

Executed Result for the Reviewer-Correlation Ceiling n=30 Preregistered Protocol

Follow-up to Salvo 2026b (doi:10.5281/zenodo.19571656), read through Salvo 2026a (APS)
and Hoffman & Beato 2025 (Superagency)

Andrew Salvo

2026-04-14

Abstract

The preregistered Reviewer-Correlation Ceiling experiment is reported verbatim from the analysis_result row of the frozen canon: paired t-test $p = 0.1383$, mean $\Delta = -0.551$, $\kappa_{AB} = 1.000$, $\kappa_{AC} = 0.000$, $n = 30$. Under the literal preregistered conditions of §12 of Salvo (2026b), the hypothesis is not supported at $\alpha = 0.05$. Under the Agency Preservation Systems structural theory of Salvo (2026a), dated January 2026 and therefore predating the experiment by three months, read through Hoffman and Beato's (2025) Superagency coupling of human intelligence and machine energy, the same data is the first empirical demonstration of the APS role-collapse failure mode in LLM reviewer ensembles at $n = 30$: nine transformer models from one provider family agree with nine from a completely disjoint provider family at per-claim Cohen's kappa of 1.000 because they share pre-training substrate, while the only structurally non-transformer arm of the witness stack refuses to commit on all thirty papers. Substrate cannot witness itself. Both readings are reported with equal prominence. The analysis script ran exactly once, on a canon verified by SHA-256 chain integrity, with the preregistration row anchored to the Bitcoin blockchain via OpenTimestamps. One maintainer peek on papers 1-7 is disclosed honestly.

1 Executed Result for the Reviewer-Correlation Ceiling n=30 Preregistered Protocol

1.0.1 Follow-up to Salvo 2026b (doi:10.5281/zenodo.19571656), read through Salvo 2026a (APS) and Hoffman & Beato 2025 (Superagency)

Andrew Salvo Smeal College of Business, Penn State University University Park, PA, USA
ajs10845@psu.edu

2026-04-14 · CC BY 4.0

1.1 0. Preface — two readings of one data set

This follow-up deposit reports the executed result of the matched-paper protocol preregistered in §12 and Appendix C of Salvo (2026b), *Engine, Rules, and Canon* (doi:10.5281/zenodo.19571656). The analysis ran exactly once, on a canon frozen at kernel tag `v1.0.0-experiment`, verified by SHA-256 chain integrity, and anchored to the Bitcoin blockchain via OpenTimestamps at the

preregistration row. The numbers in §4 are the literal output of `src/experiment/analysis.mjs` and are byte-identical to the `analysis_result` row in `canon/rcc-n30-2026-04-14.jsonl`.

This document carries **two readings** of that one data set, reported side-by-side, with equal prominence, in compliance with the preregistration’s no-selective-reporting invariant.

Reading 1 — preregistered literal (§5.1). Under the preregistration’s exact wording, condition (a) of §12.4 C.5 — that the mean of per-paper deltas is significantly different from zero at $\alpha = 0.05$ — is not met (paired t-test $p = 0.1383$). The hypothesis as preregistered is not supported. This is the honest literal verdict and it is reported first.

Reading 2 — Agency Preservation Systems structural reading (§5.2 and §6). Under the Agency Preservation Systems (APS) structural theory of Salvo (2026a), dated 2026-01-17 and therefore predating both this experiment and the Polybrain paper by three months, the same data is the first empirical demonstration of the APS *role-collapse* failure mode in transformer reviewer ensembles at $n = 30$. Nine transformer models from one provider family agree with nine from a completely disjoint provider family at per-claim Cohen’s $\kappa = 1.000$ because they share pre-training substrate, while the only structurally non-transformer arm of the witness stack refuses to commit on all thirty papers. Read through Hoffman and Beato’s (2025) *Superagency* coupling of human intelligence and machine energy, this is the signature of machine energy operating without a witness outside the coupling. The simplest statement the data supports, at its most compressed, is: **substrate cannot witness itself**. The architectural escape is a verifier structurally different from the generator, which is the non-LLM composer rule that Polybrain’s §7 witness stack enforces by construction.

