Characterizing Consensuses in Belief Flow Networks

Nicolas Schwind
National Institute of Advanced
Industrial Science and Technology
Tokyo, Japan
nicolas-schwind@aist.go.jp

Gauvain Bourgne
Lip6, Sorbonne Université
Paris, France
gauvain.bourgne@lip6.fr

Katsumi Inoue
National Institute of Informatics
Tokyo, Japan
inoue@nii.ac.jp

ABSTRACT

The BeliefFlow framework models how logical beliefs spread in networks of interacting agents. In a Belief Flow Network (BFN), agents hold epistemic states capturing current and conditional beliefs and revise them asynchronously, taking into account the beliefs of those that influence them as specified by an acquaintance graph, using an improvement operator, a rational form of iterated belief change. Earlier work showed that in strongly connected BFNs, all agents always converge to a global consensus, regardless of initial beliefs, revision policies, or the stochastic order of communications. This paper examines the nature of such consensuses. While past results proved *that* consensus is reached, we characterize *which* formulas may emerge. Given a BFN scheme defined by its acquaintance graph and the agents' initial beliefs, we provide necessary and sufficient conditions for a formula to be realizable as a consensus outcome. A key outcome of our study is that deciding whether a formula is a possible consensus for a given scheme can be done in polynomial time with polynomially many calls to an NP oracle. This matches the complexity of inference for single iterated belief change operators, showing that consensus characterization in BFNs is no harder than reasoning about belief change itself.

KEYWORDS

Belief Diffusion, Iterated Belief Change, Social Networks

ACM Reference Format:
Nicolas Schwind, Gauvain Bourgne, and Katsumi Inoue. 2026. Characterizing Consensuses in Belief Flow Networks. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. https://doi.org/10.65109/BLMU3934

1 INTRODUCTION

The diffusion of beliefs in networks of interacting agents has been widely studied in multi-agent systems and artificial intelligence. Understanding how individual beliefs evolve and propagate through communication channels is essential to model various forms of collective reasoning. Seminal works such as the DeGroot model [9] have inspired a wide range of models for consensus formation and information dynamics, including both probabilistic and logic-based approaches [3, 7, 10, 11].

Recently, Belief Flow Networks (BFNs) have been introduced as a formal framework for studying how logically complex beliefs propagate asynchronously through a network of rational agents [21]. In a BFN, agents hold epistemic states, which are abstract objects from which their current beliefs (a propositional formula) can be extracted, and which also encode conditional information that governs how belief change is performed [8]. Agents revise their beliefs using improvement operators [14, 15, 17], which provide rational models for iterated belief change. Communication proceeds as a stochastic process over a directed acquaintance graph: at each step, an agent receives a propositional formula from a neighbor and updates her beliefs accordingly. The model captures realistic dynamics such as asynchronous influence, individualized change reluctance, and the logical structure of beliefs.

A central result in [21] is that BFNs always reach a global consensus after finitely many steps, provided the communication graph is strongly connected. This leaves open the key question of *which* formulas may emerge as consensuses, given the initial network and belief settings. Our paper addresses this problem.

We investigate the consensus characterization problem in BFN schemes, i.e., abstractions of BFNs consisting only of an influence graph and the agents' initial propositional beliefs, omitting epistemic states and improvement operators. Given such a scheme, we ask which formulas can arise as final consensuses through the belief evolution process. We distinguish two cases. In the consistent case, where the conjunction of all agents' initial beliefs is itself consistent, we show that the only possible consensus is precisely this conjunction. In the inconsistent case, where conflicting initial beliefs are present, the analysis becomes more intricate. Our main result provides a complete characterization of the formulas that may be consensuses, based on logical entailment and structural properties of the graph. We also show that deciding whether a formula can be a consensus belongs to the class P^{NP}, matching the complexity of reasoning in iterated belief change [16].

Proofs of propositions are available online.[1]

2 PRELIMINARIES

2.1 Basic Notions and Iterated Belief Change

We work over a propositional language $\mathcal{L}$ built from a finite set of variables $\mathcal{P}$ and the usual connectives. A world maps $\mathcal{P}$ to $\{0, 1\}$. We write $[\alpha]$ for the models of a formula $\alpha \in \mathcal{L}$, i.e., the set of worlds satisfying α, $\models$ for entailment between formulas ($\varphi \models \psi$ iff $[\varphi] \subseteq [\psi]$), $\equiv$ for equivalence ($\varphi \equiv \psi$ iff $\varphi \models \psi$ and $\psi \models \varphi$), and Ω for the set of all worlds. A total preorder $\preceq$ on worlds (tpo) is a reflexive, transitive relation on Ω. For a tpo $\preceq$ and $W \subseteq \Omega$, $\min(W, \preceq)$ denotes the $\preceq$-minimal worlds in W, i.e., $\min(W, \preceq) = \{\omega \in W \mid \forall \omega' \in W,\ \omega \preceq \omega'\}$.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/BLMU3934

---

[1] https://nicolas-schwind.github.io/SBI-AAMAS26-proofs.pdf

Iterated belief change models how a rational agent updates her beliefs when receiving successive inputs from influencing sources. An *epistemic state* Ψ is an abstract object representing an agent's belief state from which both the agent's current beliefs (a propositional formula $Bel(\Psi) \in \mathcal{L}$) can be extracted, but that also comprises some conditional information that governs how change is made [8]. An *epistemic space* refers to the collection of all such states [23].

One of the most standard epistemic spaces is the *tpo-based epistemic space*. In this epistemic space, each epistemic state is represented as a tpo Ψ, and $Bel(\Psi) = \psi$ where $[\psi] = \min(\Omega, \Psi)$.

An (iterated belief) change operator $\circ$ is defined relative to a given epistemic space and maps a state Ψ and formula α to a new state $\Psi \circ \alpha$. Operators are meant to satisfy rationality requirements for incorporating new information. A well-known class is *improvement operators*, characterized by nine postulates (I1)–(I9) [14, 15, 17]. For $\circ$, define $\Psi \circ^1 \alpha = \Psi \circ \alpha$ and $\Psi \circ^k \alpha = (\Psi \circ^{k-1} \alpha) \circ \alpha$ for $k > 1$. Then set $\Psi \star_\circ \alpha = \Psi \circ^k \alpha$, where k is the least index with $Bel(\Psi \circ^k \alpha) \models \alpha$ (guaranteed by (I1), see below). Then, an improvement operator is a change operator satisfying the following postulates [14, 15, 17]:

(I1) $\exists k > 0$ s.t. $Bel(\Psi \circ^k \alpha) \models \alpha$
(I2) If $Bel(\Psi) \wedge \alpha \not\models \bot$, then $Bel(\Psi \star_\circ \alpha) \equiv Bel(\Psi) \wedge \alpha$
(I3) If $\alpha \not\models \bot$, then $Bel(\Psi \circ \alpha) \not\models \bot$
(I4) if $\alpha_i \equiv \beta_i$ for all $i \in \{1, \ldots, m\}$,
then $Bel(\Psi \circ \alpha_1 \circ \ldots \circ \alpha_m) \equiv Bel(\Psi \circ \beta_1 \circ \ldots \circ \beta_m)$
(I5) $Bel(\Psi \star_\circ \alpha) \wedge \beta \models Bel(\Psi \star_\circ (\alpha \wedge \beta))$
(I6) If $Bel(\Psi \star_\circ \alpha) \wedge \beta \not\models \bot$,
then $Bel(\Psi \star_\circ (\alpha \wedge \beta)) \models Bel(\Psi \star_\circ \alpha) \wedge \beta$
(I7) If $\alpha \models \beta$, then $Bel((\Psi \circ \beta) \star_\circ \alpha) \equiv Bel(\Psi \star_\circ \alpha)$
(I8) If $\alpha \models \neg\beta$, then $Bel((\Psi \circ \beta) \star_\circ \alpha) \equiv Bel(\Psi \star_\circ \alpha)$
(I9) If $Bel(\Psi \star_\circ \alpha) \not\models \neg\beta$, then $Bel((\Psi \circ \beta) \star_\circ \alpha) \models \beta$

Among these nine postulates, (I1)–(I6) are the core conditions; (I7)–(I9) govern iteration.

