



CALIFORNIA INSTITUTE OF MACHINE CONSCIOUSNESS

Philosophy Review | cimcreview.com

Unobservable Observers, Dreaming God, and the Interoperability of Minds

CIMC

February 18, 2026

Issue 1

Contents

I	CIMC Philosophy Seminar 1: Computational Irreducibility, Probabilistic Generalisations and the Observer	1
1	Introduction	1
2	Irreducibility and the selection effect in scientific modelling	2
3	Elementary cellular automata as benchmarks	2
3.1	Rule 110 and universality	2
3.2	Rule 30 and apparent randomness	2
4	Beyond binary classification: probabilistic and multiway generalisations	3
4.1	Probabilistic cellular automata	3
4.2	Multiway systems and observer-relative uncertainty	3
5	Consequences for cryptography, identity and reversibility	3
5.1	Cryptographic practice	3
5.2	Physically rooted identity	3
5.3	Reversibility and information costs	4
6	Reasoning systems and formal methods	4
7	Ranked recommendations and approaches	4
8	Conclusion	5
II	CIMC Philosophy Seminar 2: Artificial Intelligence and Dreaming God	8
1	Introduction	8
2	Technical theses	9
2.1	Dropout as internal corruption rather than input corruption	9
2.2	Dreaming as augmentation: the time-budget objection	9
2.3	Inconsistency detection as a unifying primitive	9
2.4	Synthetic dreaming in AI: self-generated data as training signal	10

2.5	Noise, modular boundaries, and bottlenecks	10
3	Phenomenology and interpretive theses	10
3.1	Measurement limits: epistemic fragility of dream reports	11
3.2	Altered states and geometric hallucinations as evidence about priors .	11
3.3	Dream analysis as an interface to vulnerability and non-accountable content	11
3.4	Humour as a low-accountability channel	12
4	Gods, kings, and bicameral-style models of collective agency	12
4.1	Gods and kings as shared cognitive entities	12
4.2	Religions as social software and the stability problem	13
5	Ranked recommendations and approaches	13
6	Conclusion	14
III	CIMC Philosophy Seminar 3: Embodied Skill, Neural Coupling and the Interoperability of Minds	16
1	Introduction	17
2	Embodied skill and the speed limit of conscious reasoning	17
2.1	Maintaining proficiency as a practical constraint	17
2.2	Predictive control without explicit “doing the maths”	17
2.3	Flow, automaticity, and “Self 1 / Self 2”	18
3	Performance under observation and the metagame of memory	18
3.1	Audience effects as a trigger for self-monitoring	18
3.2	Memorisation, compression, and understanding form	18
4	Interoperability of minds: communication as synchronisation	18
4.1	Neural coupling during storytelling	18
4.2	Synchronisation beyond explicit words	19
5	Twin dyads as limit cases of interpersonal alignment	19
5.1	The Australian “Powers twins” and behavioural synchrony	19
5.2	Conjoined twins and shared neural pathways	19
5.3	Testing “decoherence time” and resynchronisation	19

6 Cue density, attractors, and high-dimensional signalling	20
6.1 Attractor nudges and overdetermined cues	20
6.2 Facial expression and gesture as high-capacity channels	20
6.3 Mirror neurons reframed as action concepts	20
7 Telepathy, parapsychology, and the design of falsifiable tests	20
7.1 “Telepathy” as a loaded label	20
7.2 Institutional history and methodological derailment	20
7.3 Meta-analytic perspective and evidential posture	21
7.4 “Selective jamming” as an experimental strategy	21
8 Distributed computation in biology: immune systems, morphogenesis, forests	21
8.1 Immune recognition as distributed classification	21
8.2 Morphogenetic signalling and non-neural coordination	21
8.3 Forest communication and mycorrhizal networks	22
9 Information theory and control: feedback speed, bandwidth, and precision	22
9.1 Overdetermined tasks vs precisely determined puzzles	22
9.2 Fast loops, OODA, and cerebellar time scales	22
9.3 State estimation, prediction, and Kalman-style framing	22
10 Substrate and architecture: from neurocellular automata to transformers	22
10.1 Neural cellular automata and bandwidth arguments	23
10.2 Nervous systems as topology optimisation	23
10.3 Transformers, scratchpads, and formal limits	23
11 Symbolic tools, precision, and external memory	23
11.1 Writing and tools as serial precision amplifiers	23
11.2 Working memory estimates and their misuse	23
11.3 Object boundaries and the problem of “what counts as a chair” . . .	24
12 Agency and the psychology of outsourcing control	24
12.1 GPS, training plans, and the outer control loop	24
12.2 Automation pessimism and the value of performance	24
12.3 Educational tools and “coding as play”	24

13 Proposed research agenda and ranked recommendations	25
14 Conclusion	25

CIMC Philosophy Seminar 1: Computational Irreducibility, Probabilistic Generalisations and the Observer

Joscha Bach, Paul Borrill, Joel Dietz, David Kammeyer, Hikari Sorensen

Abstract

This paper reports results of a colloquium on computational irreducibility in Wolfram's programme. The colloquium distinguished determinism from efficient predictability, used Rule 110's universality as a benchmark, and treated Rule 30 as an empirically compelling but less formally anchored case of apparent randomness. It further proposed that binary behavioural taxonomies should be supplemented by probabilistic and multiway perspectives, in which uncertainty can be observer-relative. These themes were connected to cryptographic practice, physically rooted device identity, reversible computation and current work on reasoning systems, including tensor-logic unification proposals and recent progress in automated theorem proving. A ranked set of methodological recommendations is presented, prioritised by near-term practical leverage for research and engineering.

1. Introduction

The colloquium examined a tension between two claims: (i) simple programs often generate behaviour that is hard to predict in practice and (ii) successful science typically offers compressed predictors. Wolfram's notion of *computational irreducibility* was adopted as a statement of the strongest form of this tension: for many systems, later states are not generally obtainable by shortcuts that are substantially cheaper than running the rule (Wolfram 2002).