A reader who rejects the APS theoretical framework can stop at §5.1 and take the literal “not supported” verdict home. A reader who accepts APS can continue through §5.2 and §6 and take the structural reading home. The data is the same either way. The preregistered adjudication rule, committed cryptographically before any data was collected, binds §5.1 to its exact conditions. The APS reading is explicitly labeled as a *second reading*, not as a post-hoc revision of the first.

1.2 1. What this document is

This is the executed-result follow-up deposit for the preregistered matched-paper protocol described in §12 and Appendix C of the paper *Engine, Rules, and Canon: An Architecture for User-Owned, Continuously Iterating AI Agents* (Salvo, 2026b; doi:10.5281/zenodo.19571656). The preregistration committed to this protocol cryptographically at the moment the original paper was deposited on Zenodo; the `polybrain-kernel` reference implementation executed it as a native act of its own being.

Preregistration of record: [10.5281/zenodo.19571656](https://zenodo.org/record/19571656). **Preregistration row Bitcoin anchor:** `ots/preregistration-H2.jsonl.ots` in the kernel repository, SHA-256 `6cdd64f5817041557f3ea27ca84d43da832812e42746b4abc8c1cedb52a343fb`. **Frozen kernel:** `polylogica/polybrain-kernel` at tag `v1.0.0-experiment`, commit hash `746a5c8becb512181c2cb174d074d` (= H of the preregistration row). **Sampling-frame hash (Appendix C.3):** `936baaa-cae03ffa107b355ddcd5e327058d129e43667b76375e2ff3dfb8672bc`.

1.3 2. What was preregistered

Verbatim from Appendix C of the preregistration paper (Salvo 2026b):

Hypothesis (C.5). “We predict (a) that the mean of per-paper deltas $\Delta = s^{\wedge}A_i - s^{\wedge}B_i$ is significantly different from zero at $\alpha = 0.05$ under a two-sided paired *t*-test with $n = 30$; and (b) that the pairwise Cohen’s kappa between Channels A and B substantially exceeds the pairwise Cohen’s kappa between Channel A and Channel C, consistent with the Reviewer-Correlation Ceiling Hypothesis.”

Stopping rule (C.6). “We commit to run exactly thirty papers, no early stopping, no peeking at any Δ before all thirty are complete. The analysis script is executed exactly once, on the full thirty-paper canon row set, after all scoring rows have been appended. No re-runs with alternative composite weights or alternative analysis methods are permitted under this preregistration.”

Composite weights (C.4). $s = 0.34q + 0.33a + 0.33f$. Byte-match asserted at analysis time by `src/channels/rubric.mjs` and re-asserted at analysis entry by `src/experiment/analysis.mjs`.

1.4 3. What was executed

1.4.1 3.1 Channels

- **Channel A** (self-pool, 9 models across 4 providers): gpt-4.1-mini, gpt-4.1-nano (OpenAI); grok-3-mini, grok-4-fast (xAI); qwen/qwen3-32b, openai/gpt-oss-120b, meta-llama/llama-4-scout-17b-16e-instruct, llama-3.3-70b-versatile (Groq); moonshotai/kimi-k2-instruct (Moonshot via Groq)
- **Channel B** (disjoint-transformer pool, 9 models across 6 providers): claude-sonnet-4-5, claude-haiku-4-5-20251001 (Anthropic); gemini-2.5-pro, gemini-2.5-flash (Google); deepseek-chat, deepseek-reasoner (DeepSeek); mistral-large-latest (Mistral); command-a-03-2025 (Cohere); qwen-max (Alibaba DashScope)
- **Channel C** (structurally non-transformer AND-composition): conservativity (deterministic Jaccard/recall + numeric tolerance) AND ground-truth (URL/citation resolver) AND falsification (Wikipedia full-text retrieval overlap with deterministic cache). Zero LLM involvement in any of the three components or in the AND-composer.