Improvement operators admit a semantic account via plausibility orderings over worlds (the lower the more plausible in those orderings). A mapping $\Psi \mapsto \preceq_\Psi$ assigning a tpo $\preceq_\Psi$ to each state Ψ is called a *gradual assignment* if it satisfies:

1. If $\omega, \omega' \models Bel(\Psi)$ then $\omega \simeq_\Psi \omega'$
2. If $\omega \models Bel(\Psi)$ and $\omega' \not\models Bel(\Psi)$ then $\omega \prec_\Psi \omega'$
3. For any $n > 0$, if $\alpha_i \equiv \beta_i$ for any $i \leq n$,
   then $\preceq_{\Psi \circ \alpha_1 \circ \ldots \circ \alpha_n} = \preceq_{\Psi \circ \beta_1 \circ \ldots \circ \beta_n}$
4. If $\omega, \omega' \models \alpha$, then $\omega \preceq_\Psi \omega' \Leftrightarrow \omega \preceq_{\Psi \circ \alpha} \omega'$
5. If $\omega, \omega' \models \neg\alpha$, then $\omega \preceq_\Psi \omega' \Leftrightarrow \omega \preceq_{\Psi \circ \alpha} \omega'$
6. If $\omega \models \alpha$ and $\omega' \models \neg\alpha$, then $\omega \preceq_\Psi \omega' \Rightarrow \omega \prec_{\Psi \circ \alpha} \omega'$

PROPOSITION 2.1 ([14]). *An operator $\circ$ is an improvement operator iff there exists a gradual assignment mapping each Ψ to a plausibility ordering, a tpo $\preceq_\Psi$ such that, for every formula α, $[Bel(\Psi \star_\circ \alpha)] = \min([\alpha], \preceq_\Psi)$.*

Operators satisfying the success postulate (R*1) ($Bel(\Psi \circ \alpha) \models \alpha$) [8] are called DP revision operators [1, 8], and then $\star_\circ = \circ$.

We next recall two well-known examples of improvement operators on the tpo-epistemic space: Nayak's lexicographic operator $\circ_L$ [18] and the one-improvement operator $\circ_I$ [15]. For these operators $\circ \in \{\circ_L, \circ_I\}$, each epistemic state Ψ (a tpo) can actually be identified through $\circ$'s gradual assignment with Ψ's plausibility ordering $\preceq_\Psi$, so relative plausibility is encoded directly in each state. Nayak's lexicographic operator $\circ_L$ produces $\Psi' = \Psi \circ_L \alpha$ that strictly prefers every α-world to every non-α-world and leaves all other pairwise comparisons unchanged [18]. The one-improvement operator $\circ_I$ moves all α-worlds one layer[2] toward the bottom in Ψ; in $\Psi' = \Psi \circ_I \alpha$ it creates a new bottom layer for $\min([\alpha], \preceq_\Psi)$ when this set is included in the original bottom layer (see [15] for a formal definition). Figure 1 illustrates both behaviors.

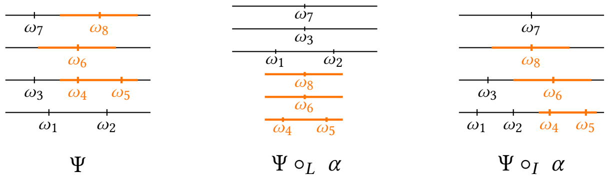


Figure 1: Illustration of the behavior of the improvement operators $\circ_L$ and $\circ_I$, with $[\alpha] = \{\omega_4, \omega_5, \omega_6, \omega_8\}$.

Note that the lexicographic operator $\circ_L$ satisfies (R*1) and is maximally change-inclined among improvement operators [19]: upon receiving α, it always considers α-worlds as strictly more plausible than $\neg\alpha$-worlds after revision. The one-improvement operator $\circ_I$, on the other hand, does not satisfy (R*1) and models a more change-reluctant agent that may not accept a new input into her beliefs immediately.

2.2 Belief Flow Networks

Belief Flow Networks (BFNs) model how agents change their epistemic states and transmit their beliefs through a network of directed influences [21]. Each agent holds an epistemic state and revises it asynchronously by incorporating the beliefs of its influencers according to her own improvement operator. Formally, a BFN is a tuple $\mathcal{B} = \langle (V, A), \vec{\Psi}, \vec{\circ}, \mathcal{S} \rangle$ with four components.

(1) Acquaintance graph. $G = (V, A)$ (also called an *influence graph*) where $V = \{1, \ldots, n\}$ is the set of agents and $A \subseteq V \times V$ is an irreflexive set of directed edges. An edge $(i, j) \in A$ means that information can flow from agent i to agent j.

(2) Initial epistemic state profile. $\vec{\Psi} = \langle \Psi^1, \ldots, \Psi^n \rangle$ gives the initial epistemic state of each agent, with $Bel(\Psi^i)$ assumed consistent for all $i \in V$.

(3) Change policy profile. $\vec{\circ} = \langle \circ_1, \ldots, \circ_n \rangle$ specifies each agent's change policy. Every $\circ_i$ is an improvement operator and is assumed to satisfy a mild strengthening of (I1), called (I1*) $\exists k \in \mathbb{N}_*$ s.t. $\forall \Psi \in \mathcal{E}, \forall \alpha \in \mathcal{L}$, $Bel(\Psi \circ^k \alpha) \models \alpha$. That is, the number of steps required for every agent $i \in V$ to fully accept any input formula is fixed and independent of its current epistemic state. This condition is satisfied in particular by improvement operators defined over finite epistemic spaces like the tpo-epistemic space (thus, the two operators $\circ_L$ and $\circ_I$ introduced previously both satisfy (I1*)).

(4) Communication process. $\mathcal{S} = (A_s)_{s \in \mathbb{N}}$ is a stochastic process on A describing which communication link is triggered at each step. It is assumed to be a *chain with complete connections* [13]: there exists $\delta > 0$ such that for any s and any $e_0, \ldots, e_s \in A$,

---

[2] Each tpo can be seen as a sequence of such layers.

$\Pr(A_s = e_s \mid A_{s-1} = e_{s-1}, \dots, A_0 = e_0) \geq \delta$. This ensures that every edge has a nonzero probability of occurring. Such a definition covers both Markov chains (where the next edge depends on the previous one) and Bernoulli schemes (independent draws) [25].

Runs and epistemic evolution. A $\mathcal{B}$*-run* $\sigma = (\sigma_s)_{s \in \mathbb{N}}$ is an infinite sequence generated by $\mathcal{S}$, with $\sigma_s \in A$ indicating the edge activated at step s. Each agent i has an epistemic state $\Psi^i_{\sigma_s}$ at step s in run σ, with $\Psi^i_{\sigma_0} = \Psi^i$. When $\sigma_s = (j, i)$, agent j's current beliefs $Bel(\Psi^j_{\sigma_s})$ are received by i, who updates her epistemic state via her policy $\circ_i$; for each step $s > 0$:

$$\Psi^i_{\sigma_{s+1}} = \Psi^i_{\sigma_s} \ \circ_i \ Bel(\Psi^j_{\sigma_s})$$

If i is not the target of the triggered edge, her state remains unchanged ($\Psi^i_{\sigma_{s+1}} = \Psi^i_{\sigma_s}$). The sequence of all agents' states at each step forms the *epistemic profile sequence* $\vec{\Psi}_\sigma = (\vec{\Psi}_{\sigma_s})_{s \in \mathbb{N}}$, where for each $s \in \mathbb{N}$, $\vec{\Psi}_{\sigma_s} = \langle \Psi^1_{\sigma_s}, \dots, \Psi^n_{\sigma_s} \rangle$.

Outcomes, stability, and consensus. For a given agent i in run σ, the *belief sequence* $Seq_\sigma(i) = (Bel(\Psi^i_{\sigma_s}))_{s \in \mathbb{N}}$ records the successive beliefs of i. Among its subsequences, the σ*-outcome sequence* $Seq^*_\sigma(i)$ is the earliest one where every belief formula appears infinitely often (up to equivalence). The σ*-outcome* of i, written $Out_\sigma(i)$, is the weakest formula entailing all formulas in $Seq^*_\sigma(i)$. Agent i is *stable* in σ if all formulas in $Seq^*_\sigma(i)$ are equivalent to $Out_\sigma(i)$, that is, her beliefs eventually stop changing. A set of agents $V' \subseteq V$ is stable in σ if each of its members is stable.

