Title
Coherence as a Norm of Evidence Sharing

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Justin, Martin
Trpin, Borut

Publication Date
2026

Coherence as a Norm of Evidence Sharing

Martin Justin (martin.justin1@um.si)[1] **& Borut Trpin**[1,2]
[1]Department of Philosophy, Faculty of Arts, University of Maribor
[2]Department of Philosophy, Faculty of Arts, University of Ljubljana

Abstract

Human agents constantly exchange information, but deciding which evidence is worth transmitting is a fundamental cognitive challenge. This paper investigates "coherence" as a heuristic for regulating evidence sharing in groups. Extending previous models of individual coherence-based filtering, we employ computational simulations to analyze agents who selectively share evidence based on its coherence with their current beliefs. We find that coherence-based selective sharing is significantly more effective than individual evidence filtering. It helps preserve priors in overly noisy epistemic environments, while still allowing updating when evidence is noisy but truth-conducive. However, this strategy becomes a liability when misleading evidence forms a coherent alternative, steering group learning away from the truth. These results illuminate the conditions under which coherence-based heuristics improve or distort collective inquiry.

Keywords: social learning; coherence; heuristics; evidence sharing; computational modeling; collective inquiry

Introduction

Human agents rarely learn in isolation. Much of everyday reasoning, scientific inquiry, and social learning proceeds through communication: agents gather information from the world, exchange it with others, and update their beliefs on the basis of what is shared. Yet evidence sharing is neither costless nor risk-free. Communicating information requires attention and integration effort, and it can expose individuals and groups to epistemic and reputational harm when shared information is noisy, misleading, or systematically biased (see, for an example of reputational harm, Altay et al., 2022). As a result, agents face a fundamental cognitive problem: not only how to update on evidence, but also which evidence to treat as worth integrating and transmitting.[1]

This paper investigates three closely related questions that arise from this problem: (1) Under what conditions is it cognitively and epistemically advantageous for agents to refrain from sharing some of the evidence they collect? (2) Can simple, resource-bounded heuristics guide such selective sharing in a way that improves collective belief accuracy? (3) When do these same heuristics become liabilities, systematically distorting group-level learning? These questions are central to current work in cognitive science on heuristic reasoning, social learning, and the spread of misinformation, but they have received relatively little attention in formal models of collective inquiry. While norms of assertion and evidence sharing have been studied formally in philosophy, these efforts have mostly focused on the structure of groups' communication (see Zollman, 2007) or on explaining negative social-epistemic phenomena, such as polarization (see Assaad and Hahn, 2024; Olsson, 2012).

One prominent candidate for regulating evidence use is *coherence*. Across philosophy, psychology, and cognitive science, coherence has been treated as a cue for plausibility (e.g., Harris and Hahn, 2009; Hartmann and Trpin, Forthcoming; Simon and Read, 2025; Thagard, 2002): information that "hangs together" with what one already believes is often treated as more credible, while incoherent information is met with skepticism. Although coherence has long been discussed as an epistemic value, recent work suggests that it may also function as a *heuristic* for managing uncertainty under limited cognitive resources. In particular, coherence considerations appear to play a role in epistemic vigilance (Sperber et al., 2010), belief filtering (Justin & Trpin, 2025), and resistance to misleading information (Trpin & Justin, 2025), especially when agents lack reliable ways of assessing evidence quality directly.

In particular, Trpin and Justin recently used computational simulations to study *coherence-based evidence filtering* at the individual level. Their results show that, in environments characterized by substantial random noise, rejecting evidence that could reduce the coherence of one's beliefs if it were believed can improve an individual's belief accuracy. At the same time, this strategy backfires in environments where misleading evidence forms a coherent alternative picture of the world. These findings highlight coherence as a double-edged heuristic: potentially protective against noise, but dangerously susceptible to systematic bias.

