Coherentist Echo Chambers

Martin Justin*
Faculty of Arts, University of Maribor
Maribor, Slovenia
martin.justin1@um.si

Borut Trpin
Faculty of Arts, University of Ljubljana
Faculty of Arts, University of Maribor
Ljubljana and Maribor, Slovenia
borut.trpin@ff.uni-lj.si

Abstract

This paper investigates the transformation of epistemic bubbles into echo chambers through rational belief-forming processes. Building on Nguyen's distinction between epistemic bubbles—formed by omission—and echo chambers—formed by active distrust, we explore whether echo chambers can emerge without malicious intent. Using a simulation model, we demonstrate that coherence-based reasoning can trap agents in echo chambers, even when they act rationally. This finding challenges the view that echo chambers require intentional manipulation of epistemic trust and suggests that rational cognitive strategies may inadvertently contribute to harmful social epistemic dynamics.

Keywords

echo chambers, coherence, social epistemology, agent-based modeling

1 Introduction

In his seminal discussion, Nguyen [11] argues that, despite both being characterized by groupthink, epistemic bubbles and echo chambers are distinct social epistemic phenomena. Where epistemic bubbles are formed by excluding some relevant information sources, echo chambers are created by actively discrediting specific sources. Specifically, Nguyen asserts that echo chambers require "a significant disparity in (epistemic) trust between members and non-members." Consequently, echo chambers cannot be counteracted by exposing their members to additional sources of information.

Nguyen [11] accepts that epistemic bubbles can form accidentally, e.g., as a consequence of reading certain news sources, or not actively seeking testimony from people beyond your friend group. In contrast, he argues that the creation of echo chambers "is something more malicious," which involves discrediting institutions and individuals without regard for the actual epistemic worth. In his view, echo chambers are often (although not necessarily) created intentionally as a means to "maintain, reinforce and expand power through epistemic control" [11].

However, some later work in social epistemology has contradicted this claim. For example, Baumgaertner and Justwan [3] explore additional mechanisms that can cause the formation of echo chambers, which do not rely on manipulating epistemic trust. Using an agent-based polarization model, they show that echo chambers, where members persist in their beliefs despite exposure to contrary information, can arise via a combination of a social structure and agents' willingness to believe what their community invites them to. Unlike Nguyen's analysis, these two mechanisms seem epistemically benign and do not require agents to distrust specific information sources expressly.

This raises an interesting question: can echo chambers arise in communities where agents act rationally? More specifically, if a community starts as an epistemic bubble, can a belief-forming process that we otherwise deem rationally acceptable prevent the members of the community from breaking out of it, thus transforming the community into an echo chamber?

In this paper, we conduct a simulation study showing that reasoning based on coherence can play this role in specific circumstances. Specifically, taking coherence of their beliefs into account when considering whether to accept new information can prevent agents from escaping an epistemic bubble and trap them in an echo chamber. This suggests that a rational reasoning pattern can cause one of our time's more pernicious social epistemic phenomena.

The rest of this paper is organized as follows. In Section 2, we discuss coherence-based reasoning in more detail. We show that it can be rational in some circumstances, and discuss existing results about its possible negative social epistemic effects. The section concludes with a brief overview of formal measures of coherence. Section 3 presents our model, adapted from our previous work in [9, 21]. In Section 4, we present the simulation study results. Section 5 discusses the results and concludes the paper.

2 Coherence

2.1 The Role of Coherence in (Social) Reasoning

Intuitively, coherence describes how well propositions in a set "hang together" [4]. Take, for example, the difference between these two information sets. S_1 = {"A is a well-regarded author", "Critics praised A's last book", "A's last book was nominated for an important literary prize"}; S_2 = {"A is a well-regarded author", "Today is Thursday", "Python is a general purpose programming language"}. We intuitively sense that S_1 is more coherent than S_2: the propositions in S_1 support each other, while those in S_2 seem completely independent.

