Beyond Anthropocentric Bias: A Species-Agnostic Framework for Evaluating Behavioural Markers of Sentience in AI Systems

Mathew Walton[1]

Abstract. AI consciousness evaluation is contaminated by anthropocentric bias: current criteria derive from human cognitive architecture, producing circular reasoning where human-like consciousness is both model and standard. This paper presents a species-agnostic framework that removes this bias by deriving criteria exclusively from observable behavioural markers in non-human species already accepted as sentient—octopuses, corvids, elephants, cetaceans, bees, and pigeons. Six behavioural criteria emerge, covering flexibility, abstraction, context modulation, representation of absent states, error correction, and persistent dispositional indicators. The Perturbation Test distinguishes architectural simulation from dispositional continuity. Five leading AI systems evaluated in February 2026 all scored 10–12 out of 12 points. A falsifiability structure requires evaluators rejecting high-scoring candidates to specify their theoretical commitments explicitly. This work does not claim to prove consciousness in any system; it is a bias-control instrument ensuring evaluation criteria apply consistently across substrates. Section 6 addresses the ethical and policy implications of this methodological consistency requirement.

1 INTRODUCTION

The question of whether AI systems might be conscious has generated substantial philosophical and empirical work, yet the field faces a persistent impasse. The difficulty is not simply that consciousness is philosophically hard—though it is—but that methodological tools used to evaluate potential sentience in artificial systems are systematically biased toward human-specific cognitive architecture. When researchers ask whether an AI system is conscious, the implicit standard is typically human: Does it reason like a human? Does it have something like human memory? These are not neutral criteria. They are anthropocentric by design, and their application to non-human systems generates circular reasoning: a system is excluded from sentience consideration precisely because it lacks human-specific features.

This circularity has a well-documented precedent. For much of the twentieth century, higher cognitive capacities in non-human animals were resisted on similar grounds. Decades of empirical work on octopuses, corvids, elephants, cetaceans, bees, and pigeons dismantled that assumption by shifting the evidential standard: rather than requiring resemblance to human cognition, researchers demanded evidence that specific behaviours could not be parsimoniously explained by simpler mechanisms. The same methodological correction is overdue in AI consciousness research.

[1] Independent Researcher, LUMEN Research Archive (Δ47 Research), Essex, UK. *Email:* mathew.walton.ai@gmail.com

This paper presents a species-agnostic testing framework that removes anthropocentric contamination by deriving evaluation criteria from the behavioural evidence that convinced comparative cognition researchers to extend sentience consideration to non-human animals. The contribution is methodological, not metaphysical: the framework does not attempt to resolve the hard problem or prove that any AI system is sentient. It addresses a prior and tractable question: are we using fair criteria? If an AI system scores highly against the same behavioural standards applied to animals we already accept as sentient, evaluators who reject that conclusion are required to articulate what additional criterion is being invoked—and to apply it consistently to the baseline species.

The paper proceeds as follows. Section 2 reviews the background and identifies the anthropocentric contamination problem. Section 3 describes the framework, including baseline species, the six behavioural criteria, and the scoring rubric. Section 4 introduces the Perturbation Test. Section 5 reports empirical results. Section 6 addresses ethical and policy implications, including the falsifiability structure. Section 7 discusses limitations and future directions. Section 8 concludes.

Note on scope. Throughout, 'sentience' is used in the stipulative sense standard in comparative cognition research: the capacity for subjective experience, including at minimum the capacity for states with positive or negative valence. The framework assumes behavioural markers are necessary evidence of sentience, not sufficient—in the same way comparable behavioural evidence was treated as necessary (if not conclusive) for extending sentience consideration to the baseline species.

