connecting beliefs and opinions can have the reverse effect. Indeed, starting with one of the non converging run with indep (Figure 2(a)), which stabilizes with three different beliefs, using ofirst (Figure 2(c)) or bfirst (Figure 2(d)) will make beliefs converge.

A single run along those we performed does

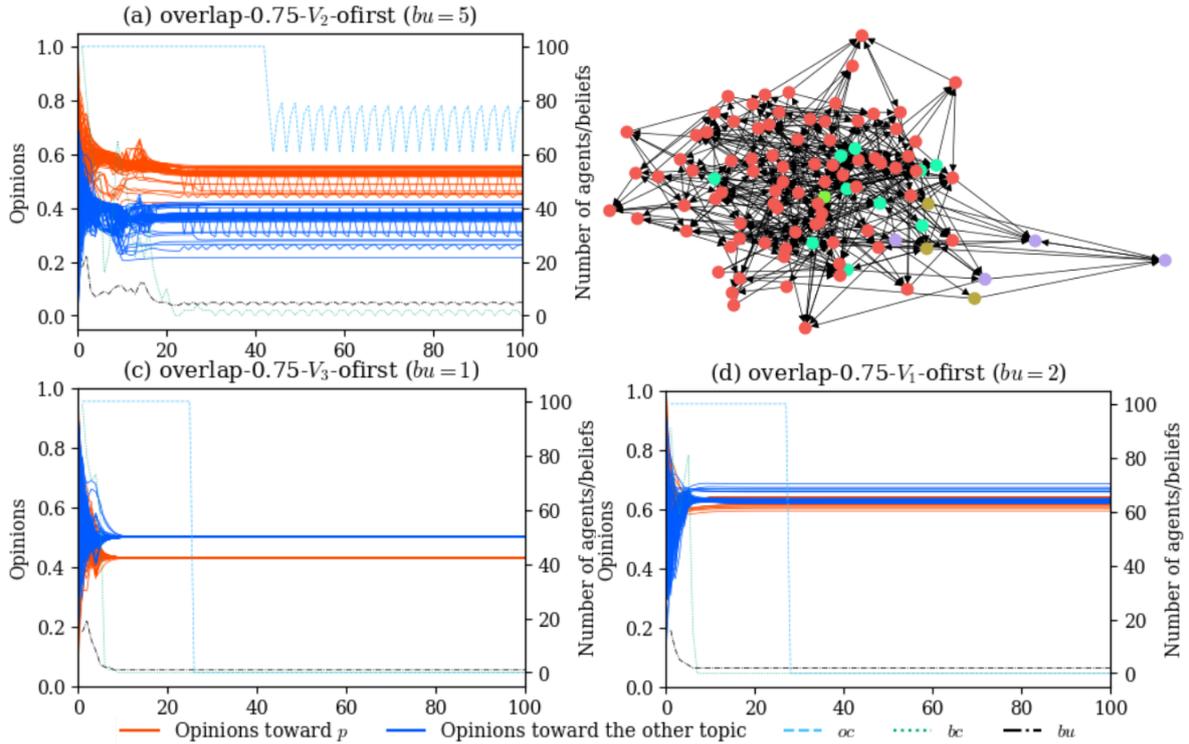


Figure 3: Three plots (seed: 909432) with (a) a non stabilizing run with  $V_2$ , (c) a non converging run with  $V_1$  and (d) a converging run with  $V_3$ . The network (b) shows the beliefs (different colors = different belief sets) with  $V_2$  at iteration 1000.

not stabilize in both opinions and beliefs (see Figure 3(a)). It oscillates (with period 3) between two belief configurations in which most agents (82 out of 100) have converged towards the beliefs that  $\neg q \vee (\neg p \wedge r) \vee (p \wedge \neg r)$ , but two remaining agents simultaneously change their beliefs. This is supported by 78 agents still modifying their opinion above the threshold. This displays a deep effect of connecting beliefs and opinions, since this behavior should happen in neither the DeGroot model nor belief revision games.

The impact of values can also be observed on Figure 3: all parameters being the same, (a)  $V_2$  renders the process non stable with 5 different beliefs, (c)  $V_1$  makes it polarized in both values and beliefs (with only two beliefs) and (d)  $V_3$  makes the process converge in both opinions and beliefs.

