Predict this: minimal commitments for testable predictive processing explanations

Samara, I.

Neuroscience and Biobehavioral Reviews 188 (2026) 106822

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Predict this: Minimal commitments for testable predictive processing explanations

Iliana Samara

Cognitive Psychology Unit, Institute of Psychology, Leiden University, Wassenaarseweg 52, Leiden 2333 AK, the Netherlands

ARTICLE INFO

Keywords:
Predictive processing
Active inference
Theory evaluation
Uncertainty
Interoception

ABSTRACT

Predictive processing (PP) is a family of frameworks that link perception, action, learning, and attention to inference under uncertainty. PP terms (e.g., priors, prediction error, precision) are becoming more common, yet commitments often remain implicit. This flexibility allows multiple mappings from PP vocabulary to predictands, measures, and mechanisms, weakening inference. Here I propose five minimal commitments (PP-MC) as a reporting and design checklist for PP-framed mechanistic claims. PP-MC asks authors to specify: (1) the predictand and timescale, (2) an operational proxy for prediction error and its expected direction, (3) the relevant uncertainty or precision construct and how it is manipulated or measured, (4) the pathway by which error is reduced, and (5) at least one discriminating prediction against an alternative. A worked example in interoception illustrates how PP-consistent pathways imply different measures and manipulations. PP-MC aims to make PP-framed explanations auditable and more readily testable in a given study.

Predictive processing (PP) is a family of frameworks about how organisms perceive, act, and learn under uncertainty. In PP, perception is often described as inference under a generative model: a structured set of assumptions about how latent causes generate sensory signals (Clark, 2013; Hohwy, 2013, 2020; see Glossary for an explanation of all terms). The difference between predicted and observed signals, or between predicted and inferred causes, is the prediction error. Active inference extends this logic to action and planning by selecting policies according to expected free energy, a prospective objective that is related to, but not identical with, variational free energy (Champion et al., 2026; Friston et al., 2017, 2023; Millidge et al., 2021).

PP vocabulary is now common across cognitive, affective, and social domains, including interoceptive accounts of emotion and selfhood (Seth, 2013; Seth and Friston, 2016). However, this flexibility introduces degrees of freedom that limit inference unless reduced by measurement and design choices. For example, the term "prediction" could refer to low-level sensory features, motor trajectories, task outcomes, or latent states such as intentions. Recent reviews highlight that support for predictive coding and active inference remains mixed, in part because underspecified constructs make several PP claims difficult to falsify (Bowman et al., 2023; Hodson et al., 2024). In addition, some commonly used neural markers of perceptual expectation show null results in specific paradigms (den Ouden et al., 2025), which complicates inference from PP terminology without explicit operationalization.

These issues are connected to continuing discussions about how empirically accurate and theoretically important PP is. Cao (2020) argues that PP labels can redescribe existing neural-signal interpretations unless the relevant causal or informational mapping is specified, and it has been similarly argued that PP should be treated as a varied framework whose empirical commitments need to be made explicit (Milkowski and Litwin, 2022). Sun and Firestone (2020) use the dark-room problem to show how unconstrained prediction-error minimization can become too permissive. The aim here is in line with those critiques: to specify what a PP-framed explanation must commit to in a given study.

This problem is not unique to PP. Broad verbal frameworks in cognitive science tend to drift unless they specify what is being predicted, what would count as error, how uncertainty is represented, and which alternatives are being ruled out. However, PP is a salient case because it has generated a shared vocabulary for hierarchical prediction, error signalling, precision weighting, and active inference (see, e.g., Clark, 2013; Friston et al., 2017; Hohwy, 2013, 2020). In verbal applications, this vocabulary is sometimes used to support mechanistic claims without an explicit mapping from constructs to observable behaviour. This mini-review focuses on study-level specification for PP-framed mechanistic claims that are presented verbally rather than as explicit computational models. I propose five minimal commitments (PP-MC) that reduce flexibility. PP-MC is intended as a minimum reporting bar

E-mail address: i.samara@fsw.leidenuniv.nl.

https://doi.org/10.1016/j.neubiorev.2026.106822
Received 3 February 2026; Received in revised form 25 May 2026; Accepted 17 June 2026
Available online 20 June 2026

and can also serve as scaffolding toward fuller formalization.