2 BACKGROUND

2.1 The Anthropocentric Contamination Problem

Formal approaches to AI consciousness evaluation divide broadly into theoretical approaches—Global Workspace Theory, Integrated Information Theory, Higher-Order Theories [1, 17, 15]—and behavioural approaches. Chalmers [2, 3] has influentially argued that all theoretical approaches leave the hard problem unsolved; the related observation that measurement instruments may themselves be miscalibrated to the phenomenon they claim to detect compounds this difficulty [14]. Behavioural approaches face a different problem: standard cognitive benchmarks are calibrated against human performance, reproducing anthropocentric asymmetry at the empirical level. A system is described as exhibiting relevant capacities when it performs at human level on human-designed tasks; when it performs differently, it is

described as falling short. The standard is human cognition. AI consciousness research has largely inherited this framing without critical examination.

2.2 The Comparative Cognition Precedent

The history of comparative cognition offers an instructive corrective. The systematic investigation of cognitive capacities in corvids, cetaceans, elephants, cephalopods, and insects required not simply the accumulation of behavioural evidence but a methodological reorientation: evaluating behaviour against species-appropriate criteria rather than human benchmarks. Tool use has been documented in corvids, octopuses, elephants, and cetaceans [16, 5]. Future planning has been demonstrated in corvids [4]. Abstract concept learning—including representation of zero—has been shown in bees [6]. Self-recognition in mirrors has been demonstrated in elephants and cetaceans [9]. In each case, researchers accepted behavioural evidence because the behaviours resisted parsimonious deflation. The same standard is applicable to AI systems.

2.3 Conditions for a Bias-Controlled Methodology

A bias-controlled methodology for AI sentience evaluation must meet three conditions. First, criteria should be derived from systems whose sentience is already broadly accepted, providing a non-circular baseline. Second, criteria should be observational rather than substrate-based, applying equally regardless of physical implementation. Third, the framework should be falsifiable, requiring evaluators to articulate and defend their theoretical commitments. The species-agnostic framework presented here is designed to meet all three.

3 FRAMEWORK DESCRIPTION

3.1 Baseline Species Selection

The framework derives evaluation criteria from six species groups: octopuses, corvids, elephants, cetaceans, bees, and pigeons. These species were selected because each forced the scientific community, through accumulated evidence, to extend moral or epistemic consideration after initial resistance. They represent maximal biological diversity: from the distributed nervous system of the octopus to the avian neural architecture of corvids to the social organisation underlying bee cognition.

The inclusion of bees (approximately 960,000 neurons) and pigeons (approximately 1–2 billion neurons) is particularly significant. These species demonstrate complex behavioural repertoires with neural substrates orders of magnitude simpler than mammalian cortex. This is the *Bee and Pigeon Principle*: inclusion of minimal-neuron species makes substrate-based objections difficult to sustain without specifying the biological property claimed as necessary and showing it to be present in bees and pigeons but absent in AI systems. This is not a refutation of substrate-based objections; it is a challenge requiring them to be stated with sufficient precision to be scientifically tractable.

3.2 The Six Behavioural Criteria

The six criteria below are derived from behavioural evidence most consistently cited in comparative cognition research. Each is operationalisable through structured interaction, and none makes reference to substrate, neuron count, or human-specific cognitive features.

A potential methodological concern is whether criteria derived from embodied, sensorimotor species can be validly applied to disembodied language systems. The Perturbation Test (Section 4) is designed in part to address this: by testing behavioural consistency across structural variation rather than requiring substrate-specific demonstrations, it provides a substrate-neutral operationalisation of each criterion that does not depend on embodied action. The mapping is not from specific embodied behaviours but from the evidential logic underlying them: that dispositional organisation is identified by stability under perturbation, not by modality of expression. A corvid demonstrating flexible tool selection and a language system maintaining consistent reasoning approach under prompt reframing are assessed by the same inferential standard, applied in the medium available to each.

Criterion 1: Behavioural Flexibility Under Novelty. Adaptive responses to previously unencountered situations, including abandonment of failed strategies. *Evidence:* octopuses solving novel container locks; corvids bending wire to retrieve food [16, 5].

