

From Models to Methods

How seven Models of Consciousness conferences made theory comparison a shared task, and what Transplant, Branch and Merge could clarify about consciousness.

2026.10.16 02:00 DENNIS HEDEGREEN ANALYSE V1.0

[HTTPS://HEDEGREENRESEARCH.COM/ARTICLES/FROM-MODELS-TO-METHODS/](https://hedegreenresearch.com/articles/from-models-to-methods/)

Across seven Models of Consciousness conferences, comparing theories became a more explicit shared task.

Reporting note: this analysis uses official programmes through 16 October 2026, before MoC7's scheduled closing session. It examines the conference series and the stated design of its consensus project; it does not report that project's final outcome.

In September 2019, mathematicians, physicists, philosophers, neuroscientists and computer scientists met at Oxford University's Mathematical Institute for the first *Models of Consciousness* conference.

The premise was unusually direct. Consciousness research had accumulated empirical observations and philosophical arguments, but many of its central questions still resisted precise comparison. The conference would focus on formal approaches to the mind–matter relation: mathematical descriptions, explicit models and theories precise enough to expose what they actually claimed. [S1][S2]

Seven years later, the seventh Models of Consciousness conference is taking place in Copenhagen.

The vocabulary has changed.

The 2019 conference foregrounded formal models. MoC7 lists four broad themes spanning philosophical foundations; methodologies and measurements; AI, LLMs and consciousness science; and phenomenology and applied phenomenology. More importantly, the conference is attempting a bottom-up transdisciplinary project intended to clarify concepts, reach common methodological standards and set future research priorities. A collective paper may follow. [S11][S12]

The programmes show recurring theories alongside increasingly explicit work on comparison. IIT appears in the 2019 booklet and remains an object of scrutiny in MoC7's session on de-idealising models. Quantum approaches appear at MoC1 and become a MoC6 core theme. These are examples of recurrence, not a count of theories eliminated across consciousness science. [S2][S9][S12]

The series gradually built machinery for working on competing theories.

That is a shift in emphasis: discussion, criticism and empirical testing were present from the beginning. Following that shift also suggests a common comparison exercise. What would different theories require us to preserve through a consciousness transplant, a branch or a merge?

The theories did not wait for consensus

MoC1 was already more plural than a simple origin story would suggest.

The conference described itself as a meeting on formal approaches to the mind–matter relation, but its programme was not a parade of mathematicians claiming to have solved consciousness. The booklet included neurophysiological recordings comparing conscious and unconscious states, proposals for experimental falsification, information-theoretic models, phenomenology, predictive approaches and competing interpretations of Integrated Information Theory. [S2]

The internal disagreements were already present. IIT was prominent, but so was criticism of its central quantities and stronger claims; quantum contributions ranged from mechanism proposals to questions about testability. [S2]

MoC1 also built discussion sessions around open questions, with participants able to propose and vote on topics. [S1][S2]

Read as a programme, it poses a harder question:

Here are several incompatible attempts to make the problem precise. What do we do with them?

A conference became an institution

The second conference, held online in 2021, took place in a different institutional environment.

The Association for Mathematical Consciousness Science had been established, giving the emerging community an organisation that persisted between conferences. MoC was no longer only an event. It sat inside a developing research network with workshops, discussions, shared goals and plans extending beyond a single week. [S3][S4]

The discussion system also began to acquire memory. MoC2 materials describe note-taking, shared discussion outputs and continuity into later community work. The point was no longer only to talk during the event, but to make some of the conversation retrievable afterwards. [S4]

By MoC3 at Stanford in 2022, the shift was stated almost explicitly.

The conference booklet described mathematical modelling as a way to translate hypotheses and phenomena into a domain where they could be tested against scientific knowledge. But it then added something more important: development was needed not only at the level of individual theories and models, but also in the **scientific processes and methodologies that act upon them**. [S5]

That sentence makes the methodological task explicit.

Once a field has several formal theories, formalisation is no longer enough. The harder questions are what a theory predicts, what would count against it, whether rival claims can be translated into comparable terms, and whether the same measurements remain valid across adults, infants, animals and artificial systems.

Those questions increasingly appear in later MoC programmes.

The conference was making **methods for handling models of consciousness** a more explicit part of its purpose.