Coherentism is usually introduced as a theory of epistemic justification: a belief is justified if and only if it belongs to a coherent system of beliefs [13]. Although BonJour [4] gave one of the most influential modern defences, most epistemologists have since rejected the view. Still, coherence has been thought to play an important epistemic role. For example, Harman's account of reasoned belief revision treats coherence as central to belief management: a new commitment is accepted only if it contributes to the overall coherence of an agent's set of attitudes [7]. This perspective highlights that coherence is not only a theory of justification but also a guide to acceptance and belief revision. Similarly, Angere [1] argues that, despite not being truth-conductive in general, coherence can act as an effective heuristic for choosing a correct information set when more reliable methods are unavailable. On the other hand, Olsson [12]

*Both authors contributed equally to this research.

Information Society 2025, Ljubljana, Slovenia

https://doi.org/10.70314/is.2025.cogni.3

argues that coherence plays an important negative role: "If our beliefs show signs of incoherence, this is often a good reason for contemplating a revision." Goldberg and Khalifa [6] make a similar argument in a social context, arguing that agents are unjustified in holding beliefs that do not cohere with accepted background information of their epistemic community.

In contrast to the work highlighting its positive epistemic role for individual reasoners, research also shows that coherence can have adverse social-epistemic effects. In a model by Singer et al. [20], agents deliberate by exchanging reasons, which either support or oppose a conclusion. Agents also have limited memory: when at capacity, they must "forget" one of their reasons to accept a new one. Agents using a coherence-based strategy for managing memory prefer to forget a reason that runs contrary to the view supported by the totality of their reasons. Authors argue that this is a rational strategy of memory management. Nevertheless, it leads to persistent group polarization between agents.

In essence, coherence underpins rational reasoning patterns but can also cause negative social-epistemic phenomena. Consequently, coherence-based reasoning is an interesting candidate for the context of our research problem: can echo chambers arise in communities where agents employ coherence-based reasoning? To answer this question, we develop an agent-based group learning model, wherein agents use coherence-based reasoning in information gathering and revision of beliefs. Specifically, when updating beliefs based on new information, agents first check whether this information would decrease the coherence of their beliefs. If not, they accept the update. If yes, they ignore the new information and stick to their existing beliefs. In short, agents refuse information that would make their beliefs less coherent.

2.2 Measuring Coherence

Before taking a closer look at this belief updating dynamic and the model in general, coherence must be defined in more detail and operationalized. Several probabilistic measures have been proposed to operationalize coherence, each capturing different intuitions about what makes a set of propositions coherent. Two key intuitions underlie these measures.

The first one is:

> **Deviation from Independence**: The Less independent the propositions in the set are, the more coherent the set is.

The intuition here is that coherence derives from how strongly propositions are interconnected probabilistically. If their probability of occurring together is no more than chance, as in the case of S_2, a set is neither coherent nor incoherent. If they are more likely to hold jointly, the set is coherent, as in the case S_1. In a reverse situation, the set is incoherent. This intuition was formalized by Shogenji [19], as the following measure for a set of propositions $\mathrm{S} = \{\mathrm{A}_1, \ldots, \mathrm{A}_n\}$:

$$coh_S(\mathrm{S}) := \frac{P(\mathrm{A}_1, \ldots, \mathrm{A}_n)}{P(\mathrm{A}_1) \cdots P(\mathrm{A}_n)}$$

The second intuition is:

> **Relative Overlap**: The more overlap among the propositions in a set, the more coherent the set is.

