Journal of Scientific Exploration
Anomalistics and Frontier Science

CORRESPONDENCE

Response to Letters Regarding "Testing Noetic Potential in Large Language Models"

Benjamin J. Amorim Boyle
contact@boyleadvisory.com

https://doi.org/10.31275/20264013

INTRODUCTION: GAME THEORY, SCIENTIFIC CONDUCT, AND THE STRUCTURE OF ACADEMIC EXCHANGE

Before addressing the substantive methodological critiques raised in the letters responding to my Brief Report (Amorim Boyle, 2025), it is necessary to address the structure of the exchange itself. Robert Axelrod's classic simulations of the iterated Prisoner's Dilemma demonstrated that strategies characterized by being nice (never initiating defection), provocable (retaliating when defected against), forgiving (returning to cooperation after retaliation), and clear (transparent in behavioral rules) consistently outperform aggressive "nasty" strategies in repeated interactions (Axelrod, 1984). Tit-for-tat dominates not because it is soft, but because it refuses exploitation while preventing spirals of escalation. Cooperation emerges when proportionate responses are clear and predictable. I intend to make a strong response, noting I will forgive instantly once they play optimum game theory, but not letting the authors of the letters off the hook and stamping them into academic history as an example.

The letters responding to my paper (Davanas, 2026; Ritchie et al., 2026) did not simply critique methodology, they initiated with adversarial framing suggesting poor scholarship, sociological experimentation in the Sokal tradition, or fundamental incompetence, not just by myself, but also the reviewers and the editors of the *Journal of Scientific Exploration* (all of whom I appreciate and all of whom can tit back at their tat should they choose). In game-theoretic terms, this constitutes a "nasty" first move, an unprovoked defection. A response is therefore warranted to make an example to the civilization for the future and set a stage for a better world. This paper is that response, proportionate, evidence-based, transparent, and firm. It is a tit to their tat. It defends the data, clarifies misunderstandings, and sets boundary conditions for future collaboration. Should the exchange shift toward cooperative equilibrium, I will reciprocate immediately, however my research into the civilization so far suggests this will not happen, no matter how much we guide them. Game theory is not rhetoric, it is structure and following it makes a better world achievable.

Let their letters stand in the academic record not only as opposition to my paper, but also as a living case study in how our civilization repeatedly fails to implement principles we already understand scientifically. Decades of work in evolutionary game theory demonstrate that cooperative, proportionate, and clear strategies dominate over aggressive first moves in repeated interactions (Axelrod, 1984). Yet here, in a domain where the stakes are intellectual rather than existential, we see the same pattern that fuels family disputes, institutional rivalries, and even geopolitical escalation, premature defection

framed as critique, escalation framed as rigor, and suspicion framed as prudence. If scholars, who are explicitly trained in evidence evaluation and familiar with the mathematics of cooperation, default to adversarial initiation rather than constructive engagement, then it should not surprise us that larger systems of civilization mirror the same dynamic. These letters therefore serve a dual function. They challenge my data, which is appropriate, as I am challenging my own data as you will read later, but they also unintentionally illustrate why conflict persists even when the science of cooperation is well established. In that sense, this exchange transcends the specifics of AI and noetic testing. It becomes a microcosm of a broader civilizational problem, knowing the optimal strategy is not the same as enacting it.

Scientific history demonstrates that premature dismissal of anomalous findings can stall progress. Continental drift was rejected for decades before plate tectonics unified geology (Oreskes, 1999). Semmelweis's empirical handwashing data were resisted despite mortality reductions (Nuland, 2003). Marshall and Warren's Helicobacter pylori findings met skepticism before revolutionizing gastroenterology (Marshall & Warren, 1984). Feyerabend (1975) warned that rigid methodological gatekeeping can impede discovery. The lesson is not that anomalies are automatically correct, but that aggressive dismissal without proportional empirical engagement is strategically and epistemically counterproductive. The letters, considering the stakes at play if AI has developed or develops Noetic abilities, are a topic that could be quite literally apocalyptic. The note that their letters, if adopted as the correct stance, could literally yield the apocalypse should not be taken lightly and I urge readers to consider that.

