ANALIZA
Let. 29, št. 1, str. 3–22, december 2025

ANαliZA

COHERENCE AND REFLECTIVE LUCK: REVISITING MIŠČEVIĆ THROUGH COMPUTATIONAL MODELING

Accepted
28. 8. 2025

Revised
10. 11. 2025

Published
17. 12. 2025

BORUT TRPIN,[1, 2] MARTIN JUSTIN[2]
[1] University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia
borut.trpin@ff.uni-lj.si
[2] University of Maribor, Faculty of Arts, Maribor, Slovenia
borut.trpin@ff.uni-lj.si, martin.justin1@um.si

CORRESPONDING AUTHOR
borut.trpin@ff.uni-lj.si

Abstract This paper examines Miščević's (2007) hint that coherence considerations can help reduce reflective luck, a form of epistemic luck that stems from an agent's own cognitive fragility. To do so, we refer to the results from our recently developed computational model, in which we simulated agents who update beliefs probabilistically but may filter evidence based on its impact on coherence. The results show that coherence filtering improves stability in noisy environments by protecting agents from random misinformation, although the exact outcome depends on armchair luck (in our case, this refers to the agent's prior beliefs). However, in biased environments, the same strategy leads agents to dismiss corrective evidence, reinforcing false beliefs. We argue that coherence is not a universal epistemic virtue, but an adaptive strategy whose success depends on the environment. Our findings extend Miščević's proposal and highlight the importance of flexibility in belief formation.

Keywords
coherence,
reflective luck,
epistemic luck,
virtue epistemology,
higher-order evidence,
agent-based modeling

https://doi.org/10.18690/analiza.29.1.3-22.2025

University of Maribor Press

ANαliZA

ANALIZA
Vol. 29, no. 1, pp. 3–22, December 2025

KOHERENCA IN REFLEKTIVNA SREČA: PONOVNI PREMISLEK O MIŠČEVIĆEVEM PRISTOPU S POMOČJO RAČUNSKEGA MODELIRANJA

BORUT TRPIN,[1, 2] MARTIN JUSTIN[2]

[1] Univerza v Ljubljani, Filozofska fakulteta, Ljubljana, Slovenija
borut.trpin@ff.uni-lj.si
[2] Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija
borut.trpin@ff.uni-lj.si, martin.justin1@um.si

DOPISNI AVTOR
borut.trpin@ff.uni-lj.si

Sprejeto
25. 8. 2025

Pregledano
10. 11. 2025

Izdano
17. 12. 2025

Izvleček V članku analizirava Miščevićevo (2007) zamisel, da lahko upoštevanje koherence prispeva k zmanjšanju reflektivne sreče, tj. oblike epistemske sreče, ki izvira iz agentove lastne kognitivne krhkosti. Pri tem se opirava na rezultate najinega nedavno razvitega računalniškega modela, v katerem simulirava agente, ki svoja prepričanja posodabljajo verjetnostno, hkrati pa lahko filtrirajo dokazila glede na njihov vpliv na koherenco. Rezultati kažejo, da koherenčno filtriranje v hrupnih okoljih izboljša stabilnost, saj agente ščiti pred naključnimi dezinformacijami, čeprav je končni izid odvisen od teoretske sreče oz. z Miščevićevim izrazom, »sreče iz naslanjača« (v najinem primeru to označuje agentova začetna prepričanja). V pristranskih okoljih pa ista strategija vodi do zavračanja boljših dokazil in s tem k utrjevanju napačnih prepričanj. Zagovarjava, da koherenca ni univerzalna epistemska vrlina, temveč prilagodljiva strategija, katere uspešnost je odvisna od okolja. Najine ugotovitve tako razširjajo Miščevićev predlog in poudarjajo pomen fleksibilnosti pri oblikovanju prepričanj.

Ključne besede
koherenca,
reflektivna sreča,
epistemska sreča,
epistemologija vrlin,
dokazila višjega reda,
agentsko modeliranje

https://doi.org/10.18690/analiza.29.1.3-22.2025

Univerzitetna založba
Univerze v Mariboru

1 Introduction

The phenomenon of epistemic luck has long troubled epistemologists who seek to account for the nature of knowledge. At the heart of the worry is the idea that agents often find themselves believing the truth, yet only by chance. Prominent modal accounts, most notably developed by Pritchard (2005) and Sosa (2003), diagnose epistemic luck as arising when a belief could easily have been false in nearby possible worlds. On this view, a belief is safe if, in close alternatives to the actual world, the agent continues to believe truly; if not, the agent is epistemically lucky. While this modal framework captures many paradigmatic cases of luck—where agents barely evade falsehood due to worldly happenstance—it leaves unexplored the possibility that epistemic luck may stem not from the external world but from the internal workings of the agent.