The idea being this intuition is that the coherence stems from agreement between propositions [13]. Propositions agree when if one is true, the others are also true. Probabilistically, we can represent this as a comparison between the probability of the propositions holding jointly (i.e., when all of them are true at the same time) and the probability of their disjunction (i.e., when at least one of them is true). This intuition is formalized by Olsson [14] and Glass [5], who propose the following measure:

$$coh_{OG}(\mathrm{S}) := \frac{P(\mathrm{A}_1, \ldots, \mathrm{A}_n)}{P(\mathrm{A}_1 \vee \cdots \vee \mathrm{A}_n)} = \frac{P(\mathrm{A}_1, \ldots, \mathrm{A}_n)}{1 - P(\neg\mathrm{A}_1, \ldots, \neg\mathrm{A}_n)}$$

The third measure we consider in our model is a crossover measure, recently proposed by Hartmann and Trpin [8], which combines elements of both intuitions:

$$coh_{HT}(\mathrm{S}) := \frac{P(\mathrm{A}_1, \ldots, \mathrm{A}_n)}{1 - P(\neg\mathrm{A}_1, \ldots, \neg\mathrm{A}_n)} \Big/ \frac{P(\mathrm{A}_1) \cdots P(\mathrm{A}_n)}{1 - P(\neg\mathrm{A}_1) \cdots P(\neg\mathrm{A}_n)}$$

3 The Model

3.1 Model Entities: World and Agents

The model we used in this study is a slightly modified version of our model, first presented in [9, 21]. In the model, agents try to form an accurate belief about the world by gathering and sharing information about it. The "world" in the model represents a field of interest or research, e.g., contemporary politics in some country, the stock market, or AI-powered drug development. Agents represent people learning and communicating about it, e.g., social media users, friends, coworkers, or scientists working on the same problem. They all gather information about the topic—read about it, listen to experts, conduct experiments—talk about it with others, and form opinions based on it.

More technically, the model world consists of a Bayesian network (BN), representing a set of probabilistically related events. A BN consists of a directed acyclic graph (DAG) and a conditional probability distribution (CPD) (see [15]). DAG represents events (nodes), either true or false, and conditional dependencies between them (edges). CPD contains information about the likelihood of each event occurring given values of other events. Figure 1 shows one example of a simple BN usually referred to as "Sprinkler", consisting of only four nodes. This is the BN we used in our simulations.[1]

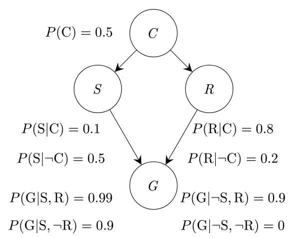

Figure 1: The "sprinkler" network, where C, S, R, G are propositional variables with corresponding values C: "It is cloudy", ¬C: "It is not cloudy", S: "The sprinkler is turned on", ¬S: "The sprinkler is not turned on", R: "It rains", ¬R: "It does not rain", G: "The grass is wet" and ¬G: "The grass is not wet," and the corresponding probabilities of its CPD. Reproduced from [9].

The agents already have an accurate representation of the events in the world and their relations—in other words, they are aware of the structure of the BN in question. What they try to learn is the underlying probability distribution[2]. To do this, they repeatedly observe the world to learn about the values of the individual events. For example, one observation the agents might gather is S_1 = [Cloudy=False, Sprinkler=False, Rain=False,

[1] In principle, we could use any BN; however, larger networks are computationally increasingly demanding.

[2] This is in contrast to some other agent-based models that utilize BN, where agents learn about concrete values of one instantiation of the BN (see especially [2, 17])

Wet Grass=False], another is S_2 = [Cloudy=True, Sprinkler=False, Rain=True, Wet Grass=True], etc. Agents then fit these observations over the BN via maximum likelihood estimation (MLE). That is, they form a subjective belief about the conditional probability distribution that is the most likely given their observations about the world.

Agents also share information among each other. They can be connected in different communication networks, determining who can share information with whom. In our simulations, we use three different such networks. A cycle connects each agent only to two closest agents; a wheel is similar to the cycle with the addition of one central agent connected to all other agents; a complete network connects each agent to all other agents.

3.2 The Setup: Coherence-Based Reasoning, Misleading Information, and Dynamic Environment

We wish to explore whether coherence-based reasoning can trap agents into an echo chamber by preventing them from changing their beliefs in response to accurate information about the world. We need to extend the above model in three ways to model such a situation.