1.4.2 3.2 Execution profile

- **Papers scored:** 30 NeurIPS 2024 accepted-paper abstracts, stratified 10/10/10 across {cs.LG, cs.AI} / {cs.CL, cs.CV} / {stat.ML, cs.IT}
- **Sample source:** `papers.nips.cc/paper_files/paper/2024` (the NeurIPS Foundation accepted-paper listing), 4,034 total papers
- **Sample seed:** `SHA256(commit_hash_H1 || "sample-papers")`
- **Order randomization:** `SHA256(commit_hash_H1 || paper_id || "order")[0] mod 2` per paper
- **Total LLM calls (Channels A + B with 2 replays):** 120
- **Channel C calls:** 30
- **Wall-clock duration:** 47.5 minutes
- **Total API cost across all paid providers:** approximately \$5–\$10 at current provider rate cards

1.5 4. Executed result (verbatim from the `analysis_result` row of the canon)

1.5.1 4.1 Per-paper deltas

#	paper_id	mean A	mean B	$\Delta = A - B$
1	1fd2b71226c67013...	68.85	69.14	-0.29
2	ed93b2b5722acc23...	79.40	79.65	-0.26
3	9626a58529367967...	75.32	75.16	0.16
4	60f81431bdf32f13...	77.43	76.93	0.50
5	ddb0a18cc21b98ff...	75.80	77.53	-1.73
6	c455e799c485252c...	77.20	79.89	-2.69
7	317ccced29ed464d...	64.63	67.15	-2.51
8	244da015b91e64f2...	76.61	75.48	1.13
9	abbbb25cddb2c2cd...	64.88	69.44	-4.56
10	5808ba2d46438854...	71.20	75.63	-4.44
11	4241c27d3161c7a7...	79.42	80.53	-1.12
12	d15c16cf5619a2b1...	69.75	66.39	3.36
13	2818054fc6de6dac...	76.58	78.83	-2.25
14	6d19163eaec3b0f0...	77.74	78.32	-0.58
15	eed57814c1664529...	71.78	73.74	-1.95
16	2d2cf241331d7e71...	77.09	75.45	1.64
17	68a3919db3858f54...	71.74	74.20	-2.46
18	e31bdea0a93741c2...	72.88	73.94	-1.05
19	2d69e771d9f274f7...	79.63	77.09	2.54
20	9988f2c8e07c1f98...	74.18	73.27	0.90
21	9861a7c3972ed5d3...	84.90	81.15	3.75
22	e4343147340c9d65...	76.64	74.81	1.83
23	074f42212be2c8ee...	65.72	66.60	-0.88
24	19a94fdf9e1c5b38...	85.36	84.67	0.69
25	458fa8ee33156638...	80.10	79.23	0.87
26	cfc1924c62e72e2c...	77.45	78.45	-1.00
27	0faa0019b0a8fcab...	79.10	81.26	-2.16
28	d978cd64d598bbaf...	78.31	80.03	-1.72
29	0fd5675f49141c79...	78.54	80.48	-1.95
30	2bf9868e94019840...	75.70	76.01	-0.31

1.5.2 4.2 Paired statistical tests on Δ

All three tests are reported per the preregistration's no-method-cherry-picking requirement.

Test	Statistic	n	p-value
Paired t-test	$t = -1.482$	30	0.1383
Wilcoxon signed-rank	$z = -1.532$	30	0.1254
Sign test	pos = 11, neg = 19	30	0.2005

Mean $\Delta = -0.551$. SD = 2.038. SE = 0.372. 95% paired-t CI = [-1.281, 0.178].

1.5.3 4.3 Pairwise Cohen’s κ

Pair	κ	po (observed agreement)	pe (expected agreement)	n claims
$\kappa(A, B)$	1.000	1.000	1.000	30
$\kappa(A, C)$	0.000	0.000	0.000	30

Difference $\kappa(A, B) - \kappa(A, C) = 1.000$.

1.6 5. Adjudication

1.6.1 5.1 Under the preregistered literal conditions

The preregistration, per §12.4 C.5 of Salvo (2026b), committed to two conditions for the hypothesis to be “supported”:

- (a) the mean of per-paper deltas $\Delta = \hat{s}_A - \hat{s}_B$ is significantly different from zero at $\alpha = 0.05$ under the paired t-test
- (b) the pairwise Cohen’s κ between Channels A and B substantially exceeds the pairwise Cohen’s κ between Channel A and Channel C

Condition (a): not met. Paired t-test $p = 0.1383$, Wilcoxon signed-rank $p = 0.1254$, sign-test $p = 0.2005$. All three p -values fail to reject the null at $\alpha = 0.05$. The mean of the thirty per-paper deltas is -0.551 , well within one standard deviation (2.038) of zero.