1. Motivating example: interoception and anxiety

Interoception, the sensing and interpretation of internal bodily states, is a prominent test case for PP (Seth and Friston, 2016). PP accounts propose that the brain generates predictions about visceral signals (cardiac, respiratory, gastric) and uses prediction errors to update beliefs about bodily states. Interoceptive dysfunction has been proposed as contributing to anxiety and depression, and active inference accounts propose mechanisms involving hyperprecise priors and belief updating (Paulus et al., 2019).

Interoception illustrates why verbal PP explanations can diverge while using the same vocabulary. Consider two PP-consistent pathways in interoceptive psychopathology accounts (Paulus et al., 2019). In Pathway A, anxiety is attributed to overly precise priors about threat-related bodily states that resist updating when visceral evidence changes. In Pathway B, interoceptive prediction errors are assigned low sensory precision, reducing their influence on belief updating and downstream emotion regulation. Both pathways are compatible with PP terminology, but they imply different predictands, different "prediction error" proxies, and different manipulations (Table 1). This matters because a study observing a group difference in interoceptive task performance could otherwise accommodate the result under either pathway. Without advance commitments about the predictand, error proxy, precision quantity, and manipulation, the same empirical pattern risks becoming post-hoc support for PP rather than a test of a particular PP mechanism.

Whether a study can actually test PP claims depends on whether the study specifies these commitments. For cardiac interoception, heartbeat perception tasks are common, but interoceptive performance and metacognitive insight are dissociable (Garfinkel et al., 2015), and heartbeat counting accuracy scores can be sensitive to nonperceptual factors (Zamariola et al., 2018). If a paper claims that "prediction errors are increased in anxiety," reviewers (and readers in general) need to know what is treated as the predictand, which measure indexes prediction error, which uncertainty construct is assumed to differ, and what pattern would count against the proposed pathway. A non-PP comparison class is also available. For example, cognitive accounts of anxiety emphasize threat-related attentional bias (Bar-Haim et al., 2007), which predicts anxiety-linked processing biases even when interoceptive precision is held constant.

2. Why PP explanations drift

Recent papers argue that PP explanations become difficult to evaluate when key commitments are left implicit, because the same PP terms can be mapped to different predictands, measures, and mechanisms (Bowman et al., 2023; Hodson et al., 2024; Milkowski and Litwin, 2022). Verbal PP explanations usually include two parts: a general architecture (hierarchical prediction, error signals, precision weighting), and a study-specific mapping from that architecture to task variables and observables. Drift arises when the architecture is treated as sufficient explanation while the mapping remains implicit.

Three drift patterns are common, and each has been identified in previous debates. First, vocabulary substitution occurs when established constructs such as expectation, attention, or cue validity are redescribed as priors or precision without specifying what the PP term adds to the mechanism. Cao (2020), for example, argues that PP descriptions of neural signals can leave unresolved what information a signal carries and how that mapping is fixed; Milkowski and Litwin (2022) made the broader point that PP needs explicit empirical commitments rather than a relabelling of familiar explanatory resources.

Second, uncommitted prediction targets occur when "prediction" can refer to a sensory feature, an outcome, an action, or a latent state. Without a primary predictand and timescale, a PP explanation can be retrofitted to many outcomes. This is one reason broad PP accounts invite challenges about theoretical scope, including the worry that unconstrained prediction-error minimization can explain too much (Hohwy, 2020; Sun and Firestone, 2020).

Third, elastic appeals to precision occur when precision weighting is invoked after the fact to accommodate an observed effect direction. Precision is central to PP, but "precision" may refer to sensory precision, prior precision, policy precision, volatility, or other uncertainty quantities. Without specifying and measuring the relevant construct, precision can absorb failed predictions rather than discipline them (Bowman et al., 2023; Hodson et al., 2024; Milkowski and Litwin, 2022).