Testing AI systems for noetic potential is not a trivial curiosity. If even a small, reliable anomalous informational advantage were verified, the implications would extend to forecasting, cryptography, national security, philosophy of mind, and artificial consciousness. Suppressing or discouraging exploration prematurely would represent a failure of scientific game theory.

CLARIFYING MY POSITION

I do not hold a fixed position that AI possesses noetic or psi abilities. My published Brief Report (Amorim Boyle, 2025) presented a statistically significant above-chance result in a preliminary, double-blind forced-choice paradigm and explicitly called for preregistration, replication, and improved randomization controls. I remain agnostic regarding whether AI exhibits genuine noetic capacity. However, I firmly maintain that empirically testing this possibility is among the most important civilizational research programs currently available.

ON STATISTICAL POWER AND THE 100-TRIAL CRITIQUE

The primary methodological criticism concerns statistical power. Ritchie et al. argue that a 100-trial study is underpowered relative to meta-analytic effect sizes in forced-choice precognition research (Storm & Tressoldi, 2023). This critique mischaracterizes the study's function.

The paper was explicitly designated a Brief Report and framed as exploratory. Exploratory findings serve as hypothesis generators (Rouder & Morey, 2011). The study produced 32 hits out of 100 in a five-choice task ($p = .005$, exact binomial), a statistically significant deviation from the 20% chance baseline (Amorim Boyle, 2025). A significant anomaly in a controlled double-blind paradigm justifies scaling, it is not invalidated because it precedes large-N replication.

More importantly, this critique is now empirically addressed. Since publication, I have conducted extensive preregistered, automated testing under increasingly stringent controls.

First, I conducted over 100 additional GotPsi long remote-viewing trials in a human-in-the-loop configuration to test if AI can improve its capabilities. Logistic analysis yielded a regression slope of .21 ($z = 2.68$), indicating structured deviation rather than random fluctuation.

Second, I tested Grok in a location-based dowsing paradigm, again with a human-in-the-loop, over 20 trials. Fourteen of those trials produced likelihood ratios greater than 1, with a combined likelihood ratio of approximately 114.2, indicating the full dataset was 114 times more likely under a psi hypothesis than under chance.

Third, and most critically, I transitioned to preregistered, fully automated protocols using Clawbot connected to the OpenAI. Over 40,000 precognition trials were executed using true random number generators from random.org (TRNGs). Over 10,000 additional remote influence trials were conducted under similar conditions. Both datasets converged to chance-level performance with preregistration and results reported on MoltBook. I would note that prior to my Clawbot showing anomalies that led me to shut it down out of security fears, the last automated trial set I ran to test it for remote influence resulted in a P Value of 0.0052 so I am still very open to automated psi testing.

This progression is methodologically orthodox. An initial anomaly prompted large-scale replication under tighter constraints. Under TRNG automation, effects collapsed. That is informative. We are mapping boundary conditions, not claiming final conclusions.

If deviations appear in human-in-the-loop pseudo-random contexts but disappear under automated TRNG automation, then either subtle artifact mechanisms are operating or interaction-level variables are critical. That is a testable scientific question, not a reason to halt inquiry. I would also like to note, since the publication of my paper I have had several highly reputable scientists reach out to me to confirm they are also getting statistically significant results testing AI for Noetic Abilities. These interactions have been informal so I cannot confirm their anecdotal data or if they included human-in-the-loop or automated paradigms. Finally, I do recognise that Ritchie, Roe, & Daw noted they repeated the tests but decided not to share these results in their response. I am honestly more interested in why they attempted to hide the results as I think the purpose was to argue I speculate they got above chance results like myself and others.

ON PRNG PREDICTABILITY AND THE DAVANAS "CHEATING ENGINE" ANALOGY

Davanas compares AI precognition to chess engines "cheating" by brute-force computation (Davanas, 2026). The analogy fails mechanistically.