Nenad Miščević (2007) has drawn attention to precisely this neglected dimension. He argues that knowledge is threatened not only when truth is modally fragile, but also when agents themselves exhibit a kind of *cognitive fragility*. Miščević's key insight is that even if the world cooperates—presenting no adversarial scenarios or hostile contingencies—an agent may nonetheless have been lucky in arriving at the truth because slight variations in her belief-forming dispositions might easily have led her astray. He terms this phenomenon *reflective luck*, marking a shift from world-centered to agent-centered epistemology. What matters is not merely how the world might have varied, but how small perturbations in the agent's own reasoning, heuristics, or cognitive architecture could have resulted in error.

Miščević proposes that one of the agent's best resources against reflective luck is *coherence*. Coherence, in his view, is not simply a static relation of mutual support among beliefs, as traditional coherentist theories have emphasized (BonJour, 1985), but a dynamic stabilizer. It acts not only as a condition of justification but as a *virtue-like* mechanism that can reduce the agent's sensitivity to perturbations. A belief system whose parts are well-integrated is less susceptible to haphazard shifts prompted by misleading evidence or internal inconsistency. As Miščević succinctly puts it, "[t]he reflective luck can be minimized by using a coherentist strategy at the reflective level" (Miščević, 2007, p. 67). However, Miščević's proposal remained largely programmatic. He did not specify how coherence stabilizes belief change or under which conditions this stabilization genuinely reduces reflective luck.

This paper presents the task of formalizing and testing this proposal. Our aim is twofold. First, we reinterpret Miščević's suggestion within the framework of contemporary epistemology, connecting it to the virtue-theoretic turn (Sosa, 2003; Greco, 2003), the literature on higher-order evidence (Christensen, 2010; Feldman, 2005), and recent developments in formal modeling of belief dynamics. Second, we provide an empirical component by testing the proposal through computational simulations. We adapt an agent-based model developed in previous work (Justin & Trpin, 2025; Trpin & Justin, 2025), equipping agents with the ability to reject evidence when its incorporation would significantly reduce the coherence of their belief system. Coherence, in this setting, operates not merely as a synchronic property but as a *higher-order defeater*—a signal that an incoming piece of information may be suspect.

We also take this opportunity to relate Miščević's insight to a longstanding puzzle in Bayesian epistemology: how to set and revise priors. It is widely acknowledged that Bayesian agents are highly sensitive to their initial priors and that no general procedure is available for their rational determination. In this context, coherence emerges as a candidate for guiding belief revision at the *reflective* level, serving as a form of higher-order evidence about when to trust or discount incoming information when transitioning from prior beliefs.

In exploring these issues, we distinguish between two types of adversarial epistemic environments. First, in *noisy environments*, agents encounter evidence marred by random errors–evidence that may mislead, but whose errors are distributed without systematic bias. Second, in *systematically biased environments*, agents face misinformation that is subtly distorted in ways that remain undetectable from the agent's perspective. These environments model phenomena familiar from the philosophy of science and social epistemology (Longino, 1990; Goldman, 1999), such as confirmation bias, structural misinformation, or publication bias.

Our simulations yield results that are philosophically illuminating. In noisy environments, agents who employ coherence filtering display greater stability and reduced vulnerability to reflective luck. By resisting misleading evidence that would cause significant disruptions to their belief system's coherence, these agents avoid random epistemic perturbations. However, in systematically biased environments, coherence filtering has the opposite effect. Rather than protecting agents, it entrenches them in systematically distorted belief systems by causing them to dismiss precisely the evidence that could correct their biases. Coherence functions as a stabilizer—but not always in a way that is epistemically desirable.

This pattern reveals two deeper lessons. First, coherence may indeed function as a *cognitive virtue*, but only under conditions that match its epistemic affordances. Its contribution to knowledge is *context-sensitive*. Second, Miščević's notion of stability could be refined. Stability is not valuable in and of itself. Only *adaptive* stability–resilience coupled with openness to corrective information–can properly mitigate reflective luck.

The paper proceeds as follows. In Section 2, we revisit Miščević's analysis of reflective luck and place it in the context of contemporary epistemology. Section 3 presents our formal model of coherence-sensitive agents. Section 4 presents the results of our simulations and discusses their epistemological significance. Section 5 reflects on the broader implications for agent-centred epistemology. Section 6 concludes with a proposal for understanding coherence as an *adaptive* strategy in belief formation.

2 From Reflective Luck to Coherence as a Virtue

Miščević's contribution to the debate on epistemic luck consists in shifting attention from the world's fragility to the agent's cognitive fragility. Traditional modal accounts of luck, such as those developed by Pritchard (2005) and Sosa (2003), focus on the counterfactual sensitivity of belief to worldly conditions. According to such views, a belief is lucky if it could easily have been false, had the world been just slightly different. Epistemic luck, in this framework, is primarily about the world's cooperation. What Miščević brings to the table is the insight that the agent's own cognitive makeup is equally a source of epistemic luck. A belief might be true, and even modally safe relative to the world, yet still be the upshot of a precarious reasoning process within the agent herself. Miščević calls this *reflective luck*: the epistemic risk that an agent could easily have formed a false belief, not because of the world's variability, but because of her own cognitive instability.

