



## AI EVALUATION OF THE *DRAKE-S* EQUATION FOR POSTMORTEM SURVIVAL AGAINST SUDDUTH'S EVIDENTIARY STANDARDS

BY JAMES HOURAN AND BRIAN LAYTHE

### ABSTRACT

We applied two advanced large-language models—OpenAI's ChatGPT-4o and GitHub Copilot—to evaluate the Drake-Survival (*Drake-S*) equation framework for postmortem survival of consciousness relative to eight previously specified evidentiary standards in the survival domain. ChatGPT-4o judged this equation approach 'Good Fit' to 'Very Good Fit' on 75% of the criteria (23/32 points, 72% overall), whereas Copilot rated it 'Good Fit' to 'Very Good Fit' on 88% of the criteria (26/32 points, 81% overall). Both programs flagged two consistent weaknesses: the use of legal benchmarks (e.g., the Daubert standard) as an evidential standard and the apparent assumption that gaps in known confounds imply the increased odds of survival, reflecting concerns about inappropriate criteria and probabilistic fallacies. Each program suggested methodological refinements—chiefly, estimating covariation among error terms, pre-registering and calibrating scoring rubrics, and embedding Bayesian updating (e.g., priors, likelihood ratios, and credible intervals)—to strengthen the framework's logical rigor and evidentiary justification. Although the *Drake-S* approach does not demonstrate ontological survival, it highlights a substantial residual (30.3%) of data that ostensibly eludes major materialist and psi-only explanations. Rigorous study of these anomalies should therefore continue in anticipation of chance- or challenge-type discoveries that can meaningfully advance our understanding of the nature or limits of human consciousness.

### INTRODUCTION

"Science has never made a discovery without being authorized and encouraged thereto by philosophy" (Thomas Mann, 1937, p. 23).

Consciousness studies is an interdisciplinary field—drawing partly on neurology, philosophy, parapsychology, cognitive science, and artificial intelligence (AI)—that aims to understand the origin, mechanisms, and boundaries of human consciousness and also perhaps non-human intelligences (Overgaard, 2017; Schurger & Graziano, 2022; Wahbeh et al., 2022). This broad territory basically condenses to ten focus areas: (1) *nature of consciousness* (i.e., defining consciousness and distinguishing it from unconscious states); (2) *levels of consciousness* (i.e., investigating and mapping different states of

consciousness, such as wakefulness, sleep, dreaming, and altered states induced by meditation or substances); (3) *neural correlates* (i.e., identifying the brain structures and neural mechanisms associated with conscious experience such as perception, thought, and emotion); (4) *philosophical questions* (i.e., addressing fundamental questions about the nature of mind and consciousness, such as theories of mind (dualism, physicalism, panpsychism) and the hard problem of consciousness or ‘why and how’ the physical processes in the brain generate subjective experience); (5) *psychological perspectives* (i.e., examining how consciousness relates to mental processes like attention, memory, or decision-making); (6) *developmental aspects* (i.e., understanding how consciousness develops and changes over a lifetime); (7) *comparative consciousness* (i.e., comparing human consciousness with the behavior and brain functions of non-human species); (8) *artificial consciousness* (i.e., exploring the possibility of creating sentience or simulations of conscious behavior in machines or software); (9) *clinical implications* (i.e., applying findings to medical and therapeutic contexts); and (10) *ethical and social implications* (i.e., considering the broader impacts of consciousness studies on society, including debates about the moral status of conscious beings, the ethical treatment of animals, and the implications of potentially conscious artificial intelligence).

