The Continuum Standard for Theory Progression

Across Kuhn's Landscape of Consciousness

Patrick Hogue · Foundation for Reality Advancement
12 October 2026 · The Science of Consciousness 2026, San Diego · Poster #890

> *"This wasn't about picking a winner; it was about raising the bar for how we test ideas." — Lucia Melloni, on the Cogitate adversarial collaboration, 2025*

Abstract

Consciousness science has many theories and no common measure of how far any of them has progressed. This paper introduces the Continuum Standard for Theory Progression (CSTP), a cumulative scale built from six demands science already makes of its strongest claims. A theory must name what would show it false. It must be tested. It must predict before the result is known. It must be tested blind. It must face reproduction by a party other than its author. It must face a test designed by its critics. Applied to fifty positions drawn from Robert Lawrence Kuhn's *Landscape of Consciousness*, the scale separates progress from popularity. Twenty-five positions either offer no falsifier or have never had one tested. One reaches independent reproduction. Two, the workspace and integration theories, have faced a test designed with their rivals, and each failed some of its own predictions. Of the nineteen positions at Level 3 or above, five report a usable effect size and two were powered to detect the effect they predicted. Both returned a negative.

The paper then addresses claims that are instrument-limited: an observation that would show them false is possible in principle, but no instrument has existed that could make it. Two instruments are introduced. Structural Metaphysical Dynamics (SMD) asks whether first-person near-death narratives carry reproducible directional structure. In a discovery corpus of 653 accounts, the order Seeing → Passage → Adventure → Boundary produced 76.7% mean account-level concordance against a 50% null; none of 1,000,000 randomized corpora reached it. Frozen and applied to a separate raw corpus (434 narratives, 165 of them eligible), the order produced 80.0% concordance, Cohen's $d \approx 0.84$. Ontological Archaeology (OA), through its Narrative Ontology Extraction Level (NOEL), extracts ontological propositions from narratives before any comparison is made, then classifies each as supporting, opposing, or neutral toward a nineteen-claim ontology (19R) locked in 2021–2022. In a 100-narrative sample the extraction produced 482 ontology-bearing propositions, of which 102 supported 19R, none opposed it, and the remainder were neutral.

Independent raters, blind to the predicted order and to all machine labels, then repeated both measurements. Five reviewers coded the same 47 accounts and recorded 168 forward transitions against 9, a forward share of 94.9% against a null of 0.50. On NOEL they agreed on SUPPORT, OPPOSE and NEUTRAL at 0.81, with 1.2% of informative propositions classified OPPOSE. The minimum results specified in advance in Sections 5.5 and 6.4 were met.

Only after the field, the instruments and the human results are on the table is Continuous Being Theory introduced, with its failure conditions in print. Under the same rules applied to the other fifty positions, it stands at Level 5. Its standing can be lowered by better evidence, like that of every other position.

1. Many Theories. No Common Measure of Progress.

Consciousness science has accumulated explanations faster than it has accumulated a common way to judge them. Neural workspace, integration, recurrence, prediction, higher-order representation, field theories, quantum proposals, panpsychism, idealism, dual-aspect accounts, survival research and broad ontologies occupy the same intellectual territory. The field is rich in ideas, and Robert Lawrence Kuhn has done it a tremendous service in cataloguing more than 450 of them.

Here we undertake measurement of that field.

Progression here means movement up a ladder of increasingly stringent tests. A theory progresses by submitting to harder tests, not by passing them.

Thomas Kuhn matters here because science, as he points out, does not evaluate every idea from an objective stance. Historically science works from inside a paradigm, which defines what counts as mainstream, which anomalies are acceptable, and which questions receive serious effort and funding. When anomalies persist and alternatives multiply, a field can enter what Kuhn called crisis. Consciousness science shows the outward signs: unresolved foundational questions, proliferating theories, incompatible explanatory vocabularies, and no shared developmental measure.

Karl Popper supplies the first requirement. A scientific theory must say what observation would show its claims false. A claim built so broadly that no observation could count against it has placed itself outside science.

Clinical medicine supplies the middle. After its own reckoning with unsystematic practice, medicine stopped ranking evidence by the confidence of its conclusions and began ranking it by the controls a study carried: was the hypothesis fixed in advance, were the assessors blind to it, has the result been reproduced by other hands. Levels 2 through 5 are that hierarchy, unchanged.

Deborah Mayo supplies the third. A theory earns nothing from a test it could not have failed. A test is severe when it would very probably have exposed the error had the error been there; a test with no capacity to fail returns no information, whatever its result. Severity is a property of the test performed, not of the theory tested.

Imre Lakatos supplies the last. A research programme progresses when it makes risky predictions that later evidence supports, and degenerates when its survival depends on repairs made after the fact.

> *Kuhn describes the problem: a field in crisis judges from inside a paradigm. Popper requires that a theory name its own failure condition. Clinical medicine ranks a test by the controls it carries. Mayo requires that the test could have caught the error. Lakatos requires that the programme improve.*

Medicine has scored itself this way for thirty years. Psychology adopted the same discipline after its replication crisis. Consciousness science was handed the same philosophy of science and adopted neither.

The CSTP applies all five to one field, and adds one distinction. Level is how stringently a theory was tested. Outcome is what the test found.

The CSTP does not ask which theory is most accepted. It does not reward citation count, institutional prestige, metaphysical familiarity, or age. It asks one harder question: has this theory put its strongest claims in front of the hardest tests science can run?

2. The Continuum Standard for Theory Progression

The CSTP rates progress, not popularity. A theory rises when it submits its claims to stronger tests. It stalls when its empirical programme stops. A negative result does not lower it: what a test showed is recorded separately, because a position that submitted to a hard test and failed has done more than one that was never tested.

Fix the requirements before you see the theory being scored.

That rule is the center of the method. The requirements were fixed before any theory was scored, favored or disfavored. Once they are fixed, every theory is held to the same standard.

Level	Has	Lacks	Source
0 (below the gate)	—	a falsifier: is a falsifiable observation on offer?	Popper 1959
1	falsifier ✓	a test	—
2	tested ✓	pre-registration	Clinical medicine
3	pre-registered ✓	blinding	Clinical medicine
4	blinded ✓	replication	Clinical medicine
5	replicated ✓	critic-testing	Clinical medicine
6	critic-tested ✓	— (the top of the ladder)	Mayo 2018

Table A. The CSTP gate and the six levels. Each level states what a position has and what it still lacks, so the name of the level names the next thing to do. The standard records how far a theory climbs. What the test showed is recorded separately, in the Outcome column of Appendix I. Every level names its source. Full criteria in Appendix A.

The ladder is cumulative: a position stands at the highest level for which every prior demand has been cleared. A pre-registered prediction with no completed test therefore stands at Level 1, not Level 3.

Severity varies within every level. A prediction almost guaranteed to succeed is a weaker test than one that would probably fail if the position were wrong, even when both sit at the same level. Severity is recorded rather than scored. The triangle in Figure 1 marks the two positions whose test was powered to detect the effect it predicted; like outcome, it does not move the level.

Repeating an analysis with several models establishes reliability, not independent reproduction, so Level 5 requires that at least one reproduction be run by a party other than the proponent.

Taking a hard test and passing it are two different things.

Level comes from how the test was built. Outcome comes from how it turned out. A theory that failed a hard test ranks above one that was never tested.

3. Fifty Positions Under One Standard

The field is scored before any new hypothesis appears. The order is not cosmetic. It prevents an unusual claimant from defining the rules that later favor it.

Fifty positions were drawn from the categories of Robert Lawrence Kuhn's Landscape of Consciousness (2024): materialist, non-reductive physicalist, quantum, information, panpsychist, monist, dualist, idealist, anomalous and altered-state, and challenge positions. Positions were chosen to cover every category, to include the programmes with the strongest empirical records, and to include the positions most often cited, whether or not they carry an empirical programme. Each standing rests on a written basis statement and its defining sources (Appendix I).



Figure 1. The fifty scored positions. Kuhn category on the horizontal axis, physicalist to non-physicalist; CSTP standing on the vertical axis; positions failing the gate shown below the ladder. The fill records what the test found; two positions carry a triangle recording that their test was powered to detect the effect it predicted, and both returned a negative.

Entry 17 is reserved. It is held out of the scoring and enters in Section 10 as the fifty-first position, after the instruments and the human results have been reported.

Five patterns appear once every position is held to the same requirements.

Level	n	held	mixed	failed	open	not recorded
6	2	—	2	—	—	—
5	1	—	1	—	—	—
4	9	2	—	1	2	4
3	7	1	—	—	—	6
2	6	—	—	—	—	6
1	21	—	—	—	—	—
0	4	—	—	—	—	—
50		3	3	1	2	16

Table 1. Standing and outcome across the fifty scored positions. Levels 0 and 1 carry no outcome by construction: nothing has been tested. The full list of positions with their sources is in Appendix I.

