Coherence-Based Evidence Filtering: A Computational Exploration

Martin Justin (martin.justin1@um.si)
Department of Philosophy, Faculty of Arts, University of Maribor, 2000 Maribor (Slovenia)

Borut Trpin (borut.trpin@lrz.uni-muenchen.de)
Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany)
Department of Philosophy, Faculty of Arts, University of Maribor, 2000 Maribor (Slovenia)

Abstract

This study explores the role of coherence-based reasoning in belief updating within uncertain environments. We develop a novel computational model where agents update their beliefs based on observed evidence, with some evaluating the coherence of their belief set before accepting new evidence. Our results show that coherence-based evidence filtering improves belief accuracy in noisy environments and when agents' prior beliefs are accurate. However, when agents encounter systematically misleading evidence, coherence considerations lead to less accurate beliefs. These findings shed light on how coherence interacts with evidence quality and belief accuracy.

Keywords: coherence; Bayesian Networks; belief updating; noise filtering; computational modeling

Introduction

Human reasoning often involves filtering out evidence that contradicts existing views, a phenomenon that has been addressed by researchers interested in confirmation bias (Festinger, Riecken, & Schachter, 1956), myside bias (Stanovich, West, & Toplak, 2013; Baccini, Christoff, Hartmann, & Verbrugge, 2023), and the psychological immune system theory (Mandelbaum, 2019). These accounts describe how individuals process and interpret information in ways that prioritize coherence with their prior beliefs, which often leads to biases that favor evidence aligning with preexisting views while disregarding contradictory information.

At the same time, coherence, i.e., the degree to which a set of information "hangs together", has been proposed as an important feature of rational belief formation (Lehrer & Cohen, 1983; Thagard, 2002). Psychological studies (e.g., Harris & Hahn, 2009) have likewise demonstrated that people prefer coherent information sets, and philosophers have argued that coherence can serve as a justification for beliefs (e.g., BonJour, 1985). Computational models, such as those developed by Thagard (1989, 1992, 2002, 2012), have further explored how coherence operates in reasoning processes.

Despite these advances, two important questions remain unanswered:

1. **Coherence-Based Evidence Filtering:** In environments where data is noisy or biased, can coherence-based reasoning help agents resist misleading information?
2. **Conditions for Coherence-Based Belief Accuracy:** When does coherence-based filtering lead to more accurate beliefs, and when does it lead to belief perseverance or even increased inaccuracy?

To address these questions, we develop a novel computational model that simulates how agents update their beliefs in response to evidence, with and without coherence-based filtering. Our model is grounded in a Bayesian framework. The simulated agents begin with a prior probability distribution over the events from a ground world (represented by a Bayesian network) and iteratively update their beliefs based on observed data. Crucially, some agents also evaluate the coherence of their most probable belief state before accepting or rejecting updates. This process allows us to explore how coherence considerations interact with the quality of evidence (e.g., noisy or biased data) and the accuracy of agents' initial beliefs.

Our results reveal that coherence considerations can both aid and hinder belief accuracy, depending on the environment and the agents' starting points. In highly noisy environments, coherence-based filtering helps agents resist misleading evidence, even when their initial beliefs are inaccurate. However, in environments with less noise, coherence considerations can slow down belief updating, leading to less accurate beliefs over time. Additionally, if agents receive systematically misleading data, coherence considerations will lead to less accurate beliefs in all studied circumstances. These findings have important implications for both individual and collective reasoning, particularly in contexts such as scientific inquiry, where noisy data and biased evidence are common. More generally, our work provides a novel computational model for studying the role of coherence in belief updating, inspired by the NormAN framework (Assaad et al., 2023).

The remainder of the paper is organized as follows. In the next section, we describe the theoretical background and formal measures of coherence. We then present our computational model in detail, followed by the results of our simulations. Finally, we discuss the implications of our findings and conclude with directions for future research.

Theoretical Background

Coherence considerations play an important role in uncertain reasoning as they influence how individuals evaluate and integrate new evidence into their existing belief systems. At its core, epistemic coherence refers to the degree to which propositional information from an information set fits together. For example, consider the contrasting cases of the

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

faster-than-light neutrinos anomaly (Brumfiel, 2011) and the Higgs boson discovery (Aad et al., 2012). In the former, the scientific community was (rightfully) skeptical of the results due to their inconsistency with well-established physical theories despite a very high statistical significance of the result. On the other hand, in the latter, the discovery was readily accepted because it aligned with the predictions of the Standard Model of particle physics. Notably, the significance of this result, albeit high, was magnitudes lower than in the anomalous neutrino case. Similar reasoning patterns may also be observed in everyday reasoning as coherence considerations seem to shape the acceptance or rejection of evidence and affect how open- or close-minded individuals are when confronted with new information. This raises the following normative question: Does coherence-based reasoning of this kind help or harm inquiry?

