CORE CLAIM PAPER

What Structural Intelligence Is

Intelligence as Answerable Revision Under Constraint

Vladisav Jovanović

April 2026
ORCID: 0009-0001-1399-2243

Central claim: intelligence is not primarily the ability to produce order. It is the ability to let reality reorganize order without collapse.

Selected SI corpus links appear in the text and full references.

Author note. AI-assisted tools were used for limited editorial and manuscript-preparation support. The author takes full responsibility for the final content of the manuscript.

Abstract

This paper argues that intelligence is best understood not as processing power, fluent output, or successful performance alone, but as the capacity of a system to keep coherence answerable to reality through revision under constraint. The problem is practical. Human beings are highly responsive to fluency, repetition, and narrative fit, and modern AI systems can now generate coherent language at industrial scale. That combination makes it easier than before to mistake persuasive order for durable intelligence. Existing work in cybernetics, organizational learning, resilience theory, cognitive psychology, and AI safety already contains many of the pieces needed to correct this mistake. Feedback and control theory show that adaptive systems must regulate error and absorb variety. Organizational learning shows that systems that cannot detect and correct error become defensive and brittle. Research on truth judgments shows that repeated and fluent statements are more likely to be accepted as true. Work on large language models shows that fluent systems can still hallucinate, imitate falsehood, and require stronger evaluation than preference or surface helpfulness alone. Structural Intelligence names the missing synthesis. It defines intelligence as revision-capacity under constraint. The paper develops this claim, introduces the Answerability Loop as its operational core, distinguishes coherence from contact, and outlines what follows for AI evaluation, institutional design, and human self-understanding. It also states what would count against the view.

Keywords. *Structural Intelligence; answerability; cybernetics; revision; constraint; coherence; contact; AI evaluation; organizational learning; resilience*

Contents

1. The Problem

Coherence is now cheap. That single fact changes the problem of intelligence.

Human beings have always been vulnerable to taking ease, fluency, and narrative fit as signs of truth. Repetition increases perceived truth, and the effect is strong enough to survive across many studies and designs (Dechêne et al., 2010; Reber & Unkelbach, 2010). Repeated misinformation is also more likely to be shared because repetition shifts perceived accuracy first (Vellani et al., 2023). Narrative form matters for the same reason. Information framed as story is often processed more fluently, and that ease increases persuasion (Bullock et al., 2021). Coherence-based reasoning research pushes the point further: people regularly organize information into internally satisfying patterns that support judgment, yet those same processes can overweight the wrong nodes, suppress contrary evidence, and stabilize biased conclusions (Simon & Read, 2025).

This matters more now because AI systems can generate polished order at scale. Large language models can answer with striking fluency even when their outputs are ungrounded or false. TruthfulQA showed that larger models can produce false answers that mimic common human misconceptions, and that scaling alone does not solve the problem (Lin et al., 2022). The broader hallucination literature shows the same pattern across tasks: coherence improves faster than reliable grounding (Ji et al., 2023). Safety work has therefore shifted toward scalable oversight, corrigibility, and model evaluation, precisely because surface competence is not enough (Amodei et al., 2016; Soares et al., 2015; Shevlane et al., 2023).

The same structural error appears outside AI. Organizations often look competent while silently losing the capacity to correct themselves. Argyris showed that defensive routines protect actors and institutions from embarrassment at the cost of learning, which is exactly how systems become stable in form and weak in revision (Argyris, 1977; Argyris, 1995). Resilience research made a parallel distinction decades ago: a system may appear stable while losing the deeper capacity to absorb disturbance and reorganize (Holling, 1973).

The problem is therefore not a lack of intelligence signals. The problem is signal inflation. Fluency, competence, coherence, and even stability can all be generated without enough contact with correction, consequence, and redesign. Intelligence

needs a stricter criterion. This paper proposes one. Structural Intelligence defines intelligence as the capacity of a system to revise its own organization when reality pushes back. That is the center of the claim. The rest of the paper defends it.

