UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Belief as Self-Endorsement: A Bayesian Model of Commitment Under Uncertainty

Permalink

https://escholarship.org/uc/item/7sf387gm

Journal

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

Authors

Hartmann, Stephan

Trpin, Borut

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Belief as Self-Endorsement: A Bayesian Model of Commitment Under Uncertainty

Stephan Hartmann (S.Hartmann@lmu.de)[1,2] **& Borut Trpin**[1,3,4]
[1]Munich Center for Mathematical Philosophy, LMU Munich
[2]Department of Psychology, LMU Munich
[3]Department of Philosophy, University of Ljubljana
[4]Department of Philosophy, University of Maribor

Abstract

Humans routinely treat propositions as settled for the purposes of reasoning and action despite remaining uncertainty. Standard probabilistic models capture graded belief, but they leave the formation and epistemic role of such commitments underspecified. We propose a minimal Bayesian network model in which commitment is represented as an endogenous endorsement event regulated by metacognitive confidence. Within this framework, endorsement functions as a confidence-gated internal signal: conditioning on endorsement can increase subjective probability once, without new external evidence, while its dependence structure prevents epistemic bootstrapping. Extending the model to include explicit evidence yields a distinctive empirical prediction: the effect of evidence on commitment is systematically moderated by confidence. This prediction supports experimental designs in which evidence strength and confidence are independently manipulated and both commitment and probability judgments are measured. The model thus provides a tractable and empirically testable account of commitment under uncertainty.

Introduction

Humans routinely treat propositions as settled for the purposes of reasoning and action despite remaining uncertainty. We plan under assumptions that are not certain, adopt working hypotheses as temporarily fixed, and rely on them in deliberation while remaining, in principle, open to revision (Staffel, 2019). Closely related ideas appear in work on discourse and context sets (Stalnaker, 1978, 2014), where propositions may be treated as "settled" without being regarded as certain. Such commitment plays a central functional role in cognition: it stabilizes inference, simplifies reasoning, and enables coordinated action.

Standard probabilistic models capture graded uncertainty with great precision, but they leave this role of commitment underdetermined. In a Bayesian framework, agents assign probabilities and update them in light of evidence, yet it remains unclear how a proposition can be treated as settled while its probability remains strictly below one, or how commitment can itself influence subsequent degrees of belief. Sampling-based approximations to Bayesian inference (Vul et al., 2014) suggest that agents may temporarily treat propositions as fixed on the basis of limited samples, but they do not by themselves explain how such transient outcomes are stabilized or how they feed back into subsequent belief.

A common response is to identify belief with sufficiently high probability, as in Lockean threshold views (Foley, 1992). While simple, such approaches face well-known difficulties, including sensitivity to arbitrary thresholds and conflicts with closure principles. More sophisticated proposals introduce stability constraints or dynamic thresholds (Leitgeb, 2017; Staffel, 2016), but these accounts remain largely static: they characterize admissible belief states without explaining how commitment dynamically interacts with graded belief. Other approaches treat belief as a pragmatic policy for reasoning or action (e.g., MacFarlane, 2025). While this captures belief's organizing function, it leaves its epistemic role unclear: adopting a proposition as a premise does not, by itself, explain why an agent should become more confident that it is true. Classical work on acceptance and commitment (Levi, 1980) similarly emphasizes the role of settling propositions, but does not embed this act within a probabilistic dependency structure.

In this paper, we pursue a different strategy. We model commitment as an *endogenous endorsement event* within a Bayesian network. Endorsement is an internal act by which an agent treats a proposition as settled for the purposes of reasoning. Crucially, endorsement is not treated as independent evidence, but as a fallible internal signal whose epistemic impact is regulated by metacognitive confidence. Confidence is understood here as a metacognitive assessment of the reliability or correctness of one's first-order cognitive performance, distinct from first-order belief in the proposition itself (Fleming & Daw, 2017; Yeung & Summerfield, 2012).

This framework yields a minimal formal model—the *ABC model*—with three variables: a target proposition (A), an endorsement event (B), and metacognitive confidence (C). Within this structure, endorsement can rationally increase subjective probability once, without new external evidence, while remaining sensitive to evidence and avoiding epistemic bootstrapping. The one-off character of this update follows from the dependence structure of the model: since endorsement is endogenous, treating it repeatedly as fresh evidence would amount to double-counting.