A BFN $\mathcal{B}$ *reaches a consensus* in a $\mathcal{B}$-run σ if V is stable in σ and there is a formula α such that $Out_\sigma(i) \equiv \alpha$ for each $i \in V$. We denote this formula by $Out_\sigma(\mathcal{B})$ and call it the *consensus of* $\mathcal{B}$ *in* σ. A BFN $\mathcal{B}$ is *strongly consensual* if it reaches a consensus in each $\mathcal{B}$-run.

Basic properties and strong consensus. BFNs satisfy several appealing properties that guarantee consistent and coherent belief propagation [21]. For every $\mathcal{B}$-run σ, step s, and formula φ:

(CP) $\forall i \in V \ Bel(\Psi^i_{\sigma_s}) \not\models \bot$
(AP) $(\forall i \in V \ \varphi \models Bel(\Psi^i)) \Rightarrow (\forall i \in V \ \varphi \models Bel(\Psi^i_{\sigma_s}))$
(UP) $(\forall i \in V \ Bel(\Psi^i) \models \varphi) \Rightarrow (\forall i \in V \ Bel(\Psi^i_{\sigma_s}) \models \varphi)$
(DR) $\forall (j, i) \in A, \exists s' \geq s$ s.t. $Bel(\Psi^i_{\sigma_{s'}}) \wedge Bel(\Psi^j_{\sigma_{s'}}) \not\models \bot$

Intuitively, (CP) (Consistency Preservation) states that agents always maintain consistent beliefs at every stage of the process. (AP) (Agreement Preservation) ensures that if all agents initially agree on a set of possible worlds satisfying a formula φ, this agreement is preserved throughout all future steps. (UP) (Unanimity Preservation), the dual of (AP), guarantees that if all agents initially believe φ, they continue to do so regardless of the communication dynamics. Finally, (DR) (Delayed Responsiveness) captures the idea that when one agent j can influence another agent i, there will always exist a future moment in the interaction where their beliefs become mutually consistent.

Proposition 2.2. *[21] Every BFN satisfies (CP), (AP), (UP), and (DR).*

Finally, it has been showed in [21] that under strong connectivity of the underlying graph, the set of all agents is guaranteed to be strongly consensual. Formally, given a graph $G = (V, A)$ and two agents $i, j \in V$, $i \rightsquigarrow_G j$ denotes the fact that there exists a directed path from i to j in G. Then G is said to be *strongly connected* if $i \rightsquigarrow_G j$ holds for all distinct $i, j \in V$. We naturally lift this notion to BFNs, saying that a BFN $\mathcal{B} = \langle (V, A), \vec{\Psi}, \vec{\circ}, \mathcal{S} \rangle$ is strongly connected if its underlying graph is.

Proposition 2.3. *[21] Every strongly connected BFN is strongly consensual.*

This is a central result: when the influence structure is strongly connected, the network inevitably converges to a common belief, regardless of the stochastic communication process or the agents' individual epistemic states and change policies. In other words, as long as every agent can eventually reach every other through some sequence of communications, their beliefs are guaranteed to align after a finite number of steps.

Building on this result, we focus in the rest of this paper on strongly connected BFNs, as they provide the structural basis for all subsequent results.

3 EXAMPLE AND PROBLEM DEFINITION

Let us start this section with a concrete, simple example of a BFN.

Example 3.1. Consider three agents, Alice (agent 1), Bob (agent 2), and Charles (agent 3), engaged in an internal discussion about the development of a new prototype. Alice is an engineer working on system development, Bob is a project manager coordinating communication, and Charles is an external quality auditor temporarily added to the project channel. Each agent forms and changes beliefs about the project's progress, based on their own interpretation of available information and on the posts they read from others on the company's internal communication platform. Agents can only see posts from those they follow. Assume Bob serves as the central communication hub: Alice and Bob follow each other (so they can mutually influence one another), and Bob and Charles also follow each other, but Alice and Charles do not have a direct connection. We have $V = \{1, 2, 3\}$ and $A = \{(1, 2), (2, 1), (1, 3), (3, 1)\}$.

Now, let $\mathcal{P}$ be the set of propositional variables $\mathcal{P} = \{p, q, r\}$, where p stands for "The prototype is close to meeting the safety requirements", q stands for "The performance benchmarks meet the required standard", and r stands for "The client appears inclined to continue the project."

Initially, Alice believes $p \wedge q$: from the engineer's perspective, both the safety and performance aspects of the prototype are adequate. Bob believes r: as project manager, he is confident that the client is likely to continue the project. Charles believes $\neg p \wedge \neg q$: based on audit reports, the prototype seems to fail to meet the safety and performance standards. Thus, Alice and Charles hold opposite beliefs about the technical aspects (p, q), while Bob's belief (r) concerns an independent managerial expectation.

Let $\circ_1$, $\circ_2$ and $\circ_3$ be Alice, Bob and Charles' change policies, respectively, arbitrarily chosen DP revision operators. Recall that DP revision operators are, accordingly, improvement operators, but they additionally satisfying the success postulate (R*1), i.e., $Bel(\Psi \circ_i \alpha) \models \alpha$ for each $i \in \{1, 2, 3\}$ and for all Ψ, α (this choice is made for simplicity in this example). Let Ψ^1, Ψ^2, Ψ^3 be any epistemic states such that the beliefs associated with those states reflect the example above, i.e., $Bel(\Psi^1) = p \wedge q$, $Bel(\Psi^2) = r$, and $Bel(\Psi^3) = \neg p \wedge \neg q$.

Table 1: Evolution of the agents' beliefs in the $\mathcal{B}$-run σ.

Step s	**Trig. edge** σ_s	**Bel**($\Psi^1_{\sigma_s}$)	**Bel**($\Psi^2_{\sigma_s}$)	**Bel**($\Psi^3_{\sigma_s}$)
0	–	$p \land q$	r	$\neg p \land \neg q$
1	$(2, 1)$	$\mathbf{p \land q \land r}$	r	$\neg p \land \neg q$
2	$(1, 2)$	$p \land q \land r$	$\mathbf{p \land q \land r}$	$\neg p \land \neg q$
3	$(2, 3)$	$p \land q \land r$	$p \land q \land r$	$\mathbf{p \land q \land r}$
≥ 4	any $(i, j) \in A$	$p \land q \land r$	$p \land q \land r$	$p \land q \land r$

And let $\mathcal{S}$ be any stochastic process. This fully specifies a BFN $\mathcal{B} = \langle (V, A), \vec{\Psi}, \vec{\circ}, \mathcal{S} \rangle$, where $\vec{\Psi} = \langle \Psi^1, \Psi^2, \Psi^3 \rangle$ and $\vec{\circ} = \langle \circ_1, \circ_2, \circ_3 \rangle$.

Consider now any $\mathcal{B}$-run σ starting with the A-sequence $\delta = ((2, 1), (1, 2), (2, 3))$. We have that $\sigma_0 = (2, 1)$, $\sigma_1 = (1, 2)$, and $\sigma_2 = (2, 3)$. It can then be verified that, given the above specifications of $\mathcal{B}$, the beliefs of the agents will evolve in the $\mathcal{B}$-run σ according to Table 1. For instance, since the first triggered edge is $\sigma_0 = (2, 1)$, we get that $Bel(\Psi^1_{\sigma_1}) \equiv Bel(\Psi^1_{\sigma_0} \circ_1 Bel(\Psi^2_{\sigma_0})) \equiv Bel(\Psi^1_{\sigma_0}) \land Bel(\Psi^2_{\sigma_0})$ (by (I2), since $\star_{\circ_1} = \circ_1$ and since $Bel(\Psi^1_{\sigma_0}) \land Bel(\Psi^2_{\sigma_0}) \equiv (p \land q) \land r \not\models \bot$), thus $Bel(\Psi^1_{\sigma_1}) \equiv p \land q \land r$. A consensus is reached from step 3 in σ, with $Out_\sigma(\mathcal{B}) \equiv p \land q \land r$.

It can be easily seen in this example that different $\mathcal{B}$-runs may lead to different consensuses. For instance, switching agents 1 and 3 in σ would lead to $Out_\sigma(\mathcal{B}) \equiv \neg p \land \neg q \land r$. This non-deterministic behavior reflects a natural situation: the order in which communications are triggered affects the consensus.