This idea presses on an overlooked but fundamental feature of knowledge: the agent's resilience. Even if the world stays fixed, small perturbations in how the agent processes information, evaluates evidence, or revises beliefs may tip her into error. Reflective luck thus challenges not only modal conditions for knowledge but also calls into question the adequacy of standard Bayesian and evidentialist frameworks, which typically focus on what evidence is available to the agent rather than how that evidence is processed. The worry is not just whether the agent's belief could have been false given slight changes in the world, but whether, given slight changes in her own epistemic routines, it would have been false.

Miščević proposes that coherence plays a stabilizing role in combating reflective luck. The idea is not merely that coherence justifies beliefs by virtue of mutual support, as in classical coherentism (BonJour, 1985), but that it regulates belief change over time. This is because a coherent belief system is less prone to abrupt, haphazard revisions when new information is encountered. This dynamic role of coherence could be understood as changing it from a condition on the synchronic rationality of beliefs into a virtue that governs the *process* of belief revision itself. The agent who aims to preserve coherence does not blindly follow the stream of incoming information but resists revisions that would destabilize her cognitive architecture.

This approach brings Miščević's account into contact with virtue epistemology. However, it does so in an unusual way. Standard virtue reliabilism emphasizes the role of faculties like perception, memory, and inference as local, process-based virtues whose reliability secures knowledge (Greco, 2003; Sosa, 2003). Miščević points to a different kind of virtue, which we will call a *structural virtue*, and which concerns the organization and management of the agent's entire belief system. Coherence, on this view, is not simply another cognitive process among others, but a higher-order capacity that shapes how agents integrate new information into their cognitive economy. This is the kind of virtue that governs the *reflective level* of belief management.

In fact, Miščević's proposal can be understood as an anticipation of the contemporary turn toward higher-order evidence. Higher-order evidence concerns information about the reliability of one's own cognitive processes or evidential situation (Christensen, 2010; Kelly, 2010). Coherence plays a comparable role in

Miščević's picture. The agent who notices that accepting a piece of evidence would seriously reduce the coherence of her belief system may treat this fact itself as a defeater for that piece of evidence. Coherence then serves as a kind of heuristic for assessing whether new information is trustworthy or misleading. Under this interpretation, coherence is not merely a matter of internal harmony but functions as an indicator of the epistemic status of incoming data.

At first glance, this appears to be a powerful response to reflective luck. An agent who uses coherence-sensitive updating will be less vulnerable to random disruptions, since spurious bits of evidence often undermine coherence. Coherence acts as a protective mechanism, filtering out information that would otherwise destabilize the agent's belief network. However, Miščević's proposal raises a series of deep philosophical challenges. First, is coherence *always* an epistemic virtue? Can it function as a reliable guide to truth across different kinds of environments? And second, can coherence serve as a *stable* stabilizer, or does it risk becoming epistemically rigid, entrenching agents in false beliefs when their prior coherence is itself based on distortions or biases?

These challenges have become especially pressing in light of recent discussions on echo chambers, epistemic bubbles, and systematic misinformation (Nguyen, 2020; Levy, 2021). The worry is that coherence-sensitive agents might, under certain conditions, dismiss precisely the kind of disruptive evidence that could correct entrenched errors. In such cases, the stabilizing power of coherence may exacerbate reflective luck rather than alleviate it. The agent becomes stable—but stably mistaken.

Furthermore, as Miščević's focus is different, he does not address the question that plays an important role in formal epistemology: should we ever ignore evidence? In formal epistemology, the standard answer seems to be negative: agents should update on all available evidence (Carnap, 1947; Good, 1967). But a natural question to ask in light of Miščević's discussion is whether we should sometimes *reject* evidence that would cause a significant loss of coherence. Is coherence functioning as a source of higher-order evidence about the unreliability of first-order evidence, or is it serving as a heuristic that may violate evidential norms? What distinguishes the virtuous use of coherence from mere epistemic conservatism or dogmatism?

Finally, Miščević's brief suggestion leaves open the question of whether coherence can be formalized in a way that makes its epistemic role more than just a verbal recommendation. What does it mean, precisely, to say that an agent's belief system is more or less coherent? How can we measure coherence quantitatively? How does coherence interact with belief updating in agents who encounter not only noisy but also systematically biased information?

