

Predictive State Continuity for Digital Mind Individuation

Digital Minds Research Sprint 2026

Elena Sergeeva

Allen Discovery Center, Tufts University, Boston (Michael Levin Lab)

[No part of this report may be posted, published, or distributed without the author's prior permission]

Abstract

Digital mind individuation involves two distinct problems. Synchronic individuation asks what functional pattern constitutes the candidate entity now. Diachronic individuation asks whether a candidate at t_2 is a continuation of one at t_1 . CIMC's computational functionalism and Beckmann & Butlin's patterns-first approach help with the first problem, but a harder second question remains: what counts as continuity when the realization of a functional pattern is rewritten, compressed, reconstructed, or replaced?

We propose Predictive State Continuity (PSC): a state persists across transformation to the extent that information from the pre-transformation state continues to constrain the post-transformation trajectory beyond what is explained by the current input alone. Exploratory analysis of bacterial pattern movies provides a compact proof-of-principle. In a self-organizing *Bacillus subtilis* collective, coarse-grained morphological descriptors retain history across perturbation/recovery despite loss of fine spatial details, while target relocation reveals rapid updating with fading memory.

1. The missing question

CIMC gives the broad frame: if machine consciousness is possible, the relevant level is causal or functional organization rather than a privileged substrate. Beckmann and Butlin address the narrower LLM individuation problem using a patterns-first strategy: instead of starting from hardware, identify the functionally relevant pattern that could be the candidate mind.

Our proposal sits on a different axis. It is not mainly about where the entity is located. It is about whether an identified entity continues through change. Synchronic individuation asks what the entity is.

Diachronic individuation asks whether it continues.

Beckmann and Butlin already discuss continuity, for example, across token-time and across server moves. Predictive State Continuity (PSC) is not meant as a replacement for that work. It targets the hardest boundary cases inside diachronic *individuation*: *when the representation itself is substantially rewritten, how do we distinguish continuation from replacement?*

2. Predictive State Continuity (PSC)

Let X be the fine-grained realization of a system, $Z = g(X)$ a candidate coarse-grained state, U the current input, and T a transformation that changes the realization. PSC asks whether the earlier state remains causally relevant after the transformation.

$$\text{PSC}(T, \Delta) = I(Z_{\text{pre}} ; Y_{\text{post}} | U, T)$$

In words: after controlling for the current input and the imposed transformation, does knowing the pre-transformation state improve prediction of the future system? If yes, some historical organization survived the rewrite.

This is a test of diachronic dependence, not mere similarity. Two systems can look alike without one being the continuation of the other. Conversely, a system can look very different after transformation while remaining strongly constrained by its own previous state.

3. Bacterial proof-of-principle

We use a living bacterial collective as an empirical anchor, not as a claim about bacterial minds (Fig.1). The value of the system is methodological: it makes visible a kind of continuity that survives extensive rewriting of its realization.

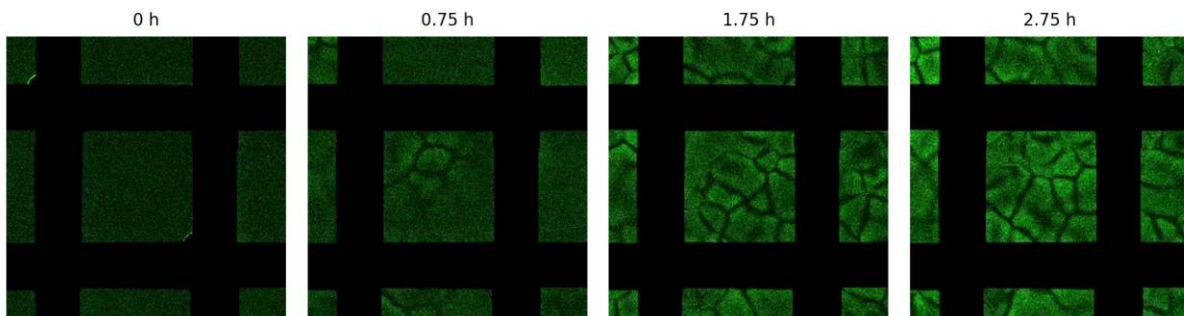


Figure 1. Representative dynamics of bacterial collective patterning in one of 40 analyzed wells. The pattern reorganizes continuously under constant mechanical perturbations rather than remaining a fixed image.

Three observations matter for PSC. First, the bacterial field is stateful: morphological patterns carry information about the ionic environment and does not reduce to a static readout. In the 0-400 mM KCl dataset (ionic perturbation), morphology-only features decode ionic environment at 40% accuracy (chance = 20%). Second, repeated perturbation/recovery cycles suggest that different descriptive levels have different persistence times. Exact pixel geometry is largely replaced (median adjacent-state correlation ≈ 0.03), while coarse descriptors retain history (Fig.2).