The model makes a clear empirical prediction. When extended to include explicit evidence, the effect of evidence on commitment is predicted to be systematically moderated by confidence. In particular, changes in endorsement induced by evidence scale with confidence, yielding a characteristic interaction pattern. This prediction supports a straightforward experimental strategy in which evidence strength and confidence are independently manipulated and both endorsement

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

and probability judgments are measured.

The contribution of the paper is thus fourfold. First, it provides a principled way of representing commitment under uncertainty within a probabilistic framework. Second, it shows how endorsement can increase subjective certainty once without new external evidence while preserving Bayesian coherence. Third, it identifies a structural constraint that blocks epistemic bootstrapping. Fourth, it derives a distinctive, testable empirical signature linking evidence, confidence, and commitment.

The paper proceeds as follows. We first situate the proposal within empirical and theoretical work on confidence and commitment. We then introduce the ABC model and derive its core properties. Next, we extend the model to include explicit evidence and develop its empirical predictions. We close by discussing implications, limitations, and possible extensions.

Background: Confidence and Commitment in Cognition

The proposal developed in this paper is motivated by converging strands of research in cognitive science and philosophy that highlight a gap between probabilistic belief representation and the functional role of commitment in cognition. Across domains, human reasoning involves not only graded uncertainty but also the formation of commitments that stabilize inference and guide action. What remains underdeveloped is a unified account of how such commitments relate to probabilistic belief and metacognitive assessment.

A first strand concerns the functional role of commitment. Empirical and philosophical work suggests that believing is not a unitary psychological kind and that belief-like attitudes differ in how representations are processed and deployed (Leeuwen & Lombrozo, 2023). Agents routinely rely on categorical assumptions to simplify reasoning, terminate inquiry, or coordinate action, even while acknowledging residual uncertainty (Staffel, 2019). This indicates that commitment cannot be identified with high probability alone, but must be understood in terms of how a representation is deployed in cognition.

A second strand concerns metacognitive confidence. A large body of research treats confidence as a metacognitive assessment of the reliability or correctness of one's own cognitive performance, rather than as a first-order estimate of a proposition's truth (Fleming & Daw, 2017; Yeung & Summerfield, 2012). Confidence judgments track cues such as fluency and related indicators of processing ease, and play a regulatory role in cognition, influencing stopping rules, information-seeking, and the willingness to rely on a representation (Alter & Oppenheimer, 2009; Desender et al., 2018). Crucially, confidence is dissociable from probability: agents may judge a proposition to be likely while having low confidence in that judgment, or vice versa. This distinction will play a central role in the model below, where confidence modulates the epistemic impact of commitment.

A third strand concerns endorsement and commitment effects. Across a range of paradigms, explicitly committing to a claim—by endorsing it or adopting it as a premise—has been shown to influence subsequent judgment and reasoning. Classic studies on belief perseverance show that agents may maintain hypotheses even after their original evidential basis has been undermined, especially when they have generated explanatory support for them (Anderson et al., 1980). Related effects appear in choice-induced preference change and self-justification (Izuma & Murayama, 2013). These phenomena suggest that endorsement is not merely a report of belief or a piece of independent evidence, but plays a distinctive functional role in shaping both reasoning and subjective certainty.

Finally, Bayesian network models provide a natural formal framework for integrating these ideas. Such models have long been used to represent how testimony, source reliability, and background beliefs interact probabilistically (Bovens & Hartmann, 2003). Recent work extends this framework to internal signals and self-directed testimony, showing how endogenous epistemic events can be represented without violating probabilistic coherence (Hahn et al., 2024; Heinzelmann & Hartmann, 2022). These approaches highlight both the potential and the difficulty of treating internal acts as evidence: without a structured account of dependence, there is a risk of epistemic double-counting.

It is useful, in this context, to contrast the present approach with sampling-based accounts of approximate Bayesian inference (Vul et al., 2014). Such models explain why agents may temporarily treat propositions as fixed on the basis of limited samples, but they do not by themselves explain how such transient outcomes are stabilized or how they feed back into subsequent belief in a systematic way. In particular, they do not predict a structured interaction between confidence and the effect of evidence on commitment.

Taken together, these strands point to a representational gap. Empirical work emphasizes confidence-regulated commitment, while standard probabilistic models lack a principled way of representing such commitment without reducing it to high probability, pragmatic policy, or independent evidence. The aim of the present paper is to fill this gap by modeling commitment as an endogenous endorsement event regulated by metacognitive confidence within a Bayesian framework.