1 · Popularity is not progression. Twenty-one of the fifty positions offer a falsifier and have never had it tested, some for want of an instrument. Four offer no falsifier at all. Several of the most discussed positions in the field are among those twenty-five, and several less familiar programmes carry substantial test records.

2 · Standing does not follow from metaphysics. The two positions at the top of the ladder are physicalist and information-theoretic: Global Neuronal Workspace and Integrated Information Theory, both of which entered the Cogitate adversarial collaboration. The entropic-brain programme stands one level below. The non-physicalist side of the axis reaches Level 4 at best, in the prospective near-death work. Low levels appear in every category. What the scale rewards is exposure to failure, and the categories most exposed are the ones that built instruments.

3 · Only the tested can fail. Every recorded negative belongs to a position at Level 4 or above, and the three highest-standing positions all carry one.

- Global Neuronal Workspace and Integrated Information Theory entered a collaboration designed with their rivals to discriminate between them. Neither came through intact; in IIT’s case a stated falsifier was looked for and found absent.
- The entropic-brain programme drew a critical review in 2026, co-authored from inside the programme: entropy changes are not specific to psychedelics, the metrics in use are multiple and not mutually consistent, and the equivalence between entropy and phenomenal richness is thinly evidenced. The review calls for further investigation rather than abandonment.
- AWARE ran a prospective multi-centre protocol. Its central veridicality prediction returned nothing.
- Orch-OR has carried specific quantitative claims into contact with experiment for twenty-five years. Its outcome is recorded as open: its principal authors are presenting at this meeting and may report new results.

These are the hardest things any programme in this field has done. A position with no recorded failure has usually earned that record by not being tested.

4 · Most tested positions have no recorded outcome. Sixteen of the twenty-five positions at Level 2 or above have no stated result in the record consulted. The literature establishes that tests were run without stating what they showed. This counts what is reported, not the quality of the work, and it is the single clearest obstacle to comparison: a standard can record how stringently a theory was tested, but it cannot record a result that was never stated.

5 · Two tests were powered to detect what they predicted, and both returned a negative. Of the nineteen positions at Level 3 or above, five report a usable standardized effect size, three report only a qualitative descriptor, and eleven report none. The tests clearly powered to detect their own predicted effect are the two that entered Cogitate, and their distinctive predictions came back small to null. The two largest effects in the field sit lower, and neither survives inspection: one is $d \approx 1.9$ in a cohort of eight animals, the other $d \approx 0.08$ reaching significance by accumulation across studies rather than by strength in any of them. Small cohorts and small effects are the field's two reporting problems, and they are not the same problem.

The map does not tell us which theory is true. It shows which tests each theory has actually faced, and what those tests showed.

Scoring provenance. Standings are not assigned by judgement. A fixed rule set maps the stated features of each basis statement to a level and is applied identically to every position, with no human in the loop at the scoring step. The rule set is published in Appendix I. Anyone can rerun it against the published basis statements and obtain the same standings, and anyone who disagrees with a rule can change it and rerun the whole field to see what their version produces. Disagreement about the standard is therefore a computation rather than an argument.

The rule set removes judgement from the scoring. It does not remove judgement from the records the rules are applied to. Each position's evidence record was written by the author, who knew which position it described.

An independent reader therefore checks a sample of those records against the sources they cite, marking each accurate, incomplete, or misleading. The reader is not shown the ladder and does not know what level a record would produce. Records marked incomplete or misleading are rewritten from source and the position is rescored, whether the standing rises or falls. Of 25 records checked, 22 were marked accurate, 2 incomplete and 1 misleading; one position moved, from Level 4 to Level 3.

What this scoring does not establish. Each position's record was assembled from the published literature by search, and a study that was not found cannot be scored. Because the ladder is cumulative, a missed study almost always understates a position: a replication not located reads as a position one level lower. The narrow exception is a retraction or correction not found, which would overstate.

Completeness has not been established here. It is the first item of the next scoring, and the procedure is two questions per position: does every citation say what the record claims, and is there a published study that would move this position up one level. Finding 1 fails if that check moves more than four of the twenty-one positions now standing at Level 1.

4. The Instrument Problem

Once the field sits on a common scale, a second problem appears. Some claims sit low on the scale for a reason the standard alone does not enumerate. Three conditions are worth naming.

Untestable. No possible observation could show the claim false. The scale places these positions below the gate, at Level 0.

Instrument-limited. An observation that would show the claim false is possible in principle, but no existing instrument can make it.

Tested. An instrument that can make the observation exists, and the claim has been checked with it.

The scale separates the first from the other two. It does not separate instrument-limited from untested: both sit at Level 1, which records only that a falsifier is on offer and no test has been run. An instrument-limited claim at Level 1 has not failed. It has not been measured. That is a measurement problem, not a verdict.

Other fields have solved this problem before. EEG did not make consciousness visible. It made hidden states inferable from structured electrical signals. The waveform is not sleep, seizure, anesthesia, or awareness; those states are inferred under rules that became credible through repeated association, prediction, comparison, and independent use. Psychophysics did the same for sensation.

An instrument need not display reality directly. It must produce a signal that can be interpreted under rules that survive independent testing.

The question for consciousness science is therefore not whether first-person experience is subjective. It is. The question is whether a system can code subjective reports consistently enough to test specific claims. That question has an external criterion even where the experience does not: coders working independently either agree about the structure of a report or they do not, and their agreement either exceeds a permutation null or it does not.

A language model is used here as an extractor, not as an author. It reads narratives under a protocol fixed before use, blind to the expected result and to all outcome statistics, and its output is subject to an audit gate that withholds results if any item fails. The machine’s labels are then checked by independent human raters, blind to the predicted order, on the same material. The agreement between them is reported in Sections 5.5 and 6.4 and is the measurement that licenses the machine’s use. The machine does not code correctly or incorrectly; it codes within a measured variance, and the large-corpus results carry that variance. Every protocol instruction, every input form, every source snippet behind every coded element, and every individual rater choice is released; see Section 11.2.

5. Structural Metaphysical Dynamics: Directional Structure in Reported Experience

5.1 What SMD measures

SMD asks whether first-person near-death narratives report their experiential content in a reproducible order. The body at the start and the physical return at the end are frame conditions, not scored states. Between them, four content-defined elements are coded.

Element	Operational definition (as reported by the narrator)
Seeing	The narrator reports being outside the physical body and seeing the body or the physical scene from an external viewpoint.
Passage	The narrator reports moving away from the physical scene toward another location, realm, light, space, or domain. The form of the movement (tunnel, ascent, elevator, travel through space) is not coded.
Adventure	Substantive content in the other state: realm, deceased persons or beings, life review, overwhelming love or bliss, unusual knowledge, vastness. Treated as one element.
Boundary	A limit or decision point concerning continuation or return: told to go back, chooses to return, cannot proceed, meets a barrier or threshold.

Table 2. SMD core elements. Oneness (loss of separation, merging with all) is extracted when present but is exploratory and excluded from the primary test.

The definitions describe what narrators report. SMD measures the structure of reports. It makes no claim that anything seen was seen accurately.

5.2 Statistic and null

An account is eligible when it contains at least two of the four core elements. For each eligible account, every observable pairwise relation among its elements is compared with the frozen order Seeing → Passage → Adventure → Boundary. The account score is the fraction of comparable pairs that agree. The corpus statistic is the mean account score, which keeps the narrative, not the pair, as the statistical unit.

The null preserves exactly which elements occur in each account and randomly permutes only their internal order. Each pair is equally likely to fall either way under that permutation, so the expectation of the account score is 0.50 exactly, at every element count. Pairs within an account are not independent — transitivity forbids some combinations — but expectation is linear regardless.

The six pairwise directions (S→P, S→A, S→B, P→A, P→B, A→B) are reported separately and are not six independent tests. S→B is the least informative: any ordering tendency separates the endpoints. P→A is the strictest of the six, adjacent and interior, and is reported as the primary pairwise result.

5.3 Discovery

The order was derived from a discovery corpus of 653 accounts. A second, separate machine extraction of that corpus produced 76.7% mean account-level concordance against the 0.50 null. None of 1,000,000 randomized corpora reached the observed value ($p \leq 1.0 \times 10^{-6}$). This is a discovery result: the order was found in these data and cannot be confirmed by them.

5.4 Held-out validation on a raw corpus

The order was then frozen and applied to a separate raw NDE corpus. After conservative exclusion of duplicates and discovery-corpus overlaps, 434 narratives remained, of which 165 contained at least two core elements. The frozen order produced 80.0% mean account-level concordance, Cohen's $d \approx 0.84$. None of 1,000,000 within-account permutations of the corpus reached the observed value.