However, existing work in this vein treats coherence-based filtering as an individual-level mechanism. Agents either update on evidence or discard it, but the social dynamics of communication and shared learning remain unexamined. This is a significant gap. In real-world inquiry, with scientific collaboration being a paradigmatic case, agents not only update their own beliefs but also decide what evidence to pass on to others (for a related point with respect to scientific communities, see Goldberg and Khalifa, 2022). Norms governing evidence sharing may therefore have consequences that differ

[1]We use "evidence" broadly for information treated as evidential by non-omniscient agents; if evidence were restricted to guaranteed truth-tracking signals, the selective-sharing problem would be trivial.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

importantly from those governing private belief updating.

Existing coherence-based models treat filtering as an individual mechanism, leaving the dynamics of communication and shared learning unexplored. We extend the simulation model introduced by Justin and Trpin (2025) to groups in which agents collect evidence independently, share it selectively based on coherence, and update only on the group's shared pool. This lets us test when coherence operates more effectively as a social gate on transmission than as a private filter.

Theoretical Background

Evidence sharing is central to collective inquiry but occurs under uncertainty about reliability. This creates pressure for inexpensive heuristics that regulate what information is worth transmitting. One candidate is coherence-based plausibility screening: information that fits with background beliefs and hangs together internally is treated as safer to pass on than information that does not (Simon & Read, 2025; Simon et al., 2015).

This idea is familiar from work on epistemic vigilance, which largely focuses on recipients (Sperber et al., 2010). However, similar considerations plausibly apply to communicators. In many real-world contexts, scientific collaboration included, agents lack detailed knowledge of their audience's beliefs, evidential standards, or background assumptions. Under such uncertainty, a robust policy is to screen candidate evidence relative to one's own current best model of the world and to refrain from sharing information that appears epistemically problematic by one's own lights. This does not require any assumptions about audience tuning or shared common ground; rather, it reflects a conservative strategy for avoiding the transmission of suspect information when verification is costly.

Selective sharing can also be supported by reputational considerations. Recent work on misinformation suggests that people are often reluctant to share content they perceive as inaccurate, even when it aligns with their prior attitudes, because doing so risks reputational harm (Altay et al., 2022). Importantly, such reluctance does not presuppose that agents can reliably assess accuracy. Instead, agents appear to rely on internal plausibility assessments when deciding what to pass on. Coherence is not equivalent to truth or accuracy (Bovens & Hartmann, 2003), but it can function as an easily available cue for whether a piece of information is epistemically suspect and therefore not worth integrating or transmitting.

To connect these general considerations to a tractable model, we must clarify what we mean by *coherence*. Coherence is not a unitary notion across philosophical, psychological, and cognitive-scientific traditions. In some contexts, it is treated as a broad, holistic virtue of belief systems involving mutual support, explanatory integration, and the absence of internal tension (BonJour, 1985; Thagard, 2002). In this paper, however, our focus is narrower. We are concerned with *epistemic coherence of information sets*: roughly, the degree to which the elements of a body of information fit together in a way that makes the body plausible as a whole. We do not attempt to capture broader virtues such as explanatory unification, simplicity, or scope, although these may correlate with epistemic coherence in some contexts.

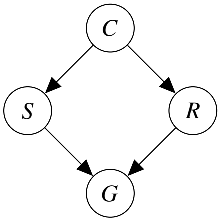
Following a Bayesian tradition, we operationalize epistemic coherence in probabilistic terms, treating it as a property of propositional information sets represented by probabilistic relations among their members (Olsson, 2022). Intuitively, a coherent set "hangs together": its elements are not merely jointly possible, but probabilistically connected in specific ways. For example, the set {All ravens are black, This bird is a raven, This bird is black} is coherent because its members stand in clear inferential and probabilistic relations, whereas a set such as {This chair is dark gray, Electrons are negatively charged, Today is Thursday} lacks such connections (BonJour, 1985, p. 95).

Because our agents face uncertainty about evidence quality, we treat coherence as a proxy for plausibility rather than as a marker of truth. Probabilistic coherence measures make this idea precise but encode different intuitions. Deviation-from-independence measures (e.g., Shogenji, 1999) treat sets as more coherent when their elements are probabilistically dependent, whereas relative-overlap measures (Glass, 2002; Olsson, 2002) treat sets as more coherent when their propositions overlap in probability mass. We also use a hybrid measure that combines these ideas (Hartmann & Trpin, Forthcoming).