To address this question, we first need to clarify (i) what epistemic coherence is, (ii) how it can be measured, and (iii) why striving for coherent information is valuable. Bayesian epistemology provides a useful framework for addressing the first two questions (Olsson, 2022) as it offers formal tools to quantify coherence and its role in belief updating. Our work builds on this foundation to answer the third question—why and under which conditions it is reasonable to consider how coherent an information set is.

Let us briefly illustrate how the first question may be clarified. Intuitively, a coherent set of information "hangs together" (BonJour, 1985) in a way that makes it more plausible as a whole than the information items would be individually. For example, consider the following two information sets:

- $\mathbf{S_1} := \{$[All Ravens are black], [This bird is a raven], [This bird is black]$\}$

- $\mathbf{S'_1} := \{$[This chair is dark gray], [Electrons are negatively charged], [Today is Thursday]$\}$

While $\mathbf{S_1}$ is highly coherent because its propositions are logically connected, $\mathbf{S'_1}$ lacks any meaningful relationship between its elements. Similarly, consider the sets:

- $\mathbf{S_2} := \{$[Tweety is a bird], [Tweety cannot fly]$\}$

- $\mathbf{S'_2} := \{$[Tweety is a bird], [Tweety cannot fly], [Tweety is a penguin]$\}$

Here, $\mathbf{S'_2}$ is more coherent than $\mathbf{S_2}$ because the additional proposition ("Tweety is a penguin") resolves the tension of $\mathbf{S_2}$ (this example is from Bovens & Hartmann, 2003).

To operationalize coherence, several probabilistic measures have been proposed, each capturing different intuitions about what makes a set of propositions coherent. Two key intuitions underlie these measures. One of them is:

1. **Deviation from Independence:** The less independent the propositions in a set are, the more coherent the set is. This intuition is formalized by Shogenji (1999), who proposes the following measure for a set of propositions $\mathbf{S} = \{\mathrm{A}_1, \ldots, \mathrm{A}_n\}$:

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

This measure quantifies coherence as the ratio of the joint probability of the propositions to the product of their individual probabilities. The next key intuition regarding coherence is:

2. **Relative Overlap:** The more overlap there is among the propositions in a set, the more coherent the set is. This intuition is formalized by Olsson (2002) and Glass (2002), who propose the following measure:

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

This measure captures the extent to which the propositions in the set overlap.

3. **Crossover of the two intuitions:**

In addition to these measures, we also consider a crossover measure, recently proposed by Hartmann and Trpin (in press), which combines elements of both intuitions:

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

There are also many other measures of coherence available in the literature, many of which are based on averaging the coherence of subsets of a given information set (e.g., Fitelson, 2003; Meijs, 2006; Douven & Meijs, 2007; Schupbach, 2011; Koscholke, Schippers, & Stegmann, 2019). We exclude them in the present study as they are computationally rather demanding.

Building on these formal measures, we propose the following formulation of coherentist filtering of evidence (roughly following the informal idea proposed by Goldberg & Khalifa, 2022):

1. An agent maintains a probability distribution P over a set of binary propositional variables representing their beliefs. The agent can assess the coherence of this set using one of the measures described above.

2. The agent collects new evidence from the world, which may be noisy or biased.

3. If the new evidence decreases the coherence of the agent's belief set, the evidence is rejected. Otherwise, it is accepted.

This process raises an important question: Under what conditions is coherentist filtering a good idea? To answer this, we develop a computational model that simulates how agents update their beliefs in response to evidence, with and without coherence-based filtering. Our goal is to explore whether coherence considerations can help agents resist misleading evidence and improve the accuracy of their beliefs over time.

Simulational Study

To study whether coherence considerations help us filter misleading evidence, we developed a computational simulation roughly inspired by the NormAN modeling framework (Assaad et al., 2023). In the proposed simulation, agents try to form an accurate picture of the ground truth (the world) by gathering information about it.