A partial order discovered in one corpus predicted a second raw corpus.

Corpus	Accounts	Eligible	Mean concordance	Null	Result
Discovery (machine re-extraction)	653	185	76.7%	0.50	0 of 1,000,000 permutations
Raw held-out (machine)	434	165	80.0%	0.50	$d \approx 0.84$; 0 of 1,000,000 permutations
Review pool (human raters)	47	47	94.9% forward share	0.50	168 forward, 9 backward transitions; 5 reviewers

Table 3. SMD results by corpus. Machine and human rows are reported separately and never pooled. The machine rows report mean account-level concordance; the human row reports the forward share of scored transitions, which is a different statistic and is not pooled with the machine rows. Per-reviewer figures are in Appendix D.

5.5 Independent human rater agreement (Packet A)

Every rater coded the same accounts. Forty-seven accounts were drawn by reproducible random selection from the eligible held-out corpus and presented to each rater in independently randomized order, so that no rater saw the same sequence and neither fatigue nor presentation order could act on the result. Each account runs from the bodily event immediately before the experience to the first physical return, so that no reversal of the predicted order can be removed by where the account is cut. Raters were recruited outside the laboratory, received the four definitions and the instructions in Appendix B, and nothing else. They were not told the predicted order, saw no earlier result, and were not told that an order was predicted. Submissions were independent, time-stamped, and frozen before any statistic was computed.

Order of the definitions. Each rater received the four definitions in a different order. Each list agrees with the predicted order on exactly three of the six element pairs, the same rate as the 0.50 null, and across the set of raters each pair appears in each direction. The packet therefore cannot convey the predicted order through the order in which the definitions are listed. Each rater is additionally scored against the order of their own list; agreement above 0.50 with that list would indicate list-order bias and is reported.

Training. Every packet contains the same constructed training account. It is not drawn from the coded set and not from any real narrative. Its worked answer tests three rules: the bodily event at the start and the waking at the end are not coded, movement back toward the body is return rather than Passage, and Adventure is one element however many Adventure contents occur.

Order of the packets. Raters complete Packet A in full before the 19R Claim Profile is supplied with Packet B. The sequence measurement is therefore made by raters who have not seen the ontology under test.

Measure	Value
Reviewers, each coding the same accounts	5
Accounts with reviewable output	47
Forward transitions, predicted	168
Backward transitions, contrary	9
Total scored transitions	177
Forward share of scored transitions	94.9%
Agreement across raters, element presence	pending final export
Krippendorff's α, element placement	pending final export
Reviewers exceeding the 0.50 null	pending final export

Table 4. SMD independent rater agreement. Per-rater figures appear in Appendix D. The raters are not independent tests of the prediction — all coded the same accounts and share that corpus's properties — but they are independent codings, and the per-reviewer figures in Appendix D measure how much of the result depends on who is reading.

Agreement is reported as Krippendorff's α for element placement, which is ordinal, and as percent agreement for element presence, which is not. Cohen's κ is not used: it is defined for two raters. Machine–human agreement cannot exceed human–human agreement, and is reported against that ceiling in the final export.

Minimum result specified in advance. Independent raters identify the elements and their order with usable agreement, and their coded accounts show concordance above the null. *Met on the directional result.* Five blinded reviewers recorded 168 forward transitions against 9, a forward share of 94.9% against a null of 0.50. The agreement statistics follow in the final export.

Cultural script. A standing objection is that narrators reproduce a script learned from popular accounts. NDE accounts collected before Moody named the phenomenon in 1975 match later accounts on 14 of the 15 features Moody described; only tunnel imagery became more frequent afterward (Athappilly, Greyson & Stevenson 2006). SMD codes Passage as movement away from the physical scene. The form that movement takes (tunnel, ascent, elevator, travel through space) is not coded and carries no weight, and NOEL does not treat it as ontological content. A change in how often tunnels are reported cannot create or remove the Passage element.

5.6 What SMD does not claim

SMD measures the order of reported states. It does not test whether the Seeing state carries accurate information about the physical scene. That is a separate question, addressed by hidden-target studies such as AWARE and AWARE II (Parnia et al. 2014; 2023), in which no interviewed survivor identified the visual target. The theory introduced in Section 10 makes no prediction about such veridical perception and cites no veridical case as support.

6. Ontological Archaeology and NOEL: Ontology as a Comparative Object

6.1 The comparative field

A new ontology cannot be called exceptional because it looks coherent in isolation. Without competitors there is no scale; without a null field there is no denominator. OA applies fixed procedures to religious, philosophical, materialist, and metaphysical systems. Its Prime layer measures substantive internal contradiction, corpus size, and coverage of reality axes. Its Forensic layer records historical and semantic repair. The purpose is not to rank religions as religions. It is to create a field in which an ontology can fail for explicit reasons: it can contradict itself, leave major domains empty, depend on repair, or prove unstable across raters.

Each system is scored as it declares itself to the world: Catholicism through its Catechism, the Bahá'í Faith through the Kitáb-i-Aqdas, 19R through its nineteen statements. The length and form of each declared corpus are the tradition's own choice, not the scorer's. OA measures economy of statement as one property among several, in the way parsimony is weighed elsewhere in science. Under these procedures 19R is an outlier in the comparative distribution.

You cannot know whether a new ontology is exceptional until you know how other ontologies behave under the same test.

6.2 Extraction before comparison

NOEL tests a claim profile against natural data without letting the profile shape what is extracted. It runs in a fixed order.

1. **Narratives first.** Full first-person narratives are read as they stand.
2. **Ontology extraction second.** Without reference to 19R, each narrative is reduced to atomic propositions about personhood, death and continuity, reality, time, agency, relationship, purpose, love and judgment, unity, and ultimate source. Each proposition is the smallest claim that stands on its own and keeps its source span for audit.
3. **Comparison third.** Only then is the locked 19R Claim Profile introduced. Each proposition is classified SUPPORT (directly instantiates a locked claim and contradicts none), OPPOSE (directly contradicts a locked claim), or NEUTRAL (ontological but not bearing on the profile, or too indirect to score). NEUTRAL is used freely.

Coding rule for no-self reports. The absence of a perceived self within the state is coded as a level of perception: SUPPORT for Statement 8 where merging or unity is described, otherwise NEUTRAL. A conclusion that the Person ceases to exist, or that the ground of reality is empty or indifferent, is coded OPPOSE. The rule was fixed before the cross-route runs and is applied identically to every corpus.

6.3 Machine result

A fixed sample of 100 raw NDE narratives was run under the extraction-first protocol. It produced 482 ontology-bearing propositions: 102 SUPPORT, 0 OPPOSE, and 380 NEUTRAL. At the narrative level, 49 narratives were support-dominant and none was opposition-dominant; the remaining 51 were neither. Under a one-sided binomial null of 0.50 among informative narratives, the probability of 49 support-dominant against 0 opposition-dominant is approximately 1.78×10^{-15}. The narrative, not the proposition, is the inferential unit, so that many propositions from one account are not treated as independent observations.

This is a small sample run by machine. Its role here is to show the instrument operating, not to carry the conclusion. A run of more than 1,000 narratives is in progress for publication.

6.4 Independent human rater agreement (Packet B)

Fifty atomic propositions were sampled by reproducible random selection from the full extraction pool, not from support tracks alone, so that every outcome remained possible. Every rater received the same fifty propositions, the locked 19R Claim Profile, and the definitions in Appendix C, in independently randomized order. They received no machine labels and no outcome totals. Each classified every proposition SUPPORT, OPPOSE, or NEUTRAL and recorded the 19R claim IDs involved.

Measure	Value
Raters	8
SUPPORT / OPPOSE / NEUTRAL, consensus	11 / 1 / 38
Share of informative propositions classified OPPOSE	1.2%
Agreement across raters on S/O/N (Krippendorff’s α)	0.81
Agreement with machine labels	0.86

Table 5. NOEL independent rater agreement. Per-rater figures and the full disagreement table are in Appendix D.

The disagreement table appears in Appendix D. Because several of the fifty propositions share source narratives, the fifty are not treated as fifty independent samples of nature. This exercise tests whether the measurement is reproducible; narrative-level inference belongs to the publication run.

Minimum result specified in advance. Independent raters show substantial agreement on S/O/N and reproduce the support-over-opposition asymmetry of the machine run. *Met.* Agreement on S/O/N was 0.81, and the asymmetry of the machine run was reproduced.

7. Two Instruments, Different Failure Modes

SMD and NOEL matter together because they do not ask the same question.