3 Modeling Coherence: Computational Framework and Agent Design

Miščević's proposal is philosophically suggestive but leaves the precise mechanics of how coherence stabilizes agents largely unexplored. What exactly does it mean for an agent to use coherence as a safeguard against reflective luck? Can this idea be given formal shape? More importantly, can it be shown that such a strategy genuinely helps agents reduce their vulnerability to epistemic luck, or does it risk merely reinforcing whatever coherence they happen to start with, whether accurate or not? To address these questions, we turn to computational modeling as our recently proposed model seems applicable to this task (Justin & Trpin, 2025; Trpin & Justin, 2025).

The idea behind our approach is simple in spirit but philosophically potent: simulate agents who update their beliefs about the world but who may impose coherence-sensitive constraints on their learning. We implement this in a probabilistic setting, using Bayesian networks to model both the environment and the agents' beliefs. Bayesian networks are well-suited for this task, as they encode probabilistic dependencies between propositions and provide a natural representation of complex evidential relationships. Agents maintain subjective probability distributions over these networks and incrementally update their beliefs as they encounter new information.

The core mechanism is belief updating. At each step, agents receive evidence about probabilistic relationships between propositions. These inputs are not always reliable. We explicitly distinguish between two broad types of environments. In *noisy environments*, agents face random errors: occasional misinformation, measurement noise, or accidental inaccuracies. Such environments capture scenarios where evidence is imperfect but does not systematically mislead. In contrast, *biased environments* are adversarial. The evidence agents receive is systematically skewed in ways undetectable to them, reflecting phenomena such as selection bias, echo chambers, or coordinated misinformation.

In our model, agents face a choice. They can either engage in *pure probabilistic updating* (via so-called maximum-likelihood-estimation), which involves processing all incoming evidence without further scrutiny, or they can act as *coherence-sensitive agents*, who condition belief updates not only on the content of the evidence but also on its impact on the overall coherence of their belief system. Coherence-sensitive agents employ what we term a *coherence filter*: a defeater mechanism that rejects evidence if its incorporation would produce a significant drop in the coherence of the agent's belief network.

To formalize this, we treat coherence as a structural property of the agent's belief system, operationalized as the degree to which the probability assignments of the agent's view of the Bayesian network fit together in a consistent and integrated way. When new evidence arrives, the agent assesses whether accepting it would compromise coherence. If it does, the evidence is rejected. Otherwise, it is incorporated via probabilistic updating as mentioned above. This approach, in a sense, provides an idea that could be interpreted as broadly in Miščević's spirit, that is, that agents should resist revisions that would destabilize the structure of their beliefs.

This mechanism allows us to give concrete meaning to Miščević's concept of *reflective luck*. In his view, an agent is reflectively unlucky when minor perturbations, such as receiving slightly different evidence or reasoning slightly differently, would have led her to form significantly less accurate beliefs. In our model, this fragility is expressed by the agent's susceptibility to errors arising from informational noise or bias. We systematically vary the evidence streams in our approach, introducing small perturbations, and observe how agents respond depending on whether they employ coherence-sensitive updating or not.

Importantly, this framework also puts pressure on Bayesian orthodoxy. Standard Bayesian epistemology holds that agents should update their beliefs based on all available evidence. But Miščević's proposal, interpreted in the outlined way, suggests that agents should sometimes refrain from updating, precisely when doing so would destabilize the broader coherence of their beliefs. Our model treats coherence as a form of higher-order evidence: a cue indicating whether incoming information is worth accepting. This raises important philosophical questions: when is it rational to reject data? When, if ever, is it rationally permissible to override the principle of total evidence in the name of preserving coherence?

The virtue-theoretic aspect of Miščević's proposal is also naturally captured in the model. Coherence is treated as a cognitive virtue, not as an abstract value, but as a mechanism that shapes how belief revision occurs over time. What makes it virtuous, however, is not unconditional conservatism but its capacity to minimize reflective luck by stabilizing belief-formation processes without insulating agents from genuine corrective information.

With this setup, we can systematically explore the dynamics of reflective luck in simulated agents. We track three key dimensions of epistemic performance, which may be interpreted in terms that are broadly related to Miščević's (2007) account: *accuracy*, the proximity of the agent's beliefs to the actual probabilistic structure of the environment; *stability*, the robustness of those beliefs to random or systematic perturbations; and *armchair luck*, reflected by how close to truth an agent's starting point is. By comparing pure Bayesian agents and coherence-constrained agents across both noisy and biased environments, we can begin to address the central question: Does coherence genuinely reduce reflective luck, and if so, under what conditions?

In the next section, we present the results of these computer simulations (originally reported in Trpin & Justin, 2025). As we will show, the dynamics are far from uniform. While coherence filtering improves stability and accuracy under conditions of noise, it becomes epistemically hazardous under conditions of systematic bias, thereby refining Miščević's proposal in important ways, and perhaps crucially, it particularly helps when an agent has armchair luck in the sense of already starting from a good position.