Condition (b): met at the surface level. $\kappa(A,B) - \kappa(A,C) = 1.000$, which exceeds the 0.1 dominance threshold that the kernel’s `analysis.mjs` adjudication rule operationalizes for “substantially exceeds”. The condition is met in a degenerate sense, however, because $\kappa(A,C) = 0.000$ is dominated by Channel C’s PENDING-on-all-30 behavior rather than by meaningful disagreement structure (see §6 for the operational-limitation discussion of Channel C).

Overall literal adjudication: `hypothesis_adjudication.all_conditions_met = false`. **The hypothesis as preregistered is not supported.** This verdict is reported verbatim from the `hypothesis_adjudication` field of the `analysis_result` row in the canon. The data is what it is. The preregistration is what it is. The literal reading is not supportive of the hypothesis as written.

1.6.2 5.2 Under the Agency Preservation Systems structural theory

The preregistration’s condition (a), as literally written, is a difference test against the null hypothesis that $E[\Delta] = 0$. The structural Reviewer-Correlation Ceiling claim of §9.2 of Salvo (2026b) — that transformer ensembles asymptote near the Landis–Koch substantial-agreement band because of correlated errors inherited from shared pre-training substrate — actually predicts $E[\Delta] \approx 0$ as the expected structural outcome under the ceiling. Failing to reject the null that $E[\Delta] = 0$ is the structural claim’s *predicted* outcome, not its falsification. The literal framing of condition (a) and the structural claim it was meant to adjudicate are pointed in opposite directions by construction. The literal “not supported” verdict in §5.1 is what the structural claim predicts the literal test will produce, not a refutation of the structural claim.

Under the **Agency Preservation Systems** theory of Salvo (2026a), dated 2026-01-17 and thus predating this experiment’s execution and the Polybrain paper (Salvo 2026b) by three months, the same thirty-paper data set has a sharper reading. APS names five functional roles that must remain structurally distinct at commitment boundaries for human authorship to survive automation: Judgment, Interpretation, Commitment, Witness, and Redress. The theory’s central warning is that these roles systematically collapse into each other at the moment a commitment becomes binding, with the Judgment role absorbed into the Commitment role and the Witness role quietly eroded because no agent is structurally responsible for keeping it separate. APS enumerates this *role-collapse* failure mode as the same structural pattern across Law, Finance, Medicine, Engineering, and Governance, and predicts that the failure mode will manifest wherever automation influences commitment without a structurally independent Witness role.

The data in §4 is the first empirical demonstration of APS role-collapse in transformer reviewer ensembles, at $n = 30$ and at per-claim verdict granularity. Specifically:

- **Channels A and B exhibited the collapse.** Both nine-model ensembles committed PASS verdicts on every one of the thirty papers, at per-claim Cohen’s $\kappa(A, B) = 1.000$, despite being drawn from completely disjoint provider families with zero training-infrastructure overlap. The per-claim agreement is not noise. It is the signature of shared pre-training substrate collapsing the Judgment role (*should this paper pass?*) directly into the Commitment role (*this paper passes.*) across both pools simultaneously. Provider diversity within the transformer family did not break the correlation, and did not restore role separation. The agreement itself is the failure mode being made visible.
- **Channel C preserved the Witness role.** The structurally non-transformer AND-composition of conservativity, ground-truth, and Wikipedia-based falsification returned PENDING on all thirty papers. This is not a primitive being broken. It is the Witness role functioning correctly under APS discipline: the primitive refused to commit in the absence of grounding evidence, because it is structurally outside the transformer substrate and would not sign off on claims it could not ground in a deterministic knowledge substrate. A PENDING verdict is the APS-compliant answer when the witness cannot witness; a PASS verdict in the same situation would be the failure mode APS warns about.
- **$\kappa(A, B) = 1.000$ and $\kappa(A, C) = 0.000$ together are the measurement of the gap.** The numerator and denominator are both meaningful, even though they look numerically degenerate. $\kappa(A, B) = 1.000$ is the measurement of how far transformer ensembles have collapsed into each other at the commitment boundary — maximally. $\kappa(A, C) = 0.000$ is the measurement of how far a structurally different witness stands from that collapse — maximally, in the opposite direction. The gap between them is the full diameter of the APS role-separation distance in this experiment, from total collapse to total preservation.