3. PP-MC: five minimal commitments

PP-MC is a five-item checklist for studies that frame a mechanistic claim in PP terms without fully specifying a formal model or comparing models quantitatively (Fig. 1). The five items correspond to recurring degrees of freedom in PP-framed explanations: what is predicted, what counts as error, which uncertainty quantity is in play, how error is reduced, and the explanation it competes with. PP-MC is intentionally minimal. It aims to increase empirical constraint and to clarify which patterns would count against the proposed mechanism.

1. **Predictand and timescale.** State what is being predicted (the predictand) and the relevant timescale. For example, specify whether predictions target sensory features (e.g., grating orientation), actions, task outcomes, or latent states, and whether the claim concerns within-trial inference or learning across trials.
2. **Prediction error proxy (and expected direction).** Define how prediction error will be indexed in the study and what direction is expected. For example, specify whether prediction error is operationalized as a deviant-minus-standard ERP, an fMRI response difference, a behavioural surprise cost, or a model residual. State which pattern corresponds to increased versus reduced error.
3. **Uncertainty or precision construct.** Specify the relevant form of uncertainty (e.g., sensory noise, ambiguity, volatility) and how it will be manipulated or measured. If precision weighting is central, state which quantity is assumed to change (sensory precision, prior precision, or policy precision) and in which condition.
4. **Error-reduction pathway.** Identify how error is reduced in the paradigm: perceptual updating, learning across time, action, or active sampling. This is crucial as different pathways imply different measures and different temporal signatures.
5. **Discriminating prediction.** Name at least one plausible comparison class (an alternative PP pathway or a non-PP account) and state a prediction where the accounts diverge, along with a design or analysis that could distinguish them. This commitment carries the most weight: the first four items make the claim explicit, but the fifth identifies the observation that would favour one pathway over

Table 1
PP-MC commitments for two PP-consistent pathways in interoceptive anxiety.

PP-MC item	Pathway A: Hyperprecise priors	Pathway B: Low-sensory precision
1. Predictand	Threat-related bodily state (latent arousal)	Cardiac signal timing
2. Error proxy	Slow updating after violated expectations	Weak prediction error signaling
3. Uncertainty	Prior precision (inflexibly high)	Sensory precision (low)
4. Pathway	Context rigidity (limited updating)	Impaired ascending error signals
5. Discriminating prediction	Stronger effect of prior manipulations	Stronger effect of signal-quality manipulation

Note. Both pathways are consistent with PP vocabulary but imply different tests.