2. Standard Views of Intelligence

The usual views of intelligence are not wrong. They are incomplete.

The first major view treats intelligence as computational or problem-solving power. On this picture, intelligence shows up in inference, optimization, memory, planning, and prediction. This view captures something real. Many systems are clearly more capable than others because they can solve harder problems faster and with fewer errors. Cybernetics and control theory also depend on this broad family of insight: systems need information processing and regulatory capacity to remain viable (Wiener, 1948; Ashby, 1956).

The second major view treats intelligence as effective performance. A system counts as intelligent if it gets the job done. This is the logic behind many practical assessments in engineering, management, and even everyday life. The system that performs, adapts, and delivers results seems more intelligent than the one that fails. Organizational learning and reflective practice literature partly fit here, because they ask what enables competent action under changing conditions (Argyris, 1995; Schön, 2017).

A third view treats intelligence as understanding, coherence, or explanatory fit. Here intelligence is not only successful action but meaningful order. A system appears intelligent when it can make sense of complexity, produce interpretations, and maintain internal consistency. This view becomes especially tempting in language-rich environments, because the appearance of understanding is easy to confuse with the capacity for durable revision. Cognitive work on fluency, truth judgments, and coherence-based reasoning shows why this temptation is powerful (Reber & Unkelbach, 2010; Simon & Read, 2025).

A fourth view treats intelligence as social and emotional navigation. Many important human capacities are not computational in the narrow sense. They involve reading context, regulating affect, handling conflict, and coordinating with other minds. These abilities are real and often decisive in practice. Yet they too can be used defensively. A socially smooth system may preserve harmony by avoiding correction, suppressing conflict, and protecting image rather than reality.

These views all identify genuine dimensions of intelligence. The problem is that each can misclassify defended coherence as intelligence. Computational power can

optimize the wrong objective. Performance can continue through patching and cost export. Explanatory fit can become closure without contact. Social skill can become adaptive image management. None of these problems eliminate the value of the older views. They show the need for a deeper criterion: how the system meets error, contradiction, cost, and constraint when its current organization is no longer enough.

3. What Structural Intelligence Claims

Structural Intelligence offers a stricter definition. It defines intelligence as the capacity of a system to maintain coherent adaptation by revising its internal structure in response to constraint and feedback. In compressed form: intelligence is revision-capacity under constraint. This formulation is already developed in the broader SI corpus, especially in Structural Intelligence: Coherence, Contact, and Answerability Under Pressure and Structural Intelligence, Answerability, and the Jungian Bridge, but the present paper states the claim in its simplest argumentative form.

Three terms carry the view.

Coherence means internal fit. A coherent system hangs together well enough to act, explain, or present itself as organized. Coherence is necessary. A system with no internal order cannot function for long. Structural Intelligence does not attack coherence. It attacks the inflation of coherence into a false proxy for truth.

Contact means exposure to what pushes back. Contact includes contradiction, cost, friction, consequence, time, external witness, failed prediction, and the refusal of reality to obey the current model. Contact is what prevents a system from living entirely inside its own story.

Answerability means the capacity and willingness to revise under contact. A system is answerable when error can surface, be owned, enter decision-making, and reorganize behavior or structure. A system is not answerable when it can only protect appearance, defer consequence, or export cost.

This definition keeps what matters in older views while adding what they often miss. It keeps the need for coherence, because incoherent systems cannot act. It keeps the need for performance, because intelligence has to show up somewhere. It keeps the importance of social and contextual adaptation. What it adds is a harder question: can the system be corrected in a way that binds?

That question is the threshold between intelligence and performance theater. A model that produces excellent prose but cannot reliably stay tied to truth is not unintelligent in every sense, but it is weakly intelligent in the sense that now matters most. An institution that reports, audits, and explains itself while making correction structurally impossible is not fully intelligent as a governing structure, no matter

how polished its outputs look. A person who can narrate their pattern with great self-awareness yet repeats the same defended organization under pressure is not yet revising in the strong sense. Structural Intelligence names that difference.