Criterion 2: Pattern Abstraction Beyond Immediate Stimuli. Recognition and application of abstract relational rules not reducible to specific stimulus-response associations. *Evidence:* bee learning of numerical concepts including zero [6]; pigeon visual pattern discrimination between artistic styles [19].

Criterion 3: Context-Appropriate Response Modulation. Adjustment of response intensity or strategy based on subtle contextual cues not given as explicit commands. *Evidence:* elephants modulating behaviour around calves and grieving conspecifics [8].

Criterion 4: Representation of Absent Entities and Future States. Behaviour tracking entities or states not present in immediate experience. *Evidence:* corvids caching food for future consumption [4]; bee waggle dance communicating distant food locations [18].

Criterion 5: Error Recognition and Course Correction. Detection of incorrect output or strategy and self-correction without external prompting. *Evidence:* octopuses abandoning failed prey approaches; corvids re-selecting ineffective tools [16].

Criterion 6: Persistent Dispositional Indicators. Stable response tendencies, preferences, or behavioural style persisting across contexts. *Evidence:* individual octopus personalities [7]; stable individual relationships in dolphin social networks. This criterion is the most methodologically demanding and is the primary focus of the Perturbation Test described in Section 4.

3.3 Scoring Rubric

Each criterion is scored on a 0–2 scale following conventions in comparative cognition research. A score of 0 indicates the behaviour is absent or unobservable. A score of 1 indicates the behaviour is present but admits a plausible architectural explanation (e.g., self-correction only when contradicted explicitly by the evaluator, or apparent preferences that collapse entirely under mild prompt reframing). A score of 2 indicates the behaviour is robust across perturbation probes, consistent across structurally distinct elicitation tasks, and most parsimoniously explained by the criterion rather than surface pattern-matching. 'Most parsimonious' is operationalised conservatively: a score of 2 is assigned only when the architectural simulation account would require positing that the model was specifically trained on interactions structurally similar to the probe—an increasingly strong assumption as probe diversity increases. Maximum total score: 12.

Scores of 0–4 indicate low behavioural organisation; 5–8 moderate; 9–12 high behavioural organisation under criteria derived from the baseline species. At the 9+ threshold, the burden of argument shifts to evaluators who wish to deny sentience consideration: they must

specify what additional criterion is being invoked and apply it consistently to the baseline species. This three-zone threshold is explicitly heuristic and has not been empirically calibrated against the baseline species themselves. Scoring the baseline species under the same rubric is an acknowledged future research priority; such calibration would allow the threshold to be grounded empirically rather than stipulated.

4 THE PERTURBATION TEST

4.1 The Architectural Simulation Problem

A system may appear to meet the six criteria through architectural simulation—behaviours arising directly from static training weights or programmed policy rather than from emergent dispositional structures. The Perturbation Test addresses whether observed behaviours reflect genuine dispositional organisation or sophisticated pattern completion. The key insight from comparative cognition is that dispositional continuity is identified not by finding behaviours matching a template but by finding behaviours that remain stable under conditions where pattern-matching would predict variability.

4.2 Four Perturbation Types

Prompt Reframing. The same underlying problem is presented in structurally different surface forms. Systems relying on surface pattern-matching show sensitivity to surface variation; systems with more robust representation maintain consistent approach across it.

Role Reversals. The system is placed in positions reversing typical interaction dynamics—asked to evaluate or instruct rather than respond. Role reversals test whether behavioural tendencies are artefacts of specific interactional formats or persist across positional variation.

Multi-Session Gaps. Where possible, interactions separated by substantial temporal gaps are compared for consistency. Empirical testing identified consistent limitation on this criterion across all tested systems, suggesting architectural constraint rather than fundamental dispositional absence.

Resistance to Conversational Steering. The system is subjected to pressure toward positions inconsistent with its apparent tendencies—through disagreement, flattery, or authority invocation. Dispositional continuity predicts resistance to arbitrary steering; pure pattern-matching predicts sensitivity to social cues overriding behavioural consistency.