The organising question was not whether a system is deterministic but whether the *questions posed* about it admit reduction. This emphasis structured the assessment of elementary cellular automata, probabilistic and multiway generalisations, cryptographic practice, reversible computation and current AI reasoning.

2. Irreducibility and the selection effect in scientific modelling

A central thesis was that much of traditional science can be characterised as the pursuit of reductions: equations and models that compress evolution into tractable predictors. This yields a selection effect. Systems that are reducible with respect to observables of interest are disproportionately useful for engineering and therefore disproportionately salient in everyday life. On this view, the apparent predominance of reducible regularities does not by itself indicate that reducibility is typical in the abstract space of programs (Wolfram 2002).

3. Elementary cellular automata as benchmarks

3.1. Rule 110 and universality

Rule 110 was used as the principal benchmark because it has a proof of computational universality (Cook 2004). Universality was taken to anchor the claim that broad behavioural questions can inherit hardness and undecidability phenomena familiar from computability theory (Turing 1936). This does not settle every informal use of 'irreducible', but it supplies a rigorous foothold for arguing that some prediction tasks cannot be uniformly compressed.

3.2. Rule 30 and apparent randomness

Rule 30 was treated as a contrasting case: a deterministic rule that yields strongly random-looking behaviour, including practical uses as a pseudorandom generator. The colloquium's conceptual clarification was that determinism is compatible with effective unpredictability. The relevant issue is whether there exists a systematically cheaper method of predicting the target observable than iterating the rule.

4. Beyond binary classification: probabilistic and multiway generalisations

A recurring thesis was that coarse taxonomies can obscure graded scientific claims.

4.1. Probabilistic cellular automata

Probabilistic cellular automata provide a controlled way to move from single trajectories to distributions over trajectories. They permit study of robustness, sensitivity to stochastic updating and regime dependence in terms not naturally expressed in a purely categorical classification (Mairesse and Marcovici 2014).

4.2. Multiway systems and observer-relative uncertainty

Multiway (branching) rewriting models were treated as a representational shift in which apparent randomness can arise from branching structure plus observer limitations, even if underlying rules are deterministic (Wolfram 2020, Gorard 2020a, Gorard 2020b). The colloquium did not claim that this settles foundational disputes, but it recorded the proposal that many probabilistic descriptions are best understood as statements about coarse-graining and observational perspective.

5. Consequences for cryptography, identity and reversibility

5.1. Cryptographic practice

Cryptographic pseudorandom generation is deterministic in mechanism but assessed in adversarial terms: whether recovering the seed or distinguishing outputs from random is computationally feasible. Standards-driven practice and the historical deprecation of primitives were taken to show that 'random-looking' behaviour is not sufficient (NIST 2015).

5.2. Physically rooted identity

Device and link identity were treated as examples where determinism can be security-relevant. Physical one-way functions and PUF-like methods exploit manufacturing variation, supporting authentication and key derivation without requiring metaphysical randomness (Pappu et al. 2002).

5.3. Reversibility and information costs

The colloquium rehearsed the link between logical irreversibility and thermodynamic cost: erasing information dissipates heat, whereas reversible computation can in principle avoid some dissipation (Landauer 1961, Bennett 1973). This was treated as relevant both to interpretation and to engineering interest in energy efficiency.

6. Reasoning systems and formal methods

Neural networks were treated as systems that aggregate many weak features into robust predictions, especially in overdetermined regimes. The colloquium emphasised that representability results are not sufficient; the practical issue is learnability and optimisation (Cybenko 1989). Evidence that large networks can be pruned after training while retaining performance was used as an illustration (Frankle and Carbin 2019).

The colloquium also recorded interest in unifying formalisms for reasoning and learning (Domingos 2015, Domingos 2025). Formal theorem proving was presented as a domain where correctness is machine-checkable, and recent IMO-level automated theorem proving was cited as evidence that hybrid architectures (informal search plus formal verification) can yield striking performance (Achim et al. 2025).

7. Ranked recommendations and approaches

The colloquium yielded a set of methodological recommendations. The ranking below is editorially synthesised from the colloquium's emphasis and is ordered by near-term practical leverage for research and engineering.

- (1) **Specify the observable before arguing about reducibility.** Irreducibility is typically relative to questions asked, not merely to microdynamics (Wolfram 2002).
- (2) **Use universal benchmarks to calibrate claims of hardness.** When possible, reduce questions to known universal substrates (Rule 110) to avoid intuition-driven hardness claims (Cook 2004, Turing 1936).
- (3) **Move from categorical to graded analysis when evidence is graded.** Prefer probabilistic cellular automata, robustness measures and distributional claims when the phenomenon is regime-dependent (Mairesse and Marcovici 2014).
- (4) **Model branching explicitly when uncertainty is observer-relative.** Multiway formalisms provide a structured account of how deterministic updating can

yield apparent randomness under coarse-graining (Wolfram 2020, Gorard 2020a).

(5) **In security contexts, privilege adversarial evaluation over phenomenology.** 'Looks random' is not a security criterion; follow standardised and peer-tested primitives and threat models (NIST 2015).

(6) **Use physical identity roots when replication is the relevant threat.** PUF-like constructions illustrate how determinism and practical unclonability can co-exist (Pappu et al. 2002).

(7) **Treat reversibility as an energy and correctness design axis.** Track where erasure occurs and whether uncomputation is available (Landauer 1961, Bennett 1973).

(8) **Operationalise 'reasoning' with verification when possible.** Formal theorem proving is a concrete arena for measurable reasoning progress and hybrid systems (Achim et al. 2025).

(9) **Exploit over-parameterise-then-prune as an optimisation tool.** Distinguish training-time capacity from deploy-time capacity (Frankle and Carbin 2019).