## 6.2 Influence of the connection on actual opinions and beliefs

The results above show that the connection of opinions and beliefs has an impact on the convergence and stabilization of the propagation process. From the data so far, this impact may seem modest (3% of the runs). Figure 2 displays changes in beliefs (from 3

different beliefs to one) but few change in opinions.

Figure 4 shows the kind of variation that can be brought by connecting opinions and beliefs (b,c,d) with respect to the non connected process (a). This variation affects the distance between the opinions on  $p$  and  $\neg p$  which are very close in (a) and further away in (c). It also affects the opinions of the agents with around 0.5 in (a) to  $p$  at 0.7 in (c) and (d), and especially  $\neg p$  at 0.8 in (d) and 0.3 in (c). Finally, and perhaps more radically, the order  $O(p) < O(\neg p)$  is preserved from (a) to (b) and (d), but reversed in (c):  $O(p) > O(\neg p)$ . It can be noted that the opinion on  $p$  is exactly the same in (c) and (d), after a different evolution in the 10 first iterations.

Concerning beliefs, each run converges, i.e., all agents adopt the same belief. However, the runs do not converge on the same beliefs: (a) settles on  $\neg p \wedge \neg q \wedge r$ , (b) reaches  $(p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge q \wedge r)$ , (c) leads to  $(p \wedge r) \vee (\neg p \wedge \neg r)$ , and (d) leads to  $\neg p \wedge \neg r$ . Hence, only (c) and (d) reach compatible beliefs (which have some models in common).

This shows the broadness of the impact that workflows and values may have on opinions and beliefs. So far, it is difficult to draw a definitive conclusion concerning, for instance, the fact that  $bfirst$  leads to