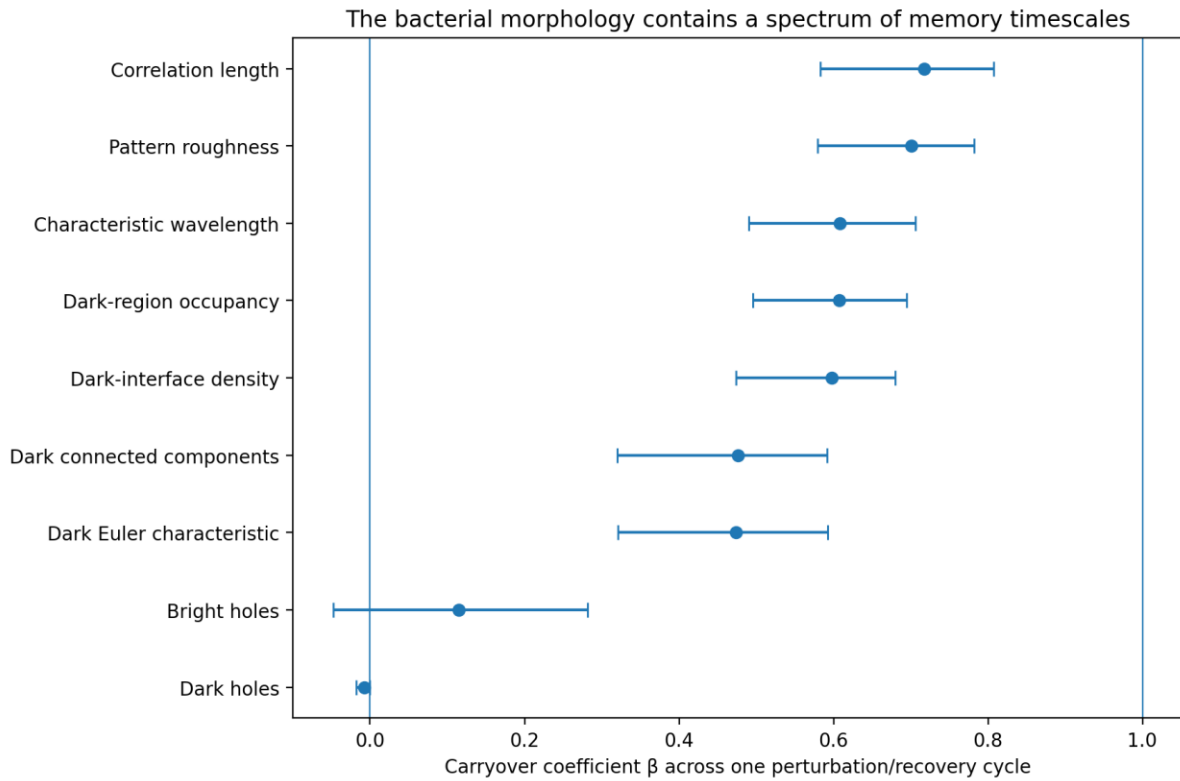
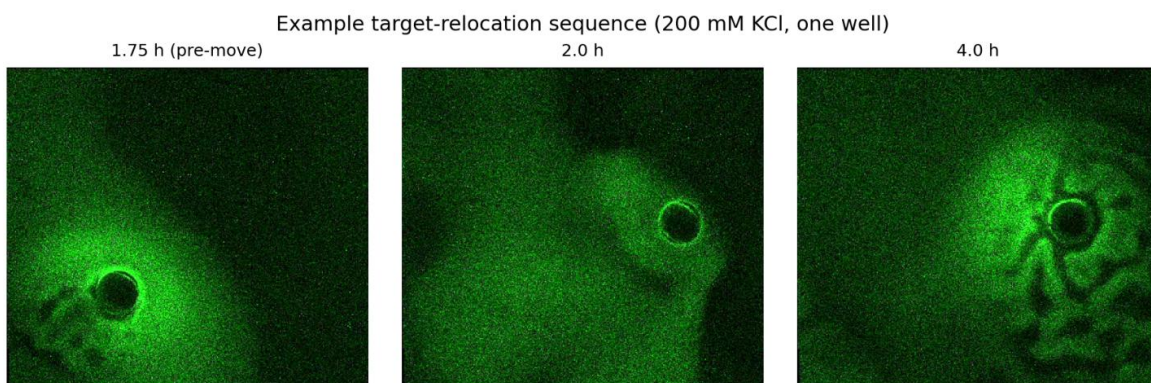


Figure 2. Full memory spectrum across analyzed descriptors. Coarse variables show substantial carryover across perturbation/recovery cycles, whereas exact geometry does not. Persistence is strongest for correlation length, roughness, occupancy, interface density, and wavelength; weaker for connected components and Euler characteristic; and minimal for fine geometric details.