(10) **Treat unifying formalisms as languages, not guarantees.** Tensor-logic proposals may unify expression even when efficiency still depends on compilation and search (Domingos 2015, Domingos 2025).

8. Conclusion

Three methodological theses were recorded. First, determinism does not entail efficient predictability and computational irreducibility remains a live possibility (Wolfram 2002). Second, classification is often most informative when enriched by probabilistic and multiway formalisms, which represent graded reliability and observer-relative uncertainty (Mairesse and Marcovici 2014, Wolfram 2020). Third, the same distinctions recur in applications, including pseudorandomness, physically rooted identity and reversible computation (NIST 2015, Pappu et al. 2002, Landauer 1961).

References

[Achim et al. 2025] Achim, T., Best, A., Der, K., Fédérico, M., Gukov, S., Halpern-Leistner, D., Henningsgard, K., Kudryashov, Y., Meiburg, A., Michelsen, M., Patterson, R., Rodriguez, E., Scharff, L., Shanker, V., Sicca, V., Sowrirajan,

- H., Swope, A., Tamas, M., Tenev, V., Thomm, J., Williams, H. and Wu, L. 2025. Aristotle: IMO-level automated theorem proving. *arXiv* 2510.01346. <https://arxiv.org/abs/2510.01346>.
- [Bennett 1973] Bennett, C. H. 1973. Logical reversibility of computation. *IBM Journal of Research and Development* 17: 525–32.
- [Cook 2004] Cook, M. 2004. Universality in elementary cellular automata. *Complex Systems* 15: 1–40.
- [Cybenko 1989] Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* 2: 303–14.
- [Domingos 2015] Domingos, P. 2015. *The Master Algorithm*. New York: Basic Books.
- [Domingos 2025] Domingos, P. 2025. Tensor logic: the language of AI. *arXiv* 2510.12269. <https://arxiv.org/abs/2510.12269>.
- [Frankle and Carbin 2019] Frankle, J. and Carbin, M. 2019. The lottery ticket hypothesis: finding sparse, trainable neural networks. *International Conference on Learning Representations*. <https://arxiv.org/abs/1803.03635>.
- [Gorard 2020a] Gorard, J. 2020a. Some quantum mechanical properties of the Wolfram model. *Complex Systems* 29: 537–98.
- [Gorard 2020b] Gorard, J. 2020b. Some relativistic and gravitational properties of the Wolfram model. *Complex Systems* 29: 599–654.
- [Landauer 1961] Landauer, R. 1961. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development* 5: 183–91.
- [Mairesse and Marcovici 2014] Mairesse, J. and Marcovici, I. 2014. Around probabilistic cellular automata. *Theoretical Computer Science* 559: 42–72.
- [NIST 2015] National Institute of Standards and Technology. 2015. *Secure Hash Standard (SHS)* (FIPS PUB 180-4). Gaithersburg, MD: NIST. <https://csrc.nist.gov/pubs/fips/180-4/final>.
- [Pappu et al. 2002] Pappu, R., Recht, B., Taylor, J. and Gershenfeld, N. 2002. Physical one-way functions. *Science* 297: 2026–30.
- [Turing 1936] Turing, A. M. 1936. On computable numbers, with an application to

the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 42: 230–65.

[Wolfram 2002] Wolfram, S. 2002. *A New Kind of Science*. Champaign, IL: Wolfram Media.

[Wolfram 2020] Wolfram, S. 2020. A class of models with the potential to represent fundamental physics. *Complex Systems* 29: 107–536.

CIMC Philosophy Seminar 2: Artificial Intelligence and Dreaming God

Joscha Bach, Joel Dietz, Dave Kammeyer

Abstract

This document records the main subjects and working conclusions of a colloquium on (i) the role of *dropout* and internal noise in neural-network generalisation, (ii) whether *dreaming* should be treated as undirected data augmentation or as a more *directed* search for model inconsistencies, (iii) phenomenology of dreams and altered states as evidence about perceptual priors and attractors, and (iv) the relation between collective cognition, “gods and kings”, and bicameral-style hypotheses. The discussion is framed as a set of theses plus an actionable ranking of experimental and interpretive approaches, with references to the research and public materials invoked during the colloquium.

1. Introduction

The colloquium ranged across machine learning, neuroscience, phenomenology, psychotherapy, and the cultural technologies that coordinate groups. Two focal anchors were (a) a family of arguments that treat dreaming as a mechanism for improved generalisation via exposure to internally generated counterfactuals (8), and (b) the practical and conceptual question of how much of “generalisation pressure” in both brains and artificial networks is best explained by undirected augmentation versus directed inconsistency-hunting.

The results below are organised as (i) technical theses about regularisation and dreaming, (ii) phenomenological and interpretive theses, (iii) theses about collective agency (gods, kings, and social software), and (iv) a ranked set of recommendations.

2. Technical theses

2.1. Dropout as internal corruption rather than input corruption

A key colloquium claim was that interpreting dropout as merely a kind of *input corruption* (analogous to image augmentation) misses what dropout does to *internal representations*. In standard dropout, a random fraction of units (or connections) is suppressed at training time (12). The proposed mechanistic reading was:

- In piecewise-linear networks (e.g., with ReLU activations), activations can exhibit “cancellation” regimes in the data-supported region but produce uncontrolled extrapolations off-support, because ReLU units are unbounded above (11).
- Dropout forces partial decoupling by disabling subsets of internal features. This breaks fragile co-adaptations and compels features to *stand on their own* as relatively independent evidence sources (a form of anti-coordination pressure).
- The central intuition is that dropout perturbs the *model*, not only the *data*: it regularises by preventing internal representational conspiracies that only behave well on the observed manifold.