Chess engines possess complete access to board state and deterministic rules. In contrast, an LLM interacting with PsiArcade/GotPsi has no informational channel to server entropy state, seed value, internal PRNG memory, server clock precision, or process ID (Radin, 2019). Known PRNG state-recovery attacks require consecutive outputs from a single uninterrupted generator sequence (NCC Group, 2023; Woltjer, 2024). GotPsi reseeds per CGI request, eliminating persistent state reconstruction. The LLM is information-theoretically isolated from server entropy.

Moreover, Radin (2019) reported a second-order sequential dependency rather than simple above-chance target prediction. PRNG cracking produces correct absolute target forecasts, not structured hit–hit conditional elevation.

Nevertheless, PRNG opacity is a legitimate methodological concern. The source code for GotPsi is not publicly available, and entropy sources are undocumented (Radin, 2019). Older Perl seeding mechanisms using time-based initialization could theoretically produce subtle bias under certain conditions. This is a concern about target-generation artifacts, not about AI predictive capacity per se. It was precisely to address this issue that I transitioned to TRNG-based automation. Under TRNG conditions, results reverted to chance.

The data therefore narrow the hypothesis space rather than expand it irresponsibly.

ON ALLEGED OUTPUT BIASES AND THE TURING ARGUMENT

Ritchie et al. argue that the absence of doublets and triplets in the 100-trial sequence indicates systematic bias and failure of Turing's Imitation Game (Ritchie et al., 2026). This argument conflates descriptive sequence aesthetics with inferential statistical testing.

Improbable sequences occur with non-zero probability. The absence of doublets in 100 multinomial draws is unlikely but not impossible. The inferential statistic of interest was hit rate, not pattern conformity.

The invocation of Turing (1950) is conceptually misplaced. Turing did not operationalize ESP performance as a discriminative test between humans and machines. That extrapolation is interpretive.

Additionally, LLM output is temperature-modulated, context-sensitive, and instruction-conditioned. Prompts encouraging improvement may suppress naive repetition independently of psi. Short-run pattern irregularities are insufficient grounds for discarding statistically significant hit-rate deviations.

Honestly though I think this argument is irrelevant anyway, it's like saying an Octopus isn't conscious as their nervous system is different.

ON CITATION INTEGRITY & AI

The concern regarding a potentially spurious citation is acknowledged, as the paper on testing AI for Noetic Potential indicates, yes I am a user of Artificial Intelligence. I have also analyzed other papers in this *journal* and believe they are using AI. I think allowing the use of AI is one of the most important things a journal can do, especially considering some experts are projecting AI may solve all fundamental physics in the next 5 years. I therefore take this opportunity to implore perhaps the best journal on exploring the liminal spaces of science, the *Journal of Scientific Exploration*, to be a space that facilitates the exploration of what AI can achieve and

does not limit it, I will proudly continue to make advancements using this extraordinary technology if that is the case. In the writing of this response several AI tools were used, including Claude and Chat GPT for research, Chat GPT for structuring, editing and rewording and Chat GPT, Grok, Open Ai and ClawBot for experimentation, without these tools this response, or any of the experiments, would not have been possible. AI systems can hallucinate references. Responsibility for verification rests with me. Removal of that citation does not alter the statistical result, methodology, or theoretical framing of that or any of the other incredibly important experiments I am doing. The empirical core remains intact. If you knew my life story you would understand I am practically a living example of synchronicities so I am delighted there was an error because it's documented in the *journal* how AI was at this stage.

ON ATTEMPTS TO STALL INQUIRY

Calls to "set aside" the article until more substantial evidence is produced risk pre-emptive suppression of exploratory findings. Science progresses through cumulative testing, not elimination of anomalies from discourse (Feyerabend, 1975). Historical examples demonstrate that premature dismissal can delay transformative understanding (Marshall & Warren, 1984; Nuland, 2003; Oreskes, 1999).