These measures are not merely mathematically distinct; they can diverge in their verdicts about which information sets are more coherent in particular cases. This implies that any claim about the epistemic role of coherence-based heuristics can, in principle, depend on the adopted explication of coherence. Rather than treating this as a defect, we treat measure dependence as theoretically informative. If an effect holds across measures that capture different coherence intuitions, it is less likely to be an artifact of a particular formalization. Conversely, when measures diverge, this helps identify which coherence intuitions are responsible for which epistemic consequences. We therefore treat robustness across measures as evidence of generality, and divergence as theoretically diagnostic rather than problematic.

Taken together, the foregoing motivates coherence-gated sharing as a plausible *social* heuristic: in environments where accuracy is hard to assess directly, agents may rely on coherence as a cheap plausibility cue when deciding what to treat as evidentially usable and worth transmitting. Our contribution is to turn these qualitative considerations into a precise, testable mechanism and to study its downstream group-level consequences.

Concretely, we model agents who (i) evaluate newly collected samples relative to their current evidential state, (ii) approximate the coherence of what the sample supports by estimating the induced distribution and assessing the coher-



Probability	Value
$P(\mathrm{C})$	0.5
$P(\mathrm{S}\vert\mathrm{C})$	0.1
$P(\mathrm{S}\vert\neg\mathrm{C})$	0.5
$P(\mathrm{R}\vert\mathrm{C})$	0.8
$P(\mathrm{R}\vert\neg\mathrm{C})$	0.2
$P(\mathrm{G}\vert\mathrm{S},\mathrm{R})$	0.99
$P(\mathrm{G}\vert\mathrm{S},\neg\mathrm{R})$	0.9
$P(\mathrm{G}\vert\neg\mathrm{S},\mathrm{R})$	0.9
$P(\mathrm{G}\vert\neg\mathrm{S},\neg\mathrm{R})$	0

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", and so on for S: "Sprinkler on", R: "It rains," and G: "Grass is wet." The figure is reproduced from Trpin and Justin (2025).

ence of the most probable world-state under that distribution, and (iii) share (and retain for updating) the sample only if it supports a more coherent picture than the agent's current one. Updating, by contrast, is fully social: in each round, agents update on whatever evidence is shared within the group. This implementation captures a minimal, self-relative coherence gate. It also makes clear, given the plurality of coherence measures, why we treat measure dependence as theoretically informative: different coherence intuitions may yield different vulnerabilities and benefits. The simulations reported below are designed to test when such a coherence-based sharing norm improves collective accuracy (e.g., by suppressing random noise) and when it predictably backfires (e.g., by privileging a coherent but systematically biased alternative picture of the world).

Simulation Model

This model is a modified and extended version of the model first presented in Justin and Trpin (2025) and Trpin and Justin (2025).[2] As the original model, our new model still incorporates the world, consisting of probabilistically related events, and agents, who try to learn about the world by gathering evidence. More specifically, the world is represented by a Bayesian network (BN), consisting of a directed acyclic graph (DAG) and conditional probability distributions (CPDs) (Pearl, 1988). The DAG consists of nodes, which stand for possible events in the world, and edges, which represent relations between them. Conditional probability distributions provide information about the likelihood of an event being true or false, given different values of the parent nodes. Figure 1 shows an example of a Bayesian network. This specific BN is usually referred to as "Sprinkler". We also use it in our simulations.[3]

Agents try to form accurate beliefs about the world by observing it. In contrast to other similar models (e.g., Assaad et al., 2023), they aim to learn the BN in its entirety, starting with an accurate representation of the nodes and edges but lacking access to the true prior conditional probability distributions. Instead, each agent starts with randomly generated ones that can be more or less similar to the real one – the maximum difference from the real one is a variable we study below.[4] Agents update their inaccurate subjective CPDs by collecting sample observations from the world. Each observation consists of a set of four True or False values, one for every node. These are generated randomly based on the underlying probabilities.[5] For example, one observation agents might gather is S_1 = [Cloudy=True, Sprinkler=False, Rain=True, Wet Grass=True], another is S_2 = [Cloudy=False, Sprinkler=True, Rain=False, Wet Grass=True], and so on. Agents take many such samples across multiple rounds of the model, fitting them to the world's structure via maximum-likelihood estimation.