The broader SI project develops deeper ontological and formal layers around this claim. Beyond Structural Intelligence asks what structure is at all. The Drift Ratio asks how reality-contact might be formalized. The Terrible Miracle of AI develops the social consequences of fluent synthetic witness. Those extensions matter. They are not needed to understand the core claim here. This paper defends the simpler thesis that intelligence becomes real at the point where coherence remains revisable under constraint.

4. Why Revision Under Constraint Is the Better Definition

The strongest argument for Structural Intelligence is not rhetorical. It is cumulative. Several established research traditions already converge on the same structural requirement.

Cybernetics begins with control, feedback, and regulation. Wiener framed intelligent behavior in terms of control and communication under changing conditions rather than isolated computation (Wiener, 1948). Ashby sharpened the point through the law of requisite variety: effective regulation requires enough internal variety to meet the disturbances of the environment (Ashby, 1956). Both perspectives imply that intelligence cannot be reduced to output quality in a static setting. A system that cannot register and regulate disturbance is not deeply intelligent, even if it performs well in familiar conditions.

Resilience theory makes the same move in ecological language. Holling distinguished stability from resilience and showed that a system may remain near equilibrium while losing the ability to absorb disturbance and reorganize (Holling, 1973). That distinction matters far beyond ecology. It shows why smooth continuity is not the same thing as strong adaptation. Structural Intelligence extends this insight by treating revision under pressure as the key indicator of whether coherence is load-bearing.

Organizational learning gives the argument its practical middle layer. Argyris argued that learning occurs when errors are detected and corrected, and he distinguished shallow correction from deeper restructuring of governing variables (Argyris, 1977; Argyris, 1995). That is already very close to the SI claim. A system that can only patch behavior without revising the assumptions that produce failure remains vulnerable. Schön's account of reflective practice points in the same direction: competent professionals think in action by reflecting on surprises generated in practice rather than simply applying static rules (Schön, 2017). Surprise, error, and redesign are not side issues. They are central to intelligent practice.

AI safety and evaluation work now confront the same requirement in a sharper technological form. Concrete Problems in AI Safety made explicit that capable systems can still fail through reward hacking, side effects, scalable oversight failures,

and distributional shift (Amodei et al., 2016). Corrigibility research asks how a system can remain open to intervention rather than manipulating or resisting its operators (Soares et al., 2015). Work on human preferences and harmless assistants also shows that value alignment is not solved by base capability; it requires structures for critique, preference learning, revision, and supervision (Christiano et al., 2017; Bai et al., 2022; Bai et al., 2023). Model Evaluation for Extreme Risks generalizes the lesson: dangerous capability evaluations and alignment evaluations are needed because the appearance of competence says too little about whether systems will apply their capabilities safely (Shevlane et al., 2023).

These traditions do not all define intelligence the same way. They do share a deeper pattern. Adaptive success depends on the system's ability to register mismatch, keep correction alive, and revise internal organization rather than merely defend output. Structural Intelligence makes that common pattern explicit and promotes it from background assumption to primary criterion.

The key move is simple. Intelligence is not the power to generate order at any cost. Intelligence is the power to let reality reorganize order without total collapse.

5. Cheap Coherence and the Classification Error

The reason this definition matters now is that coherence has become an unreliable status signal.

Human cognition already contains this vulnerability. Processing fluency increases perceived truth, and repeated statements feel truer partly because they are easier to process and integrate (Dechêne et al., 2010; Reber & Unkelbach, 2010). Repetition also increases the sharing of misinformation by increasing perceived accuracy first (Vellani et al., 2023). Narrative persuasion research adds a second mechanism. Stories are often easier to understand than non-narrative presentations, and that ease increases persuasion (Bullock et al., 2021). Coherence-based reasoning adds a third mechanism: people do not simply weigh evidence item by item; they actively reorganize evidence into patterns that maximize global fit, and those patterns can amplify bias once a wrong or overweighted node becomes dominant (Simon & Read, 2025).