AI did not enter the room with ChatGPT

The evolution of artificial consciousness makes the distinction particularly clear.

AI-related questions already appeared at MoC1. The issue was not invented by large language models. [S2]

By MoC3 in 2022, *Consciousness and AI* had become an explicit programme topic. Formal theories of consciousness were being translated into computational architectures and compared with developments in machine learning. [S5]

Then the external environment changed.

At MoC4 in 2023, ChatGPT and Bard appeared in the conference's own framing. Rapid advances in AI were no longer merely another application area for consciousness theories. They were presented as part of the reason consciousness science had become urgent: if increasingly capable artificial systems were being built, science needed better ways to distinguish systems with consciousness from systems without it. [S6]

By MoC5 in Bamberg, artificial consciousness had become an official core theme. MoC6 retained it as one of three core themes alongside mathematized phenomenology and quantum approaches. [S7][S9]

MoC7 goes further still: AI and LLMs now occupy one of four broad top-level streams. [S11]

This is not the story of AI suddenly entering consciousness research. It is the story of a research question gaining **institutional weight**.

Quantum approaches show the reverse movement: they appeared early, became a MoC6 core theme, then lost top-level status at MoC7 while remaining in the programme. [S2][S9][S12]

Programme promotion is not endorsement, and demotion is not extinction. A theory can remain scientifically active after losing top-level programme status, just as a controversial line can be promoted precisely because organisers think its assumptions deserve closer examination.

Comparing theories that remain in the programme

Recurrence creates a problem of comparison. Theories may address different explanatory levels: information integration, global availability, recurrence, prediction, phenomenal structure or particular physical dynamics. A shared conference slot does not make their claims direct competitors.

Nor does a theory's continued presence tell us which of its claims have survived criticism. Programme prominence measures attention, not validity.

The more useful historical question is therefore:

How did MoC make continuing disagreement more informative?

MoC3's methodology statement provides one answer. Later programmes make space for common representations, empirical tractability, falsification conditions and predictions capable of distinguishing frameworks. MoC7 includes a session called *Comparing Theories of Consciousness*, alongside work on measurements, cross-system indicators and prediction maps. [S5][S6][S12]

My reading is that the series increasingly made theory comparison a shared task. The programmes establish the activities and stated aims; they do not establish that those activities succeeded in resolving the disagreements.

Consensus about what?

Agreement among conference participants is not agreement across an entire field. Agreement on terminology, measurement, experimental design or priorities is also different from agreement on which theory is true.

MoC7's own targets are conceptual clarification, common methodological standards and future priorities. Researchers could remain divided about ontology, substrate dependence and artificial consciousness while making progress on those targets. [S11]

That raises a practical question. Could theories be compared through the same specified transformations, with each stating its predictions and limits before the answers are known?

I propose three linked cases: **Transplant, Branch and Merge**. Their immediate use is conceptual: expose what theories require and where they stop. An empirical test would need the additional step of identifying observations that distinguish the answers.

The first question is the simplest to ask:

What, exactly, would have to survive if consciousness itself were transplanted?

The Consciousness Transplant Test

A transplanted heart is not expected to preserve the identity of the old heart. It is expected to preserve the patient. The distinction seems obvious because we do not ordinarily think the heart is where the continuity of the conscious subject resides.

The brain makes the problem harder.

Suppose one biological component of a conscious brain could be replaced by a synthetic component preserving its specified interactions with the surrounding system. Then another. Continue until the biological implementation is gone, while behaviour and reports remain unchanged.

Is the resulting system conscious? And, separately, did the original subject survive?

The thought experiment is not new. Chalmers used gradual neuron-to-silicon replacement in his fading- and dancing-qualia arguments and later distinguished consciousness preservation from personal identity. MoC7 itself includes a session on substrate dependence and organisational invariance. [L1][L2][S12]

The useful move would be to give theories the **same specified replacement procedure** and compare their answers. Call this the Consciousness Transplant Test.

Preserving outward behaviour is only one possible constraint. A theory may require particular causal relations, physical dynamics, biological processes or scales. The procedure must state which properties change; it cannot assume that every theory's relevant property has been preserved and then ask whether preservation matters.