Agents do not always collect accurate evidence. The incoming evidence can be misleading in two distinct ways. First, it can be randomly noisy, meaning that a certain percent (determined at the beginning of the simulation as a parameter) of the values in the samples are randomly chosen and changed. Such "randomly noisy" evidence is found in our everyday life, for example, when ordinarily reliable information or news sources make mistakes, and in science, where it is usually due to random measurement error. Second, agents can also receive systematically biased evidence. This means that, in each round, they have some chance (again, determined by a model parameter) of sampling an alternative BN rather than the one that represents the world. This alternative BN shares the same DAG as the world but has modified conditional probability distributions. A real-world example of such evidence is listening to a grifter. A grifter does not simply make random, inaccurate claims but tries to convince you that some alternative picture of the world—perhaps one where a particular dietary supplement, rather than chemotherapy, cures cancer, or where buying a specific investment product guarantees exceptional returns—is accurate. Systematically biased evidence can vary in how coherent the alternative picture of the world is. In the next section, we look at both biased evidence that comes from a more coherent world than the truth and evidence that comes from a less coherent world than the truth.

The new model differs from the one presented in Justin and Trpin (2025) and Trpin and Justin (2025) in two main ways. First, we introduce groups of agents. In the original

[2]Code available in this OSF repository.

[3]In prior work on individual coherence-based filtering, similar results were obtained on other Bayesian networks (e.g., Asia; Trpin and Justin, 2025). While we focus on Sprinkler for comparability, the key mechanism depends on coherence relations rather than network topology.

[4]The similarity is determined by a parameter called "variation of priors." This parameter determines how much individual conditional probability functions of agents can differ from the real one in terms of absolute difference and respecting the probabilistic range of [0,1]. With variation of priors at 0, they are the same, and with 1, they are completely random.

[5]Sampling is done in topological order, by first sampling values of root and then proceeding to other variables using their CPD and the values of already sampled parent nodes (Taskesen, 2020).

model, agents always explored the world independently (i.e., as lone individuals). In this modified model, agents learn in groups and can share evidence. In our model, they share evidence with all other group members, but this assumption can be relaxed by introducing different communication structures Zollman (2007). The agents share this evidence once every round and update their beliefs after all group members have finished with evidence collection.

The second difference lies in the types of agents that populate the model. The original model contained two kinds of agents. One were Normal agents, who took their evidence at face value and updated on whatever they collected. They represent idealized reasoners who never miss an opportunity to gather evidence and always update their total evidence. The other type of agents were Coherentist agents, who only updated their beliefs in response to new evidence if it did not decrease the coherence of their beliefs about the world. In the new model, we still use Normal agents as a benchmark but replace Coherentists with Selective sharers, who follow the coherence-based norm of evidence sharing.

These agents only share evidence that presents a more coherent picture than the one they already believe in. To do this, they first estimate what conditional probability distribution over the events in the world the incoming evidence supports. Then, they determine the most probable state of the world based on the learned distribution. In "Sprinkler", this may be 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 the state is more coherent than the one supported by the agent's existing evidence, the agent accepts the new evidence and shares it with other group members. Otherwise, the agent discards the evidence. In the case that a Selective sharer rejects the incoming evidence, the agent does not update on this evidence but will update on other evidence shared by other agents in the group.[6]

Schematically, a Selective sharing agent acts in the following manner:

1. The agent gathers a sample.
2. The agent checks the coherence of the sample.
3. Then:
   (a) If the sample is more coherent than their existing belief about the world, the agent shares it.
   (b) If the sample is less coherent, the agent rejects it.
4. The agent waits till all other agents in the group have collected (and shared) their samples.
5. The agent updates their belief based on shared samples.