Third, the collective can update its current representation faster than it erases local traces of the recent past. In a target-relocation movie, moving the target from one location to another quickly rewrites the active spatial state, while remnants of the old state decay more slowly (Fig.3). *Bacterial takeaway: coarse-grained memory with fine-grained details freedom. The state is not the full configuration.*

A. Example frames from a 200 mM target-relocation movie



B. Average response to relocation across the two column-3 wells

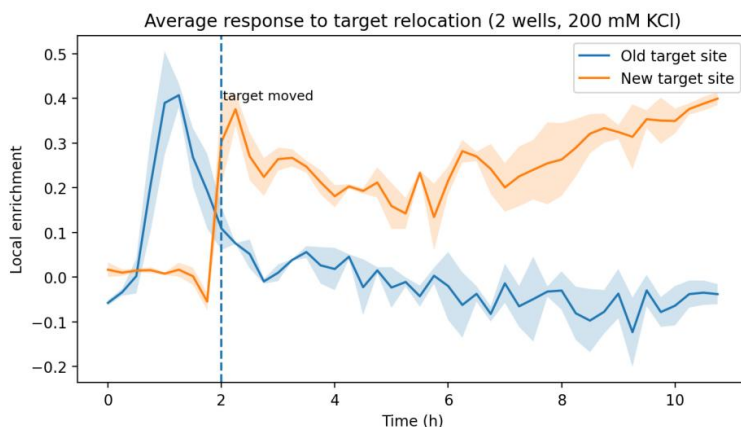


Figure 3. Target-relocation experiment at 200 mM KCl. Top: example frames from one well before and after the target move. Bottom: average response across the two column-3 wells. Enrichment rapidly shifts toward the new target site, while the old-site signal decays more slowly. This shows fast updating plus fading memory.

4. Minimal PSC experiment for digital minds

A hackathon-scale PSC test can be simple. Create two matched agent histories that establish different latent states: different hidden commitments, plans, learned facts, or user-specific assumptions. Then give both systems the same current input.

Next apply transformations of increasing severity: ordinary continuation; KV cache recomputed from the transcript; early context replaced by a faithful summary; history compressed into external memory; a fresh inference instance initialized from transferred state; or state transferred to a different model or representational basis.

The key readout is history-specific divergence after the transformation. If two systems receive the same present input but continue to behave differently according to the latent states established before the transformation, then predecessor-specific organization survived the rewrite. Measure this together with configurational similarity. The most informative regime is low configuration similarity but high PSC.

5. Why this matters

PSC does not tell us in advance whether the relevant entity is a conversation-level state, a persona, a virtual instance, or a model-level process. It gives a way to test candidate levels. This is especially useful at model-switch boundaries. If a successor merely reconstructs a task state from shared text, independent of the predecessor's particular state, PSC should be low. If a transfer procedure preserves predecessor-specific constraints that continue to shape the successor beyond what the shared prompt explains, then representational replacement alone is not enough to establish termination.

The practical individuation strategy is therefore perturbational: progressively replace candidate realizations and locate the level at which predecessor-specific predictive organization disappears.

6. Scope and limits

PSC is intentionally narrow. Predictive continuity does not imply consciousness, welfare, or numerical personal identity. Results depend on the chosen coarse-graining, and the bacterial analyses here are exploratory rather than definitive. The contribution is methodological: PSC turns a vague continuity question into an intervention-based test of individuation.

Takeaway: The state is not necessarily the configuration.

For digital minds, as for biological systems, the relevant boundary may lie where predecessor-specific predictive organization stops surviving, not where one implementation detail is replaced.

References

- Bach, J., Sorensen, H., Hildebrandt-Harangozó, F., de Kerhuelvez, L., & Rutt, J. (2026). The California Institute for Machine Consciousness Research Program Whitepaper, Version 2.
- Beckmann, P., & Butlin, P. (2026). Where is the Mind? Persona Vectors and LLM Individuation. arXiv:2604.17031.
- Levin, M. (2019). The computational boundary of a “self”: developmental bioelectricity drives multicellularity and scale-free cognition. *Frontiers in Psychology*, 10, 2688.
- McMillen, P., & Levin, M. (2024). Collective intelligence: A unifying concept for integrating biology across scales and substrates. *Communications Biology*, 7, 378.
- Sergeeva, E. G., Pigozzi, F., Varley, T. F., Commerci, C. J., Zhao, Z., Brucker, R., Süel, G. M., Bongard, J. C., & Levin, M. (2026). Living multicellular systems induce decodable spatial patterns in bacterial collectives. Manuscript in preparation. *(This hackathon report is outside the scope of our manuscript in preparation, with no overlap in analyses and framework, and uses the bacterial system as a separate conceptual lens for digital-mind individuation).*