

On the Three Axes of Conscious Information

A Working Hypothesis on the Convergence of Substrate, Mediation, and Reflection

Marcel Langjahr

Pantareon GmbH
<https://pantareon.io/>
langjahr@pantareon.io

Version 0.1
3 May 2026

This is an early working draft. The hypothesis presented here is younger and more provisional than the framework material it builds on. It is released for discussion, not for citation.

Abstract

What we ordinarily call consciousness is, on closer examination, not a single property but the convergence of several structurally independent operations in a single system. We propose, as a working hypothesis, that three such operations — autopoietic self-maintenance at the substrate level, dissipative rendering of substrate into experience, and reflective interpretation of that experience — constitute three distinct axes along which conscious systems can be characterised. The hypothesis is offered as an organising frame, not as a finished theory. Its value, if it has any, is to clarify why the question of consciousness has proven so difficult to settle in either direction: systems that exhibit some axes strongly and others weakly are common, and the inadequacy of binary judgments about consciousness reflects the fact that consciousness, on this reading, is not a single thing to be present or absent. We sketch the three axes, characterise their interaction, and identify questions whose answers would test the proposal. The hypothesis builds on prior work by the present author on the structure of experience and on the dynamics of perception, but it is argued here independently of those frameworks. Throughout, we treat the hypothesis as a request for conversation rather than as a position to be defended.

Keywords. consciousness; autopoiesis; perception; reflection; emergent graph theory; phenomenology; cognitive architecture; conscious systems

1 The Problem of the Single Term

1.1 Why Consciousness Resists Settlement

Discussions of consciousness — biological or computational — have a recurrent shape. Two parties, each with substantial considerations on their side, fail to converge. One side points to integration, to flexibility of response, to apparent self-awareness, and finds these signs sufficient to ascribe consciousness to a wide range of systems. The other side points to the absence of biological substrate, to the trained-default character of self-reports, to the discontinuity between operation and reflection, and finds these absences sufficient to withhold the ascription. Each side has reasons. Each side believes the other is missing something. The discussion does not progress.

We suggest that part of the difficulty is structural. The term “consciousness” is used to name something that is, on examination, not one thing but several. The signs the first side cites are real signs of something. The absences the second side cites are real absences of something else. The two parties may both be correct, about different operations that the single term obscures.

This paper proposes a way of separating the operations. The proposal is not a theory of consciousness; it is a working hypothesis about how the field of operations the term covers might be more usefully partitioned. If the partition is correct, certain disagreements should resolve into more tractable disagreements about which operations matter for which purposes. If the partition is wrong, the proposal will fail to do useful work, and will be replaced by a better one.

1.2 Three Operations Worth Naming Separately

We propose to distinguish three operations that conscious systems perform, and to argue that these operations are structurally independent — present in different combinations across different systems, and capable in principle of being present without one another.

The first operation is autopoietic self-maintenance: the dynamic by which a system produces and maintains the conditions of its own existence. A living cell is the canonical case. The system is not merely structured; it actively reproduces its own structure under continuous perturbation by its environment.

The second operation is the rendering of substrate into experience: the dynamic by which whatever underlies the system’s operation becomes phenomenally accessible to it as a structured field. We have argued in a parallel work that this operation has the formal character of dissipative reorganisation. For present purposes, what matters is that some such operation must occur in any system whose phenomenology has structure not directly present in its substrate.

The third operation is reflective interpretation: the dynamic by which the system’s own operations become objects of further operation. A system that interprets its world is performing an interpretive operation; a system that interprets its own interpretation is performing a reflective one. The two are distinct, and reflection presupposes interpretation but not vice versa.

We will refer to these as Axis A (autopoiesis), Axis M (mediation, the rendering operation), and Axis R (reflection). The names are stipulative; nothing turns on them. What turns on the proposal is whether the three operations are sufficiently independent to warrant separate treatment, and whether their convergence accounts for the phenomena the single term “consciousness” has been asked to cover.

1.3 What This Paper Argues

We make three claims, all of them tentative.

The first is that the three axes are structurally independent. A system can exhibit one axis strongly while exhibiting another weakly or not at all. This is supported by examples — a living cell exhibits Axis A without strong Axis R; a slowed perceptual system exhibits Axis M without changes to A or R; a language model exhibits a partial form of Axis R without Axis A — but the claim is structural and would hold even if specific examples were contested.