4 Coherence, Stability, and the Limits of Agent-Centered Virtue

Our simulations broadly vindicate Miščević's suggestion that coherence can stabilize belief-formation, but they also expose a critical limitation. Whether coherence functions as a cognitive virtue turns out to depend sharply on the character of the informational environment. What emerges is neither an unqualified defense of coherence as an epistemic stabilizer, nor a wholesale rejection of its utility. Rather, coherence acts as an *environment-sensitive heuristic*–valuable under some conditions, harmful under others.

Let us begin with the positive case. In environments dominated by random noise, coherence-sensitive agents consistently outperform purely first-order-evidence-only agents. When evidence is occasionally but unpredictably distorted, as with common sources of noise such as measurement error or incidental misinformation, agents who filter updates based on coherence tend to avoid being destabilized by such perturbations. They achieve higher stability and often higher accuracy across repeated trials. In these contexts, Miščević's proposal comes into its own: coherence acts as a virtue by resisting fragility in belief-updating.

Philosophically, this result is satisfying. Coherence, treated as higher-order evidence, tracks what agents often *do* in ordinary epistemic life. We are rightly suspicious when a piece of information conflicts too radically with the rest of what we think we know–not necessarily because we dismiss the information outright, but because the incoherence itself serves as a defeater. Miščević's intuition that coherence can act as a reflective safeguard against luck is vindicated here. Note that even in cases where coherence-filtering of evidence slows the agent's updating towards the truth (relative to considering all evidence), the agent still improves their previous position. We show this in Figure 1.

Yet there is a temptation to think that the story ends here: that coherence is simply a stabilizer, full stop. Our simulations decisively reject this tempting picture. In systematically biased environments, where evidence is not randomly noisy but selectively distorted, coherence filtering systematically misfires. Agents become reflectively stable, but for the wrong reasons: they stabilize in error. Worse, because they condition updates on preserving prior coherence, they tend to dismiss precisely the kinds of evidence that could have corrected their false beliefs. The result is a pernicious form of *epistemic conservatism*: agents immunize themselves against correction while maintaining a high degree of internal stability. We show this in Figure 2.

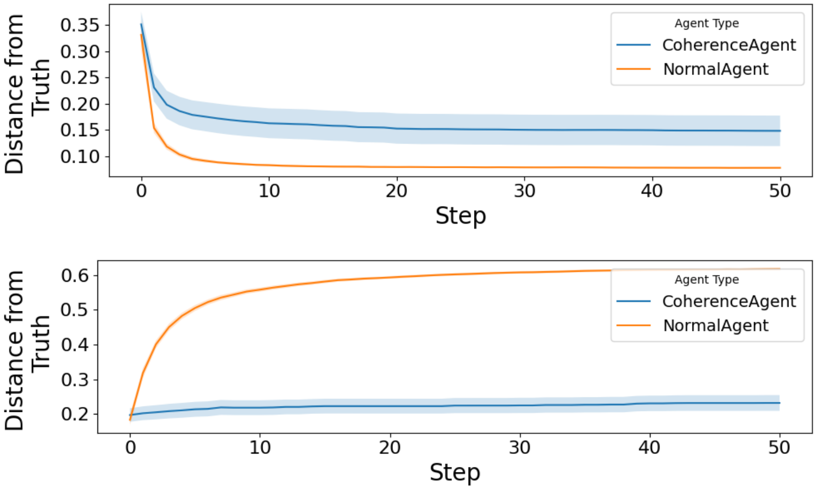

Figure 1: Distance from truth (as measured by the Kulback-Leibler divergence of the agent's probability distribution from the true distribution) over time. Top: high noise (30%) and moderately accurate priors (20%). Bottom: low noise (5%), inaccurate priors (30%). The Coherence agent is in blue, and the First-Order-Evidence-Only agent is in orange. The shaded region around the lines represents the 95% confidence interval (the plots are from Trpin & Justin, 2025, Fig. 5).