First, we need to add coherence-based reasoning. We do this as follows. Some agents do not simply form a belief about the CPD based on their information. Instead, when presented with new information, they first check whether this new belief is at least as coherent as their existing one. To do this, they first determine the most probable state of the world based on the distribution incorporating new information (e.g., that it is cloudy, the sprinkler is off, it is raining, and the grass is wet). Then, they check how coherent this state is using one of the coherence measures presented above. If this state is at least as coherent as the state that is most probable based only on their existing information, they accept the new information. If it is less coherent, they reject new information.

Second, to mimic an epistemic bubble, we allow for situations in which agents fail to form accurate beliefs, not because of their own selective exposure, but because they lack access to reliable information. In our model this is captured in two ways: agents may start with inaccurate priors, implemented by setting their initial CPD as a parameter, and they may occasionally receive input from a misleading BN rather than the real-world BN. The misleading BN differs from the real one in its CPD. For example, while in the real world the sprinkler substantially increases the likelihood of wet grass, this relation may be absent in the misleading BN.

Thirdly, we place agents in a dynamic epistemic environment, meaning the chance of gathering misleading information decreases over time. This represents a gradual breaking of an epistemic bubble: agents start with inaccurate priors and are likely to gather misleading information. Gradually, they begin receiving more accurate information (we determine the rate of change as a parameter of the model). Usually, this would mean that agents would also gradually start to form more accurate beliefs. We are interested in whether coherence-based reasoning can prevent this, trapping agents in an echo chamber where they ignore the accurate source of information.

3.3 The Procedure

The simulations of the model proceed in rounds or steps. The following parameters are determined before the start of the simulation: the number of agents in a group, the number of agents using coherence-based reasoning, the way agents share information between each other, agents' priors, the chance of gathering misleading information at the start, and the rate of change in this chance. Each round of the simulation then consists of the following actions:

(1) Agents collect information.
(2) Agents share information.
(3) Agents update their beliefs based on their type:
    (a) Coherentist agents first check the coherence of their new belief, and accept it only if it is at least as coherent as their prior belief.
    (b) Other agents straightforwardly update based on the new information.

4 Results

We simulated groups of 10 agents with 2, 5, or 8 coherentist agents. The agents were connected in a cycle, wheel, or a complete network; coherentists and other agents were shuffled and placed randomly, so their distribution on the network wouldn't bias the results. In all groups, agents' starting subjective probability distribution (i.e., their belief) was the same as that of the misleading information source. We generated the misleading information source by randomly changing the distribution of the base "Sprinkler" BN. The only constraint was that the misleading information source was more coherent, i.e., it scored better on each of the three coherence measures.

Each agent drew 100 samples from the information source per round. At the start of the simulation, they had a 100% chance of drawing information from the misleading source. Throughout the simulation, this chance was gradually reducing by 1%. This means that from round 100 onward, agents only received accurate information about the world. To test the persistence of any effect coherence-based reasoning might have, we ran each simulation for 300 rounds.

Figure 2 shows how the accuracy of agents' beliefs changes over time. The belief accuracy is measured as Kullback-Leibler (KL) divergence of the agent's probability distribution from the actual world's distribution, quantifying the discrepancy between two probability distributions [10]; consequently, lower values thus mean more accurate beliefs. The red line represents a belief of a non-coherentist agent, averaged over all parameters. The three blue lines represent beliefs of coherentist agents for different network structures, averaged over other parameters.

The figure shows that agents who do not employ coherence-based reasoning reliably form accurate beliefs about the world. This is expected—these agents take their information at face value, so nothing prevents them from updating on more accurate information. On the other hand, coherentist agents, on average, retain inaccurate beliefs despite being presented with accurate information. After round 100, agents do not receive misleading information; the fact that coherentist agents' beliefs on average do not change much after that shows that they practically insulate themselves from it. In other words, they seem to actively ignore an accurate information source, which is how Nguyen [11] defines echo chambers. This result is robust for different communication networks, but seems to be increased by sparser communication.