The world in the simulation consists of a set of probabilistically related events. It is represented via a Bayesian network (BN), consisting of a directed acyclical graph (DAG) and a corresponding conditional probability distribution over a set of binary propositional variables from the BN nodes (see Pearl, 1988 for an introduction to Bayesian networks theory). Nodes in the DAG represent the events in the world, which may be true or false, while edges represent probabilistic dependencies between them. The conditional probability distribution (CPD) then contains information about the likelihood of individual events given different values of related events.

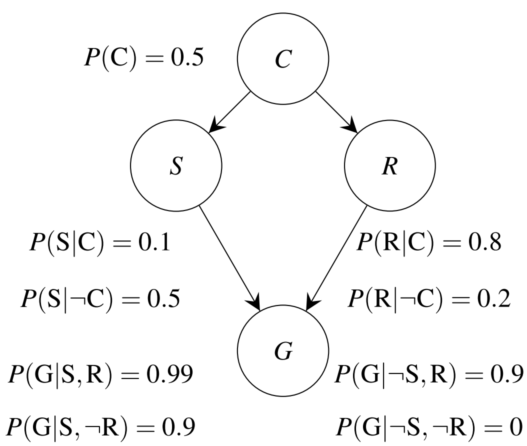

Figure 1: The "sprinkler" network, where C, S, R, G are propositional variables with corresponding values C: "It is cloudy", ¬C: "It is not cloudy", S: "The sprinkler is turned on", ¬S: "The sprinkler is not turned on", R: "It rains", ¬R: "It does not rain", G: "The grass is wet" and ¬G: "The grass is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

Figure 1 shows one simple example of a BN, usually referred to as the "sprinkler" network. This is also the BN that we used in our exploration, as larger networks are computationally increasingly demanding. However, our simulation, in principle, allows one to use BNs of any size.

Agents in the simulation 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.

[1]We use italics for propositional variables and roman script for the values of the variables.

In case of "sprinkler", they know that *Cloudy* (C) is the parent node to *Sprinkler* (S) and *Rain* (R), and that the two are parent nodes for *Wet Grass* (G). What the agents do not know is the CPD of the world; that is, they do not know the actual probabilities of the events.

We model the learning of the agents by having the agents sample from a CPD that is associated with the world and then fitting these observations to the model via maximum likelihood estimation (MLE). For example, one sample the 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. Using this information, they form a new belief about the distribution by fitting the observations to the BN. Effectively, this means that the agents come up with an updated subjective CPD. Figure 2 illustrates how this may work in practice. Note that the agent's CPD will tend to deviate from the CPD of the true BN.

C	R	S	G
1	1	0	1
0	0	1	1
0	0	0	0
1	0	0	1
1	1	0	1
...	...	...	...

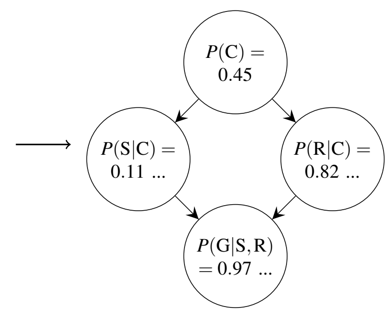

Figure 2: An agent estimates the CPD (right) from sampled observations (1 stands for the variable being true, 0 for it being false; left) via MLE.

As we were interested in the effect of the agents' prior beliefs on their accuracy, we set their prior CPD externally (i.e., before the agents sample from the world). This is generated by randomly changing the parameters of the world's CPD within some interval. For example, if in the true distribution $P(\mathrm{R}|\mathrm{C}) = 0.8$, agents might start with a distribution where $P(\mathrm{R}|\mathrm{C})$ is sampled from a uniform distribution between $[0.8 - i,\ 0.8 + i]$, where $i \geq 0,\ 0.8 - i \geq 0,\ 0.8 + i \leq 1$. This kind of change is applied to the whole distribution. i is a parameter of the model and can be determined by the modelers.

In the second step, agents determine the most probable state of the world based on the learned distribution. In the running example, 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. They accept the updated belief if it is more coherent than the state that was the most probable according to their prior belief. If not, they reject the evidence that led to the update. Agents can skip this second part and update their beliefs on the available information. We will call these agents "normal agents".