A controversial topic that arguably bridges many, if not all, of these focus areas is the profound question of survival of consciousness after permanent bodily death. This is the idea that human consciousness or its substrate involves expanded, emergent, or non-local properties (Cardeña, 2018; Parker & Brusewitz, 2003; Vernon, 2021) that are unbounded by the activity of a living brain or body (Barušs, 2023; for a recommended reading list on the survival hypothesis, see Houran et al., 2023: Appendix, p. 27). The same might likewise apply to non-human animal consciousness (Matlock et al., 2024; Matlock & Hilton, 2025). The various biological and philosophical implications of postmortem survival would certainly be paradigm-shifting if validated. Although some large-scale surveys suggest that afterlife-related beliefs in the general population have steadily declined (e.g., Duffy et al., 2023; Gallup, 2023), the overt academic interest in the survival (or afterlife) hypothesis has seemingly increased, as evidenced by the myriad of studies and commentaries appearing in psi-oriented periodicals (e.g., Beischel et al., 2021), the mainstream literature (e.g., Hashemi et al., 2023), and specialized journals (e.g., Shushan, 2024). In fact, the Bigelow Institute of Consciousness Studies’ (BICS) essay contest (Kelleher & Bigelow, 2022) on the best available evidence purportedly gathered scientific support for the survival hypothesis that was “beyond a reasonable doubt.” The data or arguments from the various authors’ winning entries nonetheless have been severely criticized on conceptual, methodological, and analytical grounds (Augustine, 2022; Sudduth, 2024; Tressoldi et al., 2022). Readers might therefore rightfully question whether there is any legitimate anomaly (or set of anomalies) in need of a scientific explanation, i.e., do any

compelling arguments or empirical evidence justify the survival hypothesis as an acceptable, or even fruitful, research topic in the field of consciousness studies?

To this point, Laythe and Houran (2022) developed a Drake-Survival (*Drake-S*) equation<sup>1</sup> as mathematical evidence that absolutist claims (pro or con) about survival-related data are presently insufficient and unwarranted. They specifically sourced hundreds of peer-reviewed studies for a probabilistic exercise that found the representative published effect sizes of several known confounds (KCs)—i.e., environmental variables, suggestion-expectancy effects (i.e., contagion, memory, and persuasion), fraud, measurement error, mental illness, and susceptibility to perceptual aberrations—explained only 61.4% of the variance in the published prevalence rates of certain anomalous experiences (AEs) that are traditionally interpreted as evidence of survival, i.e., near-death experiences, cases of the reincarnation type, haunt-poltergeist episodes, channeling or mental mediumship, and veridical anomalous experiences like after-death communications (for detailed definitions of all these phenomena, see Laythe & Houran, 2022). Rock et al. (2023) later revised this original calculation to include the contribution of putative living agent psi (LAP)<sup>2</sup> to such reports. This increased the overall explanatory power to 69.7% of the variance, but the projected upper limit of LAP failed to fill the mathematical gap. These findings suggest that both the biomedical (i.e., *extinctionist-related*) and parapsychological (i.e., *survivalist-related*) sciences require more specific or comprehensive models to adequately explain the AEs in question. Likewise, other researchers have since explored the ‘fit’ of psi and survival-type phenomena to various frameworks in modern science (see e.g., Boccuzzi, 2024).

#### SUDDUTH’S EIGHT EVIDENTIARY STANDARDS FOR SURVIVALIST PROPOSITIONS

Philosopher and researcher Michael Sudduth’s (2024) major treatise on the BICS contest sought to clarify why advancements on the survival question have stagnated. Houran (2024) outlined several logical reasons to adopt Sudduth’s perspectives, given his teaching and publishing background in the areas of critical thinking and epistemology, with a focus on theories of evidence and the justification of belief across different domains of inquiry, including general and legal epistemology, topics in the philosophy of religion (Sudduth, 2009), and, more recently, postmortem survival (Sudduth, 2016). Moreover, Sudduth (2024)

---

1. This alludes to the famous Drake equation—a probabilistic formula developed by astronomer Frank Drake (1965) to estimate the number of active, communicative extraterrestrial civilizations in the Milky Way galaxy. It combined factors like the rate of star formation, the fraction of stars with planets, and the likelihood of life developing to provide a framework for exploring the possibility of alien life. This framework is still used today, though sometimes with important modifications (e.g., Benford, 2021) or applications to broader cosmological issues (e.g., Sorini et al., 2024).