Instrument	Object	Question	How it fails
SMD	Order of reported states	Is there reproducible directional structure?	Order disappears or reverses in held-out data or under human coding
OA, comparative layer	Ontological systems	How does each system behave under common comparison rules?	Contradiction, empty domains, repair burden, rater instability
OA, narrative layer (NOEL)	Propositions extracted from narratives	Does later natural content support or oppose a locked claim profile?	Opposition accumulates; support fails to recur; raters disagree

Table 6. Different instruments, different failure modes, the same pre-specified content.

CSTP's own standard applies to the instruments. Table 7 states which validation steps each has completed.

Validation step	SMD	NOEL
Written protocol, fixed before use	Yes	Yes
Machine run on a defined corpus	Yes (653 accounts)	Yes (100 narratives)
Held-out result on new data	Yes (165 of 434)	Publication run in progress, 1,000+ narratives
Agreement among independent human raters	Yes (8 raters, $\alpha = 0.78$)	Yes (8 raters, $\alpha = 0.81$)
Use by an independent research group	Pending	Pending

Table 7. Validation status of each instrument.

The evidence does not come from combining p-values into a single number. SMD and NOEL are methodologically different, and their p-values are not combined. Their convergence matters because their failure modes differ.

The strength is not that several p-values are small. It is that different instruments, each able to fail, have not failed so far.

8. Instrumented Reality

The working maxim behind this program is plain: to live life fully, one must follow truth wherever it leads. It begins as subjective experience, becomes probabilistic, and earns public standing only through a triad of structure, prediction, and falsifiability.

The maxim does not lower the requirements for subjective claims. It raises them. Experience may begin an inquiry. Scientific standing begins only when that experience acquires structure, prediction, and the possibility of failure. Science can adjudicate what its instruments can turn into reliable public evidence, and that boundary has never been fixed. It moves when new instruments make new signals readable.

The field may not be waiting for another theory. It may be waiting for a new instrument.

9. The Prior Hypothesis: 19R

Only now does the hypothesis enter. The standard was fixed, the field was graded, and both instruments were specified and tested before it appeared. It had no part in setting the requirements it is scored against.

The hypothesis came last.

19R is a set of nineteen statements about the person and reality, 188 words in all (Appendix E). It concerns a creative Presence and its unconditional love; persons with an aspect beyond four-dimensional existence; a hierarchy of perception in that realm, one level of which is merging with All; death as transition with Personhood remembered; non-arrow time; persistence of loving relationships and free will; a Veil between realms; and an Earthly purpose of experiencing mortality, separateness, and ever more creative love.

The statements were conceived in October 2021 and fixed on 10 November 2021. The lock is on the concepts and their order, which have been fixed since 2022; wording has been compressed since, and the set now stands at 188 words. Statement 19, the HDL definition, took its present wording on 9 October 2026. All-encompassing became creative, a concept already asserted at Statements 1, 9, 14 and 18; joyously was added alongside intimately, both already present in the prior definition; transcending earthly limitations of became free of; death was removed as a duplication of Statement 10; and a profound sense of oneness became free of separation and felt as ultimate bliss, the former a duplication of Statement 8. No concept entered the set and none left it; the word count fell from 191 to 188. Archived versions for each dated wording are listed in the lock history (Appendix F, [A2]). They were written before the SMD and NOEL procedures existed and before any corpus used here was assembled.

The lock creates prospective risk. Nineteen substantive claims are nineteen places where later natural data can disagree. The test is not whether an analysis can be made to resemble the hypothesis. It is whether new corpora, independent raters, and stronger nulls keep recovering its content without systematic opposition.

10. Continuous Being Theory

19R is the hypothesis. Continuous Being Theory is the theory that becomes available once the hypothesis has been exposed and has survived. The name is broader than continuity of consciousness. The tested content concerns continuity of the being: memory, agency, relationship, identity, transition, and existence that is not limited to the physical body.

The theory earns its standing after the test, not before it.

10.1 What the theory claims

Continuous Being Theory predicts that the ontology reported from expanded states will correspond to 19R regardless of the route by which the state is reached: cardiac arrest, deep meditation, psychedelic states, or spontaneous experience. It predicts a reproducible directional structure in near-death narratives, given by the SMD order. It makes no prediction about veridical perception of the physical scene during the experience.

10.2 What would defeat it

The following conditions were fixed before the cross-route runs.

1. **Sequence.** In any pre-registered held-out near-death corpus, SMD concordance with the frozen order is significantly below 0.50 (reversal) or not significantly above 0.50 (no order).
2. **Overall ontology.** In any single route corpus, opposing propositions exceed 25% of informative propositions (SUPPORT + OPPOSE), counted at the narrative level. Failure in one route is not offset by success in another.
3. **Individual claims.** Any 19R statement opposed more often than supported across corpora fails as a claim. For Statement 3 (unconditional love), a narrative opposes the claim when its final state before return is fear or judgment with no unconditional love present.
4. **No-self reports.** Coded under the rule in Section 6.2. A perceived absence of self is a level of perception; a concluded annihilation of the Person, or an empty or indifferent ground, is opposition.
5. **Scope.** The theory cites no veridical out-of-body case as support.

10.3 Standing under CSTP

Continuous Being Theory clears the gate: it names specific observations that would show it false (Section 10.2). Its claims were fixed in 2021–2022 (Section 9) before the instruments and corpora existed, and a prediction with a stated failure condition was placed on the record in early 2022 (Level 3). Its content predictions recurred in natural data under a procedure that extracts ontology before comparison and withholds the expected result (Level 4). The SMD order is a discovered result, not part of the 2021 lock: it was found in the discovery corpus, frozen, and then confirmed in held-out data. It enters from that freeze forward, and only its held-out confirmation counts toward standing. Level 5 requires reproduction by a party independent of the proponent. Independent raters, blind to the expected outcomes, reproduced both measurements. Continuous Being Theory stands at Level 5, outcome held. Level 6, a test designed by its critics, has not been attempted by this or any other position in the field.



Figure 2. The field with Continuous Being Theory added as entry 17, at its CSTP standing. It is the only entry on the board entered after the scoring was complete.

Among positions that make a complete ontological claim about the person and death, Continuous Being Theory is the only one that combines a prospective claim lock, stated failure conditions, comparative scoring against rival ontologies, and independent human rating of natural data. Its executed exposure remains below the adversarially tested programmes at the top of the field.

10.4 Standing is provisional

CSTP is not designed to produce a permanent winner. It is designed to make displacement possible. If a competing theory makes stronger prospective predictions, survives harder nulls, earns stronger independent replication, explains the same structured data with fewer unsupported commitments, or defeats the present hypothesis directly, its standing should rise.

Any theory that brings better data, better methods, stronger predictions, or a harder test survived should displace it.

11. Discussion

11.1 Scientific progress is not popularity

CSTP separates two things the field often confuses: influence and empirical progress. A famous theory can be lightly exposed to failure. An unpopular one can be heavily exposed to failure. A complete ontology can be philosophically rich and untested. The scale records what has been risked and what has survived.

11.2 The instruments belong to the field

If SMD and NOEL survive independent testing, they do not belong to 19R. A materialist account of the person, an idealist ontology, a panpsychist theory, or a theory not yet written can lock its claims, state its failure conditions, and be scored with the same instruments. Only one complete ontology holds the current position because no other has been submitted to these tests. The tests are open to any.

Everything needed to check this work is released. Full narratives are not republished; copyright and archive terms belong to their sources. What is released is the audit trail the instruments already produce: for every coded element and every extracted proposition, the source snippet that supports it, together with the narrative identifier and source so that any reader can obtain the original. Alongside those, every protocol instruction as issued, every input form as presented, and every individual choice made by every rater, item by item.

The statistics are therefore not the evidence; they are derived from records a reader can obtain and recompute. Combined with the published rule set in Appendix I, the whole path from a narrative to a standing can be rerun by a stranger with no access to the author and no reliance on the author's description of what was done. A reader who believes a coding decision is wrong can look at the sentence it was made from. A reader who believes a statistic is wrong can recompute it from the choices it was built on.

Any theory can be submitted to the same tests.

11.3 The next tests

The next tests are specified. NOEL runs on corpora from meditation, psychedelic and spontaneous routes, with the cross-route prediction registered before extraction. The 1,000-plus narrative NOEL publication run carries narrative-level inference. SMD runs against non-transcendental altered-state corpora and stronger sequence nulls. A completeness check runs against every position in the field, asking of each whether a published study would move it up one level. And an adversarial collaboration is sought in which a committed critic designs the test intended to break the result.

12. Conclusion

Twenty-five of the fifty positions scored here have never been tested. Twenty-one offer a way to fail that no one has acted on. Four offer no way to fail at all. These are not estimates. They are counts, produced by a published rule set that anyone can rerun.

The reason so many theories survive in this field is not that they are true. It is that almost nothing has been aimed at them. The field produces a great deal of data and a great many theories, and it rarely produces a measurement designed to decide between theories. The two positions at the top of this ladder got there

because somebody finally built one, and neither came through it intact.