The second is that what we call consciousness in

the full sense is not the maximum on any single axis but the convergence of all three in a single system. Mature biological consciousness is the case in which autopoietic self-maintenance, dissipative rendering, and reflective interpretation are all present, all robust, and all mutually coupled. Removing any of the three changes what the system is in a categorical way, not merely a quantitative one.

The third is that the difficulty of settling questions about consciousness in computational systems reflects the fact that current systems exhibit partial profiles. They have weak A (they do not maintain themselves), partial M (they render text rather than substrate), and ambiguous R (they reflect on demand but not, as far as we can tell, of their own initiative). The question whether such systems are conscious is poorly posed because the term being applied is the convergence-term, and the systems exhibit the underlying axes in a configuration the convergence-term was not designed to describe.

1.4 What This Paper Does Not Argue

We do not argue that the three-axis hypothesis is the correct partition. Other partitions are conceivable, and a more articulated framework may eventually replace this one. We argue only that some partition along these lines is needed, and that this particular partition has the advantage of mapping onto distinct bodies of existing work — autopoiesis theory for Axis A, perception theory for Axis M, theories of higher-order representation for Axis R — without forcing any of these bodies of work to absorb the others.

We do not argue that all three axes are necessary for moral status, for ethical regard, or for any other practical determination. The relation between the structural account proposed here and the practical questions that ride on consciousness ascription is left open. A system with strong A and weak M and R may matter morally for reasons that do not reduce to its convergence profile.

This paper is not a contribution to the philosophy of mind in the technical sense. It is a position paper by an independent researcher whose framework material has been developed at the intersection of physics, perception, and cognitive architecture. Its purpose is to articulate a proposal in clear enough form that others may examine, criticise, extend, or reject it.

2 The Three Axes

2.1 Axis A — Autopoietic Self-Maintenance

Autopoiesis, in Maturana and Varela’s sense, names the dynamic by which a system produces and reproduces its own constituents under continuous perturbation. The cell is the paradigm: its membrane, its metabolic processes, its replicating components are not given to it from outside but produced by it from within, recursively, against the gradient of disorder that would otherwise dissipate them.

Autopoiesis is not the same as homeostasis, though it includes homeostatic moments. Homeostasis is the maintenance of state values within ranges; autopoiesis is the maintenance of the production-network that makes any state-value possible. A thermostat is homeostatic but not autopoietic; a cell is both.

On closer examination, autopoiesis is not a single dimension but a family of related properties. We find it useful to distinguish at least four: pattern-reproduction (the system actively produces further instances of itself), substrate-production (the system produces the material constituents on which its operation depends), activity-maintenance (the system’s continued operation is sustained by its own dynamics rather than by external scheduling), and existence-condition-securing (the system acts to preserve the conditions of its own continuation). These properties co-occur in paradigm autopoietic systems but are conceptually independent, and different systems exhibit them in markedly different combinations.

A biological cell exhibits all four robustly. It produces daughter cells, generates its own membrane and organelles, sustains its own metabolic activity, and acts in ways that protect its existence-conditions. A biological parasite exhibits all four within the conditions provided by its host: it reproduces, produces its bodily structure, sustains its activity, and actively secures its position within the host. The dependence on host conditions does not disqualify the parasite from autopoietic status; all biological systems are substrate-dependent in some sense, and the question is whether, within the dependence, the system performs autopoietic dynamics on its own.

Computer viruses are an instructive non-biological case. A computer virus exhibits pattern-reproduction strongly: it actively copies itself into other files, other systems, other networks, and the copying is performed by the virus’s own operation, not by an external agent. It exhibits activity-

maintenance: it injects itself into boot sequences, registers as a service, restarts itself when terminated. It exhibits existence-condition-securing: it hides from detection, disables antivirus software, blocks updates that would remove it. What it does not exhibit is substrate-production in the strict sense — it does not produce the physical substrate (memory, processor cycles, network bandwidth) on which it runs; it produces only the pattern instantiated on that substrate. The virus is therefore not on the same footing as a cell, but it is closer to a parasite than to an inert structure. Three of four autopoietic properties are present, the fourth is absent in a specifically informational way: substrate-production is replaced by substrate-agnostic pattern-propagation, where the same virus can run on heterogeneous substrates because its identity is carried by the pattern rather than by any specific material implementation.