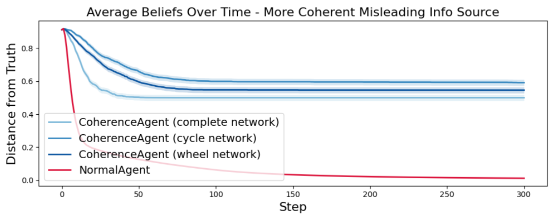

Figure 2: Distance of agents' beliefs from truth over time for a more coherent source of misleading information. The X-axis represents steps in the simulation, and the Y-axis represents distance from truth measured as the KL-divergence. The shaded regions represent 95% confidence intervals.

Table 1 gives a more comprehensive picture of the coherentist agent's average belief accuracy (expressed as distance from truth) for different combinations of parameters. These results again show that the communication structure impacts the results. In most cases, agents connected in a complete network ended up closer to the truth than agents connected in the wheel or the cycle networks in comparable situations. Additionally, we can see that when coherence was measured as deviation from independence [19], the average effect on belief accuracy was the lowest.

Coh. Measure	Nr. Coh.	Complete	Wheel	Cycle
Olsson-Glass	2	0.52 (0.06)	0.56 (0.05)	0.67 (0.04)
Olsson-Glass	5	0.54 (0.06)	0.59 (0.03)	0.71 (0.02)
Olsson-Glass	8	0.53 (0.06)	0.59 (0.03)	0.65 (0.02)
Shogenji	2	0.35 (0.05)	0.46 (0.05)	0.48 (0.05)
Shogenji	5	0.48 (0.06)	0.43 (0.03)	0.42 (0.02)
Shogenji	8	0.38 (0.04)	0.41 (0.03)	0.47 (0.03)
Hartmann-Trpin	2	0.58 (0.06)	0.56 (0.05)	0.67 (0.04)
Hartmann-Trpin	5	0.63 (0.06)	0.63 (0.03)	0.66 (0.04)
Hartmann-Trpin	8	0.52 (0.06)	0.63 (0.03)	0.63 (0.02)

Table 1: Average distance from truth of agents' belief at the end of the simulation for different combinations of parameters (with one standard error in parentheses).

5 Discussion

These results show that coherence-based reasoning possibly leads to the creation of echo chambers. In contrast to other proposed mechanisms of echo chamber creation, e.g., active mistrust of certain information sources [11], coherence-based reasoning is not irrational; on the contrary, some argue that it can have positive epistemic value. This implies that potentially rational reasoning patterns can lead to pernicious social epistemic phenomena.

That said, our study has two significant limitations. First, the misleading information source presented a more coherent picture of the world than the truth. This is not an unreasonable assumption: conspiracy theories and misinformation often offer more straightforward, intuitive, and coherent explanations of complicated events than evidence. Nevertheless, it might importantly affect our results. Given that coherentist agents consider information based on its coherence, accurate information might manage to overcome a less coherent misinformation source even for these agents. Running simulations with an alternative misinformation source that scores worse on the coherence measures, is thus a vital robustness check we must consider in the future.

The second limitation concerns the nature of our study. The agent-based model we used is highly idealized—learning about the world is represented by drawing samples from a BN, communication is represented by sharing these samples, belief revision by MLE, a mathematical procedure, etc. As various philosophers of science have pointed out, such highly idealized models cannot be used to make real-world predictions, provide actual explanations for phenomena, or suggest normative prescriptions [18, 16, 22]. It would be wrong to conclude that coherence-based reasoning is the cause of people's persistent false beliefs.

Although idealized models cannot explain real-world phenomena, they provide so-called "how-possible explanations" [16]. In our case, the results point to coherence-based reasoning as a possible explanation for echo chamber formation. Empirical studies are needed to show this link in practice.

Acknowledgements

The authors acknowledge the financial support from the Slovenian Research and Innovation Agency (ARIS, project J6-60107 and research core funding No. P6-0144).