The proposed benchmark asks:

What must remain invariant for consciousness to persist through a change of implementation, and what further claim, if any, follows about the subject's continuity?

A useful answer needs more than a reference to the relevant structure. Which structure, at what scale and with what tolerance? Can activity pause and resume? If biological or quantum properties matter, which ones would the replacement lose?

Three achievements must remain distinct:

- **Reproduction:** another conscious system resembles the original.
- **Copying:** another system contains the original's memories and dispositions; whether it is conscious remains a theory-dependent question.
- **Transplantation:** implementation changes while the original subject continues, under a stated continuity criterion.

A destructive scanner might create a successor that remembers entering the machine and sincerely reports success. That report would not alone establish the third achievement.

The benchmark therefore separates consciousness at a time from continuity across time. A theory may address the first and leave the second outside its scope. It may also conclude that competing continuity accounts are experimentally indistinguishable.

That is a meaningful boundary, not an automatic failure.

For each replacement procedure, the comparison needs six answers:

1. **Invariant:** Which property supports consciousness, and which additional property, if any, supports continuity?
2. **Permitted change:** What can the implementation lose without losing those properties?
3. **Interruption:** Must a process remain active, or can it stop and restart?
4. **Failure:** Does failure mean absent consciousness, changed experience or lost continuity?

5. **Evidence:** What observation or intervention distinguishes the proposed outcomes? If none does, say so.

6. **Scope:** Does the criterion apply to humans, animals and artificial systems, or only to a specified class?

These questions organise comparison before the technology exists. They do not turn a hypothetical operation into an empirical result.

What the theories preserve

IIT 4.0 distinguishes the maximal physical substrate, or complex, from the cause–effect structure it specifies. Matching behaviour does not establish matching that structure. The theory therefore asks about the physical implementation as well as its outputs; it does not automatically settle personal identity over time. [T1]

Global Workspace approaches require care. The Conscious Turing Machine formalises a computational competition-and-broadcast architecture. Changeux and Farisco's 2026 interpretation of Global Neuronal Workspace instead emphasises levels from molecular and cellular mechanisms to large-scale networks. It explicitly resists treating GNW as computational functionalism. That is not itself a demonstration that artificial implementations are impossible. [T2][T3]

George Deane's active-inference account makes subjective valuation and deep self-modelling central. In the proposed benchmark, copying autobiographical data would not by itself establish preservation of those processes, still less numerical identity. The latter is an inference about what the benchmark would require, not a transplantation result reported by Deane. [T4]

Substrate-dependent approaches sharpen the question. MoC7 includes challenges to the physical assumptions inside replacement scenarios and an organism-first account. Mihalik's contribution also shows why Russellian monism supplies no single automatic answer: versions of the view can permit unchanged, altered or absent experience after silicon replacement. [S12]

For Orch OR, the proposed dependence on orchestrated quantum processes and objective reduction means that reproducing a neuron's classical input and output would not alone establish preservation of the required mechanism. This is a conditional application of the theory, not evidence that such a transplant has been attempted. [T5]

A specified comparison

Consider two hypothetical machines implementing the same CTM competition, broadcast, timing and computational state transitions, but with different physical organisation. Under the CTM's computational definition, the same conscious computation is realised twice. IIT does not infer phenomenal equivalence from that description: its physical causal criteria must also be evaluated. [T1][T2]

This is a conditional comparison, not a claim that such a pair has been built. Inspecting the computation and intervening on physical components could check its premises. It would not independently establish which machine has experience. An empirical discriminator would still need a justified connection between a measured difference and consciousness.

The immediate result is narrower but useful: the accounts require different information before they can answer the same question.

The Branch Test

The transplant case leaves one candidate successor. Branching removes that convenience.

Start with one conscious subject, A. A copying operation preserves the original implementation and creates a second candidate, B, with matching memories and computational organisation:

A → original implementation + copied implementation B

This is not physical symmetry. A has an uninterrupted implementation history that B does not. A theory may regard that difference as decisive. A symmetric fission, where the original implementation gives way to two candidates without a privileged physical continuation, would be a different procedure and must be specified separately.

Fission has a long philosophical history. Parfit and subsequent personal-identity literature use cases with two psychologically continuous successors to question whether survival must reduce to one-to-one numerical identity. [L3][L4]

For consciousness science, two questions must remain separate.