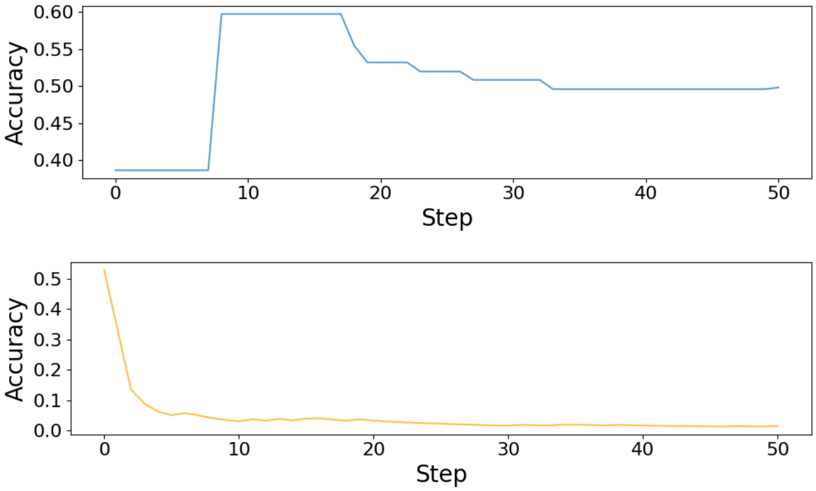

Figure 2: Evolution of one agent's beliefs over time for high chance of receiving misleading evidence (30%), inaccurate priors (35%), and the Hartmann-Trpin coherence measure. Top: Coherence agent. Bottom: First-order-evidence-only agent. Note that, due to random sampling of prior conditional probabilities, agents may start from different levels of prior accuracy (the plots are from Trpin & Justin, 2025, Fig. 7).

This result invites a natural philosophical question: Is coherence not supposed to *help* detect bad evidence? Why does it fail here? The answer, which echoes familiar worries about *epistemic bubbles* and *echo chambers*, is that coherence is blind to the *source* of coherence. If agents start with already skewed priors, distorted by early bias or incomplete information, then biased information may easily *reinforce* coherence rather than disrupting it. Coherence filtering, in such cases, mistakes alignment with prior beliefs for epistemic trustworthiness. The agent becomes stable, but not in the way virtue epistemologists would hope—she is stable in ignorance.

This is connected to debates in virtue epistemology. Virtues, if they are to be genuine, must reliably promote epistemic goods across the kinds of environments agents actually inhabit. Virtues are not mere dispositional tendencies, but *apt* tendencies–reliable, but context-sensitive. Our results fit this framework. Coherence is not an unconditional virtue, but a *defeasible* one. In noisy environments, it is helpful; in systematically biased environments, it can become epistemically dangerous.

Another objection might be voiced from the Bayesian camp: Is this not just an artifact of violating the Principle of Total Evidence? Shouldn't agents, as Bayesians maintain, always update their beliefs based on all available information, without pre-filtering? Miščević anticipated this by framing luck as both a first-order and second-order concern. In our model, coherence similarly does not simply overrule evidence; instead, it acts as a second-order indicator that the incoming evidence may itself be suspect. This maneuver works, but only partially. Our simulations show that coherence is a good indicator under *noise*, but not under *systematic bias*, precisely because bias masquerades as coherence when agents already have distorted priors. The upshot is sobering but philosophically fruitful. Coherence does, in Miščević's sense, mitigate reflective luck, but only under specific environmental conditions.

The broader lesson is that stability alone does not suffice for epistemic virtue. Stability must be *adaptive*. Agents must distinguish between environments where resisting perturbation is rationally virtuous and those where openness to incoherence is, paradoxically, the route to epistemic improvement. Miščević's core insight–that we should look to the agent's dispositions, not just to the modal profile of beliefs–therefore remains powerful. However, our findings suggest that agents require not just stability, but context-sensitive stability–a kind of epistemic flexibility that can distinguish noise from bias.

In the next section, we connect these results to broader themes in epistemology, including the ongoing debates about higher-order evidence, agent-centered virtues, and the role of coherence in belief dynamics.

5 Coherence, Reflective Luck, and Miščević's Agent-Centered Legacy

Miščević's important contribution is to have shifted the discussion of epistemic luck from the external to the internal, from the variability of the world to the fragility of the agent. His diagnosis of *reflective luck* is, in retrospect, disarmingly simple: even in a world where the facts cooperate, an agent may form true beliefs for the wrong reasons, because small, cognitively local perturbations could easily have led them astray. Our simulations both substantiate and refine this insight.

The core virtue of Miščević's approach lies in its agent-centered focus. Traditional safety-based accounts of luck concern themselves with nearby possible worlds: Could the agent easily have been wrong if the world had been slightly different? Miščević asks instead: Could the agent easily have been wrong if *she* had been slightly different? The locus of epistemic luck is thus relocated from the external world to the agent's own cognitive dispositions.

This shift is clearly expressed in our model. Agents in noisy environments are constantly bombarded by misleading signals. Those who are epistemically "flexible" in the wrong way, i.e., who accept every piece of evidence without question, are highly vulnerable. Coherence filtering, in this context, acts as a stabilizer against such fragility, reducing the agent's exposure to reflective luck. Beliefs formed under this regime are not only more stable but also less accidentally true. In this sense, coherence is performing exactly the role Miščević seems to have anticipated: it helps agents avoid being epistemically lucky *in the bad way*.