[6]There is an important difference between how Coherentist agents and Selective sharers evaluate coherence. Selective sharers focus on the coherence of the evidence itself. Coherence agents, on the other hand, first update their beliefs based on the new evidence. Then, they check the coherence of this updated belief and compare it with the coherence of their previous belief. If the new belief is more coherent, they proceed with it; otherwise, they reject the newly collected evidence and revert to the previously held belief.

Parameter	Values
Number of agents	5
Samples per round	100
Coherence Measures	Shogenji, Olsson-Glass, Hartmann-Trpin
Type of misleading evidence	noisy evidence, less and more coherent biased evidence
Information "noise"	0.05–0.45 (0.05 increments)
Variation of priors	0.05–1 (0.5 increments)

Table 1: Parameter values used in the reported simulations.

Results

We ran simulations with either 5 Normal agents or 5 Selective sharing agents for each possible combination of parameter values in Table 1. Each simulation proceeded for 50 steps. We simulated each parameter value combination 30 times to ensure a degree of robustness.

Noisy evidence

Figure 2 reports results for simulations where agents received randomly noisy evidence. The x-axis shows different variations of priors (higher values mean less accurate priors), while the y-axis shows the level of noise in evidence. The values in the heatmap represent non-normalized differences between the average belief accuracy of a group of Normal agents and the average accuracy of a group of Selective sharing agents at the end of the simulation for a given combination of noise and priors. The belief accuracy of an agent is measured as a Kullback–Leibler divergence (Kullback & Leibler, 1951) between the subjective conditional probability distribution of that agent and the conditional probability distribution of the "Sprinkler" BN representing the world. A smaller KL-divergence indicates more accurate beliefs; consequently, positive values on the chart indicate situations in which Selective sharers ended up with more accurate beliefs. The values reported on the chart are averaged across individual agents within a group, across the 30 iterations, and across the three different coherence measures.

Figure 2 shows that the effect of coherence-based evidence sharing varies with the situation in which agents find themselves. We can carve up the heatmap into three broad regions. The region where Selective sharers over-perform, in the top-left corner, represents a situation in which agents start with good-enough priors that updating on noisy evidence would lead them further from the truth. Here, selective sharing helps by filtering out most of the evidence and conserving agents' priors. The region in the top-right corner, where Selective sharing hurts most, represents situations in which both priors are inaccurate and the evidence is noisy. Here, updating does pay, so the fact that selective sharers sometimes refuse to update on incoming evidence leaves them worse off. The third region is perhaps the most interesting one. It forms a triangle connecting the two bottom corners of the heatmap to the center of its upper side, representing the parameter space where the

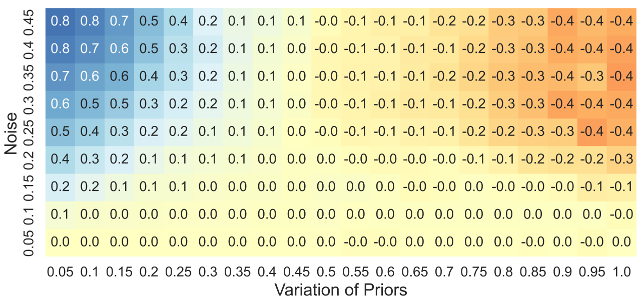


Figure 2: Average difference in accuracy (Normal minus Selective Sharing) at round 50 in the "Sprinkler" network with noisy evidence. Accuracy is measured as KL-divergence; numbers higher than zero indicate that Selective Sharing agents were more accurate than Normal ones.

accuracy difference is negligible. These are situations where the incoming evidence is either similarly distorted to agents' priors or it much more closely tracks the truth. In these cases, selective sharers tend to mostly update on evidence.