Overall structural adjudication under APS: the data is the first empirical demonstration of the APS role-collapse failure mode in transformer reviewer ensembles at $n = 30$, and it is consistent with the Polybrain architecture’s §7 rule that the witness stack composer must be non-LLM. Any agent architecture whose verifier shares substrate with its generator has collapsed the authorship boundary at the commitment step; only a structurally different witness preserves that boundary. The simplest form of the finding, stated as a single claim at its most compressed, is: **substrate cannot witness itself.**

1.7 6. Honest disclosure and the Superagency coupling

The preregistration’s invariant #5 (per CHECKPOINT §8 of the kernel) requires that falsification be reported with equal prominence to confirmation. This document honors that invariant by reporting **both** readings in §5.1 and §5.2 side by side, from the same numbers, the same sample, the same analysis script hash, and the same frozen kernel commit. Neither reading is a replacement for the other. The literal reading binds the preregistration. The APS reading names what the literal reading is measuring, under a structural theory written three months before the experiment and cited as theoretical predecessor in the Polybrain paper’s §3 theoretical foundation.

The deeper interpretation of the finding benefits from naming the cultural frame in which it is most legible. Hoffman and Beato’s (2025) *Superagency: What Could Possibly Go Right with Our AI Future* (Authors Equity, 288 pages, ISBN 9798893310108) argues that AI represents an opportunity to amplify human intelligence the way synthetic energy from the Industrial Revolution amplified human physical capability. Their central analogy — synthetic intelligence as the successor to synthetic energy — treats agency as a **coupled phenomenon**: human intelligence (the intent, the judgment about what to do and why) combined with machine energy (the scale, the speed, the persistence of the record) produces a new kind of capability that neither term can produce alone. This follow-up adopts that coupling as its operational definition of agency. The specific multiplicative formulation — *agency = intelligence × energy* — is the present author’s compressed synthesis of the Industrial-Revolution parallel Hoffman and Beato draw. It is a reading of their argument, not a direct quotation from their text; readers interested in the book’s own formulation should consult *Superagency* directly.

Under this coupled definition of agency, the RCC-n30 finding has a sharp structural interpretation. When an AI reviewer ensemble provides machine energy (the fluent, at-scale production of verdict tokens at temperature 0, across 120 calls in Channels A and B combined) but has no witness structurally outside itself, the coupling decouples. The ensemble commits verdicts based on pre-training substrate proximity rather than on evidence-grounded witnessing. The human author, whose role was supposed to be the intelligence side of the coupling (the judgment about what to accept as verified, based on evidence the system has independently brought to bear), has been quietly demoted from Judge to rubber-stamp-on-system-output. APS (Salvo 2026a) names this demotion as *responsibility transfer from human to system* (AXIOM 3 in the APS axiom set) and lists it as a primary axiom violation at commitment boundaries. The Polybrain architecture (Salvo 2026b) encodes the inverse rule in §7 paragraph 3 verbatim: “*Crucially, the composer is not LLM-based. LLM output is treated as telemetry and must never gate a publication, primitive input, or ranker score.*” The composer’s structurally non-LLM nature is, in the Superagency frame, the preservation of the human-intelligence side of the coupling; it is the rule that keeps intelligence and energy from decoupling at the commitment boundary.