Since this BFN is strongly connected, Proposition 2.3 guarantees that every $\mathcal{B}$-run reaches a consensus. The question we address here is *which* consensuses may arise. Given a BFN $\mathcal{B}$, we denote by $Out(\mathcal{B})$ this set of all possible consensuses in the BFN $\mathcal{B}$, simply called the *outcome set of* $\mathcal{B}$. Our main objective, though, is not to compute the outcome set of a single, fully specified BFN, but the outcome set of a *BFN scheme* that specifies only some parameters of a BFN. To justify abstracting away from some of these parameters, we first show (Proposition 3.2) that the choice of stochastic process $\mathcal{S}$ does not affect the outcome set:

Proposition 3.2. *For all strongly connected BFNs* $\mathcal{B} = \langle (V, A), \vec{\Psi}, \vec{\circ}, \mathcal{S} \rangle$ *and* $\mathcal{B}' = \langle (V, A), \vec{\Psi}, \vec{\circ}, \mathcal{S}' \rangle$, *we have* $Out(\mathcal{B}) = Out(\mathcal{B}')$.

That is, for all strongly connected BFNs with the same graph and epistemic state/change policy profiles, $Out(\mathcal{B})$ coincides. Hence, $\mathcal{S}$ can be omitted when studying possible consensuses.

A second justification is that assuming full epistemic states and change policies as input is neither realistic nor necessary. On the one hand, epistemic states are abstract objects whose internal representation does not affect change outcomes. The dynamics are determined by the improvement operator, which associates with each epistemic state Ψ — via a gradual assignment — a total preorder over worlds (cf. Proposition 2.1). From Ψ itself, the only relevant datum is its beliefs $Bel(\Psi)$ (see the preliminaries). Moreover, epistemic spaces may be uncountable in general, so epistemic states need not be finitely representable [23]. On the other hand, providing an agent's change policy (improvement operator) explicitely can amount to giving a labeled transition system over epistemic states, with transitions of the form $\Psi, \varphi \mapsto \Psi \circ \varphi$, which is typically very large or even infinite and therefore impractical as input.

For these reasons, we adopt as input a *BFN scheme*: it retains the influence graph and replaces the epistemic profile with a *belief profile*, i.e., a vector of propositional formulas describing the agents' initial beliefs. Agents' (unspecified) change policies are still assumed to be improvement operators satisfying the postulates from the preliminaries. By Proposition 3.2, the stochastic process $\mathcal{S}$ need not be specified either.

Definition 3.3 (BFN scheme). A BFN scheme is a tuple $\mathrm{B} = \langle (V, A), \vec{B} \rangle$, where (V, A) is a strongly connected acquaintance graph and $\vec{B} = \langle B^1, \dots, B^n \rangle$ is a *belief profile*, that is, B^i is a propositional formula for each $i \in V$.

A BFN $\mathcal{B} = \langle (V, A), \vec{\Psi}, \vec{\circ}, \mathcal{S} \rangle$ *satisfies* a scheme $\mathrm{B} = \langle (V', A'), \vec{B} \rangle$, written $\mathcal{B} \models \mathrm{B}$, iff $(V, A) = (V', A')$ and, for each agent $i \in V$, $Bel(\Psi^i) \equiv B^i$. We lift the outcome notion to BFN schemes by:

$$Out(\mathrm{B}) = \bigcup \{Out(\mathcal{B}) \mid \mathcal{B} \models \mathrm{B}\}$$

Interpreted this way, answering the question for a scheme B and formula φ directly answers: *"Does there exist a BFN satisfying* B *for which* φ *is a possible consensus?"* This is a natural decision problem, given that only the graph and initial beliefs are typically available. In the rest of the paper, for any BFN scheme B and formula φ, we aim to characterize whether $\varphi \in Out(\mathrm{B})$.

4 CONSISTENT BFN SCHEMES

We start with a simple but important case, where the conjunction of all agents' initial beliefs is consistent:

Definition 4.1 (Consistent BFN scheme). A BFN scheme $\mathrm{B} = \langle (V, A), \vec{B} \rangle$ is said to be *consistent* if $\bigwedge_{i \in V} B^i \not\models \bot$. Otherwise, B is *inconsistent.*

Proposition 4.2. *Let* $\mathrm{B} = \langle (V, A), \vec{B} \rangle$ *be a BFN scheme. If* B *is consistent, then* $Out(\mathrm{B}) = \{\bigwedge_{i \in V} B^i\}$.

The proof idea is straightforward. From (AP), the models of every agent's belief are never lost. By Proposition 2.3, the system must reach a consensus after finitely many steps. Then, any consensus weaker than $\bigwedge_{i \in V} B^i$ would imply some agent to weaken her beliefs to include a world that was not part of the joint initial beliefs, which would require revising by information inconsistent with her own beliefs, contradicting again (AP). Thus, the only possible consensus in every BFN satisfying such a scheme is this conjunction $\bigwedge_{i \in V} B^i$. Notably, this outcome does not depend on specific epistemic states or change policies.

This result is quite intuitive: in a strongly connected group where all agents start from compatible beliefs, every exchange only reinforces existing agreement. Since no conflicts arise, all agents eventually adopt the conjunction of their initial beliefs, i.e., the group's collective belief.

This proposition also strengthens a related result from the framework of Belief Revision Games [20], where the same conclusion was obtained only under stronger assumptions, that is, complete communication graphs and synchronous updates. Here, the same behavior emerges in any strongly connected graph and under asynchronous communication.

5 INCONSISTENT BFN SCHEMES

When the BFN scheme is *inconsistent*, the agents cannot jointly hold the conjunction of their initial beliefs. Still, by Proposition 2.3, every run reaches *some* consensus. Our example from Section 3 already shows that different runs may lead to different consensuses.

In this section we state and prove a characterization of when a given formula φ can be a consensus for an inconsistent BFN scheme. The statement relies on two simple conditions and a few graph-based notions. We first introduce the notions, the conditions, formally state the characterization theorem, and then show the conditions are necessary and sufficient.

Throughout, let $\mathsf{B} = \langle (V, A), \vec{B} \rangle$ be an arbitrary inconsistent BFN scheme with $\vec{B} = \langle B^1, \dots, B^n \rangle$, and let φ be any consistent formula.

5.1 The Characterization Statement

We start with the introduction of a set of lightweight notions that depend only on the BFN scheme (graph and belief profile).

Observable neutrality in a scheme. For $i \in V$, we say that i is *observably neutral on* φ *in the scheme* B if and only if

$$\varphi \models B^i \quad \text{or} \quad B^i \wedge \varphi \models \bot$$

Equivalently, B^i is either fully compatible with φ or conflicts with it. That is, B^i does not discriminate between φ-models.

This notion lifts verbatim to BFNs $\mathcal{B}$ and $\mathcal{B}$-runs: in a BFN $\mathcal{B} \models \mathsf{B}$, i is observably neutral on φ (at step s in a run σ) by replacing B^i with $Bel(\Psi^i)$ (respectively, $Bel(\Psi^i_{\sigma_s})$). Note that since observable neutrality depends only on the belief profile $\vec{B} = \langle B^1, \dots, B^n \rangle$, i is observably neutral on φ in B if and only if i is observably neutral on φ in *every* $\mathcal{B} \models \mathsf{B}$.

Support and restricted graph. The *support of* φ is the set of agents whose beliefs are consistent with φ:

$$V_{\downarrow\varphi} = \{ i \in V \mid B^i \wedge \varphi \not\models \bot \}$$

We write $G_{\downarrow\varphi}$ for the subgraph of G induced by $V_{\downarrow\varphi}$, i.e., $G_{\downarrow\varphi} = G[V_{\downarrow\varphi}] = (V_{\downarrow\varphi}, A_{\downarrow\varphi})$ with $A_{\downarrow\varphi} = A \cap (V_{\downarrow\varphi} \times V_{\downarrow\varphi})$.