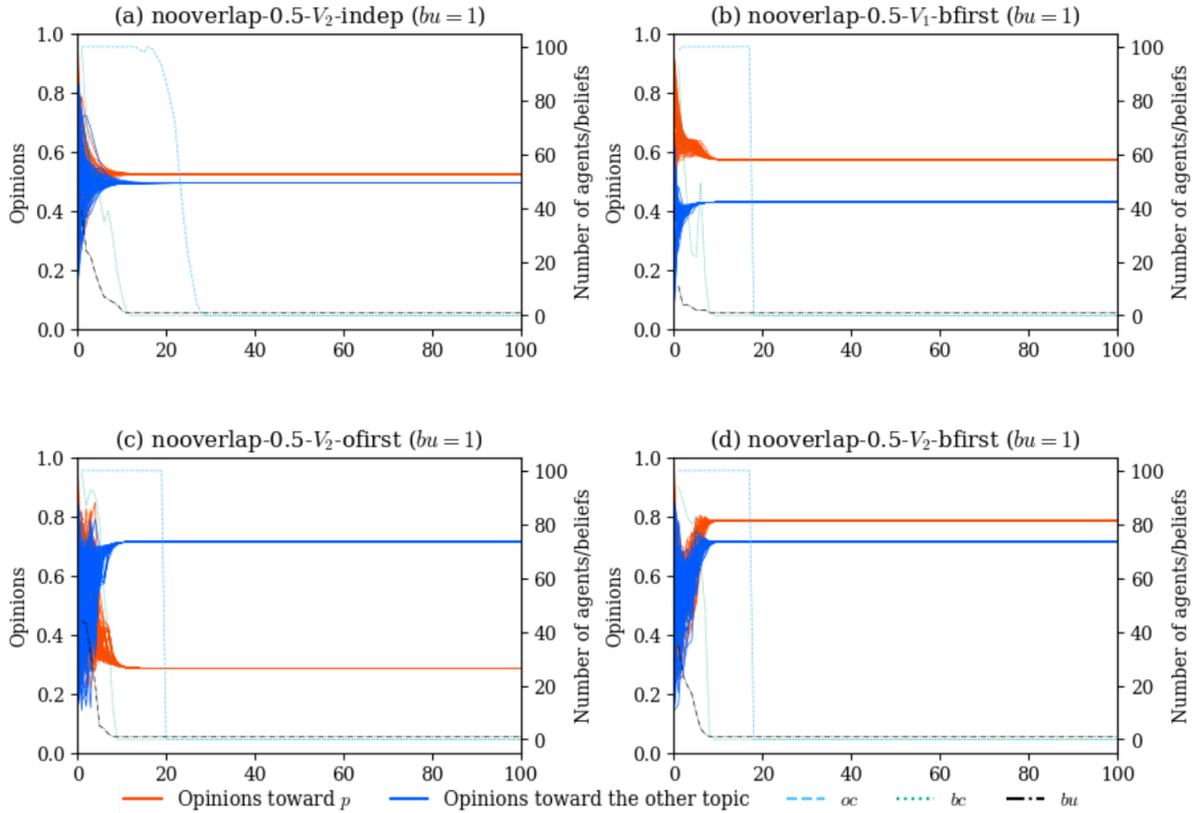


Figure 4: Four plots (seed: 501765) which show the variability of the impact of connecting beliefs and opinions generated by different workflows and values.

a larger distance between opinions. The variation is rather determined by a combination of interacting parameters, including those determining the initial situation.

### 6.3 Negative results

Contrary to expected, the initial synchronization of opinions and beliefs does not lead to the exact same results for bfirst and ofirst. Although in most cases results are very close, there are cases in which they greatly differ. This is the case of Figure 3(a): the same run with bfirst converges. What happens is that the processes start differently (due to their initial different operation) but most of the time converge towards very close values. However, in some cases this initial difference leads to divergence. This can be observed in Figure 4(c) vs. (d).

## 7 CONCLUSION

In principle, agents' beliefs should influence their opinions and vice versa. However, opinion dynamics

and belief propagation have so far been only studied in isolation. In order to model the influence they may have on each other, we used two complementary operations: opinion formation from beliefs and belief alignment with opinions.

A concrete model in which belief alignment and opinion formation rely on the same cultural values has been introduced. It aims at maintaining the internal coherence between beliefs and opinions, opinion dynamics and belief propagation being implemented by classical models.

We performed systematic experiments on a selection of parameter values to observe how they change the outcome of these processes. They showed the influence of workflow, initial opinions, graph, and values on the process stability and convergence.

We showed experimentally that connecting beliefs and opinions to ensure agent's cognitive coherence changes their social propagation. In particular, though opinion dynamics converge and belief revision games stabilize independently, this is not the case when connected: this may lead agents to stabilize with different, and even opposite, opinions and beliefs or to not stabilize at all. It seems that the complex dynam-

ics arising from the interaction between beliefs and opinions leads to maintain coherent agent groups. We also highlighted that the effects of connecting opinions and beliefs may manifest in different forms affecting the values of opinions and beliefs, their distance and even their order.

However, such effects result from the interaction between different parameters of the model. Further investigations have to be carried out to elucidate the actual influence of each parameter such as initial beliefs,  $\alpha$  and graph topology. In particular, a promising perspective is the study of which parameter configurations allow or prevent the emergence of echo chambers (Sasahara et al., 2020), the propagation of conspiracy theories (Douglas and Sutton, 2023) or the occurrence of the ‘Zollman effect’ (Zollman, 2007).

## DATA AVAILABILITY

The experiment notebooks are openly available from <https://sake.re/20250711-BROD>. They contain reference to simulator and version, processing instructions, results and jupyter notebooks analysing them.

## ETHICAL STATEMENT

The authors declare no competing interests. The authors are not aware of any ethical issues with this work.

## ACKNOWLEDGEMENTS

This work was partially supported by the Fukuro INRIA-University of Tsukuba associate team and the Echo action of the NTU-UT-UGA trilateral center.