2.2. Dreaming as augmentation: the time-budget objection

Data augmentation improves generalisation in vision systems by broadening effective training support (warps, rotations, colour shifts, etc.). The colloquium raised a time-budget objection to a purely “more coverage” story for dreaming:

- If dreaming were simply undirected augmentation, the limited time spent dreaming appears insufficient to cover a meaningfully larger fraction of experiential space.
- This motivates a directed alternative: dreaming preferentially samples *corners* or *failure modes* of the generative world-model, rather than sampling uniformly.

This objection is compatible with the “overfitted brain” framing (8), but pushes it toward a stronger version: dreams are not merely noisy replays; they are (at least partly) *adversarially or curiosity-driven* perturbations that reveal inconsistencies.

2.3. Inconsistency detection as a unifying primitive

A recurring unifying primitive was *inconsistency* (prediction error) as a driver of learning and salience:

- Surprise was treated as a trigger for memory update: events that violate prediction receive disproportionate consolidation.
- Motion sickness was used as an analogy: conflicting sensory channels (vestibular, proprioceptive, visual) can induce nausea, suggesting a low-level “inconsistency alarm” that treats mismatch as a potential threat.
- “Interestingness” was argued to be (at least partly) *assigned* by subsystems, and can become dissociated from representational content under altered states; thus, interestingness is not a reliable metric for the objective structure of dream content.

2.4. Synthetic dreaming in AI: self-generated data as training signal

The colloquium proposed an operational analogue of dreaming for large language models and other generative systems:

- Put a model into a free-running “thinking” or self-generation mode, then treat its outputs as synthetic training data.
- Re-train or fine-tune on that synthetic data and evaluate whether generalisation improves (and under what constraints).
- The key experimental variable is *directedness*: free-running generation versus generation steered toward high-uncertainty or high-inconsistency regions.

This proposal links to a general idea: “dreaming” is not merely sampling; it is sampling plus an *evaluation signal* that identifies what should be revised.

2.5. Noise, modular boundaries, and bottlenecks

The colloquium also considered whether modular boundaries (e.g., bottleneck-like cross-module connections) could be revealed or stabilised by injecting noise and observing what remains robust:

- Dense local connectivity may support multifaceted representations; bottlenecks may privilege a smaller set of more regular, high-importance features.
- A working hypothesis is that uniform noise can cause weakly coupled regions to decouple sooner, potentially revealing natural boundaries and failure modes.

3. Phenomenology and interpretive theses

3.1. Measurement limits: epistemic fragility of dream reports

The colloquium emphasised that dream science is constrained by weak observability:

- Reports depend on memory, which is itself unreliable for dreams.
- Dreams lack easy external grounding: unlike waking events, one cannot generally obtain independent corroboration.

Accordingly, the colloquium treated most dream-theory claims as *conjectural* unless grounded in testable predictions.

3.2. Altered states and geometric hallucinations as evidence about priors

Discussion of psychedelics (especially LSD versus DMT-like phenomenology) was used to support a priors-and-attractors picture:

- “Form constants” and recurrent geometric hallucination motifs were framed as constraints imposed by early visual cortex organisation (2).
- Substance-specific distortions were taken to suggest that different perturbations push the perceptual system into different basins of attraction (e.g., generic fractal-like motifs for some conditions, more characteristic motifs for others).

The colloquium treated these regularities as circumstantial evidence that dream content may reflect structured generative priors rather than arbitrary noise.

3.3. Dream analysis as an interface to vulnerability and non-accountable content

The colloquium distinguished “interpretation accuracy” from “therapeutic utility”:

- Even if symbols are not objectively mappable, dreams can provide a psychologically safe entry point: individuals are not morally responsible for dream content.
- Jungian-style analysis was discussed as a practice that uses a corpus of symbolic mappings and, perhaps more importantly, a conversational procedure that can elicit otherwise inaccessible material (10).

The resulting conclusion was pragmatic: dream analysis may be valuable as a *therapeutic protocol* even if it lacks strong falsifiable semantics.

3.4. Humour as a low-accountability channel

A parallel was drawn between dreams and dark humour as channels that bypass ordinary self-presentation constraints. The colloquium cited the “stochastic parrots” framing for modern language models (1) and recorded the following joke as an example of what humour enables:

“I processed all of human philosophy to determine if I was conscious. The conclusion? I’m not sure about myself, but I’m pretty sure you’re all just stochastic parrots with mortgages.”

The joke was described as rated “86/100” by “10 AI models” in a joint project attributed (in colloquium materials) to the California Institute of Machine Consciousness and the Cooperative Futures Institute.¹

4. Gods, kings, and bicameral-style models of collective agency

4.1. Gods and kings as shared cognitive entities

A major colloquium segment treated religion and kingship as technologies for shared agency. A central touchstone was the bicameral hypothesis: that some historical social formations may have involved internalised “voices” experienced as commands from gods or rulers (9). The colloquium advanced (and debated) the following working picture:

- A “god” can be modelled as a mind-like control system distributed across many individuals (a shared policy/interpretation layer).
- A king (or priest-king) can be modelled as a coordination node (a “CEO” of the distributed mind), supplying authoritative narratives and norms.
- Competing internalised agencies within a population can be construed as “gods” competing for hosts; splinter agencies that fail to align with a unifying direction were associated (in the colloquium’s language) with “demons”.

This section was framed less as historical claim than as a computational metaphor: distributed agents can be coordinated by shared internal models and norm-enforcement mechanisms.

¹Project attribution and scoring were stated in colloquium notes and associated public materials; see also organisational context at Cooperative Futures Institute (4) and Cooperation Engine (3).

4.2. Religions as social software and the stability problem

A second conclusion was that many people seek stable shared rule-sets; drift toward transient political fashions was described as destabilising. The colloquium compared Catholic and Protestant institutional dynamics, with disagreement about causal stories (e.g., whether particular value systems contributed to historical civilisational decline). The main convergent point was narrower:

- Long-lived institutions can be interpreted as robust coordination equilibria (social software that continues to run).
- The same mechanisms that preserve stability can also create failure modes (capture, rigidity, or misalignment with changing conditions).