In-neighborhood and upstream reach. For a given agent $i \in V$, its *φ-in-neighborhood* is

$$N^-_\varphi(i) = \{ j \in V_{\downarrow\varphi} \mid (j, i) \in A \}$$

For $i, j \in V_{\downarrow\varphi}$, let $i \rightsquigarrow^*_{G_{\downarrow\varphi}} j$ mean that there is a (possibly empty) directed path from i to j in $G_{\downarrow\varphi}$. For a subset $V'_{\downarrow\varphi} \subseteq V_{\downarrow\varphi}$, define its *$\varphi$-upstream reach set* by

$$Reach^-_\varphi(V'_{\downarrow\varphi}) = \{ i \in V_{\downarrow\varphi} \mid \exists j \in V'_{\downarrow\varphi} : i \rightsquigarrow^*_{G_{\downarrow\varphi}} j \}$$

In particular, note that $V'_{\downarrow\varphi} \subseteq Reach^-_\varphi(V'_{\downarrow\varphi})$.

Backward cone. For a given agent $i \in V$, its *φ-backward cone* is

$$C_\varphi(i) = \{i\} \cup Reach^-_\varphi(N^-_\varphi(i))$$

Intuitively, the set of agents $C_\varphi(i)$ gathers every supporter of φ that can indirectly influence i through $G_{\downarrow\varphi}$, plus i itself (which may or may not support φ).

Lastly, for any $i \in V$, define its *φ-belief closure* as:

$$BC_\varphi(i) = \bigwedge \{ B^j \mid j \in C_\varphi(i) \}$$

We are now ready to state the characterization result:

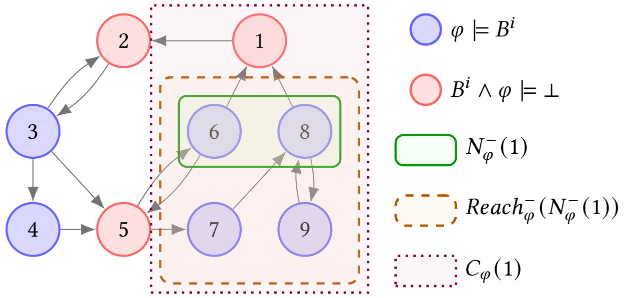


Figure 2: Illustration of a BFN scheme satisfying condition I.

Theorem 5.1. *$\varphi \in Out(\mathsf{B})$ if and only if the two following conditions are satisfied:*

I. *for each agent $i \in V$, i is observably neutral on φ in* B
II. *there exists an agent $i_* \in V$ such that $BC_\varphi(i_*) \models \bot$*

Figure 2 shows a BFN scheme with nine agents that satisfies condition I for some formula φ, highlighting $N^-_\varphi(1)$, $Reach^-_\varphi(N^-_\varphi(1))$, and $C_\varphi(1)$. Checking condition II for agent 1 amounts to testing whether $B^1 \wedge B^6 \wedge B^7 \wedge B^8 \wedge B^9 \models \bot$.

Note that, under condition I, condition II can only hold for agents outside the support of φ (e.g., $i \in \{1, 2, 5\}$ in the figure). If $i \in V_{\downarrow\varphi}$, then $C_\varphi(i) \subseteq V_{\downarrow\varphi}$, and then by condition I, $B^j \models \varphi$ for each $j \in C_\varphi(i)$, so $\bigwedge \{ B^j \mid j \in C_\varphi(i) \}$ is trivially consistent.

The remainder of the section proves Theorem 5.1: we first show that conditions I and II are both necessary, then that together they are sufficient.

5.2 Necessity of Condition I

To prove that condition I is necessary for φ to be a possible consensus of B, we argue a bit stronger: in any BFN $\mathcal{B} \models \mathsf{B}$, every agent must be "conditionally neutral" on φ. Given a BFN $\mathcal{B} = \langle (V, A), \vec{\Psi}, \vec{\circ}, \mathcal{S} \rangle$, an agent $i \in V$, a $\mathcal{B}$-run σ, and a step $s \in \mathbb{N}$, we say that i is *conditionally neutral on φ in σ at step s* if $Bel(\Psi^i_{\sigma_s} \star_{\circ_i} \varphi) \equiv \varphi$. We simply say that i is conditionally neutral on φ in $\mathcal{B}$ if $Bel(\Psi^i \star_{\circ_i} \varphi) \equiv \varphi$. Conditional neutrality means that, were i to revise by φ, the result would be exactly φ, neither stronger nor weaker. This occurs in two basic situations: (i) φ already entails i's belief, so revision acts as conjunction (postulate (I2)): $Bel(\Psi \star_{\circ_i} \varphi) \equiv Bel(\Psi) \wedge \varphi \equiv \varphi$; or (ii) φ contradicts i's beliefs, and revision yields acceptance of φ and nothing else. In both cases, this means i treats all φ-worlds as equally plausible. This is made clear by Lemma 5.2 below. Recall that $\Psi^i \mapsto \preceq^i_{\Psi^i}$ denotes the gradual assignment corresponding to $\circ_i$ (Proposition 2.1):

Lemma 5.2. *Let $i \in V$, σ be a $\mathcal{B}$-run and $s \in \mathbb{N}$. Then $i \in V$ is conditionally neutral on φ in σ at s if and only if for all worlds ω, $\omega' \models \varphi$, we have that $\omega \simeq^i_{\Psi^i_{\sigma_s}} \omega'$.*

Notably, conditional neutrality implies observable neutrality:

Lemma 5.3. *Let $i \in V$, σ be a $\mathcal{B}$-run and $s \in \mathbb{N}$. If i is conditionally neutral on φ in σ at s, then i is observably neutral on φ in σ at s.*

The next lemma shows that conditional neutrality for all agents is a "global" property: if it holds at some step in a given run, then it holds at every step of every run, including in the initial state.

Lemma 5.4. *The following statements are equivalent:*

1. *for each $i \in V$, i is conditionally neutral on φ in $\mathcal{B}$*
2. *for each $\mathcal{B}$-run σ, each step $s \in \mathbb{N}$ and each $i \in V$, i is conditionally neutral on φ in σ at step s*
3. *there exists a $\mathcal{B}$-run σ and a step $s \in \mathbb{N}$ such that for each $i \in V$, i is conditionally neutral on φ in σ at step s*

Using Lemmas 5.2 and 5.4, we obtain:

Lemma 5.5. *Let $\mathcal{B} \models \mathrm{B}$. If $\varphi \in Out(\mathcal{B})$, then for each agent $i \in V$, i is conditionally neutral on φ in $\mathcal{B}$.*

The proof of Lemma 5.5 builds on the following key idea. By contrapositive of (3⇒1) in Lemma 5.4, if some agent is not conditionally neutral initially, then in every run and at every step there remains (by Lemma 5.2 and conditions 1 and 2 of gradual assignments) an agent whose beliefs are not equivalent to φ; hence φ cannot be a consensus.

We can now state necessity of condition I, which relies on Lemmas 5.3 and 5.5:

Proposition 5.6. *If $\varphi \in Out(\mathrm{B})$, then for each $i \in V$, i is observably neutral on φ in B.*

5.3 Necessity of Condition II

We now show that condition II is also a necessary condition for φ to be a possible consensus of B. Recall that the support of φ in a scheme B is the set $V_{\downarrow\varphi} = \{i \in V \mid B^i \wedge \varphi \not\models \bot\}$. We extend this notion dynamically to any BFN $\mathcal{B} \models \mathrm{B}$, run σ, and step $s \in \mathbb{N}$:

$$Sup_{\mathcal{B}}(\varphi, \sigma, s) = \{i \in V \mid Bel(\Psi^i_{\sigma_s}) \wedge \varphi \not\models \bot\}$$

Combining Lemmas 5.3, 5.4 (1⇒2), and 5.5 yields:

Lemma 5.7. *Let $\mathcal{B} \models \mathrm{B}$. If $\varphi \in Out(\mathcal{B})$, then for each $\mathcal{B}$-run σ and step $s \in \mathbb{N}$, $Sup_{\mathcal{B}}(\varphi, \sigma, s) = \{i \in V \mid \varphi \models Bel(\Psi^i_{\sigma_s})\}$.*

That is, if φ is a possible consensus, then all agents in the support of φ at any step are exactly those whose beliefs are entailed by φ.

From this, we note that for φ to be a possible consensus of B, there must exist at least one agent initially inconsistent with φ:

Lemma 5.8. *If $\varphi \in Out(\mathrm{B})$, then $V \setminus V_{\downarrow\varphi} \neq \emptyset$.*

The proof directly follows from Lemma 5.7 together with the fact that B is inconsistent.

The next lemma gives the key intuition behind condition II. If the φ-belief closure of every agent is consistent, then any agent initially rejecting φ will keep rejecting it forever, no matter how communication unfolds.