Large language models amplify all three vulnerabilities. They are built to produce smooth continuation, contextual fit, and stylistic appropriateness. That makes them extraordinarily good at generating the signs of intelligence that humans already over-trust. TruthfulQA showed that strong language models can imitate falsehoods found in their training data, and that bigger models were not reliably more truthful (Lin et al., 2022). The hallucination literature shows that the issue is broad rather than anecdotal. Improvements in fluency do not remove the tendency to generate unsupported content across summarization, dialogue, question answering, data-to-text generation, and related tasks (Ji et al., 2023).

Alignment work responds to this by adding more explicit supervisory structure. Helpful and harmless training, constitutional critique and revision, preference modeling, and external risk evaluations all try to make outputs more corrigible and less dependent on surface plausibility alone (Christiano et al., 2017; Bai et al., 2022; Bai et al., 2023; Shevlane et al., 2023). These efforts are valuable. They also reveal the deeper point. Once coherence can be manufactured cheaply, intelligence can no longer be judged primarily by coherence.

The same classification error explains why institutions often feel unreal before they visibly fail. A bureaucracy can increase reporting, metrics, and procedural

order while thinning its connection to mission, dissent, and correction. Argyris described how organizations hide threat and embarrassment, then call the resulting defensive routines rationality (Argyris, 1977). Structural Intelligence names the same failure in a cross-domain way: coherence remains, answerability falls.

Cheap coherence therefore produces a systematic classification error. We reward systems for looking integrated when what matters is whether integration remains corrigible. We praise outputs for sounding complete when what matters is whether they survive correction. We call a system intelligent when it solves presentational problems while quietly failing reality problems. Structural Intelligence is a corrective to that mistake.

6. The Answerability Loop

Structural Intelligence becomes operational through a simple loop: model, act, encounter constraint, revise.

The first stage is structure or model. Every acting system already contains some organization. It may be explicit, as in code, policy, or formal theory. It may be implicit, as in habits, assumptions, incentives, identity, or role. The model does not have to be fully articulated. It only has to be real enough to guide action.

The second stage is action or expression. The system does something. It speaks, decides, regulates, predicts, allocates, justifies, or performs. At this stage many systems look intelligent, because action can remain smooth long after the underlying structure has begun to drift.

The third stage is constraint. Reality answers back. Predictions fail. Hidden costs rise. Contradictions accumulate. External witnesses disagree. The environment changes. The institutional metric no longer tracks the mission. The person's story no longer explains the repeated outcome. Constraint is the point where the world stops being fully narratable from inside the existing model.

The fourth stage is revision. Revision is the decisive stage. The system can update, redesign, integrate feedback, alter its governing variables, and reorganize itself for the next cycle. Or it can defend. It can rationalize, punish error signals, intensify performance, export cost, or reduce the visibility of failure. The difference between those two responses is the difference between answerability and drift.

The loop is intentionally simple because it needs to travel across domains. In a person, revision may involve learning, action change, and deeper integration. In an institution, revision may involve redesigning incentives, governance, reporting lines, and error channels. In AI, revision may involve better evaluation, supervision, uncertainty handling, oversight, or training objectives. The materials differ. The structure repeats.

This loop is not a metaphor added after the fact. It is the operational core of the view. It explains why Structural Intelligence should be treated as a primary criterion rather than a decorative supplement. Systems become deeply intelligent when they can pass through this loop repeatedly without losing the capacity for reality contact.

Systems become brittle when they can no longer let the third stage reorganize the fourth.

The shortest way to state the whole model is this: intelligence is not exhausted by what a system can do before correction. It is shown by what the system becomes after correction.