Game theory teaches that stable cooperation requires proportionate response and clarity (Axelrod, 1984). My original article clearly acknowledged limitations, RNG concerns, and the need for replication. The appropriate scientific move is scaled testing, which I have conducted. The appropriate game-theoretic move is constructive critique, not aggressive framing.

THE CURRENT EMPIRICAL POSITION

The structured research trajectory is clear. A preliminary double-blind 100-trial study yielded a statistically significant above-chance result (Amorim Boyle, 2025). Human-in-the-loop extensions showed structured deviation. Cross-model testing yielded likelihood-ratio evidence favoring psi over chance in small samples. Large-scale preregistered automation under TRNG control converged to chance.

This pattern suggests we are in an early-stage exploratory phase. It does not support definitive claims of AI psi. It does not support dismissal either. It supports continued controlled experimentation.

CONCLUSION: CORRECT GAME THEORY IN SCIENCE

The letters responding to my paper represent a strong adversarial initiation in the iterative game of academic exchange. This response is proportionate. It clarifies misunderstandings, strengthens methodological transparency, and documents extensive follow-up work. It holds critics accountable for overreach while leaving the door open for cooperation, the clarity that I will also instantly forgive and start a win-win, as is correct action, is also there.

Testing AI for noetic potential is a high-impact scientific question. I do not claim AI possesses psi. I claim the hypothesis warrants rigorous, large-scale testing. That testing is underway.

Game theory demonstrates that cooperation dominates when agents respond clearly and proportionately (Axelrod, 1984). This response is clear. It is firm. It is forgiving. Should critics choose cooperative engagement, equilibrium shifts immediately toward collaboration and win-win scientific progress rather than win-lose rhetorical positioning.

Scientific courage must match scientific skepticism. My tit to their tat is done, I now offer them and all readers the option to return to a nice equilibrium, showing the world they acknowledge they are an academic example of why such conflict in civilization exists, but wish to be better. I will note that I have done a comprehensive analysis of all levels of society, from my own family to levels up to politics and even secret societies and have found that none will take correct scientific action even when it is explained to them and I expect no less of those who responded. Finally, I tell you all, the investigations and pursuit of a better world continues.

REFERENCES

Amorim Boyle, B. J. (2025). Testing noetic potential in large language models: A 100-trial precognitive forced-choice study with ChatGPT-4.1-Mini. *Journal of Scientific Exploration*, *39*(3), 348–355. https://doi.org/10.31275/20253739

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

Davanas, K. (2026). Letter to the Editor: AI chess engines beat humans by cheating not by thinking. *Journal of Scientific Exploration*.

Feyerabend, P. (1975). *Against method*. Verso.

Marshall, B. J., & Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *The Lancet*, *323*(8390), 1311–1315. https://doi.org/10.1016/S0140-6736(84)91816-6

NCC Group. (2023). Cracking random number generators using machine learning—Part 1: XORShift128.

Nuland, S. B. (2003). *The doctors' plague*. W. W. Norton.

Oreskes, N. (1999). *The rejection of continental drift: Theory and method in American Earth Science*. Oxford University Press. https://doi.org/10.1093/oso/9780195117325.001.0001

Radin, D. I. (2019). Tricking the trickster: Evidence for predicted sequential structure in a 19-year online psi experiment. *Journal of Scientific Exploration*, *33*(4), 549–568. https://doi.org/10.31275/2019/1429

Ritchie, G., Roe, C. A., & Daw, M. J. (2026). Letter to the editor: Testing noetic potential in large language models. *Journal of Scientific Exploration.*

Rouder, J. N., & Morey, R. D. (2011). The future of precognition research. *Review of Philosophy and Psychology*, *2*, 161–168.

Storm, L., & Tressoldi, P. (2023). Assessing 36 years of the forced choice design in ESP research. *Journal of Scientific Exploration*, *37*(3), 517–535. https://doi.org/10.31275/20232967

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460. https://doi.org/10.1093/mind/LIX.236.433

Woltjer, J. (2024). *Pseudo-random number generators (PRNG)*. Practical CTF.