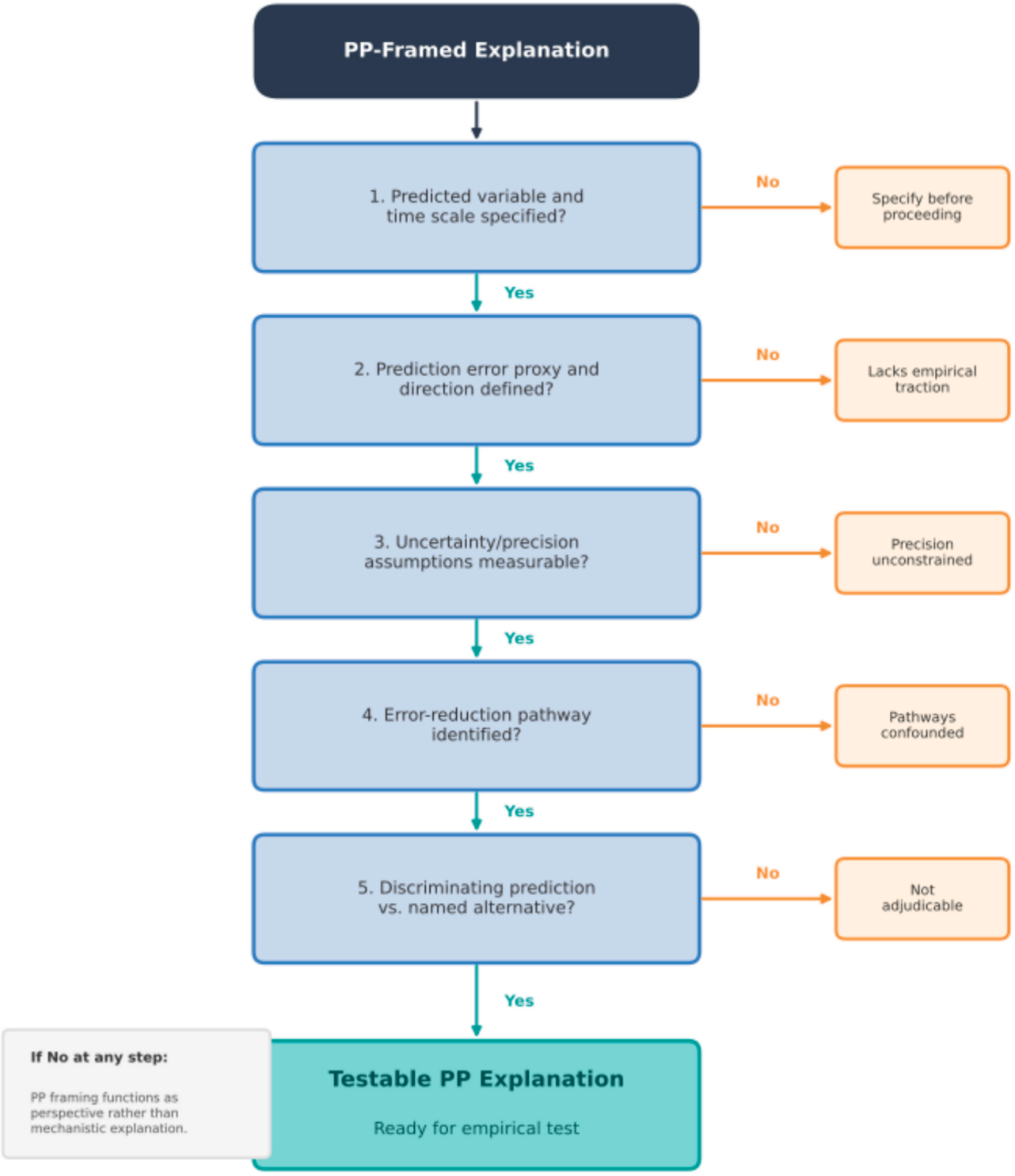

Fig. 1. PP-MC decision sequence. A PP-framed explanation becomes testable when it specifies five minimal commitments. If any commitment is left unspecified, PP often functions mainly as an interpretive perspective.

another or make the proposed PP account lose against a comparison account.

4. PP-MC in practice: worked examples from recent studies

Recent work illustrates that PP-MC items can be operationalized without committing to a single formalization. For Item 2 (prediction error proxy and direction) and Item 5 (discriminating prediction), Tabas and von Kriegstein (2024) designed two fMRI experiments that induced two orthogonal sources of expectations in audition: stimulus statistics (via repetition) and task rules (via instructions). They tested three competing scenarios, namely that prediction error responses in the auditory pathway reflect stats-informed expectations, task-informed expectations, or a combination of both. Across auditory midbrain, thalamus, and cortex responses were consistent with the combination scenario. This is an example of a PP-framed claim that becomes testable

because it specifies the prediction target (the next sound in a sequence), the error proxy (responses to deviants relative to standards), the relevant uncertainty source (contradictory predictions), the pathway (within-trial inference), and a contrast that distinguishes alternatives.

For Items 3 and 4 (uncertainty or precision, and pathway), learning work shows how different forms of uncertainty constrain what should count as prediction error versus learning signals. For example, the hierarchical Gaussian filter formalizes learning under volatility and distinguishes updating under expected noise from updating under environmental change (Mathys et al., 2014). PP-MC does not require adopting a specific model. It asks authors to specify which uncertainty construct the design targets and whether the intended pathway is perceptual updating, learning, action, or sampling.