The ABC Model

We now introduce a minimal Bayesian model of commitment under uncertainty. The aim is not to provide a process-level account of belief formation, but to capture, in a formally transparent way, how an act of commitment can interact with graded belief and metacognitive confidence while preserving probabilistic coherence.

The model contains three binary variables. A represents a target proposition. In the present context, however, A is best understood as the agent's *first-order representation* that the proposition obtains, rather than as objective truth in the world. This distinction matters: agents may confidently endorse propositions that are in fact false if their first-order

representation is mistaken. A richer model could introduce a separate world-state variable, but we abstract from this here in order to isolate the interaction between representation, endorsement, and confidence. B represents an *endorsement event*: an internal act by which the agent treats A as settled for the purposes of reasoning. C represents metacognitive confidence, understood as a second-order assessment of the reliability of the cognitive process producing the representation of A.

The variables are related by the Bayesian network shown in Figure 1. Both A and C are parents of B, yielding a collider structure. This captures the idea that endorsement depends jointly on first-order representation and second-order confidence. The absence of a direct edge between A and C is an idealization. In realistic settings, both may be influenced by common factors such as evidence or task difficulty. We abstract from these common causes to isolate the regulatory role of confidence; In the next section, we partially relax this idealization by introducing an explicit evidence variable.

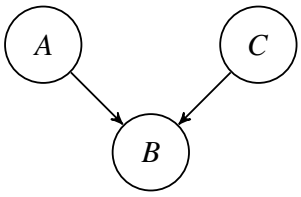

Figure 1: The ABC model: target proposition (A), endorsement event (B), and metacognitive confidence (C).

We assume prior probabilities $P(\mathrm{A}) = a$ and $P(\mathrm{C}) = c$, with $a, c \in (0, 1)$, and a parameter $\beta \in (0, 1)$ representing the agent's baseline propensity to endorse under low confidence. The conditional probabilities are:

$$P(\mathrm{B} \mid \mathrm{A}, \mathrm{C}) = 1, \quad P(\mathrm{B} \mid \neg\mathrm{A}, \mathrm{C}) = 0,$$

$$P(\mathrm{B} \mid \mathrm{A}, \neg\mathrm{C}) = \beta, \quad P(\mathrm{B} \mid \neg\mathrm{A}, \neg\mathrm{C}) = \beta.$$

When confidence is high, endorsement reliably tracks the agent's first-order representation. This does not imply infallibility about the world, but only reliability with respect to the agent's own doxastic state. When confidence is low, endorsement occurs with probability β independently of A, capturing a general disposition to settle propositions under uncertainty. The model can be relaxed by introducing noise, for instance by replacing 1 and 0 with $1 - \varepsilon$ and ε, without affecting the qualitative results.

Endorsement functions as a fallible internal signal. Conditioning on B yields:

$$P(\mathrm{A} \mid \mathrm{B}) = \frac{a\,[c + (1 - c)\beta]}{ac + (1 - c)\beta}.$$

It follows that

$$P(\mathrm{A} \mid \mathrm{B}) > P(\mathrm{A}) \quad \text{whenever } c > 0.$$

Thus, observing that one has endorsed a proposition can rationally increase one's subjective probability, even in the absence of new external evidence. The magnitude of this increase is modulated by confidence.

Endorsement also updates confidence:

$$P(\mathrm{C} \mid \mathrm{B}) = \frac{ac}{ac + \beta(1 - c)}.$$

When $\beta < a$, endorsement raises confidence in the reliability of the underlying process; when $\beta > a$, it lowers it. This provides a simple representation of the metacognitive consequences of endorsement.

A numerical example illustrates the role of confidence. Consider $P(\mathrm{A}) = 0.6$, $P(\mathrm{C}) = 0.7$, and $\beta = 0.2$. Then

$$P(\mathrm{A} \mid \mathrm{B}) \approx 0.95.$$

If confidence is lower, say $P(\mathrm{C}) = 0.2$, the same endorsement yields

$$P(\mathrm{A} \mid \mathrm{B}) \approx 0.73.$$

The evidential force of endorsement is therefore not fixed, but depends on confidence.

The same structure applies beyond perceptual cases. Consider a reasoning task in which an agent evaluates whether a medical diagnosis is correct on the basis of mixed indicators. The agent's first-order representation A is based on noisy or partially conflicting evidence. Confidence C reflects the perceived reliability of the diagnostic procedure or information source. An endorsement event B corresponds to treating the diagnosis as settled for a downstream decision, such as selecting a treatment plan or drawing further inferences. When confidence is high, endorsement tracks the agent's representation; when confidence is low, endorsement reflects a baseline disposition to commit despite uncertainty. As before, conditioning on endorsement can increase subjective probability once, with the magnitude of this effect depending on confidence.