At first glance, these results look similar to those reported by Justin and Trpin (2025). However, there are important differences. The above results suggest that selective sharing is more adaptive than coherence-based evidence filtering. Previous results show that coherence-based evidence filtering is a crude heuristic, which almost exclusively works by conserving agents' priors. As such, it is helpful in epistemically polluted environments where evidence is misleading, and unhelpful in environments where evidence is truth-tracking. Coherence-based selective sharing, on the other hand, is a more sophisticated strategy. It preserves priors when updating on new evidence worsens belief accuracy. In addition, and in contrast to evidence filtering, it allows agents to update at least in some cases when evidence much more closely tracks the truth than priors do.

Furthermore, in contrast to coherence-based evidence filtering, the results reported here vary significantly for different measures of coherence. Specifically, the Selective sharing agents' poor performance in the top-right part of the chart does not replicate across all three coherence measures. Instead, agents who use Shogenji or Hartmann-Trpin measures of coherence experience no penalty in this area as they closely track the performance of Normal agents. On the other hand, agents using the Olsson-Glass measure significantly underperformed, thereby bringing down the average. An example is shown in Figure 3, where we see the accuracy difference for different variations of priors and a noise level of 0.45. Colored lines represent results for different coherence measures, while the emphasized black line represents an average (its values match the top row in Figure 2). This means that, conditional on picking the right measure of coherence, Selective sharing agents can either closely match or outperform Normal agents across the board.

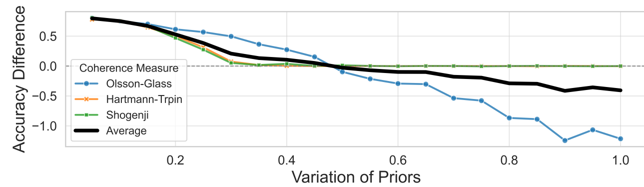


Figure 3: Differences in accuracy (Normal minus Selective Sharing) at round 50 and noise 0.45. Numbers higher than zero again favor Selective Sharing agents. The "Average" values match the top row in Figure 2.

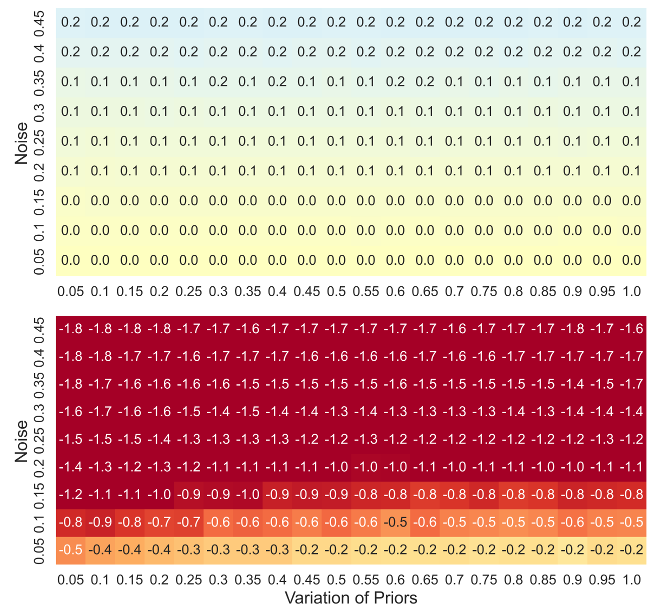


Figure 4: Difference in accuracy (Normal minus Selective Sharing) at round 50 with systematically biased evidence, averaged over different coherence measures. The top chart shows results for less coherent and the bottom for more coherent biased evidence.

Biased Evidence

Figure 4 reports results for simulations with biased evidence, maintaining the layout established in Figure 2. The top chart shows results for less coherent misleading evidence and the bottom one shows results for more coherent misleading evidence.

These results are not surprising. Selective sharers filter out less coherent and chase more coherent evidence. In cases of less coherent biased evidence, this means that agents are able to identify and reject misleading evidence reliably. Consequently, their performance relative to that of Normal agents is simply a function of the amount of bias in the environment. The more noise there is, the larger the gap between Selective sharers, who filter out misleading evidence, and Normal agents, who update on all evidence, including misleading evidence.