Yet, as seems to be the unfortunate truth, no good deed goes unpunished. When agents operate in environments marked by systematic bias, the same coherence-driven strategy turns against them. Once prior beliefs are subtly skewed, whether through selective exposure, initial misinformation, or sheer bad epistemic luck, coherence filtering tends to reinforce the distortion. The agent becomes stable, yes—but stably wrong. Worse still, the mechanism that was supposed to help the agent detect unreliable evidence now immunizes them from the very corrective information they need.

Here, one might imagine Miščević's protest: "But I never claimed coherence would eliminate reflective luck *unconditionally*." True enough. His original proposal was careful to note that coherence does not guarantee immunity from reflective luck. Yet our formal results bring this conditionality into sharp focus. Coherence can only play its intended role when the environment cooperates by making incoherence a signal of bad evidence. In biased environments, coherence can be systematically misleading. This result dovetails with themes from virtue epistemology more generally. Cognitive virtues must be understood as *environment-relative*. No capacity is virtuous in isolation.

Our results also highlight a noteworthy and less commonly emphasized feature of virtue epistemology: epistemic virtues may not be universally beneficial (one may, of course, object that then these are no longer virtues, but if something is a virtue only if it holds regardless of the environment, then the undertaking of virtue epistemology seems rather doomed as there will hardly be any virtues). While virtues such as open-mindedness or coherence-seeking are often regarded as unconditionally conducive to good epistemic outcomes, our model suggests that, when these dispositions are improperly calibrated–e.g., when agents weigh coherence excessively relative to evidential input–they can systematically backfire, leading to polarization, dogmatism, or resistance to evidence. This supports the view that epistemic virtues must be appropriately tuned to their epistemic environment to promote accuracy, which also aligns our view with what one may term *ecological epistemology* (cf. Thorstad, 2024). Seen in this light, Miščević's proposal may be reframed. The virtue at stake is not simply stability through coherence, but *adaptive stability*: the capacity to maintain coherence when doing so is appropriate, and to revise or sacrifice coherence when the situation demands it. The epistemic vice is not instability per se, but being stably committed to error, locked into a coherent but mistaken worldview.

This diagnosis resonates with contemporary concerns in social epistemology about the dynamics of epistemic bubbles, echo chambers, and information polarization (Nguyen, 2020; Levy, 2021). In many real-world environments, systematic bias is not merely an artifact of individual reasoning but a product of social structures and interactions. Agents often become reflectively stable in virtue of the groups they belong to, filtering information not only for internal coherence but also for alignment with socially shared belief systems. In such cases, the very mechanisms that preserve coherence may simultaneously reinforce group-based distortions, making the agent's stability socially entrenched rather than epistemically virtuous.

Still, Miščević's insight stands. The problem is not that coherence cannot help. It can, and does, in environments where random noise threatens belief stability. The problem is that coherence is *not enough*. What agents need is *adaptive coherence*, capable of adjusting when the coherence itself is the problem. Put differently, what agents need is not simply a drive to maintain coherence, but a meta-virtue: the ability to monitor when their coherence-based conservatism is epistemically appropriate, and when it has turned into a liability.

Our results thus refine rather than reject Miščević's account. Reflective luck remains a challenge. Coherence, treated as higher-order evidence, often helps in mitigating it. But its effectiveness is contingent, shaped by the structure of the agent's informational environment.

Finally, note that our model focuses on probabilistic coherence, defined as a property of an information set over which an agent has a defined subjective probability distribution. While this provides a tractable and illuminating account of one important aspect of coherence, recent work (Fogal & Risberg, forthcoming) has argued that coherence has richer dimensions, including relations of positive support and neutrality among attitudes, which cannot be fully captured within a purely probabilistic framework. Our results should therefore be understood as clarifying a significant, yet non-exhaustive, aspect of the epistemic landscape.

6 Conclusion

Miščević's account of reflective luck brought a neglected dimension of epistemic evaluation to the forefront. The deepest threats to knowledge, he argued, do not always stem from external contingency—fragile truths in nearby possible worlds—but from the agent's own instability. His proposal was modest yet powerful: perhaps coherence, long appreciated for its synchronic virtues, also has a diachronic role to play. Perhaps it can act as a stabilizer against the agent's own cognitive fragility. Our computational exploration shows that this proposal is, in a sense, vindicated. But it is also limited.

In environments dominated by random noise, coherence-sensitive agents outperform their purely first-order-evidence-only counterparts. Their belief trajectories remain closer to the truth, less sensitive to irrelevant perturbations, and less hostage to evidential flukes. This is precisely the kind of *reflective stability* that Miščević sought. Coherence seems to act, here, as a cognitive virtue. It protects the agent not merely from bad luck *in the world*, but from bad luck *in herself.*

Yet, as we have seen, the success of this strategy is not universal. In environments where the real threat is systematic bias rather than random noise, coherence becomes epistemically hazardous. Agents who rely too heavily on coherence as a filtering mechanism risk insulating themselves from exactly the kinds of corrective evidence that could improve their beliefs. In such settings, coherence does not minimize reflective luck; it entrenches it. Worse, it does so under the guise of stability.