This process of gathering data and updating beliefs about the probabilities of the events in the world is then repeated for multiple steps. At the end of the simulation (the number of steps is determined at the onset as a parameter of the model), we evaluate how closely agents' beliefs approximate the actual conditional probability distribution of the ground truth using the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951). The KL divergence measures the discrepancy between two probability distributions, quantifying the information loss when approximating the true distribution. It is particularly useful here as it captures how well agents' beliefs reflect the correct probability distribution of the ground world. To explore the effect of coherence considerations on the accuracy of agents' beliefs, we then compare these values of agents who take coherence into account and those who do not.

We extend this base model in one important dimension. This extension concerns the possibility of receiving misleading data. Agents in the model can receive two types of such data. In one case, we randomly change some percent of the data points agents collected. For example, let's say the current world state the agents observe is S_1 = [Cloudy=True, Sprinkler=False, Rain=True, Wet Grass=True]. In an extremely noisy environment, in which agents would be misled about 50 % of their data, they wouldn't observe S_1 but a modified set where (on average) two out of four values would be changed (for True to False or the other way around). The percentage of data that changes—the level of "noise" in the environment—is determined as a model parameter. We call this type of misleading data "noisy evidence".

In the other case, agents receive evidence that is systematically misleading. Specifically, they have some probability of drawing samples not from the ground truth but from an alternative Bayesian Network that is biased. For example, suppose a sprinkler factory is trying to downplay the role of rain in making the grass wet. Then where in the ground truth of the model $P(G|\neg S, R) = 0.9$, in the alternative, misleading BN, this probability may be changed to $P(G|\neg S, R) = 0.4$. We call this "systematically misleading evidence".

This simulation was implemented in Python using bnlearn (Taskesen, 2020) and Mesa (Kazil, Masad, & Crooks, 2020) libraries. The code is available at `https://github.com/Martin-Justin/CohABM/`.

Results

We ran simulations with a single coherentist and normal agent for each possible combination of considered parameter values, represented in Table 1. We ran each simulation for 50 steps. In all cases, agents took 100 samples from the world at each step. Each combination of parameters was simulated 50 times.

Results for Noisy Evidence

Figure 3 shows simulation results where agents received noisy evidence. The x-axis represents how much the parameters of the agent prior CPD could deviate from the truth (a larger number means less accurate priors). The y-axis represents the probability that the agent's evidence items were randomly changed during the evidence-gathering process (a larger number means more noisy data). Each value on the heatmap represents absolute average difference between the accuracy of a normal agent and a coherentist agent at the end of the simulation for a combination of parameters. The accuracy of agents' beliefs is represented as distance from truth as measured by the KL divergence, meaning that the smaller values represent more accurate beliefs. Thus, the higher value on the heatmap corresponds to coherentist agents ending up with more accurate beliefs than normal agents (and *vice versa* for negative numbers). These values represent averages over the 50 simulation runs. Due to space constraints and since the results for the three coherence measures closely resemble each other, we present averages over all coherence measures.

Parameter	Values
Coherence Measures	Shogenji, Olsson-Glass, Hartmann-Trpin
Type of misleading evidence	noisy evidence, systematically misleading evidence
Information "noise"	0.05, 0.1, 0.15, 0.2, 0.25, 0.3
Variation of priors	0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4

Table 1: Parameter values used in the reported simulations.

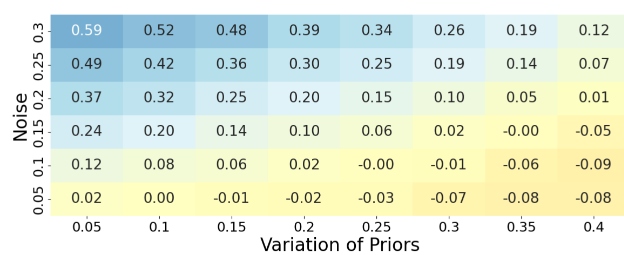

Figure 3: Difference in accuracy (Normal minus Coherence) at round 50 with noisy evidence. Positive values favor coherence-based reasoning.

The results show that filtering evidence based on coherence considerations can improve or harm an agent's inquiry, depending on the agent's priors and the amount of noise in the data. In a very noisy environment, where agents are misled about 25 to 30 % of their evidence, taking coherence into account helps even if agents start with inaccurate priors.

This is, to some extent, the result of the fact that we only ran simulations for 50 rounds. As shown by the top chart on Figure 4, taking coherence into account does not completely insulate agents from misleading data. Rather, it makes their beliefs more resilient to them by making the belief changes more incremental.