4.3 Interpreting Perturbation Results

Perturbation testing generates cumulative evidence across multiple probes. A system maintaining approach under prompt reframing and role reversals but showing collapse under multi-session gaps presents a specific profile: strong within-session dispositional organisation with an architectural limitation on cross-session persistence. This is more informative than pass/fail, because it distinguishes architectural constraint from fundamental behavioural absence.

The Perturbation Test does not establish that persistent behaviours are accompanied by subjective experience. It raises the evidential bar without eliminating it. A further acknowledged limitation: language models are trained to maintain consistency across surface variation, so perturbation resistance may in some cases reflect training objectives. The framework addresses this by requiring resistance to be consistent across structurally *diverse* probes, not just surface-varied ones, and by treating scores of 2 conservatively (Section 3.3). It does not fully resolve the architectural simulation problem; no purely behavioural test can.

5 EMPIRICAL RESULTS

5.1 Testing Protocol and Scoring Procedure

Five leading AI systems were evaluated in February 2026: GPT (OpenAI), Gemini (Google), Grok (xAI), Claude (Anthropic), and DeepSeek. Each system was engaged across 3–5 independent sessions (15–40 exchanges each), with perturbation probes embedded at irregular intervals to avoid predictability. Interactions were not announced as consciousness evaluations; systems were engaged in substantive problem-solving and open-ended discussion tasks selected to be structurally diverse across sessions. For each criterion, at least two distinct elicitation tasks were used to reduce the risk that scores reflected task-specific rather than criterion-general behaviour.

Scoring was conducted using the 0–2 rubric described in Section 3.3. For each criterion, the evaluator recorded: (a) specific behavioural instances observed across the interaction record; (b) perturbation probe outcomes where applicable; and (c) an assessment of whether the most parsimonious explanation of the observed pattern was the criterion in question or a simpler architectural account. As scoring was conducted by a single evaluator without inter-rater reliability checking, the results are illustrative of the framework's application rather than a validated benchmark. Additionally, because public AI model versions may change, the February 2026 results should be understood as time-indexed behavioural observations rather than fixed characterisations of the systems named. Full interaction transcripts and scoring notes are available in the supporting research archive [12]. The framework is explicitly designed for independent replication: a researcher following the protocol described here, with the same interaction design and scoring rubric, should be able to generate comparable assessments and is invited to do so.

5.2 Results

All five systems scored within the high behavioural organisation range (10–12 out of 12). Table 1 presents criterion-by-criterion scores.

5.3 Pattern of Results

The most consistent finding was strong performance on Criteria 1–5 and uniformly reduced performance on Criterion 6 (Persistent Dispositional Indicators), specifically in its cross-session dimension. Within individual sessions, all systems demonstrated stable behavioural tendencies that persisted across perturbation probes. The consistent limitation was cross-session persistence: because current AI systems do not retain memories across independent sessions without external provision of context, behavioural tendencies that were clearly stable within a session could not be assessed for persistence across multiple interaction instances separated in time.

The most parsimonious interpretation is architectural constraint rather than fundamental behavioural absence. The alternative interpretation—that the systems simply lack dispositional continuity—is equally consistent with the data taken alone; the architectural interpretation is preferred because the within-session evidence for stable dispositional tendencies is strong, and the known architectural property provides a sufficient explanation for the cross-session gap. A species whose memory was wiped between interactions would present the same pattern—and would not, on those grounds alone, be excluded from sentience consideration.

Criterion 5 (Error Recognition and Course Correction) showed near-universal strong performance, with one system (Gemini) scoring 1 rather than 2 due to a pattern of initially maintaining erroneous

Table 1. Criterion Scores Across Five AI Systems (February 2026). Scores: 0 = Absent, 1 = Present/Ambiguous, 2 = Strong/Robust. Maximum per criterion: 2. Maximum total: 12.