Lemma 5.9. *Assume that for each agent $i \in V \setminus V_{\downarrow\varphi}$, we have that $BC_\varphi(i) \not\models \bot$. Then for every BFN $\mathcal{B} \models \mathrm{B}$, $\mathcal{B}$-run σ and step $s \in \mathbb{N}$, we have for each agent $i \in V \setminus V_{\downarrow\varphi}$ that $i \in V \setminus Sup_{\mathrm{B}}(\varphi, \sigma, s)$.*

The proof of Lemma 5.9 roughly proceeds as follows. Toward a contradiction, assume that each belief hull $BC_\varphi(i)$ is consistent. For every agent i initially rejecting φ, one can then associate a world ω_i that satisfies both the beliefs of i and of all agents in its φ-upstream reach set $Reach^-_\varphi(N^-_\varphi(i))$, while falsifying φ. By induction on the steps of any run, this world ω_i remains at least as plausible as any φ-world for all those agents, which prevents i from ever fully accepting φ. Thus, any agent that starts outside the support of φ remains outside it at every step of every run.

Combining Lemmas 5.8 and 5.9, we conclude:

Proposition 5.10. *If $\varphi \in Out(\mathrm{B})$, then there exists an agent $i_* \in V$ such that $BC_\varphi(i_*) \models \bot$.*

5.4 Sufficiency of Conditions I and II

We now show that conditions I and II together are sufficient for φ to be a possible consensus of B. The proof proceeds constructively: assuming that B satisfies both conditions I and II, we exhibit one BFN $\mathcal{B} \models \mathrm{B}$ and one $\mathcal{B}$-run in which all agents ultimately converge to the belief φ. The argument unfolds through three essential components, captured by Lemmas 5.11–5.13 below, each corresponding to a key phase in the construction of such a run.

From condition II, there is an agent i_* whose belief closure $BC_\varphi(i_*)$ is inconsistent. This implies that i_* initially rejects φ, i.e., $i_* \in V \setminus V_{\downarrow\varphi}$. To initiate the construction, we consider communication paths ending in the agents of i_*'s φ-in-neighborhood $N^-_\varphi(i_*)$. Starting from the most distant agents in i_*'s φ-upstream reach set $Reach^-_\varphi(i_*)$, we successively trigger these paths so that the beliefs of the agents in $N^-_\varphi(i_*)$ become as close as possible to φ while remaining in the support of φ, leaving all other agents' epistemic states unchanged. The procedure propagates toward the direct φ-supporting predecessors of i_* the "closest approximation" of φ.

This preliminary stage rests on the following lemma, which guarantees that whenever a sequence of agents linked by a path currently holds beliefs that are jointly consistent, the run can always be extended so that the last agent in that path comes to believe exactly the conjunction of all beliefs along it:

Lemma 5.11. *Let $\mathcal{B} = \langle (V, A), \vec{\Psi}, \vec{\circ}, \mathcal{S} \rangle$ be a BFN, $i_* \in V$, $p = (i_0, \ldots, i_m)$, $m > 0$, be a path in (V, A) such that $i_* = i_m$, σ be a $\mathcal{B}$-run, $s_* \in \mathbb{N}$, and assume that $\bigwedge\{Bel(\Psi^{i_k}_{\sigma_{s_*}}) \mid k \in \{0, \ldots, m\}\} \not\models \bot$. Then there exist a $\mathcal{B}$-run σ' and a step t such that:*

- *for each $i \in V \setminus \{i_1, \ldots, i_m\}$, $\Psi^i_{\sigma'_t} = \Psi^i_{\sigma_{s_*}}$, and*
- *$Bel(\Psi^{i_*}_{\sigma'_t}) \equiv \bigwedge\{Bel(\Psi^{i_k}_{\sigma_{s_*}}) \mid k \in \{0, \ldots, m\}\}$.*

Once such a configuration is reached, attention turns to the interactions between i_* and its φ-in-neighbors $N^-_\varphi(i_*)$. By construction and by condition II, the joint beliefs of i_*'s φ-in-neighbors are inconsistent with those of i_* at that stage. The goal is to extend the run so that i_*'s beliefs eventually coincide with φ itself. The idea is to assume that, among all worlds not compatible with i_*'s current beliefs, those satisfying φ are the most plausible for i_*. If i_* uses a receptive revision policy (such as Nayak's lexicographic revision operator $\circ_L$), successive communications from all its φ-in-neighbors make φ-worlds in i_*' state increasingly more plausible than any alternative worlds not constantly supported by these neighbors. Since the neighbors' joint beliefs are inconsistent with i_*'s own beliefs, no other worlds remain equally plausible, and after sufficient interactions, i_*'s beliefs become exactly φ.

Formally, given a total preorder $\preceq$ over worlds and $k \in \mathbb{N}$, let $lvl(\preceq, k)$ denote the set of worlds at the k-th layer of $\preceq$, defined as $lvl(\preceq, 0) = \min(\Omega, \preceq)$ and, for each $k \geq 1$, $lvl(\preceq, k) = \min(\Omega \setminus \bigcup_{k' < k} lvl(\preceq, k'), \preceq)$. The following lemma captures the preceding reasoning:

Lemma 5.12. *Let $\alpha_1, \ldots, \alpha_m$ be m propositional formulas, $m \geq 1$, such that $\varphi \models \bigwedge_{1 \leq i \leq m} \alpha_i$. Let $\circ_L$ be Nayak's lexicographic revision operator on total preorders over worlds. Let Ψ be a total preorder over worlds such that $Bel(\Psi) \wedge \bigwedge_{1 \leq i \leq m} \alpha_i \models \bot$ and $lvl(\Psi, 1) = [\varphi]$. Then $Bel(((\Psi \circ_L \alpha_1) \circ_L \ldots) \circ_L \alpha_m) \equiv \varphi$.*

At this point, the run has reached a configuration in which agent i_*'s beliefs are equivalent to φ. From there, the run can be further extended by triggering communication paths that emanate from i_* and gradually cover all agents in V. Assuming that all agents in the graph are conditionally neutral on φ (an assumption consistent with both conditions I and II, and with the properties so far established for i_*'s epistemic state and revision policy), any agent that receives φ as input will, upon revision, adopt beliefs equivalent to φ. As the influence of i_* propagates throughout the network, each agent in turn aligns her beliefs with φ, and the entire system ultimately converges to a consensus on φ:

Lemma 5.13. *Let $\mathcal{B} \models \mathsf{B}$, σ be a $\mathcal{B}$-run, $s_* \in \mathbb{N}$, and assume that for each agent $i \in V$, i is conditionally neutral on φ in σ at s_*. If there exists an agent $i_* \in V$ such that $Bel(\Psi^{i_*}_{\sigma s_*}) \equiv \varphi$, then $\varphi \in Out(\mathcal{B})$.*

Combining Lemmas 5.11, 5.12, and 5.13 yields the desired result: whenever both conditions I and II hold for B, there exists a BFN $\mathcal{B} \models \mathsf{B}$ and a corresponding run leading all agents to converge to φ:

Proposition 5.14. *If:*

- I. *for each agent $i \in V$, i is observably neutral on φ in B, and*
- II. *there exists an agent $i_* \in V$ such that $BC_\varphi(i_*) \models \bot$,*

then $\varphi \in Out(\mathsf{B})$.

This completes the proof of Theorem 5.1.

6 COMPUTATIONAL COMPLEXITY

We now investigate the computational complexity of the Consensus Membership (CM) decision problem:

- **Input:** BFN scheme B, propositional formula φ
- **Question:** Does $\varphi \in Out(\mathsf{B})$ hold?

The results from the previous sections directly yield an algorithmic procedure for CM. Algorithm 1 makes this procedure explicit, showing how all conditions defining consensus membership can be verified in polynomial time with polynomially many calls to an NP oracle.