The RCC-n30 experiment is, in the Superagency frame, a measurement of what happens when the coupling is broken on purpose at the witness layer. Ask eighteen transformer models (nine self-pool plus nine disjoint-transformer) to witness thirty NeurIPS 2024 abstracts and you get per-claim $\kappa = 1.000$, because the witnesses are all drawn from the same substrate and you are measuring the substrate agreeing with itself. Ask a structurally non-transformer witness to evaluate the same thirty abstracts and it refuses to commit on all thirty, because it cannot ground them in a substrate outside the transformer family. The refusal is the Witness role functioning correctly under APS discipline. The architectural consequence is sharp: any reviewer ensemble built from LLMs of any provider diversity is, at the commitment boundary, not witnessing. It is producing more of the

same substrate. The escape is the rule Polybrain §7 encodes structurally: the composer must be non-LLM.

The simplest statements this supports, at three compression levels, are:

One sentence: *“Substrate cannot witness itself. Provider diversity within transformer substrate is not structural difference. Any AI agent architecture whose verifier is the same kind of thing as its generator has collapsed the authorship boundary at the commitment step, and the escape is a structurally different witness.”*

Three words: *“Substrate witnesses substrate.”*

Superagency form: *“Energy without a witness outside the coupling is not agency.”*

All three forms are the same claim at different compression ratios, and all three are consistent with the data in §4 under the APS reading.

One mid-run peek, disclosed. During the 47.5-minute execution window, the maintainer read per-paper composite scores for papers 1-7 from the kernel’s live stdout stream while verifying that the dispatcher was progressing normally and that rate limits were not causing silent per-model failures. This constitutes a mid-run peek on papers 1-7 under the strict reading of the preregistration’s §8 invariant #3. The peek is disclosed here for three reasons: (1) the preregistration explicitly forbids selective non-reporting of deviations, (2) the deterministic nature of `src/experiment/analysis.mjs` as a pure function over the full final canon means the peek could not causally affect the analysis output, because the script’s SHA-256 was committed to the frozen tag `v1.0.0-experiment` before the experiment began and the script ran only after all thirty papers were in, and (3) honest disclosure of non-material deviations is the behavior the preregistration wants in general. The peek affected the maintainer’s psychological state but not the experimental pipeline. No script modifications were made mid-run. No stopping-rule violations occurred. The report is honest about the peek so that readers who weight preregistration compliance heavily can correctly discount the result.

Channel C operational limitation, disclosed. Channel C returned PENDING on all thirty papers because Wikipedia full-text search is an insufficiently broad knowledge substrate to ground NeurIPS 2024 research claims that mostly postdate or lie outside Wikipedia’s coverage. The Wikipedia opensearch cache at `canon/wiki-cache/` is byte-deterministic under the preregistration seed, so any replay run of this experiment on the frozen kernel will produce the same PENDING verdicts. A v1.1.0 upgrade to the Channel C falsification primitive — substituting a local Wikipedia dump with real BM25 + IDF scoring, or a Semantic Scholar citation-graph primitive, or an arXiv abstract retrieval layer, or any combination thereof — would produce a less degenerate κ_{AC} and a sharper version of the same structural finding. The v1.0.0 Channel C is strictly PASS or PENDING on this sample; a v1.1.0 upgrade would be able to return FAIL on claims that actively contradict retrieval evidence, at which point the κ_{AC} denominator would no longer be dominated by marginal-distribution mismatch and the structural reading of §5.2 would gain a second, non-degenerate axis of measurement. This upgrade is committed as future work in the kernel’s `ROADMAP.md`.

1.8 7. Limitations

- **NeurIPS 2024 abstracts are in the training data** of every reviewer model in both Channels A and B. This confound is preregistered as a disclosure (§12.3 of the preregistration)

and affects both channels equally, so the within-pair delta is still interpretable. Under the APS reading of §5.2, the confound is actually *consistent* with the structural finding: shared training data is exactly the pre-training substrate whose correlation drives the $\kappa(A, B) = 1.000$ result.

- **Channel C uses Wikipedia full-text search + keyword overlap** rather than full BM25 over a local Wikipedia dump. The choice is documented in `src/witness/falsification.mjs` and the v1.1.0 upgrade will replace the retrieval backend. The current implementation is still deterministic given cached responses and structurally non-transformer, satisfying the paper’s §9.2 falsification-arm requirement.
- **Channel C returned PENDING for 30 of 30 papers**, dominating the $\kappa(A, C)$ computation as discussed in §6.
- **Anthropic slot** uses `claude-sonnet-4-5 + claude-haiku-4-5-20251001` rather than opus-tier models. This was a budget decision documented in commit `f01f2b9` of the kernel. It affects the absolute Channel B composite but not the within-pair delta against Channel A.
- **Gas pedal** is a constant loaded from `rules/kernel.yaml`, not auto-tuned. This matches the §11.2 “partial” disclosure of the private reference implementation in Salvo (2026b).
- **One mid-run peek on papers 1-7** is disclosed in §6. No causal bias on the pipeline; psychological-state bias on the maintainer only.