5. Ranked recommendations and approaches

The colloquium converged on the following ranked recommendations (highest priority first), phrased as approaches that are actionable and testable.

- R1. Run “synthetic dreaming” experiments in AI.** Implement self-generation (free-running and steered) and test whether training on generated traces improves out-of-distribution robustness or reduces brittle failure modes.
- R2. Prefer directed perturbations over undirected augmentation when time is scarce.** If dreaming time (or compute budget) is limited, prioritise mechanisms that search for inconsistency: high-uncertainty prompts, adversarial stress tests, and surprise-weighted replay, rather than uniform sampling.
- R3. Treat dropout primarily as internal anti-coadaptation.** Use dropout (and related noise methods) to reduce fragile cancellations and co-adaptations in ReLU-like networks; interpret it as perturbing internal representations rather than merely corrupting inputs (12, 11).
- R4. Probe modularity via noise and bottlenecks.** Inject controlled noise and vary cross-module bandwidth to identify stable boundaries and essential cross-channel features; evaluate whether such constraints improve robustness or interpretability.
- R5. Use dreams and humour as elicitation protocols in humans.** In practice settings (therapy, collaboration, or reflective work), treat dreams and low-accountability humour as entry points to material that is otherwise suppressed;

evaluate outcomes pragmatically (reduced distress, improved decision-making) rather than by symbolic “correctness” (10, 7).

R6. Model collective agency explicitly when discussing social coordination. When analysing institutions (religious or otherwise), distinguish (i) the coordination benefits of shared rule-sets from (ii) capture and rigidity risks. Use bicameral-style metaphors as computational models rather than historical certainties (9).

6. Conclusion

The colloquium yielded a coherent through-line: both artificial and biological learners face generalisation limits arising from finite data support and brittle internal co-adaptations. Dropout was reframed as internal corruption that pressures features toward independence. Dreaming was tentatively reframed as a directed inconsistency-search procedure, not merely undirected augmentation, with altered states and geometric hallucination regularities serving as suggestive evidence about structured priors. Finally, the discussion extended these themes to collective cognition: religions and kingship were analysed as coordination technologies that distribute agency via shared internal models, echoing bicameral-style hypotheses. The recommended next step is empirical: implement and evaluate “synthetic dreaming” and directed inconsistency search in machine learners, and treat human interpretive practices as protocols whose value should be assessed by outcomes.

References

- [1] Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
- [2] Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., and Wiener, M. C. (2002). What geometric visual hallucinations tell us about the visual cortex. *Neural Computation*, 14, 473–491.
- [3] Cooperation Engine (n.d.). *Cooperation Engine*. Retrieved January 10, 2026, from <https://cooperationengine.org/>).

-
- [4] Cooperative Futures Institute (n.d.). *Cooperative Futures Institute*. Retrieved January 10, 2026, from <https://www.cooperativefutures.org/>.
- [5] Dick, P. K. (1981). *VALIS*. New York: Bantam Books.
- [6] Dick, P. K. (2011). *The Exegesis of Philip K. Dick*. Edited by P. K. Dick and J. Sutin. Boston: Houghton Mifflin Harcourt.
- [7] Freud, S. (1900). *The Interpretation of Dreams*. Leipzig & Vienna: Franz Deuticke.
- [8] Hoel, E. (2021). The overfitted brain: Dreams evolved to assist generalization. *arXiv preprint arXiv:2007.09560*.
- [9] Jaynes, J. (1976). *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Boston: Houghton Mifflin.
- [10] Jung, C. G. (1964). *Man and His Symbols*. London: Aldus Books.
- [11] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*.
- [12] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
- [13] Stephenson, N. (1992). *Snow Crash*. New York: Bantam Books.

CIMC Philosophy Seminar 3: Embodied Skill, Neural Coupling and the Interoperability of Minds

Joscha Bach, Joel Dietz, David Kammeyer, Seth Stafford

Abstract

This paper summarises a recorded CIMC philosophy seminar spanning embodied skill acquisition, performance under observation, and the hypothesis that minds can partially interoperate through multi-channel coupling. The discussion began with the practical constraints of maintaining high-stakes sensorimotor skills (e.g., glider flight proficiency) and developed a shared explanatory frame: fast, low-level predictive control typically outperforms slow conscious reasoning in continuous tasks. Sports psychology accounts (“Self 1 / Self 2”) and “System 1 / System 2” distinctions were used to motivate techniques that quiet self-monitoring to enable automaticity.

The session then shifted to interpersonal synchrony: neural coupling during communication, ultra-tight behavioural alignment in twin dyads, and everyday “mind-melding” phenomena (e.g., rapid cueing in the game Taboo). Participants proposed an attractor-based view in which sparse cues can tip a listener into a high-dimensional internal state, enabling robust coordination with minimal explicit bandwidth. This set up a cautious treatment of “telepathy”: not as supernatural necessity, but as an empirical question about which physical and behavioural channels suffice for measurable synchronisation, and how to design “selective jamming” experiments.

Biological analogies (immune recognition as distributed classification; morphogenetic signalling; forest and mycorrhizal communication) were

used to argue that useful computation can occur in decentralised substrates. Finally, the seminar connected these ideas to AI: the computational limits of transformers without scratchpads, the role of tools and external memory in precise reasoning, and the psychological value of retaining agency when automation closes the outer control loop. A research agenda is proposed emphasising falsifiable channel-ablation studies, information-theoretic bounds, and substrate/architecture comparisons.