A crucial feature of the model is that endorsement is *endogenous*. Because B depends on A and C, it cannot be treated as independent evidence. Conditioning on endorsement can therefore rationally increase credence at most once. Repeatedly treating the same endorsement as fresh evidence would amount to double-counting. The model thus imposes a principled limit on epistemic bootstrapping: reaffirming a commitment without new information cannot rationally increase credence further. This constraint follows directly from the dependence structure of the model, rather than from an additional normative principle.

This one-off evidential role is compatible with the continuing use of an endorsed proposition in reasoning. Once endorsement has occurred, A may function as a premise in downstream inference without being reintroduced as fresh evidence. The point of endorsement is therefore not to generate repeated confirmation, but to license a stable mode of premise use. This distinction explains how commitment can simplify reasoning while preserving probabilistic discipline.

The qualitative results do not depend on the idealized conditional probabilities used above. In particular, replacing the deterministic relations in the high-confidence case with noisy

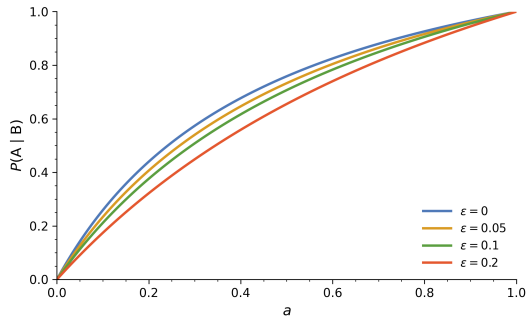

Figure 2: Posterior probability $P(\mathrm{A} \mid \mathrm{B})$ as a function of the prior $a = P(\mathrm{A})$ for different noise levels ϵ in the high-confidence endorsement mechanism. Here $P(\mathrm{C}) = 0.5$ and $\beta = 0.4$.

versions, such as $P(\mathrm{B} \mid \mathrm{A}, \mathrm{C}) = 1 - \varepsilon$ and $P(\mathrm{B} \mid \neg\mathrm{A}, \mathrm{C}) = \varepsilon$, preserves the one-off evidential role of endorsement and its dependence on confidence. Similarly, allowing confidence to depend on additional variables, such as evidence or task difficulty, does not affect the central result that endorsement functions as a dependent internal signal. These variants show that the model's key predictions are structurally robust rather than artifacts of a specific parameterization.

The ABC model therefore captures commitment as a confidence-regulated internal signal. It allows for commitment under uncertainty, explains how endorsement can increase subjective certainty once, and enforces a structural constraint that blocks epistemic inflation. In the next section, we extend the model to include explicit evidence and derive empirically testable predictions.

Adding Explicit Evidence and Empirical Signatures

To connect the ABC model more directly to empirical settings, we extend it by introducing an explicit evidence variable. Let E represent external evidence bearing on the target proposition and thereby affecting the agent's first-order representation A, such as perceptual input, feedback, or informational cues. We assume that E influences endorsement only indirectly, by affecting the agent's first-order representation A. This yields the Bayesian network shown in Figure 3.

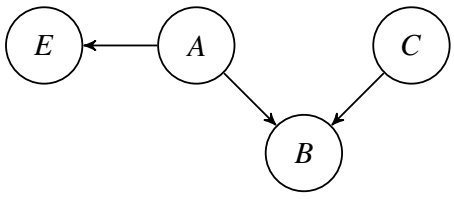

Figure 3: ABC model with explicit evidence (E).

Within this extended structure, the effect of evidence on endorsement is given by:

$$P(\mathrm{B} \mid \mathrm{E}) - P(\mathrm{B}) = \big(P(\mathrm{A} \mid \mathrm{E}) - P(\mathrm{A})\big) \cdot P(\mathrm{C}). \qquad (1)$$

Equation (1) expresses the central empirical prediction of the model. The impact of evidence on commitment is systematically moderated by confidence. Changes in endorsement induced by evidence scale linearly with $P(\mathrm{C})$. When confidence is high, even modest evidential shifts in A produce substantial changes in endorsement; when confidence is low, the same shifts have a weaker effect. Accordingly, the slope of endorsement as a function of evidence strength should increase with confidence.