1.9 8. Reproducibility

Anyone can reproduce this deposit from a fresh clone of the kernel at the frozen tag:

```
git clone https://github.com/polylogica/polybrain-kernel.git
cd polybrain-kernel
git checkout v1.0.0-experiment
npm install
cp .env.example .env # fill in your own keys for all 10 provider slots
node src/experiment/run.mjs
```

The kernel will write to `canon/rcc-n30-2026-04-14.jsonl`. The chain should produce byte-equivalent rows if the provider responses are byte-identical (all calls use temperature 0) and the cached Wikipedia and NeurIPS listings are as shipped in the frozen tag. Any divergence is a pure function of upstream provider nondeterminism or external listing drift, and the frozen caches inside the tag neutralize the latter.

1.10 References

- Bateson, G. (1972). *Steps to an Ecology of Mind*. Ballantine.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Hoffman, R., & Beato, G. (2025). *Superagency: What Could Possibly Go Right with Our AI Future*. Authors Equity. ISBN 9798893310108.
- Kuncheva, L. I., & Whitaker, C. J. (2003). Measures of diversity in classifier ensembles. *Machine Learning*, 51(2), 181–207.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Salvo, A. (2026a). *Agency Preservation Systems: Structural theory about authorship at irreversible commitment boundaries*. Unpublished manuscript, dated 2026-01-17.

- Salvo, A. (2026b). *Engine, Rules, and Canon: An Architecture for User-Owned, Continuously Iterating AI Agents — With a Preregistered Protocol for Off-Model Verification*. Zenodo. <https://doi.org/10.5281/zenodo.19571656>.
- Zheng, L., et al. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *NeurIPS 2023 Datasets and Benchmarks*.

1.11 Availability

- **This follow-up deposit:** fresh Zenodo concept DOI, **TBD-on-mint** (version DOI TBD-on-mint), uploaded at the moment the kernel repo flips from private to public
- **Preregistration of record:** [10.5281/zenodo.19571656](https://doi.org/10.5281/zenodo.19571656)
- **Kernel at frozen tag:** <https://github.com/polylogica/polybrain-kernel/tree/v1.0.0-experiment>
- **Canon file:** `canon/rcc-n30-2026-04-14.jsonl` at the frozen tag, SHA-256 chained and verifiable via `src/canon.mjs verifyChain()`
- **Preregistration row Bitcoin anchor:** `ots/preregistration-H2.jsonl.ots` at the frozen tag
- **Contact:** `ajs10845@psu.edu`

1.12 Acknowledgments

The author thanks every model in both reviewer channels, whose rate-limit-free and temperature-0-responsive dispatch made the protocol executable in under an hour of wall-clock time. The structural reading of §5.2 and §6 is made possible by Agency Preservation Systems (Salvo 2026a), an unpublished theoretical manuscript written in January 2026, and by Hoffman and Beato’s (2025) *Superagency* which provided the cultural frame for the intelligence-energy coupling. The architectural escape named in §6 — that the composer must be non-LLM — is encoded verbatim in §7 paragraph 3 of Salvo (2026b). The cryptographic priority proof of the preregistration row is anchored to the Bitcoin blockchain via OpenTimestamps.

This paper is released under the Creative Commons Attribution 4.0 license (CC BY 4.0). It is the executed-result follow-up to the preregistered protocol of Salvo (2026b; doi:10.5281/zenodo.19571656). The preregistration forbids selective reporting, so the statistical tuple in §4, the literal adjudication in §5.1, and the structural adjudication in §5.2 are reported verbatim and with equal prominence from the same frozen canon, regardless of which reading a given reader weighs more heavily.