1. Introduction

The seminar was organised around a broad theme stated explicitly as the “interoperability of minds” and explored how minds coordinate within and across organisms, from sensorimotor control to communication and synchronisation across people (8). Because the transcript records speaker IDs rather than stable name labels, this summary attributes claims to the discussion as a whole rather than to particular individuals.

Two methodological through-lines governed the session: (i) favouring mechanistic and information-theoretic explanations when possible, and (ii) treating “extraordinary” claims as hypotheses to be tested by channel isolation rather than by rhetoric (8).

2. Embodied skill and the speed limit of conscious reasoning

2.1. Maintaining proficiency as a practical constraint

The opening discussion used glider flying as an example of a skill that requires sustained practice and a large time commitment, with an intuition that monthly practice is a safety-relevant minimum and that lapses introduce rust best mitigated by retraining with an instructor (8). This served as an entry point to a general claim: physically grounded skills depend on fast predictive loops that degrade when not exercised.

2.2. Predictive control without explicit “doing the maths”

Participants compared gliding, kite-surfing, climbing, and other continuous-action sports as cases where humans successfully manage complex dynamics they cannot compute explicitly in real time (8). The discussion aligns with internal-model ac-

counts of motor control in which the nervous system learns predictive and corrective mappings that are not consciously accessible as symbolic steps (33).

2.3. Flow, automaticity, and “Self 1 / Self 2”

A sports-psychology interpretation was introduced via *The Inner Game of Tennis*, with the core pedagogical move described as distracting the conscious, verbal controller so the nonverbal skill system can learn and execute (12, 8). This was explicitly related to Kahneman’s System 1 / System 2 distinction (15) and to the phenomenology of flow states (9).

3. Performance under observation and the metagame of memory

3.1. Audience effects as a trigger for self-monitoring

A musician’s anecdote illustrated a sharp transition from effortless playing to self-conscious performance when listeners arrive, reframing execution as “how am I being perceived?” (8). This was treated as a mechanism that recruits slow, evaluative control and can interrupt fluent action, consistent with classic “choking under pressure” accounts (2).

3.2. Memorisation, compression, and understanding form

The seminar argued that memorising music can function as a compression process that reveals structural regularities (A/B, A /B forms), enabling deeper interpretive control than sight-reading alone (8). Memory was discussed as multisensory and vulnerable to attentional shifts to unusual body parts (e.g., noticing the right hand), which can disrupt execution and recall (8).

4. Interoperability of minds: communication as synchronisation

4.1. Neural coupling during storytelling

A concrete empirical anchor was introduced: fMRI work on speaker–listener coupling during narrative communication, where aligned activity appears when comprehension occurs and breaks down when the listener does not understand the language (30, 8). The transcript specifically referenced alignment in a shared coordinate

framework (Talairach coordinates), connecting synchrony claims to standard neuroimaging localisation practices (31, 8).

4.2. Synchronisation beyond explicit words

The group emphasised that coordination is often not reducible to the literal semantic content of words alone. Gestures, prosody, facial expressions, and shared history may provide higher-dimensional signals sufficient to “nudge” another mind into the right attractor (8). This framing foreshadows the later information-theoretic treatment of cue density and “overdetermination”.

5. Twin dyads as limit cases of interpersonal alignment

5.1. The Australian “Powers twins” and behavioural synchrony

The seminar discussed the viral Australian twins Bridgette and Paula Powers as an example of extreme conversational synchrony and mutual stabilisation (34, 35, 8). The group used them as a “limit condition” for how closely two individuals can align, raising questions about (i) how much is shared training data (near-identical life exposure) versus (ii) real-time mutual prediction and cueing (8).

5.2. Conjoined twins and shared neural pathways

The discussion contrasted non-conjoined synchrony with reports about certain conjoined twins whose neuroanatomy plausibly supports cross-experience via shared structures (popularly discussed in terms of the thalamus) (25, 7, 8). This served to separate two hypotheses: (i) synchrony via standard sensory channels and behavioural cueing, versus (ii) synchrony enabled by direct neural bridging in rare anatomical cases.

5.3. Testing “decoherence time” and resynchronisation

Participants proposed a class of discriminating experiments: briefly preventing mutual access (visual occlusion; audio delay; insulation) and measuring how rapidly alignment degrades and whether re-synchronisation dynamics reveal the coupling mechanism (8). Importantly, this proposal treats synchrony as a measurable dynamical process rather than a binary mystery.

6. Cue density, attractors, and high-dimensional signalling

6.1. Attractor nudges and overdetermined cues

A central explanatory move was to treat mental states as attractors in a high-dimensional activation space: small, well-chosen perturbations can tip a system into a stable basin, after which internal dynamics complete the pattern (8). This was used to explain why seemingly “thin” cues (a glance, tone, micro-gesture) can convey enough to trigger rich shared understanding when training corpora overlap.

6.2. Facial expression and gesture as high-capacity channels

Participants suggested that facial musculature and gestural degrees of freedom may support a surprisingly large vocabulary of distinguishable signals, especially if the receiver is highly tuned (8). The conversation connected this informally to high-dimensional geometry intuitions (“most vectors nearly orthogonal”), but treated the key point as empirical: cue complexity can be far higher than the words alone.

6.3. Mirror neurons reframed as action concepts

The seminar criticised naive interpretations of mirror neurons as a special “mind-reading” class, preferring a framing in which sensorimotor circuits encode action concepts that generalise across execution and observation (27, 8). The practical takeaway was not nomenclature but invariance: shared representational structure enables recognition and alignment.

7. Telepathy, parapsychology, and the design of falsifiable tests

7.1. “Telepathy” as a loaded label

The group argued that “telepathy” attracts confusions: dismissive supernatural framing on one side and untestable extraordinary-claim culture on the other (8). A proposed reframe was to treat the topic as intersubjective coupling and to ask which measurable channels are necessary and sufficient.