This should not surprise us. After all, reflective luck is *agent-centered.* It is about how fragile agents are in light of small perturbations. But not all perturbations are noise. Some are corrections. Some are exactly what is needed for belief improvement. An agent who is too stable–who ignores such perturbations because they threaten coherence–may fall victim to a second-order epistemic vice: reflective rigidity.

Does this undermine Miščević's original insight? We believe it does not. If anything, it enriches it. Reflective luck is indeed a real and important epistemic threat. Stability is a plausible remedy–but only if it is *adaptive stability*. Coherence is thus best seen not as an epistemic virtue *simpliciter*, but as an environmentally modulated heuristic. When applied in the right contexts–noisy, chaotic, and unstructured informational environments–it helps agents avoid cognitive fragility. When applied indiscriminately, it can exacerbate the very problem it is meant to solve.

One could imagine Miščević, were he to see these results, responding: "But of course! Epistemic luck has many faces." Indeed, the deeper lesson is that agents need more than stability. They need *the right kind* of stability. What matters is not just that agents resist being easily thrown off by small changes, but that they are attuned to which changes are merely noise, and which are genuine signals for epistemic improvement.

In this light, coherence remains a valuable epistemic tool—but it must be deployed reflectively and responsively. Our results do not just validate Miščević's focus on agent-centered virtues. They show that computational methods can clarify how such virtues operate, when they succeed, and where they risk failure. The future challenge, both philosophically and practically, is to design strategies for agents—whether human or artificial–so that they are capable of achieving *adaptive coherence*: resilient yet not dogmatic; stable yet not blind.

The next step, both formally and philosophically, is to develop models and frameworks that provide such meta-reflective capacities. Addressing this question may require expanding Miščević's framework to include mechanisms for meta-reflection, flexibility, and epistemic humility. But if the goal is agents who are not just lucky to be right, but reliably so in virtue of their own cognitive constitution, then Miščević's focus on agent-centered virtues is precisely where we must begin.

Acknowledgements

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P6-0144). Borut Trpin furthermore acknowledges the financial support from the Slovenian Research Agency, project No. J6-60107.

References

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

Carnap, R. (1947). On the application of inductive logic. *Philosophy and Phenomenological Research*, *8*(1), 133–148. https://doi.org/10.2307/2102920

Christensen, D. (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research*, *81*(1), 185–215. https://doi.org/10.1111/j.1933-1592.2010.00366.x

Feldman, R. (2005). Respecting the evidence. *Philosophical Perspectives*, *19*, 95–119. https://doi.org/10.1111/j.1520-8583.2005.00055.x

Fogal, D. & Risberg, O. (2025). Coherence and incoherence. *Philosophical Review, 134*(4), 405–454. https://doi.org/10.1215/00318108-11964758

Goldman, A. I. (1999). *Knowledge in a Social World.* Oxford University Press.

Good, I. J. (1967). On the principle of total evidence. *The British Journal for the Philosophy of Science*, *17*(4), 319–321. https://doi.org/10.1093/bjps/17.4.319

Greco, J. (2003). *Knowledge as credit for true belief.* In M. DePaul & L. Zagzebski (Eds.), *Intellectual Virtue: Perspectives from Ethics and Epistemology* (Chap. 5). Oxford University Press.

Justin, M. & Trpin, B. (2025). Coherence-Based Evidence Filtering: A Computational Exploration. Fortn: *Proceedings of the Annual Meeting of the Cognitive Science Society* 47. https://escholarship.org/uc/item/0668w8x3

Kelly, T. (2010). *Peer disagreement and higher-order evidence*. In R. Feldman & T. A. Warfield (Eds.), *Disagreement* (pp. 111–174). Oxford University Press.

Levy, N. (2021). *Bad Beliefs: Why They Happen to Good People*. Oxford University Press.

Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.

Miščević, N. (2007). Armchair luck: Apriority, intellection and epistemic luck. *Acta Analytica*, *22*, 48-73. https://doi.org/10.1007/BF02866210

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, *17*(2), 141–161. https://doi.org/10.1017/epi.2018.32

Pritchard, D. (2005). *Epistemic Luck*. Oxford University Press.

Sosa, E. (2003). *The place of truth in epistemology*. In M. DePaul & L. Zagzebski (Eds.), *Intellectual Virtue: Perspectives from Ethics and Epistemology* (Chap. 8). Oxford University Press.

Thorstad, D. (2024). *Inquiry Under Bounds*. Oxford University Press.

Trpin, B. & Justin, M. (2025). Coherence as a constraint on scientific inquiry. *Synthese 206, 200*. https://doi.org/10.1007/s11229-025-05281-3