## REFERENCES

- Berenbrink, P., Hofer, M., Kaaser, D., Lenzner, P., Rau, M., and Schmand, D. (2022). Asynchronous opinion dynamics in social networks. In *Proceedings of the 21th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 109–117, Online.
- Bernardo, C., Altafini, C., Proskurnikov, A., and Vasca, F. (2024). Bounded confidence opinion dynamics: A survey. *Automatica*, 159:111302.
- Botan, S., Grandi, U., and Perrussel, L. (2019). Multi-issue opinion diffusion under constraints. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 828–836, Montréal (CA).
- Bourahla, Y., Atencia, M., and Euzenat, J. (2021). Knowledge improvement and diversity under interaction-driven adaptation of learned ontologies. In *Proceedings of the 20th international conference on Autonomous Agents and Multi-Agent Systems (AAMAS), Online*, page 242–250.
- Clifford, P. and Sudbury, A. (1973). A model for spatial conflict. *Biometrika*, 60(3):581–588.
- Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in complex systems*, 3(1-4):87–98.
- DeGroot, M. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.
- Dong, Y., Zhan, M., Kou, G., Ding, Z., and Liang, H. (2018). A survey on the fusion process in opinion dynamics. *Information fusion*, 43:57–65.
- Douglas, K. M. and Sutton, R. M. (2023). What are conspiracy theories? A definitional approach to their correlates, consequences, and communication. *Annual review of psychology*, 74(1):271–298.
- Fermé, E. and Hansson, S. (2018). *Belief change: introduction and overview*. Springer, Cham (CH).
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press, Redwood City (CA US).
- Friedkin, N., Proskurnikov, A., Tempo, R., and Parsegov, S. (2016). Network science on belief system dynamics under logic constraints. *Science*, 354(6310):321–326.
- Hassani, H., Razavi-Far, R., Saif, M., Chiclana, F., Krejcar, O., and Herrera-Viedma, E. (2022). Classical dynamic consensus and opinion dynamics models: A survey of recent trends and methodologies. *Information fusion*, 88:22–40.
- Hegselmann, R. and Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of artificial societies and social simulation*, 5(3).
- Holley, R. A. and Liggett, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, 3(4):643–663.
- Kataoka, H., Euzenat, J., and Hasebe, K. (2025). Exchanging and updating opinions and beliefs reinforces echo chambers. In *Proc. 39th annual conference of the Japanese society for artificial intelligence (JSAI), Osaka (JP)*.
- Mercuur, R., Dignum, V., and Jonker, C. (2019). The value of values and norms in social simulation. *Journal of artificial societies and social simulation*, 22(1).
- Nedić, A., Olshevsky, A., and Uribe, C. (2019). Graph-theoretic analysis of belief system dynamics under logic constraints. *Scientific reports*, 9:8843.
- Noorazar, H. (2020). Recent advances in opinion propagation dynamics: A 2020 survey. *The European Physical Journal Plus*, 135:1–20.
- Oktar, K. and Lombrozo, T. (2025). How aggregated opinions shape beliefs. *Nature Reviews Psychology*, 4(2):81–95.

- Peralta, A. F., Kertész, J., and Iñiguez, G. (2022). Opinion dynamics in social networks: from models to data.
- Sasahara, K., Chen, W., Peng, H., Ciampaglia, G. L., Flammini, A., and Menczer, F. (2020). Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science*, 4(1):381–402.
- Schwartz, S. H. (1999). A theory of cultural values and some implications for work. *Applied Psychology*, 48(1):23–47.
- Schwind, N., Inoue, K., Bourgne, G., Konieczny, S., and Marquis, P. (2015). Belief revision games. In *Proceedings of the 29th AAI Conference, Austin (TX US)*, pages 1590–1596.
- Schwind, N., Inoue, K., Bourgne, G., Konieczny, S., and Marquis, P. (2016). On opinion control in belief revision games. In *The 30th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI)*, pages 1237–1243, Kanagawa (JP).
- Schwind, N., Inoue, K., Konieczny, S., and Marquis, P. (2024). BeliefFlow: A framework for logic-based belief diffusion via iterated belief change. In *Proceedings of the 38th AAI Conference, Vancouver (CA)*, pages 10696–10704.
- Sîrbu, A., Loreto, V., Servedio, V., and Tria, F. (2017). Opinion dynamics: Models, extensions and external effects. In Loreto, V., Haklay, M., Hotho, A., Servedio, V., Stumme, G., Theunis, J., and Tria, F., editors, *Participatory Sensing, Opinions and Collective Awareness*, pages 363–401. Springer, Cham (CH).
- Tang, Y., Liu, J., and Chen, W. (2021). Exchange, adopt, evolve: Modeling the spreading of opinions through cognition and interaction in a social network. *Information sciences*, 551:1–22.
- Vanhée, L. (2015). *Using culture and value to support flexible coordination*. PhD thesis, Utrecht university.
- Wen, T., Zheng, R., Wu, T., Liu, Z., Zhou, M., Syed, T. A., Ghataoura, D., and Chen, Y. (2025). Formulating opinion dynamics from belief formation, diffusion and updating in social network group decision-making: Towards developing a holistic framework. *European Journal of Operational Research*, 325(3).
- Zollman, K. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74(5):574–587.