2. LAP is the idea that psi-type phenomena are caused by the mind or consciousness of a living person (i.e., the agent). Thus, putative psi abilities, such as influencing objects or gaining information beyond normal sensory means, come from living agents rather than external entities or spirits.

supplemented his critique with eight guidelines (denoted here as ‘evidentiary standards’) for bolstering current arguments and future studies (pp. 516–518):

- (1) *Logical architecture*. Do not pile up facts without clearly structured arguments. Arguments should present premises, conclusions, and sub-arguments in standard forms to avoid confusion.
- (2) *Critical assessment of objections*. It is essential to distinguish between outright dismissal of the survival hypothesis and claims that survivalists have failed to provide adequate justification. Different types of skepticism require distinct responses.
- (3) *Logical fallacies*. Do not make weak inferences, such as assuming that non-survival hypotheses must be false if they do not account for certain observations, or asserting the survival hypothesis is ‘most probable’ based on a thin evidential base.
- (4) *Distinction of evidence and belief*: Applying Royall’s (1997) insight, separate what the evidence implies from what one believes. Evidence may favor a hypothesis without mandating belief in it.
- (5) *Conceptual clarity*. All conclusions require clear criteria, such as explaining what is meant by ‘evidence’ and how strong it is considered to be.
- (6) *Inappropriate evidential criteria*. Do not rely on legal standards (e.g., ‘beyond a reasonable doubt’), which have an epistemic basis that is rigorous and inappropriate to survival-related evidence.
- (7) *Revisiting inference to best explanation (IBE)*. Strengthen IBE arguments by merging them with Bayesian probability or avoid IBE, as traditional IBE is self-defeating and lacks probability value.
- (8) *Adopting proper Epistemic principles*. Apply epistemic frameworks like the law of likelihood or Bayesian reasoning, which provide structure for evaluating evidence and making belief-based assessments.

These eight points or standards essentially define a structured evaluation of pro-survival arguments or studies, focusing on their logical coherence, evidential support, and interpretive clarity. They serve as a flexible and valuable tool that can be applied either (a) retrospectively through a content analysis to gauge the robustness of past research or (b) prospectively to support quality control in the development of new studies or arguments.

#### THE PRESENT STUDY

We heed Altenmüller et al.’s (2021) call for more critical reflection of one’s own research findings and conclusions. This view draws on the principles of research transparency (Beaty, 2024) and intellectual humility (Bak et al., 2022), i.e., recognizing that personal assumptions or viewpoints might be wrong or incomplete and therefore remaining open to revising those views in light of new information. Accordingly, we conducted an AI-based content analysis of our latest presentations of the *Drake-S* equation (Houran et al., 2023; Rock et

al., 2023). This exercise essentially constitutes a self-initiated, post-publication review (O'Sullivan et al., 2021) that assesses the underlying premise (i.e., data, methodology, and main conclusion) of our probabilistic approach to the survival question relative to Sudduth's (2024) evidentiary standards. We predicted no particular outcome(s) from this self-critique (Part 1), but similar to Houran et al. (2023), we aimed to modify our prior probability calculations in accordance with Part 1's results or insights as appropriate (Part 2). This study's design, analysis, and research materials were not pre-registered, but we strived to follow the Journal Article Reporting Standards (Kazak, 2018) in describing how we determined our research samples, all data exclusions (if any), specific research questions, applicable manipulations, and all measures and data abstractions.

#### **PART 1: AI ANALYSIS OF THE *DRAKE-S* EQUATION**

AI has emerged as a valuable tool for conducting content and rating analyses (Gupta et al., 2024; Kar et al., 2023; Platt & Platt, 2023), offering unique advantages that extend beyond human evaluators. One primary benefit is its consistency and objectivity. Unlike humans, who can bring personal biases and varied interpretations to the evaluation process, AI systems tend to apply criteria uniformly across all items, reducing subjective influence and ensuring more reliable outcomes (Kaur et al., 2022; Lee et al., 2020). When supplied with explicit rating rubrics (e.g., four-point rating scales), these programs indeed can reliably map text onto such anchors, minimizing drift in standards over time (Hamilton et al., 2023). This is particularly valuable for projects where consistency in ratings is critical. Another advantage of these programs is their speed and scalability. AI can analyze large volumes of data quickly, making it an ideal solution for projects involving thousands of documents or ratings. Tasks that might take days or weeks for human evaluators to complete can be accomplished by AI in a fraction of the time. This utility enables researchers to analyze datasets that are otherwise impractical to review manually (Lee et al., 2020). As a result, AI is generally cost effective and eliminates the potential for human fatigue or cognitive load in evaluation tasks that can subsequently lead to inconsistencies or errors.