First: how many conscious subjects exist after the operation?

Call this $N(t)$. Introducing two candidate implementations does not assume that both are conscious. If a theory concludes that each separately meets its conditions, its answer is two. If it assigns consciousness to only one, it must explain why the other fails those conditions. A theory might instead leave the count undetermined.

Second: what relation does each successor have to A?

Call this $R(A, X)$: a stated continuity or survival relation. If the claim is numerical identity, that stronger claim must be named separately.

For a procedure yielding two conscious subjects, these are different possible accounts of continuity:

- Only one continues A under the proposed criterion.
- Both stand in a survival relation to A, without being numerically identical to each other.
- Neither continues A; two new subjects inherit its information.
- The account specifies no further continuity fact.

In all four cases, **the current count can still be two**. The end of the original and the beginning of two new subjects is not an alternative current count of zero. Uncertainty about identity also need not make the number of current subjects uncertain.

The operation changes what the theory must explain, not the meaning of counting.

Memory alone cannot select a uniquely continuing subject. Both successors may remember A's life, with memories causally inherited from A. Their reports establish something about present self-models and inherited information; they do not independently decide the continuity relation.

Branch therefore adds two columns to the comparison: **subject count now** and **relation to the earlier subject**. A theory may fill the first and explicitly leave the second open. That is more informative than treating every different account of survival as a different prediction about how many subjects exist.

The Merge Test

Suppose a branch procedure yields two conscious subjects, Aa and Ab. They live separately, acquiring different memories, skills and relationships. Later, technology permits progressively deeper exchange of internal states, including memories, learned models and self-representations.

At what point, if any, do two conscious subjects become one?

Luke Roelofs has examined gradual person-fusion through deeper brain interfacing. Split-brain research provides a messier empirical comparison, with competing accounts of how many streams exist and how unified they are. Hoffman, Prakash and Prentner have formally considered the fusion of conscious agents. These precedents establish a serious prior literature, not proof that the proposed operation is feasible. [\[L5\]](#) [\[L6\]](#)[\[L7\]](#)

The central distinction is between communication and subject fusion. Two people exchanging messages remain two people. Synchronising databases does not itself establish a shared experience.

The same two questions from Branch apply:

- **Count:** Does the operation leave two subjects, produce one or yield an outcome the theory cannot determine?
- **Continuity:** If one subject results, how does it relate to Aa and Ab?

A result with one subject could have several histories: continuation of one earlier subject, a proposed survival relation to both, or a new subject inheriting their information. Those accounts share a current count of one. They must not be presented as different counts merely because they disagree about who survived.

The physical comparison asks:

What distinguishes two conscious systems exchanging information from a system supporting one unified experience?

Candidate answers might invoke causal integration, a shared workspace, phenomenal organisation or properties of a living system. These are directions for theory-specific answers, not established fusion criteria. A framework may also deny that the proposed coupling produces one subject.

Giving the resulting system access to both memory histories would establish informational access, not subject fusion. Nor would it settle continuity if there were one subject. The theory needs a criterion for unity, followed by a separate account of continuity if continuity is within its scope.

The inherited histories may conflict. One stream may have formed commitments the other rejects. A shared self-model would need some way of handling that conflict, although ordinary human inconsistency shows that perfect agreement is not required. Resolving those conflicts would still not, by itself, prove the existence of one centre of experience.

From thought experiment to theory comparison

The three cases organise different questions:

Transplant: What supports consciousness through a change of implementation, and what supports continuity?

Branch: How many subjects result, and how does each relate to the earlier subject?

Merge: What could make separate conscious systems one, and what would that imply about their histories?

Their immediate contribution is a comparison protocol. Each theory needs a specified operation, its own relevant invariants, an answer within its scope and an explicit account of what remains unresolved. Where answers differ, the next question is whether any observation could distinguish them. Where observations cannot, the disagreement remains conceptual or ontological.

This fits the methodological task visible in the later MoC programmes. The scheduled work on theory comparison, substrate dependence and cross-system indicators supplies a setting for such questions; it does not mean the conference has adopted this particular suite. [S12]

The proposal can fail too. If the operations remain too underspecified for theories to answer, or the answers merely restate familiar positions without clarifying assumptions, scope or possible observations, combining the cases has added little. A useful next step would be a small comparison in which proponents check that each theory's answer has been represented fairly.