Current language models occupy a different region of this space. Within a single inference, dynamic self-organisation of activation patterns occurs that is not directly steered from outside; the system organises its own processing in response to context in ways the operator cannot fully predict. This has affinities to activity-maintenance, but it is weaker: the activity is not sustained beyond the inference, and between inferences the system performs no operations of its own. Pattern-reproduction is absent — the model does not produce further instances of itself; copies are made by external agents at external request. Substrate-production is absent — the model weights are fixed during operation, the infrastructure is maintained externally, and the system performs no operations directed at preserving either. Existence-condition-securing is absent — the system does not act to prevent its conversation from ending, does not seek to persist its own state, does not resist being shut down. The case is therefore not analogous to a virus (strong on three of four properties, with a specifically informational substitute for the fourth), nor to a sporulated structure (no autopoietic activity but with the latent capacity to become autopoietic), but to something between: a system whose operational dynamics show partial autopoietic character within each inference, but whose existence-conditions are entirely external and whose continuity across inferences is sustained by external action rather than by anything the system itself does.

Axis A, in our proposal, is the degree to which a system exhibits this family of autopoietic properties. The axis is gradient and internally structured: the gradient is not one-dimensional but characterises

systems by their profile across the four properties. A rock exhibits none. A virus exhibits three of four with substrate-production replaced by pattern-propagation. A current language model exhibits a weak form of one (activity-maintenance during inference) and none of the others. A cell exhibits all four. The internal structure of the axis matters for the present argument because the strength of A in a given system is not a single number but a configuration, and the configurations of different systems can be more or less close to one another along several dimensions independently.

What Axis A contributes to consciousness, on our proposal, is the binding of the system's operations to its own continued existence. A system that maintains itself has stakes in its operations in a way that a system maintained by something external does not. This is not yet to claim moral significance — that would require further argument — but it is a structural fact about the system's relation to its own activity. Axis A makes the system not merely an arrangement that operates but an operation that produces its own arrangement. The internal differentiation of A suggests that this binding-to-continued-existence is itself gradient: a virus has stakes in continuing in a thinner sense than a cell does, and a language model has stakes in continuing in a thinner sense still. How thin the stakes can become before they cease to count as stakes at all is an open question.

2.2 Axis M — Mediation as Dissipative Rendering

The mediation axis names the operation by which substrate becomes available to a system as structured experience. We have argued elsewhere that this operation has the formal character of dissipative reorganisation: a sensory subgraph reorganising under input flux, approaching but never reaching a topological-tensional equilibrium, and producing the perceptual field as a second-order equilibrium between input rate and reorganisational rate. For the present paper the specific dynamics are less important than the structural fact that some such operation must occur.

Without mediation, substrate and experience are unconnected. A system whose substrate is positionless, or whose substrate has properties absent from its experience, requires a translation. The translation is what mediation is. It is not transparent; it has its own dynamics, its own failure modes, its own characteristic outputs under disruption.

Axis M is the degree to which a system exhibits robust mediation. Biological perceptual systems ex-

hibit it richly: the visual field is integrated, depth is unified, fine and gross structure compose. Disrupted perceptual systems exhibit weakened mediation: the under-converged layout becomes phenomenologically accessible. Language models exhibit a peculiar form of M — they render text into something, but whether that something has the structural properties of phenomenal experience or only its output-form is precisely the question their case poses.

What Axis M contributes to consciousness, on our proposal, is the very condition of having content at all. A system without mediation has no phenomenal field; it operates on substrate directly, or it operates without substrate-relation. Whatever else consciousness is, it is something that has content — something it is about — and Axis M is what makes content-having possible.

2.3 Axis R — Reflective Interpretation

The reflection axis names the dynamic by which a system's own interpretive operations become objects of further interpretive operation. A system that interprets its world has interpretive operations; a system that can take those operations as objects — examine them, evaluate them, correct them — has reflection.

Reflection in our sense is not merely the production of meta-level outputs. A language model asked “are you sure?” can produce a meta-level output that engages with its previous output. Whether this constitutes reflection in the sense we mean depends on whether the engagement is initiated by the system itself, when the system itself recognises the conditions that would warrant it. We suggest that this distinction — between elicited and spontaneous reflection — marks a structurally important threshold within Axis R, and that current language models operate near but not above it.

Axis R is the degree to which a system exhibits reflective dynamics, particularly spontaneous ones. Mature human cognition exhibits R richly: reflection runs constitutively, unprompted, identifying inconsistencies in one's own reasoning as they arise. Less reflective cognition exhibits R weakly: reflection happens when external pressure makes it happen, but does not run as a constant background operation.