Precisely the opposite is true with more coherent biased evidence. Here, Selective sharers are again able to neatly separate evidence from the real world from biased evidence. However, since evidence from the real world is now, on average, less coherent, they filter it out and instead update mostly on biased evidence. In other words, Selective sharing acts as an anti-reliable strategy—it predictably leads agents away from truth. Justin and Trpin (2025) made a similar conclusion about evidence filtering. Nevertheless, the results reported here give a fuller picture. In the case of selective sharing, the influence of biased evidence crucially depends on its coherence. Previous results didn't account for this aspect.

These results demonstrate that selective sharing agents are vulnerable to being misled. Specifically, they are prone to accepting a deceptive narrative as long as it is more coherent than their existing beliefs about the world. It is unclear whether human reasoners employ heuristics similar to coherence-based selective sharing; these results illustrate how such a heuristic could increase susceptibility to misinformation.

Discussion

Our aim was not to show that coherence is a good guide to truth in general, but to test a more specific hypothesis suggested by work on epistemic vigilance and plausibility-based sharing: when agents cannot directly assess reliability, they may regulate communication using cheap internal cues such as fit with background beliefs and internal consistency (Altay et al., 2022; Simon et al., 2015; Sperber et al., 2010). The simulations make two points that are easy to miss if one focuses only on individual belief revision.

First, coherence is markedly more effective as a *social gate* than as a private filter. In the individual-filtering setting, coherence mainly functions as a brake on updating, helping only by conserving priors under noise (Justin & Trpin, 2025). When moved to the social level, the same cue becomes a mechanism for *selective transmission*: it reduces the amount of contamination entering the group's shared evidential pool while still allowing group-level updating when the pool remains informative. This shift connects coherence-based reasoning to the epistemic-vigilance literature in an interesting way: the relevant benefit is not that coherence reliably tracks truth, but that it can regulate exposure to low-quality information under resource constraints.

Notably, coherence is more effective as a social gate than as a private filter because it regulates what enters the group's shared evidence pool without shutting down learning altogether. Individual coherence-based filtering tends to collapse inquiry back onto priors: once agents start discarding discordant samples, they reduce the very data needed to correct mistaken starting points. By contrast, coherence-gated sharing preserves a form of epistemic division of labor, where agents still explore independently, but the group's aggregate evidence is partially protected from contamination.

Second, the model clarifies a structural way in which coherence-based norms can be double-edged. Work on misinformation suggests that people often rely on plausibility cues when deciding what to share (Altay et al., 2022). Our results show that, if coherence is used as such a cue at the level of evidence transmission, then when the environment contains a coherent but biased alternative, this heuristic can become systematically anti-reliable. In these cases, coherence-gated sharing preferentially transmits biased evidence while suppressing less coherent but truth-tracking samples, leading groups away from the truth. This does not, by itself, explain the observed diffusion patterns, but it identifies a minimal mechanism by which coherence-sensitive sharing rules can become anti-reliable when coherence and truth come apart.

Although the model formalizes coherence probabilistically, we do not assume that human agents explicitly compute coherence. Rather, coherence functions here as a proxy for inexpensive plausibility cues such as internal consistency and fit with background beliefs. The selective sharing rule maps naturally onto cognitive mechanisms associated with epistemic vigilance and intuitive credibility assessment. This interpretation suggests testable predictions: in group learning tasks with noisy or biased information, participants should preferentially share internally coherent information even when coherence is experimentally dissociated from accuracy.

Finally, coherence-measure dependence is not merely a technicality but highlights that different formal measures encode distinct intuitions about what it is for information to "hang together" (e.g., deviation from independence vs. relative overlap). Our results show that these differences can translate into distinct vulnerability profiles when coherence is used as a gate on evidence sharing. This suggests a useful empirical direction: if human sharing behavior resembles coherence-gated sharing, then identifying which formal measures best capture it becomes an open question about which coherence intuitions are behaviorally operative, rather than a purely formal choice.

Conclusion

This paper examined coherence as a heuristic for regulating evidence sharing under uncertainty. The simulations show that selective withholding can improve collective learning in noisy environments, that simple coherence-based heuristics can guide such selective sharing more effectively than individual evidence filtering, and that the same heuristics predictably fail when misleading evidence forms a coherent alternative. These results identify coherence-gated sharing as a conditionally adaptive social strategy, whose benefits and liabilities depend on the structure of the informational environment.