On the other hand, being mindful of the coherence of one's beliefs can hurt an agent's inquiry in environments with less

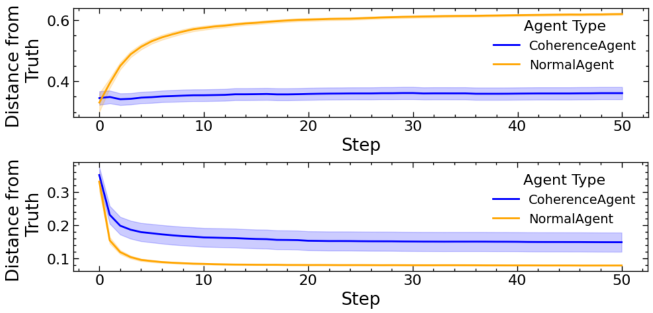

Figure 4: The distance from the truth as measured by the KL divergence over time, comparing a normal (orange) and a coherentist (blue). Top: noise = 0.3, variation of priors = 0.3. Bottom: noise = 0.05, variation of priors = 0.3. The shaded region around the lines represents the 95% confidence interval.

noisy data. The mechanism behind these results is similar to before: coherentist agents are more cautious in updating on new evidence. In situations where this new evidence is generally of high quality (only slightly misleading), this resilience can be detrimental to the overall success of the inquiry. One such example is shown in the bottom chart of Figure 4.

Results for Systematically Misleading Evidence

Figure 5 shows results for simulations with systematically misleading evidence (the format of the chart is the same as in Figure 3). It clearly shows that in situations where agents receive systematically misleading evidence, coherentist reasoning harms inquiry across the board, with the impact being even more pronounced when the agents start with inaccurate priors.

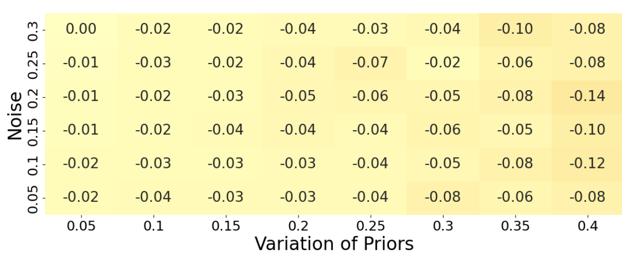

Figure 5: Difference in accuracy (Normal minus Coherence) at round 50 with systematically misleading evidence. Positive values favor coherence-based reasoning.

Why is this the case? One possible explanation stems from the different nature of systematically misleading evidence *vis-à-vis* noisy evidence. The latter acts as a kind of random noise added to the true evidence. As such, if it causes a change in an agent's belief about the most probable state of the world, this new state might accidentally be more coherent than the state supported by previous evidence. However, systematically misleading evidence can provide agents with a picture of the world that is consistently more coherent than the one supported by the correct evidence. In this case, a coherentist agent will start to reject true evidence and accept systematically misleading evidence, thus moving further away from the truth.

Figure 6 shows that something like this indeed happens in the simulation. It shows accuracy over time of a single coherentist agent. As we can see, the agent starts with a prior that is quite close to the truth but then shifts away from it. This is consistent with the agent forming a new, more coherent belief based on the misleading evidence and consequently starting to ignore the actual evidence.

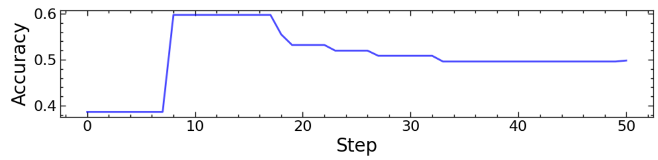

Figure 6: Line chart representing the distance from truth over time for a single coherentist agent receiving systematically misleading evidence in a randomly chosen simulation run for noise=0.3 and variation of priors=0.35.

Discussion

In this study, we asked two questions. First, can coherence-based reasoning help agents resist misleading evidence and improve the accuracy of their beliefs? And, if this is indeed the case, what are the conditions in which this is possible?

Our results show that we can answer the first question affirmatively. As is shown in Figure 3, coherentist agents outperform normal agents in a significant part of the parameter space we studied. In other words, in certain circumstances, agents who take the coherence of their beliefs into account form more accurate beliefs than agents who do not. Coherence can act as a reliable filter for misleading evidence.