#### *Method*

##### AI programs

To promote scoring consistency and minimize rater bias, the content evaluations were conducted using two advanced large language models (LLMs—advanced AI systems trained to understand and generate human-like text), i.e., ChatGPT-4o and GitHub Copilot. ChatGPT-4o, a state-of-the-art multimodal model developed by OpenAI (2024), which some researchers contend is 'currently the most powerful generative AI model available for public use' (Pérez-Núñez, 2023, p. 355), was contextually primed with Sudduth's (2024) full article describing his evaluation criteria. Subsequently, GitHub

Copilot (Microsoft, 2025), an AI assistant powered by OpenAI's Codex model (Chen et al., 2021), was employed to independently review and replicate the scoring process, providing a complementary perspective based on its distinct architecture and training data.

This dual-model approach enabled cross-validation of outputs, facilitating the detection of consensus or discrepancies arising from model differences. We should note that although both ChatGPT-4o and GitHub Copilot are developed and maintained by Microsoft, they are optimized for distinct purposes—i.e., general-purpose reasoning and conversational synthesis versus code generation and technical assistance, respectively. While architectural overlap is possible, their divergent fine-tuning and deployment contexts provide a credible basis for independent evaluations. The use of both models serves as a pragmatic form of AI corroboration under shared scoring constraints, allowing for internal cross-validation of theory ratings.

### Literature set

We used four sets of textual data for the analysis: (a) three peer-reviewed journal articles that specified the rationale, construction, refinement, and results of the *Drake-S* equation (Houran et al., 2023; Laythe & Houran, 2022; Rock et al., 2023), as well as (b) Sudduth's (2024) extensive philosophical critique of evidentiary reasoning in the survival domain that culminated in his eight standards.

### Standardized rating system

To systematically evaluate the comparative fit of the *Drake-S* equation to eight evidentiary standards, we developed a study-specific rating system administered via the ChatGPT-4o LLM. Each domain is rated on a four-point Likert scale ranging from 1 (Very Poor Fit) to 4 (Very Good Fit). Detailed anchor descriptions were provided for each scale point within every domain to ensure consistent interpretation and application. This standardized rating approach enabled a replicable and transparent AI-driven evaluation aligned with established epistemic criteria, facilitating structured comparison of theoretical models within the study's framework:

#### (1) *Logical architecture:*

- 1 (Very Poor Fit). The argument is scattered, with premises and conclusions unclear or jumbled. It seems like a list of loosely related observations without a clear progression or logical structure.
- 2 (Poor Fit). The main argument is partially organized, but sub-arguments are not well distinguished from the main points, creating confusion about the argument's actual structure.
- 3 (Good Fit). The study has a reasonably clear structure, with main premises and conclusions stated. However, some parts could use more clarity, especially in distinguishing sub-arguments.

- 4 (Very Good Fit). The study presents a well-organized argument, with clear premises, conclusions, and sub-arguments that follow a logical flow, making the argument easy to follow.
- (2) *Critical assessment of objections:*
- 1 (Very Poor Fit). The study dismisses objections outright without addressing them substantively or confusingly conflates different forms of skepticism, suggesting a lack of engagement with counterarguments.
  - 2 (Poor Fit). Objections are mentioned but inadequately addressed, or the study often conflates dismissal of rival hypotheses with failure to establish the survival hypothesis as true.
  - 3 (Good Fit). The study acknowledges and reasonably addresses major objections, with some effort to distinguish skepticism types, but could improve in depth.
  - 4 (Very Good Fit). The study thoroughly engages objections, carefully distinguishing between different types of skepticism and addressing each in a thoughtful, structured way.
- (3) *Logical fallacies:*
- 1 (Very Poor Fit). The study is riddled with fallacious reasoning, including major errors like assuming rival hypotheses are false because they lack explanatory power, or making sweeping generalizations about survival likelihood.
  - 2 (Poor Fit). Some logical fallacies are present, such as overstated conclusions or reliance on weak probabilistic reasoning, undermining the argument's credibility.
  - 3 (Good Fit). The study avoids major fallacies and presents logical reasoning, though there are minor issues in how it handles probability and inference.
  - 4 (Very Good Fit). The study is free from logical fallacies, presenting cogent reasoning and careful probabilistic assessments without overstating conclusions.
- (4) *Distinction of evidence and belief:*
- 1 (Very Poor Fit). The study fails to distinguish between what the evidence suggests and what we should believe, treating any evidence as a reason for full belief in the survival hypothesis.
  - 2 (Poor Fit). The study sometimes confuses evidence with belief, implying that supporting evidence mandates belief without clarifying the level of support required.
  - 3 (Good Fit). The study generally distinguishes evidence from belief, though it could benefit from clearer guidance on the implications of the evidence relative to belief.