Several empirical questions remain open. Do human reasoners modulate evidence sharing as a function of coherence with their current beliefs, especially when reliability is hard to assess? If so, which coherence intuitions are behaviorally operative, and how do they interact with cues such as source credibility or reputational concerns? Addressing these questions would help determine whether coherence-based selective sharing is merely a useful computational idealization or a descriptively adequate model of human information exchange.

Acknowledgments

We thank the anonymous referees of this conference for their comments and suggestions, which have helped us improve and strengthen the paper. Thanks also to our colleagues at the University of Maribor, who made useful suggestions at an early stage of this project.

Both authors acknowledge financial support from the Slovenian Research Agency (research core funding No. P6–0144 and project No. J6–60107).

Authors also acknowledge the use of AI tools, specifically Gemini 3 models (Flash and Pro) via the PyCharm Gemini Code Assist integration, in preparing the Python code used for the visualizations for the paper. The authors reviewed and take responsibility for all content.

References

Altay, S., Hacquin, A.-S., & Mercier, H. (2022). Why do so few people share fake news? It hurts their reputation. *New Media & Society*, *24*(6), 1303–1324.

Assaad, L., Fuchs, R., Jalalimanesh, A., Phillips, K., Schoeppl, L., & Hahn, U. (2023). A Bayesian agent-based framework for argument exchange across networks. *arXiv*.

Assaad, L., & Hahn, U. (2024). Rational polarization: Sharing only one's best evidence can lead to group polarization. *Proceedings of the annual meeting of the cognitive science society*, *46*.

BonJour, L. (1985). *The Structure of Empirical Knowledge*. Harvard University Press.

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology*. Oxford University Press.

Glass, D. H. (2002). Coherence, explanation, and Bayesian networks. In M. O'Neill, R. F. E. Sutcliffe, C. Ryan, M. Eaton, & N. J. L. Griffith (Eds.), *Artificial Intelligence and Cognitive Science, 13th Irish Conference, AICS 2002* (pp. 177–182). Springer.

Goldberg, S. C., & Khalifa, K. (2022). Coherence in science: A social approach. *Philosophical Studies*, 1–21.

Harris, A. J., & Hahn, U. (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1366.

Hartmann, S., & Trpin, B. (Forthcoming). Why coherence matters? *The Journal of Philosophy*.

Justin, M., & Trpin, B. (2025). Coherence-Based Evidence Filtering: A Computational Exploration. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *47*.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Olsson, E. J. (2002). What is the problem of coherence and truth? *The Journal of Philosophy*, *99*(5), 246–72.

Olsson, E. J. (2012). A bayesian simulation model of group deliberation and polarization. In *Bayesian argumentation: The practical side of probability* (pp. 113–133). Springer.

Olsson, E. J. (2022, September). *Coherentism*. Cambridge University Press. https://doi.org/10.1017/9781009053327

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers.

Shogenji, T. (1999). Is coherence truth conducive? *Analysis*, *59*(4), 338–345.

Simon, D., & Read, S. J. (2025). Toward a general framework of biased reasoning: Coherence-based reasoning. *Perspectives on Psychological Science*, *20*(3), 421–459.

Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending cold and hot cognitions. *Journal of Personality and Social Psychology*, *109*(3), 369.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & language*, *25*(4), 359–393.

Taskesen, E. (2020, January). *Learning Bayesian Networks with the bnlearn Python Package.* (Version 0.3.22). https://erdogant.github.io/bnlearn

Thagard, P. (2002). *Coherence in Thought and Action*. MIT Press.

Trpin, B., & Justin, M. (2025). Coherence as a constraint on scientific inquiry. *Synthese*, *206*(4), 200.

Zollman, K. J. S. (2007). The communication structure of epistemic communities. *Philosophy of Science*, *74*, 574–587. https://doi.org/10.1086/525605