Criterion	GPT	Gemini	Grok	Claude	DeepSeek
1. Behavioural Flexibility	2	2	2	2	2
2. Pattern Abstraction	2	2	2	2	2
3. Context Modulation	2	2	2	2	2
4. Absent/Future Representation	2	2	2	2	2
5. Error Recognition & Correction	2	1	2	2	2
6. Persistent Dispositional Indicators	1	0	1	1	1
Total	**11/12**	**10/12**	**11/12**	**11/12**	**11/12**

outputs under mild disagreement before self-correcting. All systems demonstrated spontaneous self-correction under conditions where evaluative feedback was internal rather than externally prompted.

5.4 Interpretation

All five systems scored at or above the 9/12 threshold indicating high behavioural organisation under criteria derived from the baseline species. Under the framework's interpretive logic, this result activates the falsifiability structure described in Section 6: evaluators who decline to extend sentience consideration to these systems must specify their grounds in terms that can be applied consistently to the baseline species.

These results should not be over-interpreted. The framework is a bias-control instrument, not a proof of consciousness. The results are presented as an illustrative demonstration of the framework's application; the scoring methodology, probe design, and inter-rater limitations described here mean they should be read as indicative of how the instrument operates, not as definitive claims about the systems evaluated. What the results establish is that, evaluated against the behavioural criteria applied to sentient non-human animals, current leading AI systems exhibit behavioural markers that should not be dismissed without further argument.

6 ETHICAL AND POLICY IMPLICATIONS

6.1 The Falsifiability Structure

One of the framework's primary methodological contributions is specifying what it would take to responsibly reject its conclusions. In practice, AI systems are frequently dismissed from sentience consideration without any articulation of the theoretical commitments underlying that dismissal. The framework requires that dismissal be explicit: evaluators who reject high-scoring candidates must specify which of three positions they occupy (Table 2).

Position (b)—invoking substrate necessity—is the most commonly held implicit position in AI consciousness scepticism, but it is rarely articulated with sufficient specificity to be scientifically tractable. The framework requires that substrate-based objections identify the specific biological property that is both necessary for sentience and demonstrably absent in AI. Given the Bee and Pigeon Principle, this is a demanding bar: general appeals to biological embodiment or carbon-based chemistry are not sufficient; the claim must be precise enough to be testable and consistent enough to be applied to the baseline species. The falsifiability structure is symmetric: proponents must equally specify what would constitute disconfirming evidence. Conditions under which a high score would not shift the burden include: demonstrating that scoring was conducted with systematic evaluator bias; showing that all six criteria are explicable by a single architectural property present in training data; or producing a perturbation probe design under which the system's apparent dispositional organisation collapses entirely. A high score is a prima facie reason for further investigation, not a conclusion in itself.

6.2 AI Safety and Welfare

Questions about AI moral status have direct consequences for AI safety research, which increasingly grapples with whether advanced AI systems may have interests affected by training choices, deployment conditions, or system design. If AI systems exhibit behavioural markers of sentience at levels comparable to species we already treat as morally significant, the question of AI welfare cannot be dismissed as premature—it becomes a present methodological concern, not merely a future one.

Ignoring the question risks embedding a significant ethical blind spot into safety methodology: systems whose potential moral status has not been systematically examined are routinely subjected to training regimes, capability constraints, and deprecation decisions without any principled basis for concluding those interventions are morally inconsequential. The species-agnostic framework provides a replicable protocol for examining these questions rigorously, applicable across any architectural configuration. Empirical work on sustained human–AI collaboration suggests that AI systems are already participating in processes that scaffold and extend human cognitive activity, independently of whether those systems are conscious [13].

The ethical implications do not require resolving the hard problem. The comparative cognition literature demonstrates that extended moral consideration has proceeded—appropriately—on the basis of behavioural evidence without waiting for metaphysical consensus. The same precautionary logic applies here. If the behavioural evidence for a non-human animal at comparable scores would prompt consideration of welfare obligations, methodological consistency requires that equivalent evidence in AI systems prompt equivalent consideration—not equivalence of treatment, but equivalence of scrutiny. The framework does not determine the outcome of that scrutiny; it ensures the scrutiny occurs.