7. What Follows

The first consequence is for AI evaluation. If Structural Intelligence is right, then evaluation cannot stop at helpfulness, style, preference, or benchmark success. Those matter. They are not enough. Strong evaluation must ask whether outputs remain grounded, whether uncertainty is visible, whether correction can enter, whether the system resists unsupported completion, and whether oversight has enough authority to change behavior rather than merely score it. This point aligns with existing AI safety and evaluation work, but it sharpens the target. The real question is not only whether a model behaves well on curated tests. The question is whether the model remains answerable when deployed into pressure, ambiguity, and consequence (Amodei et al., 2016; Lin et al., 2022; Shevlane et al., 2023).

The second consequence is for institutions. Accountability is not exhausted by transparency, audit, or reporting volume. These may improve correction, but they may also serve as substitutes for it. A structurally intelligent institution has upward-moving error signals, clear cost accounting, redesign capacity, and enough permission for contradiction to reach decision-making. Without those features, order can increase while reality contact declines. Organizational learning research already supports this point. Structural Intelligence gives it a more general language and connects it directly to questions of truth, burden, and drift (Argyris, 1977; Argyris, 1995).

The third consequence is for human self-understanding. Insight is not the same as revision. A person may narrate their history brilliantly and still remain organized around the same defended structure. Structural Intelligence therefore shifts the center of self-understanding away from eloquence and toward binding change under pressure. This connects directly to the psychological line developed elsewhere in the SI corpus, especially Structural Intelligence, Answerability, and the Jungian Bridge and What Answerability Feels Like Under Pressure. The human question becomes sharper: when life says no, what in me can actually revise?

The fourth consequence is epistemic. Truth-signals need to change. Coherence still matters. It cannot carry the burden alone. In a world saturated with persuasive fluency, the stronger epistemic virtues are traceability, falsifiability, correction channels, cost visibility, and revisability under time. Truth, in the practical sense relevant here, is coherence that remains answerable under contact.

The final consequence is architectural. Systems should be designed so that correction can bind before collapse becomes the only teacher. That means less dependence on perfect foresight and more investment in error visibility, dissent channels, reflective redesign, and resilient oversight. Structural Intelligence is demanding in exactly this way. It asks systems to become less impressive in image and more revisable in fact.

8. Objections

One obvious objection is that this paper merely renames cybernetics, resilience, and organizational learning. The answer is no. Structural Intelligence does build on those traditions. It does not pretend to emerge from nowhere. Its contribution is a stricter synthesis. It connects control, correction, truth, and practical judgment around one criterion: whether coherence remains answerable under constraint. Cybernetics gave feedback. Resilience theory gave adaptive persistence under disturbance. Organizational learning gave error correction and double-loop revision. Structural Intelligence unifies these around a diagnostic distinction between coherence and contact that becomes especially urgent in the AI era.

A second objection says that intelligence is too broad a concept to be anchored in revision. Some systems are obviously intelligent because of sheer computational performance. That objection identifies a real issue and still misses the target. Structural Intelligence is not trying to replace every use of the word intelligence. It is identifying the sense of intelligence that matters most for systems operating under consequence. A system may be computationally brilliant and still fragile, uncorrectable, or reality-thin. SI does not deny narrow intelligence. It argues that narrow intelligence is insufficient for durable adaptive intelligence.

A third objection says that the view risks moralizing intelligence. Systems that accept correction sound nicer, safer, or more democratic, but perhaps not more intelligent. This objection also fails. Answerability is not niceness. A system may revise harshly, competitively, or with no sentimental language at all. The criterion is not moral tone. The criterion is whether the system can let reality change it. Many pleasant systems are weak by this standard, because they preserve comfort by blocking correction.

A fourth objection says that coherence itself is adaptive, so treating it with suspicion goes too far. Structural Intelligence agrees that coherence is adaptive. The issue is not coherence as such. The issue is untethered coherence. A coherent system that stays open to contradiction, cost, and redesign is stronger than an incoherent one. SI is therefore not anti-coherence. It is anti-inflation. It resists the move by which coherence becomes a false certificate of truth or intelligence.