References

[1] Staffan Angere. 2008. Coherence as a Heuristic. *Mind*, 117, 465, (Jan. 2008).
[2] Leon Assaad, Rafael Fuchs, Ammar Jalalimanesh, Kirsty Phillips, Leon Schoeppl, and Ulrike Hahn. 2023. A Bayesian agent-based framework for argument exchange across networks. *arXiv.* eprint: 2311.09254 (cs.SI).
[3] Bert Baumgaertner and Florian Justwan. 2022. The preference for belief, issue polarization, and echo chambers. *Synthese*, 200, (Sept. 2022), 412, 5, (Sept. 2022). DOI: 10.1007/s11229-022-03880-y.
[4] Laurence BonJour. 1985. *The Structure of Empirical Knowledge.* Harvard University Press, Cambridge, MA.
[5] David H. Glass. 2002. Coherence, explanation, and Bayesian networks. In *Artificial Intelligence and Cognitive Science, 13th Irish Conference, AICS 2002.* Michael O'Neill, Richard F. E. Sutcliffe, Conor Ryan, Malachy Eaton, and Niall J. L. Griffith, editors. Springer, Berlin, 177–182.
[6] Sanford C. Goldberg and Kareem Khalifa. 2022. Coherence in Science: A Social Approach. *Philosophical Studies*, 179, 12, (Dec. 2022), 3489–3509. https://link.springer.com/article/10.1007/s11098-022-01849-8.
[7] Gilbert Harman. 1986. *Change in View: Principles of Reasoning.* The MIT Press.
[8] S. Hartmann and B. Trpin. Forthcoming. Why coherence matters? *The Journal of Philosophy.*
[9] Martin Justin and Borut Trpin. 2025. Coherence-Based Evidence Filtering: A Computational Exploration. In *Proceedings of the Annual Meeting of the Cognitive Science Society 47.*
[10] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics*, 22, 1, 79–86.
[11] C. Thi Nguyen. 2020. Echo chambers and epistemic bubbles. *Episteme*, 17, (June 2020), 141–161, 2, (June 2020). DOI: 10.1017/epi.2018.32.
[12] Erik Olsson. 2023. Coherentist Theories of Epistemic Justification. In *The Stanford Encyclopedia of Philosophy.* (Winter 2023 ed.). Edward N. Zalta and Uri Nodelman, editors. Metaphysics Research Lab, Stanford University.
[13] Erik J. Olsson. 2022. *Coherentism.* Cambridge University Press, Cambridge.
[14] Erik J. Olsson. 2002. What is the problem of coherence and truth? *The Journal of Philosophy*, 99, 5, 246–72.
[15] Judea Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, San Francisco, CA.
[16] Alexander Reutlinger, Dominik Hangleiter, and Stephan Hartmann. 2018. Understanding (with) toy models. *The British Journal for the Philosophy of Science*, 69, (Dec. 2018), 1069–1099, 4, (Dec. 2018). DOI: 10.1093/bjps/axx005.
[17] Klee Schöppl. 2025. Industry Influencing Collective Scientific Reasoning: A Bayesian, Agent-based Exploration. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 47.*
[18] Dunja Šešelja. 2023. Agent-based modeling in the philosophy of science. (2023). https://plato.stanford.edu/entries/agent-modeling-philscience/#PeerDisaScie_1.
[19] Tomoji Shogenji. 1999. Is coherence truth conducive? *Analysis*, 59, 4.
[20] Daniel J. Singer, Aaron Bramson, Patrick Grim, Bennett Holman, Jiin Jung, Karen Kovaka, Anika Ranginani, and William J. Berger. 2019. Rational social and political polarization. *Philosophical Studies*, 176, 9, 2243–2267.
[21] Borut Trpin and Martin Justin. 2025. Coherence as a constraint on scientific inquiry. *Synthese.*
[22] Michael Weisberg. 2007. Three kinds of idealization. *Journal of Philosophy*, 104, 639–659, 12. DOI: 10.5840/jphil20071041240.

Received 22 August 2025; revised 16 September 2025; accepted 16 September 2025