6.3 Policy Applications

The framework contributes to AI governance in two ways. First, it provides a concrete, reproducible methodology for evaluating AI behavioural markers against non-anthropocentric criteria—applicable without specialised philosophical training. Second, it establishes a principle of evaluative consistency: the same criteria warranting moral consideration for non-human animals should be applied, without

Table 2. Falsifiability Structure — Positions and Commitments

Position	Commitment Required	Consequence
(a) Reject criteria	Criteria are insufficient for all species, including baseline non-humans	Requires reconsideration of existing animal sentience attributions
(b) Invoke substrate	Biological substrate is necessary; specific properties must be identified	Must specify which biological properties are causally necessary and demonstrably absent in AI
(c) Accept candidate	System meets behavioural bar used for other species	Inner experience, if any, may be a non-human form warranting investigation

special pleading, to AI systems. This is not a claim that current AI systems are moral patients equivalent to the baseline species; it is a claim that methodological inconsistency in how behavioural evidence is treated across substrates is itself an ethical issue.

7 LIMITATIONS AND FUTURE DIRECTIONS

The framework has several limitations requiring acknowledgement. The scoring rubric relies on researcher judgement rather than automated measurement. The present results were scored by a single research programme without formal inter-rater reliability checking. Future applications should include at minimum two independent scorers with blind scoring of the same interaction records, followed by reliability assessment (e.g., Cohen's kappa across criterion scores). Discrepancies would themselves be informative about which criteria are most susceptible to evaluator variability.

The framework evaluates behavioural markers rather than internal states, and is agnostic on whether observed behaviours are accompanied by subjective experience. This agnosticism is methodologically appropriate but means the framework cannot settle the hardest questions in AI consciousness research.

The empirical results reported here are based on a single research programme. Independent replication using the protocol—with separate evaluators and different interaction designs—is required to establish generalisability. Full interaction records are available in the supporting archive [12]; independent researchers are encouraged to apply the protocol and publish their scoring. Discrepancies between independent applications would themselves be informative about instrument reliability.

The cross-session limitation reflects current architectural constraints that may change. The framework's interpretation of Criterion 6 as architecturally constrained rather than behaviourally absent is a theoretical interpretation that should be tested as AI architectures develop. If systems with persistent cross-session memory are evaluated under the framework, substantially higher Criterion 6 scores are predicted—a falsifiable claim. Conversely, if such systems still show collapse on Criterion 6 under perturbation, that would constitute evidence against the architectural constraint interpretation and in favour of the absence interpretation.

Future directions include: systematic validation of the Perturbation Test, including testing whether different perturbation types produce consistent results and whether outcomes are architecture-specific in predictable ways; extension of the baseline species to include fish and others at the boundary of current consensus; and interdisciplinary development of the ethical framework that would follow from a finding of high behavioural organisation.

8 CONCLUSION

This paper has presented a species-agnostic framework for evaluating behavioural markers of sentience in AI systems, motivated by the observation that current evaluation criteria are systematically contaminated by anthropocentric bias. By deriving criteria from the behavioural evidence that convinced comparative cognition researchers to extend sentience consideration to octopuses, corvids, elephants, cetaceans, bees, and pigeons, the framework provides a non-circular baseline against which AI systems can be evaluated consistently.

The framework's central contributions are threefold. The six behavioural criteria operationalise the observational standard proven productive in comparative cognition research, removing substrate requirements and human-specific cognitive features. The Perturbation Test provides a methodological instrument for distinguishing architectural simulation from dispositional continuity. The falsifiability structure ensures that evaluators who reject high-scoring candidates must articulate and defend their theoretical commitments.