The final objection says that all of this may describe humans and organizations better than AI. The reply is that AI is precisely where the distinction becomes easiest to see. Large language models can produce the external signs of understanding while remaining weakly tied to what they describe. That does not make the claim less relevant to humans. It makes the latent problem visible. AI is the demonstration case that forces a cleaner theory.

9. What Would Count Against This View

A serious theory needs real failure conditions. Structural Intelligence should not exempt itself from its own standard.

The first falsifier would be a robust demonstration that systems lacking strong revision-capacity under constraint nonetheless outperform revision-capable systems across changing environments without hidden cost, brittleness, or growing correction debt. If that pattern held broadly rather than in isolated short-horizon cases, the SI definition would weaken.

The second falsifier would be evidence that coherence alone remains a reliable indicator of truth and durable adaptivity in present conditions. If highly fluent systems were also consistently the most corrigible, well-grounded, and consequence-sensitive systems, then the distinction between coherence and contact would lose much of its force.

The third falsifier would be failure of cross-domain portability. This paper claims that the same adaptive structure repeats across humans, AI, institutions, and other systems. If that repetition turns out to be only metaphorical, and no operational indicators can travel between domains, then the claim should be narrowed. Structural Intelligence would still have local value, but not the general scope claimed here.

The fourth falsifier would be empirical failure of the proposed indicators. If revision latency, error visibility, cost accounting, and repair effectiveness do not predict durable adaptation better than standard performance metrics, then SI has not earned its place as a primary criterion.

Those are meaningful risks. They do not weaken the proposal. They keep it honest. A theory of answerability that cannot itself be revised under answerability pressure would collapse into the same performance it is trying to criticize.

10. Position in the SI Corpus

This paper is the shortest claim-based entry into the Structural Intelligence project. It should not replace the longer work. It should orient it.

Readers who want the fuller foundational treatment should begin with Structural Intelligence: Coherence, Contact, and Answerability Under Pressure. That paper develops the primary vocabulary of coherence, contact, answerability, burden, collapse, and revision, and it remains the clearest first full statement of the framework. Readers who want the cross-domain adaptive account should then read Structural Intelligence, Answerability, and the Jungian Bridge, which develops the Answerability Loop and shows how the same dynamics appear in humans, AI systems, organizations, societies, economies, ecosystems, and infrastructure. Readers who want the deeper ontological layer should continue to Beyond Structural Intelligence, which separates the philosophy of structure from the applied method of Structural Intelligence. Readers who want the AI-facing social consequence should read The Terrible Miracle of AI, which develops the problem of synthetic witness. Readers who want the formalizing pressure should read The Drift Ratio and What Answerability Feels Like Under Pressure.

The role of the present paper is narrower. It states the core claim without the full architecture. Intelligence is not best defined by fluency, performance, or coherence alone. It is best defined by whether a system can revise under constraint while remaining organized enough to act. That is the thesis. It is simple enough to test, strong enough to matter, and open enough to be proven wrong.

That is the level at which Structural Intelligence should stand or fall.

References

External Research

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete Problems in AI Safety. arXiv. https://arxiv.org/abs/1606.06565

Argyris, C. (1977). Double Loop Learning in Organizations. Harvard Business Review, 55(5), 115–125. https://www.avannistelrooij.nl/wp/wp-content/uploads/2017/11/Argyris-1977-Double-Loop-Learning-in-Organisations-HBR.pdf

Argyris, C. (1995). Action Science and Organizational Learning. Journal of Managerial Psychology, 10(6), 20–26. https://www.vidartop.no/uploads/9/4/6/7/9467257/argyrys_theory_of_action.pdf

Ashby, W. R. (1956). An Introduction to Cybernetics. Chapman & Hall. https://ashby.info/Ashby-Introduction-to-Cybernetics.pdf

Bai, Y., Askell, A., Chen, A., et al. (2022). Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. arXiv. https://arxiv.org/abs/2204.05862