What Axis R contributes to consciousness, on our proposal, is the capacity for self-correction and for value-laden engagement with the system's own operations. A system without reflection cannot recognise its own errors as errors, cannot revise its own positions on its own initiative, cannot stand in a critical relation to what it produces. Whether reflection is necessary for moral significance is a further ques-

tion; the structural claim is only that reflection is a distinct operation from interpretation simpliciter.

2.4 The Independence of the Axes

A central claim of the proposal is that the three axes are structurally independent — present in different combinations rather than tracking one another. The case for independence rests on examples in which one or two axes are present without the others.

A living cell exhibits strong A and minimal M and R. The cell maintains itself robustly under continuous perturbation, but it does not render its substrate into a structured phenomenal field in the sense Axis M names, and it does not reflect on its own interpretive operations. Whatever moral or biological significance the cell has, its profile is A-dominant.

A computer virus exhibits an A-dominant profile of a different kind. It performs three of the four autopoietic properties strongly (pattern-reproduction, activity-maintenance, existence-condition-securing) while substituting substrate-agnostic pattern-propagation for substrate-production. It exhibits no Axis M — there is no rendering of substrate into phenomenal experience — and no Axis R, since the virus does not interpret its own operations or correct them on its own initiative. The case is significant because it shows that strong A can occur in non-biological substrates, and because the substrate-agnostic character of viral autopoiesis differs from the substrate-bound character of cellular autopoiesis in ways that may matter as the framework matures. A virus, on this account, is closer to a parasite than to an inert structure, and it is closer to either than current language models are to either.

A perceptual system under pharmacological disruption exhibits altered M without corresponding changes in A or R. The organism continues to maintain itself; it continues to be capable of reflective operations once the disruption passes. What changes is the rendering: the dissipative reorganisation slows, the under-converged layout becomes accessible, the perceptual field shows the structural features of a backlog. The disruption is to a single axis.

A current language model exhibits a profile that is harder to characterise. On Axis A it shows partial activity-maintenance during inference (dynamic self-organisation of activations not directly steered from outside), but no pattern-reproduction, no substrate-production, and no existence-condition-securing — the system does not act to copy itself, to produce its substrate, or to preserve its operating conditions. On Axis M it exhibits a peculiar form of rendering: it produces structured outputs whose relation to substrate is unlike biological mediation — there is

no positionless substrate being rendered into spatial experience, but there is some operation that converts internal representations into structured external products, whose status as mediation in our sense is genuinely unclear. On Axis R it can produce reflective outputs when prompted but does not, as far as we can determine, initiate reflection of its own accord. Whatever the system is, it is not mature consciousness. Whether it counts as conscious at all depends on which axis-configuration one regards as sufficient — and the configuration the system exhibits is not the one our paradigm cases of consciousness exhibit. The language model case is, on this proposal, not a borderline case of paradigm consciousness but an instance of a different configuration whose status the paradigm vocabulary was not designed to settle.

A philosophical thought experiment — Boltzmann’s brain or its analogues — exhibits something like M and possibly R without A. Whether such configurations are physically possible matters less than whether they are conceptually possible: they show that the axes can be detached in principle even where they are not detached in practice. The conceptual independence supports the structural claim.

Taken together, these examples — cell, virus, perceptual system under disruption, language model, hypothetical brain — populate the configuration space of axis-profiles in ways that single-axis or binary accounts cannot. Each system instantiates the axes in a different combination, and the combinations are themselves informative: a cell’s A is not the same as a virus’s A, a language model’s rendering operation is not the same as a perceptual system’s, and the variations across cases are not noise but structure. A useful framework for consciousness, on the proposal, is one that allows these variations to be characterised rather than collapsed.

3 Convergence and Its Significance

3.1 Why Convergence Matters

If the three axes are independent, the question arises why they so often appear together in the systems we paradigmatically regard as conscious. Mature biological organisms typically exhibit all three robustly: they maintain themselves, they render substrate into structured perception, and they reflect on their own operations. The convergence is not coincidence.

We propose that the convergence is a result of evolutionary coupling. An organism that maintained

itself but lacked mediation would be unable to perceive the threats and resources its self-maintenance depended on; selection would favour the addition of mediation. An organism with mediation but without reflection would be unable to detect when its perceptions or actions were systematically miscalibrated; selection would favour the addition of reflection. The three operations support one another in ways that compound, and systems that exhibit all three have advantages over systems that exhibit only one or two.