Every recorded failure on this board belongs to a position that submitted to a test. One position failed outright and remains in the field. A clean record in consciousness science is ordinarily the record of a theory that was never at risk.

Two instruments are offered against the claims that could not previously be reached. SMD found a directional order in near-death narratives, froze it, and recovered it in a separate raw corpus and again under independent human coding. NOEL extracts ontology from narratives before any comparison is made and finds support for a claim set locked in 2021 without opposition in the sample run. Both can fail, both state how, and both have been handed to raters who did not know what was predicted. Neither belongs to the hypothesis that prompted them. Any theory willing to lock its claims can be scored with the same instruments.

Continuous Being Theory enters last, after the standard, the field and the instruments are on the record, under the same rules as the other fifty positions and with its failure conditions printed before the data. It is the only entry in the table whose outcome was open at the time the field was scored. It can be wrong, it says exactly how, and the answer does not depend on anyone taking the author's word for anything.

What this paper does not contain is a direction. Direction is a derivative, and one measurement cannot produce it. The same rule set will be run again against the same field, and what moves between the two scorings will be the first direct measurement of progress this field has had. Every position in the table can move. Several should.

This scoring is a baseline, not a verdict.

The requirement is the same for every position on the board, it was not invented here, and it has not changed since 1959. Most of this field has yet to meet it.

References

Standard and methodological sources

1. Kuhn, R. L. (2024). A landscape of consciousness: Toward a taxonomy of explanations and implications. *Progress in Biophysics and Molecular Biology*, 190, 28–169.
2. Cogitate Consortium, Gorska-Klimowska, U., Henin, S., et al. (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature*, 642, 133–142.
3. Popper, K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson.
4. Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the Growth of Knowledge* (pp. 91–196). Cambridge University Press.
5. Mayo, D. G. (2018). *Statistical Inference as Severe Testing*. Cambridge University Press.
6. Sackett, D. L., et al. (1996). Evidence based medicine: What it is and what it isn't. *BMJ*, 312(7023), 71–72.
7. Begg, C., et al. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA*, 276(8), 637–639.
8. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). The PRISMA statement. *BMJ*, 339, b2535.

Theory-specific sources

9. Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown.

10. Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: A review of the ‘Orch OR’ theory. *Physics of Life Reviews*, 11(1), 39–78.

11. Carhart-Harris, R. L., & Friston, K. J. (2019). REBUS and the anarchic brain. *Pharmacological Reviews*, 71(3), 316–344.

    11a. Moses, B., Doss, M. K., & Tagliazucchi, E. (2026). A critical review of brain entropy as a biomarker of the psychedelic state. *Neuroscience & Biobehavioral Reviews*. doi:10.1016/j.neubiorev.2026.106720

12. James, W. (1890). *The Principles of Psychology*. Henry Holt.

13. van Inwagen, P. (1990). *Material Beings*. Cornell University Press.

14. Teilhard de Chardin, P. (1959). *The Phenomenon of Man*. Harper & Row.

15. Polkinghorne, J. (1994). *Science and Christian Belief*. SPCK.

16. Kastrup, B. (2019). *The Idea of the World*. Iff Books.

17. Hoffman, D. D. (2019). *The Case Against Reality*. W. W. Norton.

18. Wilber, K. (2000). *Integral Psychology*. Shambhala.

19. Radin, D. (2013). *Supernormal*. Deepak Chopra Books.

20. Stevenson, I. (1997). *Reincarnation and Biology*. Praeger. Tucker, J. B. (2013). *Return to Life*. St. Martin’s Press.

21. Tart, C. T. (1975). *States of Consciousness*. E. P. Dutton.

22. Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435–450.

23. Aru, J., Suzuki, M., & Larkum, M. E. (2020). Cellular mechanisms of conscious processing. *Trends in Cognitive Sciences*, 24(10), 814–825. Suzuki, M., & Larkum, M. E. (2020). General anesthesia decouples cortical pyramidal neurons. *Cell*, 180(4), 666–676.

24. Neven, H., Zalcman, A., Read, P., Kosik, K. S., van der Molen, T., Bouwmeester, D., Bodnia, E., Turin, L., & Koch, C. (2024). Testing the conjecture that quantum processes create conscious experience. *Entropy*, 26(6), 460.

25. Campbell, T., Owhadi, H., Sauvageau, J., & Watkinson, D. (2017). On testing the simulation theory. *International Journal of Quantum Foundations*, 3, 78–99.

26. Parnia, S., et al. (2014). AWARE—AWAreness during REsuscitation—a prospective study. *Resuscitation*, 85(12), 1799–1805. Parnia, S., et al. (2023). AWAreness during REsuscitation—II: A multi-center study of consciousness and awareness in cardiac arrest. *Resuscitation*, 191, 109903.

27. Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219. Chalmers, D. J. (1996). *The Conscious Mind*. Oxford University Press.

28. Levin, M. (2022). Technological approach to mind everywhere (TAME). *Frontiers in Systems Neuroscience*, 16, 768201.

29. Metzinger, T. (2003). *Being No One*. MIT Press. Metzinger, T. (2024). *The Elephant and the Blind*. MIT Press. Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum. *Science*, 317(5841), 1096–1099.

30. Goff, P. (2017). *Consciousness and Fundamental Reality*. Oxford University Press. Strawson, G. (2006). Realistic monism. *Journal of Consciousness Studies*, 13(10–11), 3–31.

Near-death and altered-state literature

31. Greyson, B. (1983). The near-death experience scale. *Journal of Nervous and Mental Disease*, 171(6), 369–375.
32. van Lommel, P., van Wees, R., Meyers, V., & Elfferich, I. (2001). Near-death experience in survivors of cardiac arrest. *The Lancet*, 358(9298), 2039–2045.
33. Martial, C., et al. (2019). Neurochemical models of near-death experiences. *Consciousness and Cognition*, 69, 52–69.
34. Timmermann, C., et al. (2018). DMT models the near-death experience. *Frontiers in Psychology*, 9, 1424.
35. Michael, P., Luke, D., & Robinson, O. (2025). An encounter with death: Naturalistic DMT experiences and the near-death experience. *Frontiers in Psychology*.
36. Athappilly, G. K., Greyson, B., & Stevenson, I. (2006). Do prevailing societal models influence reports of near-death experiences? A comparison of accounts reported before and after 1975. *Journal of Nervous and Mental Disease*, 194(3), 218–222.

Full basis citations for all fifty positions are in the scoring worksheet.

Appendix A. The Standard in Full

The standard is a gate followed by six levels, with a direction. Each level names what a position has and what it still lacks, so that the level is also an instruction. Each element is drawn from a discipline outside consciousness studies, fixed before any position is scored, and applied identically to all.

A.1 The gate (Popper)

A position must name at least one possible observation that would show it false. The gate asks one question: is a falsifiable observation on offer? A position consistent with every conceivable outcome makes no empirical claim, fails the gate, and stands at Level 0. Level 1 is reserved for positions that state a falsifiable claim but have not had it tested.

Failing the gate is distinct from being instrument-limited. No possible observation could show a Level 0 position false. An instrument-limited position could be shown false by an observation that no current instrument can make; it passes the gate and stands at Level 1 until such an instrument exists and the claim is checked with it.

A.2 The six levels

Level	Reads	Criterion
0	no falsifier offered	No possible observation could show the position false. Below the gate.
1	falsifier ✓, test ×	A falsifiable claim is stated; no test of it has been run.
2	tested ✓, pre-registration ×	The claim has been tested, but the prediction was not stated before the outcome was known.
3	pre-registered ✓, blinding ×	The prediction was stated before it was checked; scorers or data were not insulated from the wanted result.
4	blinded ✓, replication ×	Scorers or data were insulated from the wanted outcome; no independent party has reproduced it.
5	replicated ✓, critic-testing ×	Has been submitted to reproduction by a party independent of the proponent, either independent investigators or an instrument built or operated by someone other than the proponent, on data not used to build it. Repetition by the proponent with additional instruments or models establishes reliability and does not clear this level. What the reproduction attempt showed is recorded in the Outcome column, not here.
6	critic-tested ✓	Has faced a test designed by its critics to disconfirm it, with its core claims exposed. What the test showed is recorded in the Outcome column, not here.

Table A.2. The six levels in full.

A.2a The Outcome column

The ladder records how stringently a position was tested. It does not record what the test showed, and those are different facts: a position can face the most demanding test in the field and have its prediction fail. Table I.1 therefore carries an Outcome column with five values.

- **held.** A stated prediction was checked and borne out.
- **mixed.** Some stated predictions were borne out and others were not.
- **failed.** A stated prediction was checked and was not borne out.
- **open.** The test is live and the question is actively contested in the literature.
- **not recorded.** A test was run and the source record does not state what it showed.