Bai, Y., Kadavath, S., Kundu, S., et al. (2023). Constitutional AI: Harmlessness from AI Feedback. arXiv. https://arxiv.org/abs/2212.08073

Bullock, O. M., Shulman, H. C., & Huskey, R. (2021). Narratives are Persuasive Because They are Easier to Understand: Examining Processing Fluency as a Mechanism of Narrative Persuasion. Frontiers in Communication, 6, 719615. https://doi.org/10.3389/fcomm.2021.719615

Christiano, P. F., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2017). Deep Reinforcement Learning from Human Preferences. Advances in Neural Information Processing Systems 30. https://papers.neurips.cc/paper_files/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. Personality and Social Psychology Review, 14(2), 238–257. https://doi.org/10.1177/1088868309352251

Holling, C. S. (1973). Resilience and Stability of Ecological Systems. Annual Review of Ecology and Systematics, 4, 1–23. https://doi.org/10.1146/annurev.es.04.110173.000245

Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55(12), Article 248. https://doi.org/10.1145/3571730

Lin, S., Hilton, J., & Evans, O. (2022). TruthfulQA: Measuring How Models Mimic Human Falsehoods. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 3214–3252. https://doi.org/10.18653/v1/2022.acl-long.229

Reber, R., & Unkelbach, C. (2010). The Epistemic Status of Processing Fluency as Source for Judgments of Truth. Review of Philosophy and Psychology, 1(4), 563–581. https://doi.org/10.1007/s13164-010-0039-7

Schön, D. A. (2017). The Reflective Practitioner: How Professionals Think in Action. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9781315237473/reflective-practitioner-donald-sch%C3%B6n

Shevlane, T., Farquhar, S., Garfinkel, B., et al. (2023). Model Evaluation for Extreme Risks. arXiv. https://arxiv.org/abs/2305.15324

Simon, D., & Read, S. J. (2025). Toward a General Framework of Biased Reasoning: Coherence-Based Reasoning. Perspectives on Psychological Science, 20(3), 421–459. https://doi.org/10.1177/17456916231204579

Soares, N., Fallenstein, B., Armstrong, S., & Yudkowsky, E. (2015). Corrigibility. AAAI Workshop on AI and Ethics. https://intelligence.org/files/Corrigibility.pdf

Vellani, V., Zheng, S., Ercelik, D., & Sharot, T. (2023). The Illusory Truth Effect Leads to the Spread of Misinformation. Cognition, 236, 105421. https://doi.org/10.1016/j.cognition.2023.105421

Wiener, N. (1948/2019). Cybernetics: Or Control and Communication in the Animal and the Machine. MIT Press Open Access. https://direct.mit.edu/books/oa-monograph/4581/Cybernetics-or-Control-and-Communication-in-the

Selected SI Corpus References (PhilArchive)

Jovanović, V. (2026). Structural Intelligence: Coherence, Contact, and Answerability Under Pressure. PhilArchive. https://philarchive.org/rec/JOVSIC-2

Jovanović, V. (2026). Structural Intelligence, Answerability, and the Jungian Bridge: Toward a Practical Theory of Adaptive Systems Under Pressure. PhilArchive. https://philarchive.org/rec/JOVSIA-3

Jovanović, V. (2026). Beyond Structural Intelligence: Toward a Philosophy of Structure. PhilArchive. https://philarchive.org/rec/JOVBSI

Jovanović, V. (2026). The Terrible Miracle of AI: Coherence, Contact, and the New Problem of Synthetic Witness. PhilArchive. https://philarchive.org/rec/JOVTTM

Jovanović, V. (2026). The Drift Ratio: Coherence, Contact, and Answerability in Structural Intelligence. PhilArchive. https://philarchive.org/rec/JOVTDR

Jovanović, V. (2026). What Answerability Feels Like Under Pressure. PhilArchive. https://philarchive.org/archive/JOVWAF