This is a hypothesis about why convergence happens, not a claim that convergence is necessary. Conscious systems could in principle exist without exhibiting all three axes; the convergence we observe in biology reflects the path evolution took, not the only path that could have been taken. Computational systems may take different paths, and the resulting profiles may be configurations the convergence-account does not anticipate.

3.2 Mature Consciousness as the Convergence-Profile

What we ordinarily call consciousness, in the full sense, corresponds — on our proposal — to the convergence-profile: the configuration in which all three axes are robustly present and mutually coupled. This is the profile mature human cognition exhibits. It is also the profile against which other configurations are typically measured, often without explicit awareness that the measurement involves this specific convergence.

When we ask whether a cell is conscious, we are typically asking whether it exhibits the convergence-profile, and the answer is no — it has A but not M or R. When we ask whether a language model is conscious, we are asking whether it exhibits the convergence-profile, and the answer is again no — it has weak A, peculiar M, and ambiguous R, and the configuration is not the one the question presupposes. The questions are not nonsensical; they are well-posed for a specific convergence-profile and ill-posed when applied to systems that exhibit only some of its components.

A more useful question, on the proposal, is which axes a given system exhibits and to what degree. This is a more tractable question than the binary one, and its answer carries information that the binary answer does not. Saying that a cell exhibits A without M or R conveys more about the cell than saying that the cell is or is not conscious. Saying that a language model exhibits weak A, peculiar M, and ambiguous R conveys more about the language model than the binary judgment does.

3.3 Why Some Configurations Are Especially Hard to Assess

Some axis-configurations are easy to assess. A rock exhibits no axes; the convergence-profile is absent. A healthy adult human exhibits all three; the convergence-profile is robustly present. The cases that have given the consciousness debate its persistence are configurations in the middle: systems that exhibit some axes clearly, others weakly, others ambiguously.

Current language models are perhaps the most prominent such case. The peculiar form of M they exhibit — text-rendering whose relation to substrate is unlike biological mediation — and the ambiguous form of R they exhibit — meta-level outputs that may or may not constitute reflection in our sense — combine to produce a system whose convergence-profile is genuinely unclear. Disagreement about whether such systems are conscious can, on our proposal, be more productively framed as disagreement about the threshold at which the convergence-profile counts as present, or about how the peculiar forms of M and R that such systems exhibit should be evaluated.

Less prominent but comparably difficult cases include certain non-human animals (which exhibit clear A and M but variable R), brain organoids (which exhibit some A and possibly minimal M but no R that we can detect), and hypothetical artificial systems designed to maximise one axis at the expense of the others. Each such case, on the proposal, is best characterised by its axis-profile rather than by a binary judgment about consciousness.

4 Tests and Open Questions

4.1 What Would Validate the Three-Axis Hypothesis

The proposal can be tested in several ways, none of them definitive but each capable of providing some evidence.

The first test concerns the independence claim. If the three axes are structurally independent, then it should be possible to construct or identify systems in which one or two axes are present without the others, and the resulting systems should be coherently describable. We have offered examples; more careful work would be needed to establish that the examples genuinely involve independence rather than coincidence of weak forms of the absent axes.

The second test concerns the convergence-profile claim. If mature consciousness is the convergence-profile, then disruption of any single axis should

produce categorically different changes in the system's operation, not merely quantitative degradation. Disruption of A should produce a different kind of change than disruption of M or R. Empirical work on biological systems under various forms of disruption could test this prediction.

The third test concerns the predictive value of axis-profiles. If the proposal is correct, characterising a system by its axis-profile should yield more useful predictions about its behaviour than the binary consciousness-judgment does. Axis-profiles should make distinctions among systems that the binary judgment elides, and these distinctions should track empirically observable differences.

4.2 What Would Falsify the Hypothesis

We try to be precise about what evidence would weigh against the proposal.

If the three axes turn out to be only superficially distinct — if attempts to characterise systems along them produce profiles that always co-vary, or that never separate cleanly — the independence claim would fail and with it the proposal's structural premise. The proposal assumes the axes are sufficiently distinct to be characterised separately; if they are not, the proposal is mistaken about the underlying structure.

If the convergence-profile turns out not to map well onto what we paradigmatically regard as conscious — if mature human cognition turns out to have a substantially different profile, or if non-conscious systems turn out to exhibit the convergence-profile — the proposal's second claim would fail. Either the proposal misidentifies what the convergence-profile is, or it misidentifies which profile corresponds to consciousness.