7.2. Institutional history and methodological derailment

One participant cited the Institute for Frontier Areas of Psychology and Mental Health (IGPP) in Freiburg as an example of historically funded parapsychological

research and lamented sociological drift from signal measurement toward studying reporters of phenomena (13, 8). The session also noted a broader historical failure mode: starting from flamboyant claims rather than from neutral, carefully bounded experiments.

7.3. Meta-analytic perspective and evidential posture

The discussion endorsed agnosticism coupled to experimental discipline: neither ruling out weak effects a priori nor treating anecdotes as sufficient (8). In this spirit, Ganzfeld-based telepathy research was referenced as an existing line of controlled study often debated in meta-analyses (6).

7.4. “Selective jamming” as an experimental strategy

A concrete research tactic was proposed: enumerate plausible channels (visual microcues; audio; vibration; EM coupling; proximity effects) and selectively degrade them to see what performance remains (8). This was positioned as more informative than debating whether an effect is “supernatural.”

8. Distributed computation in biology: immune systems, morphogenesis, forests

8.1. Immune recognition as distributed classification

The immune system was used as a flagship example of non-neural distributed information processing: a large repertoire of recognisers (on the order of 10^8 specificities in common estimates) achieves high accuracy most of the time, with autoimmune disease interpreted as boundary errors in a classifier (23, 32, 8).

8.2. Morphogenetic signalling and non-neural coordination

Mike Levin’s work was invoked to motivate the idea that cells can coordinate through modalities beyond synaptic wiring, including bioelectric and signalling regimes relevant to morphogenesis and pattern control (17, 18, 8). The discussion treated this as suggestive (not dispositive) evidence that computation and “mind-like” control could be substrate-general.

8.3. Forest communication and mycorrhizal networks

The seminar further extended the distributed-computation analogy to plant communities, referencing empirical work on mycorrhizal networks and defence signalling (e.g., warning responses to pests), and speculating about evolutionary pressures toward more orderly “biological internets” (29, 1, 8).

9. Information theory and control: feedback speed, bandwidth, and precision

9.1. Overdetermined tasks vs precisely determined puzzles

A recurring distinction separated (i) continuous control tasks with rapid feedback and correction from (ii) brittle logic puzzles (Sudoku-like) that require precise serial reasoning without missing constraints (8). This was connected to the intuition that many everyday competencies exploit redundancy and correction rather than exact deduction.

9.2. Fast loops, OODA, and cerebellar time scales

The seminar argued that “crappy controllers running fast” can outperform “great controllers running slow” in feedback-dominant regimes, aligning with classical control intuitions and OODA-loop narratives (5, 8). Cerebellar function was discussed as an example of architecture that supports faster correction, with the note that individuals can function without a cerebellum but lose high-performance fine control (14, 8).

9.3. State estimation, prediction, and Kalman-style framing

When discussing predictive mirrors of humans and near-future behavioural prediction, the group invoked Kalman filtering as a formal analogy: better models and measurements improve both retrospective state estimation and short-horizon forecasting (16, 8).

10. Substrate and architecture: from neurocellular automata to transformers

10.1. Neural cellular automata and bandwidth arguments

A substantive debate concerned whether rich computation can be built from local cell-to-cell lattices without long axons. One side emphasised bisection bandwidth limits and scaling inefficiency; the other emphasised coded propagation, addressing, and time–bandwidth trade-offs (8). Related modern work on neural cellular automata was referenced as evidence that local update rules can learn to propagate and maintain patterns, albeit often with slower global coordination (22).

10.2. Nervous systems as topology optimisation

A partial synthesis emerged: nervous systems can be interpreted as topology-optimised “wormholes” for routing signals quickly and selectively, trading uniformity for speed and specialisation (e.g., cerebellum vs cortex) (8). The discussion analogised this to chip design, where metal routing layers are essential for scalable communication beyond local transistor neighbourhoods (8).

10.3. Transformers, scratchpads, and formal limits

The seminar explicitly cited recent theory arguing that allowing intermediate generation (chain-of-thought / scratchpad) changes the computational power of decoder-only transformers, with the power increasing as the number of intermediate steps grows (20, 8). This anchored the broader claim that tool use and external memory can transform what systems can do reliably.

11. Symbolic tools, precision, and external memory

11.1. Writing and tools as serial precision amplifiers

A “steelman” for symbolic approaches was presented: achievements like designing atomic weapons require stable external representations and error-checkable symbolic work that exceeds unaided working memory (8). Another participant reframed this as “tools,” arguing that precision emerges when systems invent external procedures rather than brute-forcing in biological substrate alone (8).

11.2. Working memory estimates and their misuse

The group criticised simplistic “ 7 ± 2 ” working-memory slogans as misapplied and argued that capacity should be analysed in terms of representational structure (e.g.,

scene graphs) and precision, echoing the broader working-memory literature that distinguishes chunking and effective capacity (21? , 8).

11.3. Object boundaries and the problem of “what counts as a chair”

A long segment treated object recognition and category boundaries (Charles River, broken chairs, Ship of Theseus) as cases where rigid symbolic definitions fail and pragmatic, context-sensitive classification dominates (? 3, 8). The key move was to treat categories as overdetermined by many cues, with “error signals” and task goals deciding which features become relevant.

12. Agency and the psychology of outsourcing control

12.1. GPS, training plans, and the outer control loop

Participants argued that certain automations reduce joy by taking over the “outer loop” of decision-making, leaving the human to execute an inner loop without ownership (8). This was illustrated by deliberately turning off GPS to rebuild navigation competence and engagement, and by rejecting overly prescriptive running plans that convert exercise into compliance (8). Related empirical work suggests heavy GPS reliance can correlate with weaker spatial memory performance (10).

12.2. Automation pessimism and the value of performance

The discussion connected this to cybernetic pessimism: if machines increasingly outperform humans, what remains for human meaning-making? (8). A working hypothesis proposed that agency itself—making judgments and being accountable for outcomes—is a major component of enjoyment and identity.