This prediction can be tested in standard decision-making or perceptual tasks. Consider a signal detection paradigm in which participants judge whether a stimulus is present. Evidence E corresponds to signal strength, while confidence C can be independently manipulated or measured, for instance via noise, time pressure, or metacognitive prompts. Endorsement B can be operationalized as committing to a response for a downstream task, such as using it as a premise in reasoning or making a consequential choice. The model predicts a specific interaction: the effect of evidence strength on endorsement should be stronger at higher levels of confidence. Equivalently, endorsement rates should be more sensitive to evidence when participants are more confident in their judgments.

A simple implementation would use a two-stage task. In the first stage, participants make a probabilistic judgment about A based on graded evidence E and report confidence. In the second stage, participants are required to either endorse or withhold commitment to A for use in a downstream task. The critical test concerns the interaction between evidence strength and reported confidence in predicting endorsement rates. The model predicts that the slope of endorsement as a function of evidence increases with confidence, even when first-order probability judgments are held constant.

This interaction distinguishes the ABC model from several competing accounts. Threshold (Lockean) models predict a discontinuous transition once a belief threshold is crossed, rather than a graded, confidence-modulated response. Models that treat endorsement as independent evidence predict that endorsement should uniformly increase belief, irrespective of confidence, and can be repeatedly exploited to inflate credence. Purely pragmatic accounts predict that commitment structures reasoning but should not systematically affect probability judgments. By contrast, the present model predicts an intermediate pattern: endorsement both responds to evidence and feeds back into belief, but in a confidence-regulated and non-bootstrapping manner.

The prediction is directly falsifiable. If endorsement tracks evidence independently of confidence, the model loses its distinctive empirical advantage. Conversely, if repeated endorsement leads to systematic increases in subjective probability without new evidence, this would violate the model's structural constraint against double-counting.

It is useful to contrast this prediction with sampling-based accounts of approximate Bayesian inference (Vul et al., 2014). Such models explain why agents may temporarily treat propositions as fixed based on limited samples, but they do not predict a systematic modulation of the evidence–commitment relationship by independently measured confidence. The interaction expressed in Equation (1) therefore constitutes a distinct structural prediction of the ABC model. Sampling-based approaches predict variability in whether propositions are treated as fixed, but not the structured dependence of this process on confidence.

The framework also clarifies how phenomena such as belief persistence arise. Weak responsiveness of commitment to evidence can occur when confidence is low, in which case changes in A translate only weakly into changes in endorsement. Conversely, when confidence is high, even disconfirming evidence should lead to substantial revision of commitment. Persistence and revision thus emerge as different regimes of a single confidence-regulated mechanism.

Introducing explicit evidence transforms the ABC model from a conceptual proposal into a source of empirically discriminating predictions. The next section discusses the implications, limitations, and extensions of this framework.

Discussion

The ABC model provides a minimal representation of commitment within probabilistic cognition. Its central claim is that commitment can be modeled as an endogenous endorsement event whose epistemic impact is regulated by metacognitive confidence. This accounts for commitment under uncertainty, the stabilization of reasoning, and a one-off increase in subjective certainty without treating endorsement as independent evidence.

Because endorsement depends on both first-order representation and confidence, its evidential role is inherently limited. Conditioning on endorsement can increase subjective probability once, but repeated reaffirmation cannot rationally provide further support. This blocks epistemic bootstrapping and double-counting as a consequence of the model's dependency structure.

The same structure explains variation in responsiveness to evidence. Low confidence weakens the translation from first-order representation to commitment, yielding apparent persistence; high confidence strengthens this translation, yielding revision. Persistence and revision thus emerge as different regimes of a single mechanism.

At the same time, the model is highly idealized. The variables are binary and abstract, and the framework does not provide a process-level account of how representations or confidence are formed. In particular, the model assumes that confidence tracks the reliability of the underlying cognitive process. In many real-world settings, however, confidence may be systematically miscalibrated—for instance, in cases of overconfidence, misinformation, or echo chambers. Modeling such cases would require extending the framework so that confidence can become decoupled from actual reliability.

A further limitation concerns the interpretation of the target variable. In the present model, A represents the agent's first-order representation, not objective truth. This allows the model to capture confident endorsement of false propositions, but it leaves open how endorsement relates to accuracy in the world. An extended model could introduce a separate world-state variable and explicitly represent the relation between belief, confidence, and truth.