The interoception example (Table 1) shows how PP-consistent pathways that differ mainly in their precision assumptions yield different discriminating tests. A paper that frames anxiety as "hyper-precise priors" should predict stronger effects of manipulations that target prior beliefs (e.g., threat-expectation inductions). A paper that frames anxiety as "low sensory precision" should predict stronger effects of manipulations that change signal quality. Both should state what patterns would count against the proposed pathway and which alternative account is being ruled out.

5. Practical implementation

One implementation of PP-MC is a short "PP commitments" box in the Methods or Theory section of a manuscript. This can be written as five sentences:

1) Predictand and timescale: what is predicted, and over what horizon?
2) Error proxy: how prediction error will be indexed, and what direction is expected?
3) Uncertainty: which uncertainty construct is relevant and how it is manipulated or measured?
4) Pathway: whether the explanation relies on updating, learning, action, or sampling?
5) Discriminating prediction: what alternative account is targeted and what pattern would favour one account over the other?

These commitments are not a substitute for formal modelling or quantitative model comparison. Formal models are the stronger route when the goal is to make a PP hypothesis precise: they can specify the generative model, priors, precision terms, policy objective, and parameter settings that produce a predicted pattern. PP-MC instead helps clarify the intended mapping from PP terms to observables and requires at least one contrast that could, in principle, disconfirm the proposed mechanism. For reviewers, PP-MC can function as an evaluation prompt: if a paper uses PP terminology but leaves commitments unspecified, the PP framing contributes interpretive context rather than a mechanistic explanation.

A related modelling risk is that a highly flexible PP model can be refitted to each new phenomenon. Such fits may show that PP can describe a result, but they do not by themselves test a single hypothesis unless the relation among fittings is constrained. Stronger evidence would come from models whose core architecture and parameter settings carry across tasks or datasets, and from quantitative comparisons showing that the PP model outperforms plausible non-PP or alternative-PP models (Bowman et al., 2023).

6. Limitations and scope

PP-MC targets PP-framed explanations that are presented verbally without a fully specified generative model and formal model comparison. When a study specifies competing computational models and compares them quantitatively, PP-MC should be mostly redundant: the same commitments should be visible in the model structure, parameterization, and comparison set. PP-MC is also not intended to enforce a single formalization of PP. It functions as a minimal set of commitments that reduces avoidable flexibility and supports discriminating tests of specific PP mechanisms. Meeting PP-MC does not guarantee that a study provides a strong test. It increases constraint and clarifies which additional design features would be needed to adjudicate between accounts.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the author used ChatGPT (OpenAI) to proofread the manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Funding

No funding was received for this work.

Declaration of Competing Interest

The author declares no conflict of interest.

Acknowledgment

I thank Fabiola Diana for thoughtful feedback on an earlier version of this manuscript.

References

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M.J., Van IJzendoorn, M. H., 2007. Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. Psychol. Bull. 133 (1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1.

Bowman, H., Collins, D.J., Nayak, A.K., Cruse, D., 2023. Is predictive coding falsifiable? Neurosci. Biobehav. Rev. 154, 105404. https://doi.org/10.1016/j.neubiorev.2023.105404.

Cao, R., 2020. New labels for old ideas: Predictive processing and the interpretation of neural signals. Rev. Philos. Psychol. 11 (3), 517–546. https://doi.org/10.1007/s13164-020-00481-x.

Champion, T., Bowman, H., Marković, D., Grześ, M., 2026. Reframing the expected free energy: Four formulations and a unification. Neural Comput. 38 (3), 439–469. https://doi.org/10.1162/NECO.a.1491.

Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36 (3), 181–204. https://doi.org/10.1017/S0140525X12000477.

Friston, K., Da Costa, L., Sakthivadivel, D.A.R., Heins, C., Pavliotis, G.A., Ramstead, M., Parr, T., 2023. Path integrals, particular kinds, and strange things. Phys. Life Rev. 47, 35–62. https://doi.org/10.1016/j.plrev.2023.08.016.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., Pezzulo, G., 2017. Active inference: a process theory. Neural Comput. 29 (1), 1–49. https://doi.org/10.1162/NECO_a_00912.