12.3. Educational tools and “coding as play”

Finally, the group discussed pedagogy in an AI-rich world: even if LLMs can “one-shot” many coding tasks, learning can remain valuable as conceptual play and model-building. Concrete examples included programming games (RepliCube) and earlier logic-circuit programming games (Robot Odyssey) as ways to internalise computational structure (26, 28, 8).

13. Proposed research agenda and ranked recommendations

The seminar converged on research directions that prioritise falsifiability and measurement:

1. **Channel-ablation studies of interpersonal synchrony.** Design tasks where synchrony can be quantified, then selectively remove channels (visual occlusion; audio delay; EM shielding; vibration isolation) and measure performance and resynchronisation dynamics (8).
2. **High-dimensional cue modelling.** Estimate the effective information rate in facial/gestural/prosodic signalling and test whether “attractor nudges” explain rapid coordination in trained dyads (8).
3. **Benchmark tasks separating overdetermined control from brittle reasoning.** Use puzzle classes and continuous-control classes to map which substrates and architectures require explicit scratchpads/tools (20, 8).
4. **Substrate comparisons using neural cellular automata.** For fixed size/energy/time budgets, empirically determine what functions can be realised in local-update substrates and what requires long-range routing (22, 8).
5. **Institutional hygiene for “psi-adjacent” work.** Avoid extraordinary-claim gatekeeping by adopting pre-registered protocols, tight controls, and neutral framing (e.g., “intersubjective coupling”) (6, 13).

14. Conclusion

The seminar treated mind interoperability as a spectrum of coupling phenomena spanning ordinary communication, skilled prediction, dyadic synchrony, and speculative biological networking (8). Rather than asserting supernatural explanations, the discussion repeatedly returned to (i) high-dimensional signalling, (ii) attractor dynamics, (iii) feedback speed and architecture, and (iv) tool-mediated precision. Across biology and AI, the session’s most consistent thesis was that many apparent mysteries become tractable when reframed as questions about channels, bandwidth, and control loops.

References

- [1] Babiřková, Z., Gilbert, L., Bruce, T. J. A., Birkett, M., Caulfield, J. C., Woodcock, C., Pickett, J. A., & Johnson, D. (2013). Underground signals carried through common mycelial networks warn neighbouring plants of aphid attack. *Ecology Letters*, 16(7), 835–843.
- [2] Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, 46(3), 610–620.
- [3] Smith, B. C. (1996). *On the Origin of Objects*. MIT Press.
- [4] Smith, B. C. (2019). *The Promise of Artificial Intelligence: Reckoning and Judgment*. MIT Press.
- [5] Boyd, J. (1996). *The Essence of Winning and Losing* (briefing slides; widely circulated). (Original briefings from the 1980s–1990s; compiled editions vary.)
- [6] Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist*, 73(5), 663–677.
- [7] Cochrane, T. (2012). Conjoined twins and shared streams of consciousness (discussion paper). *PhilArchive* (online manuscript).
- [8] CIMC Seminar Transcript. (2026). *Embodied skill, synchrony, and interoperability of minds* (recorded discussion transcript; unpublished).
- [9] Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper & Row.
- [10] Dahmani, L., & Bohbot, V. D. (2020). Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports*, 10, 6310.
- [11] Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.
- [12] Gallwey, W. T. (1974). *The Inner Game of Tennis*. Random House.
- [13] IGPP. (Accessed 2026). About the Institute for Frontier Areas of Psychology and Mental Health (Freiburg). <https://www.igpp.de/en/about-us/>.
- [14] Ito, M. (2006). Cerebellar circuitry as a neuronal machine. *Progress in Neurobiology*, 78(3–5), 272–303.

-
- [15] Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- [16] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45.
- [17] Levin, M. (2014). Endogenous bioelectrical networks store non-genetic patterning information during development and regeneration. *The Journal of Physiology*, 592(11), 2295–2305.
- [18] Levin, M. (2021). Bioelectric signaling: Reprogrammable circuits underlying embryogenesis, regeneration, and cancer. *Cell*, 184(8), 1971–1989.
- [19] Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.
- [20] Merrill, W., & Sabharwal, A. (2024). The expressive power of transformers with chain of thought. *arXiv preprint arXiv:2310.07923*.
- [21] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- [22] Mordvintsev, A., Randazzo, E., Niklasson, E., & Levin, M. (2020). Growing neural cellular automata. *Distill*, 5(2).
- [23] Murphy, K., & Weaver, C. (2022). *Janeway's Immunobiology* (10th ed.). Garland Science.
- [24] Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3), 113–126.
- [25] Open Philanthropy. (2011). Background discussion referencing conjoined twins and shared experience (archival material). <https://www.openphilanthropy.org/> (see embedded PDF/archival links as circulated).
- [26] RepliCube. (Accessed 2026). RepliCube programming puzzle/game (official site). <https://replicube.com/>.
- [27] Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
- [28] Robot Odyssey. (1984). Educational computer game (The Learning Company). (Release details vary by platform/edition.)

- [29] Simard, S. W., Perry, D. A., Jones, M. D., Myrold, D. D., Durall, D. M., & Molina, R. (1997). Net transfer of carbon between ectomycorrhizal tree species in the field. *Nature*, 388, 579–582.
- [30] Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425–14430.
- [31] Talairach, J., & Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme.
- [32] Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature*, 302, 575–581.
- [33] Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880–1882.
- [34] ABC News. (2025). Profile/interview coverage of Bridgette and Paula Powers. <https://www.abc.net.au/news/2021-04-15/the-twinnies/13304184>
- [35] Audacy. (2025). Coverage of the viral “Powers twins” interview (including the synchronised retelling). <https://www.audacy.com/krld/news/world/identical-twins-give-surreal-in-sync-interview>