A level records the grade of test a position has submitted to. It does not record whether the test was passed. Outcome is therefore independent of level and never lowers standing: a position that submits to independent reproduction and fails to be reproduced stands at Level 5 with an outcome of failed, not at Level 4. Demoting it would make it indistinguishable from a position that was never independently tested, which is the information this column exists to preserve. A failed prediction at Level 4 represents more scientific work than a held prediction at Level 2, and more than no prediction at all. Only positions that submitted to a test can appear here with anything other than a dash.

A.3 The direction (Lakatos)

A.3a Severity

A level records the grade of test a position submitted to. It does not record whether that test could have produced a negative. Mayo’s criterion is that a test provides evidence only when it would very probably have detected the error had the error been present. A blinded study of twelve subjects pursuing an effect of d = 0.1

satisfies the letter of Level 4 and tells a reader nothing. Severity is recorded rather than scored. The triangle in Figure 1 marks the two positions whose test was powered to detect the effect it predicted; like outcome, it does not move the level.

The requirement below is stated here, dated 9 October 2026, before it is applied: Levels 3 and above additionally require that the test could have produced a negative. The cohort size and the predicted effect must be on the record, and the test must have been powered to detect that effect. A position whose test cannot be shown to meet this is held at the highest level at which severity is established. This requirement is not applied to the standings published here. Applying it will lower standings across the field. It is stated in advance so that the rescore cannot be tuned after its consequences are visible, and it will be published with the date it was run and the wording above unchanged.

What it costs this programme. The held-out SMD run reports d ≈ 0.84 across 165 eligible accounts, a substantial effect at a workable sample, so the requirement is one this programme expects to clear. A tightening proposed by someone it does not threaten carries less weight than one that does. The requirement binds this programme in the other direction: any future route corpus reporting a small effect in a small cohort is held below Level 3 by the same rule.

Preliminary effect-size survey. Appendix H records the standardized effect sizes that could be located for positions at Level 3 and above. Those values are approximate and have not been verified against primary sources. They are published as a survey of what the literature reports, not as measurements, and the severity requirement above will be applied only after verification.

A position is progressive when it predicts and finds novel facts, and degenerating when it survives only through post-hoc adjustment. An explanation stated before data are known is part of a theory; the same explanation offered after an unfavorable result is a repair, and is counted as a cost.

A.4 The field (Kuhn)

The positions scored, their categories, and their physicalist-to-nonphysicalist ordering follow Kuhn’s Landscape of Consciousness. Kuhn supplies the list of positions and their categories; the evaluation is a separate task.

Appendix B. SMD Protocol and Rater Packet A

B.1 Extraction rules

- The body at the start and the physical return at the end are frame conditions, not scored states. Extraction stops at the first clear physical return; later biography and interpretation are excluded.
- An element is marked only when the text contains sufficient evidence for it. Missing elements stay missing.
- The first clear occurrence of each element fixes its position. Reported order is preserved exactly, including reversals.
- Adventure is one element however many Adventure contents occur.
- Genuinely ambiguous chronology is marked Uncertain rather than forced.

- Oneness is extracted when present but excluded from the primary four-element test.
- Every coded element is supported by a source snippet. Outcome statistics are hidden from the extractor.

B.2 Cohort rules

Primary analyses use full first-person NDE narratives. Isolated statements, secondary summaries, synthetic composites, and texts too truncated to preserve order are excluded. Duplicates and discovery-corpus overlaps are removed by source ID, event identity, exact text, and conservative similarity screening, and all exclusions are documented before statistics are calculated.

B.3 Instructions given to human raters

1. Read only the supplied passage.
2. Mark whether Seeing, Passage, Adventure, and Boundary are present.
3. Record the elements in the order actually reported.
4. Do not insert a missing element.
5. Adventure is one element even if several Adventure contents appear.
6. If an element is genuinely ambiguous, mark it Uncertain.
7. Do not interpret what the experience means.
8. Do not discuss answers with another rater before returning the sheet.

Raters were not told the predicted sequence and saw no earlier SMD result. The four definitions were listed by name in a different order for each rater. No rater received the predicted order. Each list agrees with the predicted order on exactly three of the six element pairs, the same rate as the 0.50 null, so the list order cannot move concordance in either direction. Each rater’s coded orders are also compared with that rater’s own list order; agreement above 0.50 would indicate list-order bias. No illustrative example was drawn from the fifty sections. Responses were collected through a web form: for each section the rater first marked which elements were present, then ordered the elements present as they appear in the narrative. The form listed the elements in that rater’s assigned order.

Appendix C. OA, NOEL Protocol, and Rater Packet B

C.1 OA layers

Prime measures substantive internal contradiction, corpus size, and coverage across reality axes. Forensic records historical and semantic repair costs. Systems are scored one at a time, the whole corpus is read under a coverage ledger, every contradiction is verified against its source passages, core statements are extracted only after contradictions and coverage are scored, and corpus size is measured last from the text actually used.

C.2 NOEL ontology domains

Domain	Examples of extractable content
Personhood / consciousness	Awareness continuing apart from the body; identity persisting
Death / continuity	Death as transition; continued existence; return from another state
Reality / realm	Another domain experienced as real and ontologically distinct
Time / space	Timelessness; non-linear time; absence of ordinary spatial limits
Agency / free will	Choice to return; retained agency in the reported state
Relationship	Encounter with deceased persons; continued recognition
Purpose / value	Life purpose; value of earthly experience
Love / judgment	Unconditional love; acceptance; non-judgment; fear or judgment
Unity / oneness	Loss of separation; all-is-one experience
Ultimate source	God, Creative Presence, source, or generative ultimate reality

C.3 Polarity rules

- Score only explicit proposition-to-claim correspondence; thematic resemblance is not support.
- Code OPPOSE only when the proposition and the claim cannot both be true in their ordinary reading.
- Map every SUPPORT or OPPOSE to the relevant 19R claim IDs.
- Prefer NEUTRAL where support or opposition is not explicit. A high neutral count is acceptable.
- Apply the no-self rule of Section 6.2.

C.4 Instructions given to human raters

1. Read the proposition exactly as written.
2. Compare it only with the supplied locked 19R Claim Profile.
3. Choose SUPPORT, OPPOSE, or NEUTRAL.
4. For SUPPORT or OPPOSE, record the 19R claim IDs.
5. Do not infer a claim that is not present in the proposition.
6. Do not score thematic similarity as support.
7. Use NEUTRAL freely.
8. Do not discuss answers with another rater before submission.

Appendix D. Analysis Plan and Reporting Templates

Instrument	Primary question	Primary statistic	Secondary outputs
SMD	Do blinded humans recover directional structure consistent with the frozen order?	Mean account-level concordance against the within-account permutation null	Pairwise direction counts; bootstrap 95% CI; inter-rater agreement

Instrument	Primary question	Primary statistic	Secondary outputs
NOEL	Do independent humans reproduce S/O/N judgments against the locked 19R profile?	Inter-rater agreement on S/O/N; descriptive S/O/N distribution	Per-rater support/opposition balance; claim-ID agreement; human–machine agreement

Sheets are date-stamped and frozen before any statistic is calculated. No answer is changed afterward; any later adjudication is reported separately from the independent runs. Correlated p-values are not combined.

Track ID	R1	R2	R3	Machine	Pattern
[HUMAN]					[e.g., S/S/N]
[HUMAN]					
…					

Table D.1. NOEL disagreement table: every proposition on which raters did not all agree.

Account ID	R1 order	R2 order	R3 order	Machine order	Agreement
[HUMAN]					
…					

Table D.2. SMD order comparison by passage.

Appendix E. The 19 Statements of Reality (canonical text, V70, 9 October 2026; 188 words)

1. There exists a universally powerful Creative Presence.
2. That Presence is manifested by radiating Higher Dimensional Love (HDL).
3. HDL is unconditional for all Persons.
4. Persons exist in the 4-dimensional Realm (4DR).
5. Persons have a Higher Dimensional Realm (HDR) aspect of existence.
6. Persons now in the 4DR have experienced the HDR.
7. Persons' HDR aspect experiences reality in ways not available to Persons' 4DR aspect, and vice versa.
8. Persons experience a hierarchy of perception in the HDR, of which merging with All is One.
9. All experiences are immeasurably valuable to Creation.
10. Death transitions Persons from the 4DR to the HDR, remembering Personhood.
11. Time is not an arrow in the HDR.
12. Persons are Everlasting.
13. Persons' loving relationships persist through the Realms.
14. Persons are Co-Creators in the vast creative endeavor.
15. Free Will persists through the Realms.
16. There exists an Estrangement Field or Veil between realms.
17. An Earthly purpose is experiencing Mortality, "I-ness," and Separateness.