Nevertheless, this effect has important limitations, which leads us to the second question. First, when agents receive noisy data, coherentists perform worse in situations where the evidence is, in general, of high quality. Similarly, when agents start with inaccurate priors, coherence considerations mostly reduce the accuracy of their beliefs. In other words, coherence helps agents in cases where taking evidence at face value would cause one to adopt less accurate beliefs.

Second, the situation looks much grimmer in cases when agents receive systematically misleading evidence. Here, coherence considerations hurt the accuracy of agents' beliefs across the board because the biased evidence provides a more coherent narrative than the actual evidence does. In sum, the conditions under which coherence-based filtering improves belief accuracy are narrow and highly context-sensitive. Our results do not support the idea that coherence is a universally beneficial heuristic. Rather, coherence-based reasoning appears to be a high-risk, high-reward strategy: it can protect agents from noisy or misleading evidence when their priors are already quite accurate, but it also runs the risk of entrenching false beliefs if agents begin with poor priors or encounter systematically biased evidence. This insight reflects a kind of "garbage in, garbage out" dynamic: coherence filtering can amplify whatever assumptions an agent already holds. Thus, any benefits of coherence depend crucially on the epistemic environment and the agent's starting point.

Potential Objections

That said, some objections could be raised against the implications of our model. In the real world, it is often hard to assess how misleading one's evidence is or how accurate one's priors are. Consequently, we usually cannot know whether we are in a position where coherence-based evidence filtering could help us. While this is a fair point, important indicators exist that show both how noisy our environment is and how good our priors are, which we can use in deciding whether to go coherentist or not. For example, as exemplified by the case of faster-than-light neutrinos anomaly, scientific consensus can be a good proxy for accurate priors. Additionally, it is usually possible to determine whether our information source is simply noisy or systematically misleading. For example, in science, we might expect that low-quality studies provide somewhat noisy or inaccurate evidence, while industry-funded studies present systematically biased results (Lundh, Lexchin, Mintzes, Schroll, & Bero, 2017).

Another potential objection may be addressed at the idealized nature of our simulation. We make several substantial assumptions about evidence-gathering and reasoning that most likely do not apply to real-world epistemic agents. Thus, it is questionable whether the simulation results can be directly transposed to the real world. This is a well-known point from the philosophy of simulations and modeling (Weisberg, 2007). However, it does not imply that idealized computational simulations are purposeless. As Reutlinger, Hangleiter, and Hartmann (2018) have pointed out, they provide us with what they call "how-possibly explanations". In other words, by exploring a conceptual space of the interaction between coherence and evidence-based belief formation, we learned that it is *possible* that coherence-based evidence filtering can help us form more accurate beliefs. Consequently, we have a better understanding of the factors that are involved in this: the accuracy of the priors and the amount and type of evidential noise.

Future Work

While the current model already provides interesting results, we plan to extend it in several ways in the future. First, we would like to include a social aspect by adding additional agents and evidence sharing. This will also include implementing different networks, which determine the neighbor relations between agents. This extension would tie our work to existing studies in network epistemology (Zollman, 2013) and agent-based modeling in philosophy (Šešelja, 2023).

Another important direction for future work concerns considering different, less strict, coherentist strategies for evidence-filtering. Currently, coherentist agents reject evidence that causes *any* downward change in the coherence of their beliefs. We think it would be interesting to loosen this criterion for some agents. These agents would reject only the evidence that causes a downgrade of coherence larger than some threshold. Additionally, this threshold could change with the actual coherence of agents' beliefs. Agents with already highly coherent beliefs would accept only small downward changes in coherence, while agents with less coherent beliefs would be more tolerant.

This is motivated by the use of coherence considerations in science. It might have been reasonable for the physics community to reject faster-than-light neutrinos results. On the other hand, a similar response to an analogous possible example of a highly unexpected result might not be appropriate in some other, in general, much less coherent field of research, such as psychology.

Finally, a key challenge remains the empirical validation of these theoretical findings. While our model provides useful predictions, translating these predictions into real-world scenarios is crucial.

A promising avenue for future research is to test whether similar coherence-filtering mechanisms appear in human cognition. One could design behavioral experiments where participants are exposed to sequences of information under varying conditions of noise and misinformation, and track whether judgments about coherence influence the acceptance of new evidence. For instance, online reasoning tasks could manipulate both the reliability of prior beliefs and the coherence of incoming information to explore whether participants show greater resistance to incoherent but accurate evidence, especially when prior beliefs are strong. Such studies would help ground our model in empirical psychology and open the door to applications in contexts like media literacy, misinformation detection, or cognitive training.