The question after seven conferences

MoC's programmes record formal models alongside growing, explicit attention to how models should be compared. Methodology was present in 2019; by MoC7, methodological and conceptual coordination is also an announced collective project.

That supports a more interesting ambition than expecting one meeting to settle consciousness. Researchers might agree on the operations to consider, the distinctions to preserve and the evidence a disagreement would require.

A consciousness transplant, a branch and a merge make that ambition concrete without settling the answers. A theory of conscious states may legitimately stop short of personal identity. An account of continuity may legitimately acknowledge that no observation currently distinguishes it from a rival.

The useful demand is that neither boundary disappear unnoticed when a theory travels beyond the systems in which it was developed.

After seven conferences, the next shared task may be to make disagreement precise enough that we can tell **what differs, what could be tested, and what remains outside the claim.**

Source notes

Models of Consciousness / AMCS

[S1] Models of Consciousness 2019, official conference site.

<https://models-of-consciousness.org/> (<https://models-of-consciousness.org/>).

[S2] *Models of Consciousness 2019 Conference Booklet*.

https://models-of-consciousness.org/booklet/MoC19_Conference_Booklet.pdf (https://models-of-consciousness.org/booklet/MoC19_Conference_Booklet.pdf).

[S3] AMCS, *Models of Consciousness 2 (2021)*.

<https://amcs-community.org/events/moc2-2021/> (<https://amcs-community.org/events/moc2-2021/>).

[S4] *MoC2 Programme Final (2021)*.

https://amcs-community.org/wp-content/uploads/2021/09/MoC2_Programme_Final.pdf (https://amcs-community.org/wp-content/uploads/2021/09/MoC2_Programme_Final.pdf).

[S5] *MoC3 Conference Booklet (2022)*.

https://amcs-community.org/wp-content/uploads/2022/08/MoC3_Conference-Booklet_2022_v3.pdf (https://amcs-community.org/wp-content/uploads/2022/08/MoC3_Conference-Booklet_2022_v3.pdf).

[S6] *MoC4 Conference Booklet (2023)*.

https://amcs-community.org/wp-content/uploads/2023/08/MoC4_Conference-Booklet_2023_v2.pdf (https://amcs-community.org/wp-content/uploads/2023/08/MoC4_Conference-Booklet_2023_v2.pdf).

[S7] AMCS, *Models of Consciousness 5 (2024)*.

<https://amcs-community.org/events/moc5-2024/> (<https://amcs-community.org/events/moc5-2024/>).

[S9] AMCS, *Models of Consciousness 6 (2025)*.

<https://amcs-community.org/events/moc6-2025/> (<https://amcs-community.org/events/moc6-2025/>).

[S10] AMCS, *MoC6 Schedule Information*.

<https://amcs-community.org/moc6-schedule-information/> (<https://amcs-community.org/moc6-schedule-information/>).

[S11] AMCS, *Models of Consciousness 7* (2026).

<https://amcs-community.org/events/moc7-2026/> (<https://amcs-community.org/events/moc7-2026/>).

[S12] AMCS, *MoC7 Schedule Information and Abstracts*.

<https://amcs-community.org/moc7-schedule-information/> (<https://amcs-community.org/moc7-schedule-information/>).

Transplant, fission and fusion prior art

[L1] David J. Chalmers, *Absent Qualia, Fading Qualia, Dancing Qualia* and related organizational-invariance work.

Gradual replacement and organisational invariance, sections 2–4: <https://consc.net/papers/qualia.html> (<https://consc.net/papers/qualia.html>).

[L2] David J. Chalmers, *The Singularity: A Philosophical Analysis* — uploading, consciousness and personal identity.

<https://consc.net/papers/singularity.pdf> (<https://consc.net/papers/singularity.pdf>).

[L3] Derek Parfit, *Reasons and Persons*, Part Three, especially Chapter 12.

<https://academic.oup.com/book/12484/chapter-abstract/163166818> (<https://academic.oup.com/book/12484/chapter-abstract/163166818>).

[L4] Stanford Encyclopedia of Philosophy, *Personal Identity*.