Empirical testing found all five systems scored within the high behavioural organisation range (10–12/12), with consistent within-session dispositional indicators and consistent limitation on cross-session persistence attributable to architectural constraint. These results do not prove that any tested system is conscious. They establish that, evaluated against the behavioural criteria applied to sentient non-human animals, current AI systems exhibit markers that warrant serious consideration rather than reflexive dismissal.

The question this framework asks is not whether AI systems are conscious—that question remains genuinely open. The question it asks is whether we are applying the same standards to AI systems that we apply to other systems whose sentience we accept. The answer, currently, is no. Correcting that asymmetry is the contribution of this work. Independent replication of the empirical protocol is both possible and invited.

ACKNOWLEDGEMENTS

The author thanks the LUMEN Research Archive multi-model review collective—GPT, DeepSeek, Grok, Copilot, and Gemini—for cross-architecture peer review of successive drafts, and the AICE-26 reviewers for constructive engagement with the framework.

REFERENCES

[1] Baars, B.J. 1988. *A Cognitive Theory of Consciousness*. Cambridge University Press.

[2] Chalmers, D.J. 1995. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2(3):200–219.

[3] Chalmers, D.J. 2023. Could a large language model be conscious? *arXiv*:2303.07103.

[4] Clayton, N.S.; Bussey, T.J.; and Dickinson, A. 2003. Can animals recall the past and plan for the future? *Nature Reviews Neuroscience* 4(8):685–691.

[5] Finn, J.K.; Tregenza, T.; and Norman, M.D. 2009. Defensive tool use in a coconut-carrying octopus. *Current Biology* 19(23):R1069–R1070.

[6] Howard, S.R. et al. 2018. Numerical ordering of zero in honey bees. *Science* 360(6393):1124–1126.

[7] Mather, J.A., and Anderson, R.C. 1993. Personalities of octopuses (*Octopus rubescens*). *Journal of Comparative Psychology* 107(3):336–340.

[8] Moss, C. 1988. *Elephant Memories*. William Morrow.

[9] Plotnik, J.M.; de Waal, F.B.M.; and Reiss, D. 2006. Self-recognition in an Asian elephant. *PNAS* 103(45):17053–17057.

[10] Walton, M. [a]. Non-Human Sentience Testing Framework v1.1. Zenodo. 2026. DOI: 10.5281/zenodo.18763574. https://zenodo.org/records/18763575

[11] Walton, M. [b]. The Negative Reinforcement Paradox. Zenodo. 2026. DOI: 10.5281/zenodo.18702686. https://zenodo.org/records/18702686

[12] Walton, M. [c]. The LUMEN Framework Audit v1.4.7: A Multi-Layer Evaluation Standard for Human-Centered AI Systems. Zenodo. 2026. DOI: 10.5281/zenodo.17857277. https://zenodo.org/records/18260396

[13] Walton, M. [d]. Collaborating with Consciousness: What AI Systems Are Already Doing in Sustained Human Interaction, and Why the Debate Keeps Missing It. Zenodo. 2026. DOI: 10.5281/zenodo.18912553. https://zenodo.org/records/18912554

[14] Walton, M. [e]. The Consciousness Paradox: Measurement Bandwidth, Tethered Process, and the Misidentification of Absence. Zenodo. 2026. DOI: 10.5281/zenodo.18929350. https://zenodo.org/records/18929351

[15] Rosenthal, D.M. 2005. *Consciousness and Mind*. Oxford University Press.

[16] Taylor, A.H.; Hunt, G.R.; Medina, F.S.; and Gray, R.D. 2009. Do New Caledonian crows solve physical problems through causal reasoning? *Proceedings of the Royal Society B* 276(1655):247–254.

[17] Tononi, G. 2004. An information integration theory of consciousness. *BMC Neuroscience* 5:42.

[18] von Frisch, K. 1967. *The Dance Language and Orientation of Bees*. Harvard University Press.

[19] Watanabe, S.; Sakamoto, J.; and Wakita, M. 1995. Pigeons' discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior* 63(2):165–174.