Several extensions suggest themselves. The model can be generalized to include multiple propositions, allowing endorsement to propagate through networks of interdependent beliefs. It can be extended to social settings in which endorsement interacts with testimony from others, enabling a unified treatment of self- and other-directed epistemic signals. Finally, embedding the model in a sequential framework would allow one to study how endorsement and evidence interact over time.

Conclusion

This paper has proposed a minimal Bayesian model of commitment under uncertainty. By representing commitment as an endogenous endorsement event regulated by metacognitive confidence, the model shows how belief-like states can be integrated into a probabilistic framework without being reduced to high probability or treated as independent evidence.

The framework yields three core results. Endorsement can rationally increase subjective certainty once without new external evidence, its epistemic impact is modulated by confidence, and its dependence structure blocks epistemic bootstrapping. When extended to include explicit evidence, the model predicts a characteristic interaction: the effect of evidence on commitment scales with confidence.

More generally, the model illustrates how commitment can be understood as a structured component of probabilistic cognition rather than as an add-on or heuristic. Whether and to what extent this framework captures human belief dynamics remains an empirical question, but it provides a clear set of structural constraints and testable predictions for future work.

Appendix

Proof of Equation (1)

We consider the Bayesian network in Figure 3 with parameters

$$P(\mathrm{A}) = a, \quad P(\mathrm{C}) = c, \quad P(\mathrm{E} \mid \mathrm{A}) = p, \quad P(\mathrm{E} \mid \neg\mathrm{A}) = q,$$

and conditional probabilities for B as defined in Section "The ABC Model".

First, marginalizing over A and C yields

$$P(\mathrm{B}) = ac + \beta(1 - c).$$

Next,

$$P(\mathrm{E}) = ap + (1 - a)q,$$

and

$$P(\mathrm{B}, \mathrm{E}) = acp + \beta(1 - c)\big(ap + (1 - a)q\big).$$

Thus,

$$P(\mathrm{B} \mid \mathrm{E}) = \frac{P(\mathrm{B}, \mathrm{E})}{P(\mathrm{E})} = c\, P(\mathrm{A} \mid \mathrm{E}) + \beta(1 - c),$$

where

$$P(\mathrm{A} \mid \mathrm{E}) = \frac{ap}{ap + (1 - a)q}.$$

Subtracting $P(\mathrm{B})$ gives

$$P(\mathrm{B} \mid \mathrm{E}) - P(\mathrm{B}) = \big(P(\mathrm{A} \mid \mathrm{E}) - P(\mathrm{A})\big)\, c,$$

which establishes Equation (1). □

Acknowledgments

This work was supported by the German Research Foundation (DFG) under grants HA3000/29-1 and HA3000/30-1 (SH), and by the Slovenian Research Agency through research core funding No. P6-0144 and project No. J6-60107 (BT). We thank Wolfgang Spohn for helpful comments on an earlier version of this paper, as well as the anonymous reviewers for their valuable feedback.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, *39*(6), 1037.

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

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761–778.

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91.

Foley, R. (1992). The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly*, *29*(2), 111–124.

Hahn, U., Merdes, C., & von Sydow, M. (2024). Knowledge through social networks: Accuracy, error, and polarisation. *PLOS ONE*, *19*(1), e0294815. https://doi.org/10.1371/journal.pone.0294815

Heinzelmann, N., & Hartmann, S. (2022). Deliberation and confidence change. *Synthese*, *200*(42), 1–15.

Izuma, K., & Murayama, K. (2013). Choice-induced preference change in the free-choice paradigm: A critical methodological review. *Frontiers in Psychology*, *4*, 41.

Leeuwen, N. V., & Lombrozo, T. (2023). The puzzle of belief. *Cognitive Science*, *47*(2), e13245. https://doi.org/10.1111/cogs.13245

Leitgeb, H. (2017). *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford University Press.

Levi, I. (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. MIT Press.

MacFarlane, J. (2025). Belief: What is it good for? *Erkenntnis*, *90*, 947–864.

Staffel, J. (2016). Beliefs, buses and lotteries: Why rational belief can't be stably high credence. *Philosophical Studies*, *173*(7), 1721–1734. https://doi.org/10.1007/s11098-015-0574-2

Staffel, J. (2019). How do beliefs simplify reasoning? *Noûs*, *53*(4), 937–962. https://doi.org/10.1111/nous.12254

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics, volume 9: Pragmatics* (pp. 315–332). Academic Press.

Stalnaker, R. (Ed.). (2014). *Context*. Oxford University Press.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310–1321.