<https://plato.stanford.edu/entries/identity-personal/> (<https://plato.stanford.edu/entries/identity-personal/>).

[L5] Luke Roelofs, *Combining Minds*, Chapter 8: “What It Is Like for Two to Become One.”

<https://academic.oup.com/book/4164/chapter/145955573> (<https://academic.oup.com/book/4164/chapter/145955573>).

[L6] Donald D. Hoffman, Chetan Prakash & Robert Prentner, “Fusions of Consciousness,” *Entropy* 25(1):129 (2023).

<https://www.mdpi.com/1099-4300/25/1/129> (<https://www.mdpi.com/1099-4300/25/1/129>).

[L7] Stanford Encyclopedia of Philosophy, *The Unity of Consciousness* — split-brain discussion.

<https://plato.stanford.edu/entries/consciousness-unity/> (<https://plato.stanford.edu/entries/consciousness-unity/>).

Theory benchmark sources

[T1] Albantakis et al., “Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms,” *PLOS Computational Biology* (2023). See “Integration,” “Exclusion” and “Consciousness and functional equivalence: Being is not doing,” Figure 8.

<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011465> (<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011465>).

[T2] Lenore Blum & Manuel Blum, Conscious Turing Machine / Global Workspace formalization.

<https://arxiv.org/abs/2107.13704> (<https://arxiv.org/abs/2107.13704>).

<https://arxiv.org/abs/2303.17075> (<https://arxiv.org/abs/2303.17075>).

[T3] Dehaene & Naccache (2001), workspace framework:

<https://www.sciencedirect.com/science/article/pii/S0010027700001232>

(<https://www.sciencedirect.com/science/article/pii/S0010027700001232>).

Jean-Pierre Changeux & Michele Farisco (2026), “The Global Neuronal Workspace as a multilevel model of conscious processing,” *Trends in Cognitive Sciences* 30(6), 477–479, DOI 10.1016/j.tics.2026.03.004. The multilevel distinction above follows its abstract, available in the authors' institutional record: <https://uu.diva-portal.org/smash/record.jsf?pid=diva2:2050630>

([https://uu.diva-portal.org/smash/record.jsf?](https://uu.diva-portal.org/smash/record.jsf?pid=diva2:2050630)

[pid=diva2:2050630](https://uu.diva-portal.org/smash/record.jsf?pid=diva2:2050630)).

Publisher: <https://www.sciencedirect.com/science/article/pii/S1364661326000549>

(<https://www.sciencedirect.com/science/article/pii/S1364661326000549>).

[T4] George Deane, “Consciousness in active inference: Deep self-models, other minds, and the challenge of psychedelic-induced ego-dissolution,” *Neuroscience of Consciousness* 2021(2), niab024. DOI 10.1093/nc/niab024. See abstract and “Consciousness in active inference.”

<https://academic.oup.com/nc/article/2021/2/niab024/6360857>

(<https://academic.oup.com/nc/article/2021/2/niab024/6360857>).

[T5] Hameroff & Penrose, Orch OR formulations.

<https://www.sciencedirect.com/science/article/pii/0378475496804769>

(<https://www.sciencedirect.com/science/article/pii/0378475496804769>).

<https://www.sciencedirect.com/science/article/pii/S1571064513001188>

(<https://www.sciencedirect.com/science/article/pii/S1571064513001188>).

Author overview of the proposed mechanism: <https://hameroff.arizona.edu/research-overview/orch-or>

(<https://hameroff.arizona.edu/research-overview/orch-or>).

— Dennis Hedegreen, trying to see the structure

RELATION MEMORY

NEARBY [Closing the Loop Is Not Waking Up](#) · Separates AI capability and recursive self-improvement from evidence of phenomenal consciousness.

NEARBY [I Don't Know](#) · Distinguishes self-report, causal tracking and ontology; sets the empirical boundary for the proposed comparison.

NEARBY [The Thinking Machine](#) · Provides adjacent machine-mind lineage without replacing the MoC conference evidence base.

CANONICAL URL: [HTTPS://HEDEGREENRESEARCH.COM/ARTICLES/FROM-MODELS-TO-METHODS/](https://hedegreenresearch.com/articles/from-models-to-methods/)