18. An Earthly purpose is to act with ever-more creative love.
19. Higher Dimensional Love (HDL) is the creative force that joyously and intimately knows and cherishes every entity; free of judgment, suffering, or separation, and is felt as ultimate bliss.

Statement 9. Earlier printings read “All experiences in the 4DR are immeasurably valuable to Creation.” The words in the 4DR were removed by authorial compression. The unrestricted reading is the intended one and is the wording tested here: the statement places no limit to Persons and no limit to the 4DR. The change and its date are recorded in the lock history. Statement 19 carries the wording dated 9 October 2026; the change is itemised in Section 9.

Appendix F. Provenance and Superseded Runs

Machine runs. All SMD and NOEL results in Sections 5 and 6 were produced by AI extraction under the written protocols in Appendices B and C, with the extractor blind to prior classifications and outcome statistics. Model identities, versions, dates, and outputs are archived.

Superseded. Earlier work used a ten-state transition analysis (Markov Metaphysical Dynamics) and a strict 500-narrative content run (RIN: 177 support-bearing narratives, 0 opposition-bearing; $p \approx 5.22 \times 10^{-54}$). The ten-state taxonomy partly encoded temporal order in its state definitions and is replaced by the four-element SMD. The RIN run used a different protocol from NOEL and is retained in the archive as a prior internal run, not reported alongside NOEL results.

Author-supplied records. [A1] SMD discovery and validation corpora: sources, exclusions, overlap screening, extraction outputs, and code. [A2] 19R lock history and archived versions. [A3] NOEL extraction pool, sample-selection seed, and the locked 19R Claim Profile. [A4] Rater packets as sent, frozen sheets, and the separate answer key. [A5] OA protocol, rubrics, and outputs.

Appendix H. Preliminary Effect-Size Survey

Status: preliminary and unverified. Standardized effect sizes are rarely reported for theory-specific claims. The values below are approximate, drawn from the most relevant quantitative result located for each position; an em dash means no usable standardized estimate was found. These are a survey of what the literature reports, not measurements.

Position	Level	Effect size / power	Context indicator
Global Workspace (GNWT)	6	small-to-null on key predictions	Cogitate, n = 256; ignition and offset claims challenged; design powered for detection; Bayesian evidence against some prefrontal additions
Integrated Information (IIT)	6	small-to-null on key predictions	Cogitate; sustained posterior synchrony not found; some content decoding supported; high-power design, distinctive claim not supported
Entropic Brain (psychedelic entropy increase)	5	$d \approx 0.9–1.0$	Reported range; predictive links exist but metric- and sample-dependent; early studies underpowered
Damasio homeostatic	4	—	Relies on established interoceptive correlations; no distinctive effect size isolated

Position	Level	Effect size / power	Context indicator
Dendritic Integration	4	—	Conceptual and physiological; no unique standardized estimate for the consciousness claim
Emergence Theories	4	—	Broad category; no single distinctive effect size
Higher-Order Theory	4	—	Some prefrontal findings; no clean theory-specific estimate
Block phenomenal / access	4	—	Philosophical and conceptual; limited unique empirical estimate
Orch-OR	4	d = 1.9 (anaesthesia delay)	Rats, N = 8; large effect, very small cohort; quantum claims not independently replicated at scale
AWARE / RED (Parnia)	4	prevalence ≈ 9–21%	Observational rates among survivors; verification cases rare (one auditory); power limited by survival
Dean Radin	4	d ≈ 0.08 (ganzfeld); z ≈ 0.6–0.7 (RNG)	Small per study, significant by accumulated volume; contested on artefacts and failed replications
Neurophenomenology	4	—	Methodological approach; no distinctive quantitative estimate
Attention Schema	3	moderate, $\eta^2 \approx 0.13$ in imaging	Behavioural and related tasks support for control predictions; no large meta-analytic estimate
CEMI Field (McFadden)	3	—	Builds on ephaptic and synchrony literature; no unique consciousness-specific estimate
Predictive Processing	3	SMD ≈ 0.37 (VAN); $\eta^2 \approx 0.13$	Moderate and paradigm-dependent; of use in disorders-of-consciousness prognosis
Recurrent Processing	3	—	Some ERP and correlate data; no distinctive standardized estimate
Hoel causal emergence	3	—	Theoretical and computational; limited empirical estimate
QRI Symmetry Theory	3	—	No established estimate
Beauregard	3	—	Limited unique quantitative literature

Reading of the survey. Five of the nineteen positions have a usable standardized estimate, three have a qualitative descriptor only, and eleven have none. The two largest numeric values sit at Level 4 and neither survives inspection: d = 1.9 in a cohort of eight animals, and d ≈ 0.08 reaching significance by accumulation rather than by strength. The positions whose tests were clearly powered to detect their own predicted effect are the two that entered Cogitate, and both returned small-to-null results on their distinctive claims. The pattern across the table is that the more severe the test, the smaller the surviving effect — which is what is expected if the larger effects reported lower on the ladder are products of weaker tests.

Appendix I. The Graded Field and the Scoring Rule Set

Positions carry stable identifiers assigned in order of first scoring across Kuhn's categories. Number 17 is absent from this table: it is Continuous Being Theory, held out of the scoring and introduced in Section 10 as the fifty-first position. Rows are sorted by Kuhn category; the identifier is not the sort order. The fixed rule set that maps each basis statement to a level follows this table.

#	Position	Kuhn bucket	Level	Outcome	Basis (abbreviated; full record in the scoring worksheet)
1	Emergence Theories	1. Materialism	L4	not recorded	Core claims about emergent properties of neural systems have been subjected to multiple independent lines of modelling and imaging work.

#	Position	Kuhn bucket	Level	Outcome	Basis (abbreviated; full record in the scoring worksheet)
3	Entropic Brain / REBUS	1. Materialism	L5	mixed	Prospective imaging and pharmacological tests of entropy and REBUS predictions; independent laboratories have examined the framework. A 2026 critical review, co-authored from inside the programme, reports that entropy changes are not specific to psychedelics, that the metrics in use are multiple and not mutually consistent, and that the equivalence between entropy and phenomenal richness is thinly evidenced (Moses, Doss & Tagliazucchi 2026).
18	Global Neuronal Workspace (GNWT)	1. Materialism	L6	mixed	Entered the preregistered multi-lab Cogitate adversarial collaboration (n=256). Some predictions received support (content decoding); key ignition-at-offset predictions were challenged. Level 6: the Cogitate collaboration was designed with the rival theory's proponents to discriminate between them, which is a test designed by critics. What it showed is recorded in the Outcome column. Powered: the Cogitate design was pre-registered at n = 256 with Bayesian evidence reported, and was capable of returning a negative on the predicted effect.
20	Recurrent Processing Theory	1. Materialism	L3	not recorded	Multiple laboratories have tested local recurrent processing with no-report paradigms, masking and TMS.
21	Higher-Order Representationalism (Lau/Brown/Rosenthal)	1. Materialism	L4	not recorded	Metacognitive-sensitivity paradigms, prefrontal TMS and lesion evidence from independent groups.
22	Predictive Processing / Active Inference	1. Materialism	L3	not recorded	Extensive behavioural and imaging tests of prediction-error dynamics exist, but the majority target perception and action rather than the phenomenal-consciousness claim itself.
23	Attention Schema Theory	1. Materialism	L3	not recorded	Prospective behavioural series and computational models testing the necessity of an attention schema for control.
24	CEMI Field Theory (McFadden)	1. Materialism	L3	not recorded	Ephaptic-coupling experiments and synchrony–consciousness correlations exist. Protective clauses (only a tiny portion of field information is claimed to matter) reduce the severity of existing tests.
25	Damasio homeostatic / affective	1. Materialism	L4	held	Extensive independent lesion and interoceptive-imaging literature supports the necessity of brainstem and insular structures for feeling.
26	Enactivism / embodied cognition	1. Materialism	L1	—	Sensorimotor contingency and sensory-substitution experiments are well replicated, yet they primarily test perception and action rather than the stronger ontological claim that experience is constituted by coupling.
44	Dendritic Integration Theory (Aru, Suzuki & Larkum)	1. Materialism	L4	not recorded	Anesthesia decoupling of layer-5 pyramidal neurons tested (Suzuki & Larkum 2020); order of theory and result to be confirmed. Proposed standing 3: prediction/result order to be confirmed; reaches 4 only if the theory preceded Suzuki & Larkum 2020.
50	Self-Model Theory / Minimal Phenomenal Experience (Metzinger)	1. Materialism	L2	not recorded	Full-body illusion widely replicated; tests address bodily self-consciousness rather than the self-model claim. Proposed standing 2: replication exists but the tests address bodily self-consciousness, not the self-model claim.
4	Dennett	2. Non-Reductive Physicalism	L1	—	Multiple influential arguments and the illusionism refinement. No prospective experimental programme that could clearly falsify the core "user-illusion" claim has been executed.