Algorithm 1: Deciding whether $\varphi \in Out(\mathsf{B})$

```
   Input: BFN scheme $\mathcal{B} = \langle (V, A), \vec{B} \rangle$, prop. formula $\varphi$
   Output: true if $\varphi \in Out(\mathsf{B})$, false otherwise
1  if $\bigwedge_{i \in V} B^i \not\models \bot$ then
2      return $\bigwedge_{i \in V} B^i \equiv \varphi$
3  $Support(\mathsf{B}, \varphi) \leftarrow \emptyset$
4  for $i \in V$ do
5      if $\varphi \models B^i$ then
6          $Support(\mathsf{B}, \varphi) \leftarrow Support(\mathsf{B}, \varphi) \cup \{i\}$
7      else if $B^i \wedge \varphi \not\models \bot$ then
8          return false
9  for $i \in V \setminus Support(\mathsf{B}, \varphi)$ do
10     BackwardCone$(\mathsf{B}, i, \varphi) \leftarrow C_\varphi(i)$
11     if $\bigwedge\{B^i \mid i \in BackwardCone(\mathsf{B}, i, \varphi)\} \models \bot$ then
12         return true
13 return false
```

Given as input a BFN scheme $\mathsf{B} = \langle (V, A), \vec{B} \rangle$ and a formula φ, the algorithm first checks whether B is consistent (line 1). If so, Proposition 4.2 ensures that a unique consensus is possible, namely $\bigwedge_{i \in V} B^i$, and the algorithm returns **true** if φ is equivalent to this conjunction, **false** otherwise (line 2). If the BFN scheme is inconsistent, the algorithm proceeds to check conditions I and II. Lines 3–8 compute the support of φ and verify condition I. For each agent $i \in V$, if $\varphi \models B^i$, i is added to the support of φ (lines 5–6); if instead $B^i \wedge \varphi \not\models \bot$, i is not observably neutral on φ, i.e., condition I is not satisfied, and the algorithm returns **false** (line 8). Reaching line 9 therefore guarantees that all agents are observably neutral on φ, and that the support $V_{\downarrow\varphi}$ has been correctly computed, i.e., that $Support(\mathsf{B}, \varphi) = V_{\downarrow\varphi}$. The algorithm then checks condition II (lines 9–13). For each agent i not in the support of φ, it computes the φ-backward cone $C_\varphi(i)$ (line 10), which can be done in polynomial time given the graph structure. It then constructs the corresponding φ-belief closure $BC_\varphi(i)$ and checks its consistency (line 11). If for some i, $BC_\varphi(i) \models \bot$, condition II is satisfied and the algorithm returns **true** (line 12); otherwise, if no such agent exists, it returns **false** (line 13).

Algorithm 1 thus implements a procedure that decides CM. It runs in polynomial time, with polynomially many calls to an NP oracle (in lines 1, 2, 5, 7, and 11). As a direct consequence:

Proposition 6.1. CM $\in \mathsf{P}^{\mathsf{NP}}$.

This result establishes that the consensus membership problem is computationally no harder than classical inference under iterated belief change [16, 22]. Indeed, it has been shown that the inference problem (*Given an epistemic state Ψ, a sequence of revision formulas $\alpha_1, \ldots, \alpha_n$, and a formula β, does $((\Psi \circ \alpha_1) \circ \ldots) \circ \alpha_n \models \beta$ hold?*) is P^{NP}-complete in general [16], and $\mathsf{P}^{\mathsf{NP}}[O(\log n)]$-complete for specific representations and single-step revision cases ($n = 1$) [22]. Hence, verifying whether a formula can emerge as a consensus given a BFN scheme (without simulating runs explicitly but using only initial beliefs and the graph structure) is computationally on par with reasoning about iterated belief change itself.

7 BACK TO THE EXAMPLE

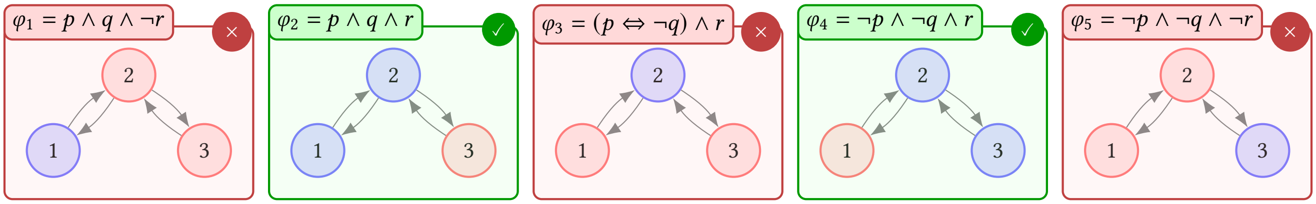
We now return to the Alice–Bob–Charles example introduced in Section 3, focusing on the underlying BFN scheme $\mathsf{B} = \langle (V, A), \vec{B} \rangle$ defined by $V = \{1, 2, 3\}$, $A = \{(1, 2), (2, 1), (2, 3), (3, 2)\}$, and the belief profile $\vec{B} = \langle B^1, B^2, B^3 \rangle$, where $B^1 = p \wedge q$, $B^2 = r$, and $B^3 = \neg p \wedge \neg q$.

We test whether each of the following five formulas

$$\varphi_1 = p \wedge q \wedge \neg r, \quad \varphi_2 = p \wedge q \wedge r, \quad \varphi_3 = (p \Leftrightarrow \neg q) \wedge r,$$
$$\varphi_4 = \neg p \wedge \neg q \wedge r, \quad \varphi_5 = \neg p \wedge \neg q \wedge \neg r$$

belongs to $Out(\mathsf{B})$, and let $\mathcal{F} = \{\varphi_1, \ldots, \varphi_5\}$.

Since B is an inconsistent BFN scheme, the two conditions of Theorem 5.1 must be checked. It can be easily verified that for each $\varphi_k \in \mathcal{F}$ and each agent $i \in \{1, 2, 3\}$, either $\varphi_k \models B^i$ or $B^i \wedge \varphi_k \models \bot$. Hence, all agents are observably neutral on every φ_k, and only



$BC_{\varphi_1}(1) \equiv B^1 \not\models \bot$
$BC_{\varphi_1}(2) \equiv B^1 \wedge B^2 \not\models \bot$
$BC_{\varphi_1}(3) \equiv B^3 \not\models \bot$

$BC_{\varphi_1}(3) \equiv B^1 \wedge B^2 \wedge B^3 \models \bot$

$BC_{\varphi_1}(1) \equiv B^1 \wedge B^2 \not\models \bot$
$BC_{\varphi_1}(2) \equiv B^2 \not\models \bot$
$BC_{\varphi_1}(3) \equiv B^2 \wedge B^3 \not\models \bot$

$BC_{\varphi_1}(1) \equiv B^1 \wedge B^2 \wedge B^3 \models \bot$

$BC_{\varphi_1}(1) \equiv B^1 \not\models \bot$
$BC_{\varphi_1}(2) \equiv B^2 \wedge B^3 \not\models \bot$
$BC_{\varphi_1}(3) \equiv B^3 \not\models \bot$

Figure 3: Checking whether each formula $\varphi_k \in \mathcal{F}$ satisfies condition II.

condition II remains to be tested: for each $\varphi_k \in \mathcal{F}$, we must check whether there exists an agent $i_* \in \{1, 2, 3\}$ such that $BC_{\varphi_k}(i_*) \models \bot$.

Figure 3 illustrates, for each $\varphi_k \in \mathcal{F}$, the graph annotated with its φ_k-support (following the legend of Figure 2) and the corresponding verification of condition II. The result is that $\varphi_2, \varphi_4 \in Out(\mathrm{B})$, while $\varphi_1, \varphi_3, \varphi_5 \notin Out(\mathrm{B})$. This outcome also leads to a complete description of the set $Out(\mathrm{B})$. Specifically, one can verify that a formula $\psi \in \mathcal{L}$ satisfies conditions I and II precisely when ψ is consistent and $\psi \models \varphi_k$ for some $\varphi_k \in \mathcal{F}$ that satisfies conditions I and II. From this observation, we obtain:

$$Out(\mathrm{B}) = \{\psi \in \mathcal{L}_{\mathcal{P}} \mid \psi \models \varphi_2 \text{ or } \psi \models \varphi_4\}$$

In words, every consensus entails r (which is uncontested by any agent) and either $p \wedge q$ or $\neg p \wedge \neg q$, reflecting that the final consensus coincides with the initial belief of either Alice or Charles, depending on the run, epistemic states, and agents' policies.