Garfinkel, S.N., Seth, A.K., Barrett, A.B., Suzuki, K., Critchley, H.D., 2015. Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. Biol. Psychol. 104, 65–74. https://doi.org/10.1016/j.biopsycho.2014.11.004.

Hodson, R., Mehta, M., Smith, R., 2024. The empirical status of predictive coding and active inference. Neurosci. Biobehav. Rev. 157, 105473. https://doi.org/10.1016/j.neubiorev.2023.105473.

Hohwy, J., 2013. *The predictive mind* (First edition). Oxford University Press.

Hohwy, J., 2020. New directions in predictive processing. Mind Lang. 35 (2), 209–223. https://doi.org/10.1111/mila.12281.

Mathys, C.D., Lomakina, E.I., Daunizeau, J., Iglesias, S., Brodersen, K.H., Friston, K.J., Stephan, K.E., 2014. Uncertainty in perception and the Hierarchical Gaussian Filter. Front. Human. Neurosci. 8 (825). https://doi.org/10.3389/fnhum.2014.00825.

Milkowski, M., Litwin, P., 2022. Testable or bust: theoretical lessons for predictive processing. Synthese 200 (6), 462. https://doi.org/10.1007/s11229-022-03891-9.

Millidge, B., Tschantz, A., Buckley, C.L., 2021. Whence the expected free energy? Neural Comput. 33 (2), 447–482. https://doi.org/10.1162/neco_a_01354.

den Ouden, C., Kashyap, M., Kikkawa, M., Feuerriegel, D., 2025. Limited evidence for probabilistic cueing effects on grating-evoked event-related potentials and orientation decoding performance. Psychophysiology 62 (5), e70076. https://doi.org/10.1111/psyp.70076.

Paulus, M.P., Feinstein, J.S., Khalsa, S.S., 2019. An active inference approach to interoceptive psychopathology. Annu. Rev. Clin. Psychol. 15 (1), 97–122. https://doi.org/10.1146/annurev-clinpsy-050718-095617.

Seth, A.K., 2013. Interoceptive inference, emotion, and the embodied self. Trends Cogn. Sci. 17 (11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007.
Seth, A.K., Friston, K.J., 2016. Active interoceptive inference and the emotional brain. Philos. Trans. R. Soc. B Biol. Sci. 371 (1708), 20160007. https://doi.org/10.1098/rstb.2016.0007.
Sun, Z., Firestone, C., 2020. The dark room problem. Trends Cogn. Sci. 24 (5), 346–348. https://doi.org/10.1016/j.tics.2020.02.006.
Tabas, A., von Kriegstein, K., 2024. Multiple concurrent predictions inform prediction error in the human auditory pathway. J. Neurosci. 44 (1), e2219222023. https://doi.org/10.1523/JNEUROSCI.2219-22.2023.
Zamariola, G., Maurage, P., Luminet, O., Corneille, O., 2018. Interoceptive accuracy scores from the heartbeat counting task are problematic: evidence from simple bivariate correlations. Biol. Psychol. 137, 12–17. https://doi.org/10.1016/j.biopsycho.2018.06.006.

Glossary

Active inference: A family of models in which perception, action, learning, and policy selection are treated as inference under a generative model, with action guided by expected free energy over possible policies.

Expected free energy: A prospective objective used in active inference to evaluate possible policies, often combining pragmatic value (preferred outcomes) and epistemic value (information gain).

Generative model: A structured set of assumptions about how latent causes produce observable sensory signals. In PP, it supports prediction and belief updating.

Predictand: The target of prediction in a PP-framed claim (e.g., a sensory feature, an outcome, an action, or a latent state).

Prediction error: A mismatch between predicted and observed signals (or between predicted and inferred causes, depending on the formalization) that can drive updating.

Precision: A weighting on prediction error that determines its influence on updating. Often formalized as inverse variance and linked to uncertainty (e.g., noise, ambiguity, volatility).

Variational free energy: A quantity used in variational inference that bounds surprise for current observations. In active inference, variational free energy is central to perceptual inference, while expected free energy is a prospective objective for policy selection; the relation between them depends on the formal formulation.