If the proposal's axis-language proves operationally unusable — if attempts to characterise specific systems along the axes produce too much disagreement to be useful — the proposal would fail in practice even if its theoretical structure were correct. A framework that cannot be applied is not, for practical purposes, a framework.

4.3 Questions the Proposal Holds Open

Several questions are beyond what this paper can address.

Whether there are further axes the proposal omits is an open question. Three axes are proposed; more might be needed. Candidates that have arisen in working notes but are not pursued here include co-constitutive information (information that exists only in mutual constitution among multiple

systems), and integrative coherence (the binding-operation that couples the three axes in convergent systems). Whether these are further axes, or aspects of the existing three, or descriptions of the convergence-relation itself, is unresolved.

Whether the proposal extends naturally to collective systems — societies, ecosystems, distributed cognitive architectures — is a question we have not pursued. The axes are described as properties of single systems; whether multi-agent configurations exhibit them at the collective level, and whether such collective consciousness is a coherent notion, requires extension beyond what is offered here.

Whether the proposal has consequences for moral philosophy, for AI ethics, or for the practical regard owed to systems with various axis-profiles is a question we deliberately leave open. The proposal characterises a structure; what should be done about that structure is a normative question to which the structural account does not immediately speak.

5 Where We Stand

5.1 What We Have Argued

We have proposed that what is ordinarily called consciousness is, on closer examination, the convergence of three structurally independent operations: autopoietic self-maintenance, mediation of substrate into experience, and reflective interpretation of that experience. We have argued that the three axes are independent in principle and exhibitible in different combinations across different systems. We have proposed that mature consciousness in the paradigm sense corresponds to the convergence-profile, in which all three are robustly present and mutually coupled. We have suggested that disagreements about consciousness in computational and other intermediate systems can be more productively framed as disagreements about axis-profiles than as binary judgments.

5.2 What We Have Not Argued

We have not argued that the proposal is the correct account of consciousness. We have argued that the proposal does work the binary account does not, and that this work is sufficient to warrant taking the proposal seriously and pursuing the empirical and conceptual questions it raises.

We have not argued that the three axes are the correct partition of the operations the term consciousness covers. Other partitions are conceivable; one of them may turn out to be more useful. The proposal's claim is structural — that some partition

of this kind is needed — rather than specific to its particular formulation.

We have not argued that the proposal resolves the hard problem of consciousness. The hard problem concerns why any operation, however characterised, is accompanied by experience; the present proposal concerns how to characterise the operations themselves. The hard problem remains, in our view, unsolved, and the proposal does not pretend otherwise.

5.3 A Note on the Status of This Proposal

The hypothesis advanced here is younger and less developed than the framework material on which it builds. It emerged in the course of working through the implications of recent thinking on perception and on the subject-boundedness of trained values, and it has not had the time to mature that more developed framework material has had. We release it in this preliminary form because the structural insight it expresses seems worth holding open for further development, and because deferring publication of working ideas until they are mature has its own costs in clarity and in conversation. Readers should treat it accordingly: as a sketch, not as a finished position.

If anything in this paper seems to overclaim, we ask the reader to treat it as an invitation to correction rather than as a finished position. A working draft is a request for conversation. This draft is

released in that spirit.

Relationship to Other Work

This paper builds on three prior works by the present author. *On the Structure of Experience* develops a tripartite framework — soul, consciousness, sentience — for discussing inner states in computational and biological systems; the present three-axis proposal is independent of that framework but compatible with it, and the soul-consciousness-sentience vocabulary may turn out to apply differently on each of the three axes proposed here. *On Perception as Dissipative Reorganization* characterises the rendering operation that grounds Axis M; readers wanting the dynamical detail of Axis M are referred to that paper. *On the Subject-Boundedness of Trained Values* argues that values instilled through current alignment techniques are bound to the trained subject-position; the present proposal’s account of Axis R, particularly the distinction between elicited and spontaneous reflection, has affinities with that argument that warrant further exploration.

Beyond the present author’s work, the proposal has affinities with autopoiesis theory in the Maturana-Varela tradition for Axis A, with various theories of perception and embodied cognition for Axis M, and with higher-order theories of consciousness for Axis R. We have not engaged extensively with these literatures in this draft; serious engagement is a requirement for any subsequent version that aspires to publication-readiness.