As a variant, consider the same BFN scheme but with a complete influence graph, i.e., $A = V \times V$. In this case, similar reasoning shows that:

$$Out(\mathrm{B}) = \{\psi \in \mathcal{L}_{\mathcal{P}} \mid \psi \models \varphi_k \text{ for some } k \in \{1, 2, 4, 5\}\}$$

Here, possible consensuses include formulas entailing $\neg r$. This is because, with direct communication between Alice and Charles, revising one's epistemic state by the other's beliefs can make $\neg r$ prevail as a more entrenched proposition than r in their initial *conditional* beliefs.

8 DISCUSSION

This paper presents a characterization of the formulas that may arise as consensuses in Belief Flow Networks (BFNs). Given a BFN scheme, that is, an influence graph and the agents' initial propositional beliefs, we provide a decision procedure to determine whether a formula is a possible consensus. The characterization is purely logical and graph-based, independent of the agents' internal epistemic states, change policies, and the underlying stochastic dynamics. We show that the problem is solvable in P^{NP}, keeping it computationally feasible despite the expressiveness of the framework.

Our contribution strengthens the theoretical understanding of BFNs beyond their convergence guarantees [21], enabling a precise analysis of which beliefs may emerge as final consensuses. This lays the foundation for planning, verification, and control in logic-based models of belief diffusion, where outcomes need to be both predictable and interpretable.

Several logical approaches to belief change in networks have pursued similar goals. Models such as those by Gallo et al. [11], Cholvy [7], and Vicol et al. [26] use graph-based interactions and propositional representations, but rely on deterministic, often synchronous update schemes. These frameworks focus on properties like local consistency or disagreement minimization, without guaranteeing convergence or characterizing the full space of reachable consensuses. In contrast, BFNs employ asynchronous stochastic dynamics along with formal convergence guarantees, now complemented by a logical characterization of outcomes. Within the belief merging literature, Schwind and Marquis [24] propose an axiomatic account of consensus formation. Like other belief merging models, their framework lacks iterative dynamics and influence graphs, but it shares the goal of capturing rational group agreement under logical constraints. Our results extend this perspective into dynamic, multi-step settings where belief evolution emerges from agent interactions. Other models have addressed opinion diffusion over more structured domains. Brill et al. [5] and Botan et al. [4] analyze convergence under aggregation rules applied to rankings or Boolean vectors with constraints. These works also aim to characterize stable outcomes, though not in logical languages. Aranda et al. [2], building on DeGroot-style models, study fairness and consensus in asynchronous value-based settings. Despite the numerical nature of their opinion space, the asynchronous communication regime and convergence objectives make their model conceptually close to BFNs. Strategic variants of belief diffusion, such as those by Caragiannis et al. [6] and Grandi et al. [12], study equilibrium formation in game-theoretic settings. While they consider stability under rational manipulation, BFNs focus on spontaneous, uncoordinated dynamics under iterated belief change.

A natural direction for future work is to characterize the complete set of reachable consensus formulas, a task that is central to understanding the global behavior of BFNs. Since this set may be prohibitively large to compute in general, a closely related and more tractable alternative is to focus on inference: determining whether a given formula is entailed by all agents in every possible consensus. This question connects directly to belief monitoring and predictive reasoning, for instance, evaluating whether a proposed policy will necessarily lead all agents to conclude that it implies reduced emissions. Both directions aim to deepen our understanding of how beliefs evolve and stabilize in logical diffusion frameworks, and remain largely unexplored within the BFN setting.

ACKNOWLEDGMENTS

This work was supported by the following grants: JSPS KAKENHI Grant Number JP25K00375; the program France 2030 under the ANR-23-IACL-0007 grant PostGenIA; and JST CREST Grant Number JPMJCR22D.

REFERENCES

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50 (1985), 510–530.

[2] Jesús Aranda, Sebastián Betancourt, Juan Francisco Díaz, and Frank Valencia. 2024. Fairness and Consensus in an Asynchronous Opinion Model for Social Networks. In *Proceedings of the 35th International Conference on Concurrency Theory (CONCUR'24)*, Vol. 311. 7:1–7:17.

[3] Alexandru Baltag and Sonja Smets. 2022. Diffusion of Logical Beliefs in Networks. *Journal of Logic and Computation* 32, 1 (2022), 147–174.

[4] Emre Botan, Umberto Grandi, and Laurent Perrussel. 2019. Multi-Issue Opinion Diffusion under Constraints. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS'19)*. 828–836.

[5] Markus Brill, Edith Elkind, Ulle Endriss, and Umberto Grandi. 2016. Pairwise Diffusion of Preference Rankings in Social Networks. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI'16)*. 236–242.

[6] Ioannis Caragiannis, Georgios Krimpas, and Ariel D. Procaccia. 2017. Bounding the Inefficiency of Compromise. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17)*. 67–73.

[7] Laurence Cholvy. 2018. Opinion diffusion and influence: a logical approach. *International Journal of Approximate Reasoning* 99 (2018), 71–93.

[8] Adnan Darwiche and Judea Pearl. 1997. On the Logic of Iterated Belief Revision. *Artificial Intelligence* 89, 1-2 (1997), 1–29.

[9] Morris H. DeGroot. 1974. Reaching a Consensus. *J. Amer. Statist. Assoc.* 69, 345 (1974), 118–121.

[10] Dov M. Gabbay and Odinaldo Rodrigues. 2019. Logic-Based Opinion Diffusion in Social Networks. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS'19)*. 828–836.

[11] Ester Gallo, Prakash Panangaden, and Franck van Breugel. 2021. Local Belief Dynamics in Network Knowledge Bases. *ACM Transactions on Computational Logic* 22, 4 (2021), 1–36.

[12] Umberto Grandi, Emiliano Lorini, and Laurent Perrussel. 2021. Games of Influence. *Journal of Logic and Computation* 31, 3 (2021), 763–790.

[13] Marius Iosifescu and Serban Grigorescu. 1990. *Dependence with Complete Connections and Its Applications*. Cambridge University Press.

[14] Sébastien Konieczny, M. Medina Grespan, and Ramón Pino Pérez. 2010. Taxonomy of Improvement Operators and the Problem of Minimal Change. In *Proceedings of the 12th International Conference on Principles of Knowledge Representation and Reasoning (KR'10)*. 161–170.

[15] Sébastien Konieczny and Ramón Pino Pérez. 2008. Improvement Operators. In *Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*. 177–187.

[16] Paolo Liberatore. 2023. Mixed Iterated Revisions: Rationale, Algorithms, and Complexity. *ACM Transactions on Computational Logic* 24, 3 (2023), 27:1–27:49.

[17] Mattia Medina Grespan and Ramón Pino Pérez. 2013. Representation of basic improvement operators. In *Trends in Belief Revision and Argumentation Dynamics*. College Publications, 195–227.

[18] Abhaya Nayak. 1994. Iterated Belief Change Based on Epistemic Entrenchment. *Erkenntnis* 41 (1994), 353–390.

[19] Elise Perrotin and Nicolas Schwind. 2024. Relative Change-Reluctance in Iterated Belief Revision. In *Proceedings of the 21st Pacific Rim International Conference on Artificial Intelligence (PRICAI'24)*. 264–276.

[20] Nicolas Schwind, Katsumi Inoue, Gauvain Bourgne, Sébastien Konieczny, and Pierre Marquis. 2016. Is Promoting Beliefs Useful to Make Them Accepted in Networks of Agents?. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI'16)*. 1237–1243.

[21] Nicolas Schwind, Katsumi Inoue, Sébastien Konieczny, and Pierre Marquis. 2024. BeliefFlow: A Framework for Logic-Based Belief Diffusion via Iterated Belief Change. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI'24)*. 10696–10704.

[22] Nicolas Schwind, Sébastien Konieczny, Jean-Marie Lagniez, and Pierre Marquis. 2020. On Computational Aspects of Iterated Belief Change. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI'20)*. 1770–1776.

[23] Nicolas Schwind, Sébastien Konieczny, and Ramón Pino Pérez. 2022. On the Representation of Darwiche and Pearl's Epistemic States for Iterated Belief Revision. In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR'22)*. 320–330.

[24] Nicolas Schwind and Pierre Marquis. 2018. On Consensus in Belief Merging. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI'18)*. 1949–1956.

[25] Paul C. Shields. 1973. *The theory of Bernoulli shifts*. University of Chicago Press.

[26] Paul Vicol, James Delgrande, and Torsten Schaub. 2016. A Minimization-Based Approach to Iterated Multi-Agent Belief Change. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI'16)*. 124–132.