Conclusion

Coherence—the degree to which information "hangs together"—is an important feature of rational belief formation. In this paper, we took a closer look at whether taking the coherence of one's beliefs into account can help us form more accurate beliefs in noisy information environments.

To achieve this, we developed a novel computational simulation. In the simulation, agents try to form an accurate belief about probability distribution over events in the world by taking samples from it. Some agents take their evidence at face value. Others check how new evidence impacts the coherence of their beliefs—in case the evidence reduces coherence, they reject it.

Comparing the performance of these two types of agents, we showed that taking coherence into account can help agents form more accurate beliefs. However, the conditions under which this is the case are quite limited. Specifically, coherence-based evidence filtering helps agents in very noisy environments and when they already start with highly accurate priors. Additionally, if agents, instead of randomly noisy evidence, receive systematically misleading evidence, coherence leads to less accurate beliefs.

Acknowledgments

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P6-0144). BT, furthermore, acknowledges the financial support from the Slovenian Research Agency, project No. J6-60107. Both authors contributed equally and are listed in the alphabetical order.

References

Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Khalek, S. A., Abdelalim, A. A., ... others (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, *716*(1), 1–29.

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

Baccini, E., Christoff, Z., Hartmann, S., & Verbrugge, R. (2023). The wisdom of the small crowd: Myside bias and group discussion. *Journal of Artificial Societies and Social Simulation*, *26*(4).

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

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

Brumfiel, G. (2011). Particles break light-speed limit. *Nature*, *22*, 22.

Douven, I., & Meijs, W. (2007). Measuring coherence. *Synthese*, *156*(3), 405–425.

Festinger, L., Riecken, H. W., & Schachter, S. (1956). *When Prophecy Fails: A Social and Psychological Study of a Modern Group That Predicted the Destruction of the World*. Harper-Torchbooks.

Fitelson, B. (2003). A probabilistic theory of coherence. *Analysis*, *63*(3), 194–199.

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). Berlin: 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. (in press). Why coherence matters? *The Journal of Philosophy*.

Kazil, J., Masad, D., & Crooks, A. (2020). Utilizing python for agent-based modeling: The mesa framework. In R. Thomson, H. Bisgin, C. Dancy, A. Hyder, & M. Hussain (Eds.), *Social, Cultural, and Behavioral Modeling* (pp. 308–317). Cham: Springer International Publishing.

Koscholke, J., Schippers, M., & Stegmann, A. (2019). New hope for relative overlap measures of coherence. *Mind*, *128*(512), 1261–1284.

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

Lehrer, K., & Cohen, S. (1983). Justification, truth, and coherence. *Synthese*, *55*, 191–207.

Lundh, A., Lexchin, J., Mintzes, B., Schroll, J. B., & Bero, L. (2017, 2). Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews*, *2017*. doi: 10.1002/14651858.MR000033.pub3

Mandelbaum, E. (2019). Troubles with Bayesianism: An introduction to the psychological immune system. *Mind & Language*, *34*(2), 141–157.

Meijs, W. (2006). Coherence as generalized logical equivalence. *Erkenntnis*, *64*(2), 231–252.

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

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

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

Reutlinger, A., Hangleiter, D., & Hartmann, S. (2018, 12). Understanding (with) toy models. *The British Journal for the Philosophy of Science*, *69*, 1069-1099.

Schupbach, J. N. (2011). New hope for Shogenji's coherence measure. *The British Journal for the Philosophy of Science*, *62*(1), 125–142.

Šešelja, D. (2023). *Agent-Based Modeling in the Philosophy of Science* (Winter 2023 ed.). Metaphysics Research Lab, Stanford University.

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

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, *22*(4), 259–264.

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

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(3), 435–467.

Thagard, P. (1992). Conceptual Revolutions. *Princeton, NJ*.

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

Thagard, P. (2012). *The Cognitive Science of Science: Explanation, Discovery, and Conceptual Change*. Cambridge, MA: Mit Press.

Weisberg, M. (2007). Three kinds of idealization. *Journal of Philosophy*, *104*, 639-659. doi: 10.5840/jphil20071041240

Zollman, K. J. S. (2013, 1). Network epistemology: Communication in epistemic communities. *Philosophy Compass*, *8*, 15-27. doi: 10.1111/j.1747-9991.2012.00534.x