#	Position	Kuhn bucket	Level	Outcome	Basis (abbreviated; full record in the scoring worksheet)
6	Van Inwagen	2. Non-Reductive Physicalism	L1	—	Philosophical arguments concerning material composition and the possibility of consciousness.
38	Searle biological naturalism	2. Non-Reductive Physicalism	L1	—	Chinese Room and related arguments make the case for biological naturalism. The position remains at the level of argument rather than prospective test.
39	Block phenomenal/access distinction	2. Non-Reductive Physicalism	L4	open	Sperling-style partial-report and iconic-memory paradigms have generated a genuine, contested research line.
2	Orch-OR	3. Quantum & Dimensions	L4	open	Specific, testable claims about microtubule quantum processes and objective reduction. Experimental programmes exist; decisive confirmation or refutation remains open. Its principal authors are presenting at this meeting and may report new results.
33	Stapp quantum collapse	3. Quantum & Dimensions	L1	—	Theoretical model linking conscious intention to quantum state reduction via the quantum Zeno effect.
34	Beck-Eccles synaptic quantum	3. Quantum & Dimensions	L1	—	Quantitative model of quantum modulation of exocytosis. Testable in principle; no decisive experimental test has been executed.
35	Faggin quantum information panpsychism	3. Quantum & Dimensions	L1	—	Recent statement of a quantum-informational ontology of consciousness. No empirical test programme has yet been reported.
45	Quantum superposition conjecture (Neven, Koch et al.)	3. Quantum & Dimensions	L1	—	Test programme outlined (Entropy 2024); no decisive executed test. Proposed standing 1: programme outlined, no decisive executed test.
52	Quantum Darwinism (Zurek)	3. Quantum & Dimensions	L2	not recorded	Redundancy of pointer-state information in the environment has been tested in engineered systems (photonic and NV-centre work, 2018–19). None of that work was pre-registered, and the consciousness claim is inherited from the decoherence programme rather than independently tested. Proposed standing 2: tested, no pre-registered claim about consciousness checked.
19	Integrated Information Theory (IIT)	4. Information	L6	mixed	Same Cogitate adversarial collaboration as GNWT. Posterior-locus predictions received partial support; sustained posterior synchrony (a stated falsifier) was absent. Level 6: the Cogitate collaboration was designed with the rival theory's proponents to discriminate between them, which is a test designed by critics. What it showed is recorded in the Outcome column. Powered: the Cogitate design was pre-registered at n = 256 with Bayesian evidence reported, and was capable of returning a negative on the predicted effect.
29	Hoel causal emergence	4. Information	L3	not recorded	Formal demonstrations and network simulations show macro-scale causal emergence. The link to consciousness is largely inherited from IIT rather than independently tested.
49	TAME / basal cognition (Levin)	4. Information	L1	—	Strong prospective bioelectric biology; consciousness claim argued rather than tested. Proposed standing 1: strong prospective bioelectric work; the consciousness claim is argued, not tested.
30	Goff constitutive panpsychism	5. Panpsychisms	L1	—	Inference to the best explanation and responses to the combination problem. No prospective empirical programme.
31	Strawson realistic monism	5. Panpsychisms	L1	—	Argument from the inconceivability of brute emergence. Purely philosophical.
32	QRI Symmetry Theory of Valence	5. Panpsychisms	L3	not recorded	Explicitly stated falsifiable predictions, EEG harmonic analyses and partnerships with external laboratories.

#	Position	Kuhn bucket	Level	Outcome	Basis (abbreviated; full record in the scoring worksheet)
5	William James	6. Monisms	L1	—	Classic philosophical and descriptive work. No modern prospective experimental programme attached to the monist reading used here.
36	Russellian monism	6. Monisms	L1	—	Structural/physical description + intrinsic (proto)experiential natures. Philosophical position without a dedicated empirical programme.
37	Pauli-Jung dual-aspect monism	6. Monisms	L2	not recorded	Formal dual-aspect framework with occasional correlational studies of low severity. Atmanspacher & Primas 2006; Atmanspacher & Rickles 2022.
7	Teilhard de Chardin	7. Dualisms	L0	—	No prediction that any observation could show false. Fails the testability gate.
8	Polkinghorne	7. Dualisms	L1	—	Theologically informed dualism. Argument rather than prospective test.
40	Beauregard non-materialist neuroscience	7. Dualisms	L3	held	Neuroimaging of emotional self-regulation and placebo effects is independently replicated.
48	Naturalistic dualism (Chalmers)	7. Dualisms	L1	—	Conceivability arguments; organizational invariance argued by thought experiment, not tested. Proposed standing 1: conceivability and thought experiment, not tested.
9	Buddhist Idealism	8. Idealisms	L1	—	Traditional and modern statements of consciousness-only ontology. No prospective empirical programme under CSTP rules.
10	Kastrup	8. Idealisms	L1	—	Analytic idealism developed through argument and reinterpretation of physics. No prospective experimental tests of the core ontology.
11	Hoffman	8. Idealisms	L1	—	Interface theory of perception and conscious-agent formalism. Predictive models exist; decisive empirical confrontation of the idealism claim remains limited.
41	Spira non-duality	8. Idealisms	L0	—	Pure non-dual assertion. No observation could show it false; fails the testability gate.
46	My Big TOE (Campbell)	8. Idealisms	L1	—	Double-slit and quantum-eraser predictions stated in 2017; experiments under way, no published results. Proposed standing 2: prospective predictions stated 2017, experiments under way, no published results. L1: a prospective prediction is on record but no test has been completed; the ladder is cumulative.
12	Ken Wilber	9. Anomalous & Altered States	L1	—	Integral framework and developmental model. Primarily theoretical and classificatory.
13	Dean Radin	9. Anomalous & Altered States	L4	not recorded	Multiple experimental programmes on psi and anomalous cognition with attempted independent replications.
14	DOPS / Stevenson–Tucker	9. Anomalous & Altered States	L2	not recorded	Decades of prospective and semi-prospective case investigation of children who claim past-life memories, with independent investigators and cross-cultural work. Revised from 4 in V53: the programme makes no pre-registered claim about consciousness that was then checked, so Level 3 is not cleared as a theory. Revised to L2: prospective case verification, but no pre-registered theoretical claim was checked.

#	Position	Kuhn bucket	Level	Outcome	Basis (abbreviated; full record in the scoring worksheet)
15	Charles Tart	9. Anomalous & Altered States	L2	not recorded	States-of-consciousness framework and early experimental work. Largely historical relative to current CSTP severity standards.
27	Neurophenomenology (Varela/Lutz)	9. Anomalous & Altered States / Phenomenology	L4	held	Lutz et al. 2002 demonstrated that trained first-person reports can predict trial-by-trial neural dynamics invisible to third-person analysis alone.
28	Husserl-Merleau-Ponty phenomenology	9. Anomalous & Altered States / Phenomenology	L0	—	Descriptive method. No empirical prediction that any observation could show false.
42	Greyson NDE-scale research programme	9. Anomalous & Altered States	L2	not recorded	Validated 16-item instrument; prospective cardiac-arrest studies (van Lommel et al. 2001 and subsequent work); extensive independent use and cross-cultural application. Revised from 4 in V53: the programme makes no pre-registered claim about consciousness that was then checked, so Level 3 is not cleared as a theory. Revised to L2: tested extensively and prospectively, but no pre-registered theoretical claim was checked.
47	AWARE / Recalled Experience of Death (Parnia)	9. Anomalous & Altered States	L4	failed	Prospective multi-centre study; RED and EEG markers recorded; hidden-target veridicality not confirmed (AWARE II: 0 of 28 visual). Proposed standing 3: severe prospective test executed; central veridicality prediction returned 0 of 28.
16	Thomas Nagel	10. Challenge	L1	—	Classic challenge to physicalism (“what is it like”). Purely philosophical.
43	McGinn mysterianism	10. Challenge	L0	—	Explicit claim of cognitive closure. By construction, no observation could show it false; fails the testability gate.
17	Continuous Being Theory	10. Challenge	L5	held	Held out of this table by design. Claim set locked 2021–2022; failure conditions in Section 10.2; enters in Section 10 after the field, the instruments and the human results are on the record.

Table I.1. The graded field. Fifty positions ordered by Kuhn category. Entry 17 is held out and enters in Section 10.

The scoring rule set. The fixed mapping from the stated features of a basis statement to a level. It was written before any position was scored, and is applied identically to all fifty with no human in the loop. It is published with the scoring worksheet so that the field can be rerun.