- 4 (Very Good Fit). The study makes a clear distinction between evidence and belief, explaining what the evidence suggests independently of what we might believe about survival.
- (5) *Conceptual clarity:*
- 1 (Very Poor Fit). The study's definitions and key concepts are vague or missing, making it difficult to understand the conclusions or evidentiary standards being used.
  - 2 (Poor Fit). Some conceptual clarity is present, but key terms like 'evidence' and 'explanatory power' are inconsistently defined or applied, leading to interpretive ambiguity.
  - 3 (Good Fit). Concepts are mostly clear and defined, with some ambiguity in certain terms, but overall conclusions are understandable.
  - 4 (Very Good Fit). The study is precise in defining concepts and terms, making its conclusions easy to interpret and its use of terms like 'evidence' and 'explanatory power' transparent.
- (6) *Inappropriate evidential criteria:*
- 1 (Very Poor Fit). The study heavily relies on irrelevant standards (like legal proof standards) to evaluate evidence, creating a misleading impression of evidentiary support.
  - 2 (Poor Fit). The study includes some inappropriate criteria, such as vague comparisons to legal standards, although it attempts to make epistemic points.
  - 3 (Good Fit). The study mostly uses appropriate evidential standards, though there are minor references to non-relevant criteria that could be omitted.
  - 4 (Very Good Fit). The study relies solely on relevant and well-justified evidential standards without any inappropriate comparisons to legal or other unrelated standards.
- (7) *Revisiting IBE:*
- 1 (Very Poor Fit). The study uses IBE heavily but without any probabilistic grounding, ignoring key philosophical critiques, leading to a weak foundation for its conclusions.
  - 2 (Poor Fit). IBE is used without sufficient support, and the argument does not address the limitations of IBE or include Bayesian elements to strengthen the analysis.
  - 3 (Good Fit). IBE is reasonably employed, with some acknowledgment of limitations, though it could be improved by integrating Bayesian or probabilistic frameworks.
  - 4 (Very Good Fit). The study uses a well-supported form of IBE, incorporating Bayesian analysis or addressing philosophical critiques to strengthen its inferential framework.
- (8) *Proper epistemic principles:*



- 1 (Very Poor Fit). The study disregards or misapplies epistemic principles, relying on ad hoc reasoning rather than a structured approach like the law of likelihood or Bayesian reasoning.
- 2 (Poor Fit). The study references relevant principles but applies them inconsistently or without proper justification, reducing the argument's credibility.
- 3 (Good Fit). Epistemic principles are applied fairly well, with some gaps in consistency or rigor, but the framework remains generally sound.
- 4 (Very Good Fit). The study consistently applies well-chosen epistemic principles, like Bayesian or likelihood reasoning, giving a solid foundation to its conclusions.

### Procedure

Rather than formal model training, we employed contextual priming whereby the LLMs were supplied with selected literature (described above) to inform their evaluations of the *Drake-S* equation relative to Sudduth's (2024) evidentiary standards for survival research. The evaluation procedure involved a five-step protocol:

(1) *Document ingestion and segmentation.* Source materials were segmented into blocks of approximately 1,500–2,500 tokens to accommodate the LLM's input constraints. Segments were grouped by thematic relevance (e.g., KCs, inferential gaps, Bayesian modelling, and LAP) and annotated with brief metadata tags to facilitate targeted prompting.

(2) *Prompt construction and rating procedure.* For each of Sudduth's (2024) eight evidentiary standards, Each LLM was prompted using a standardized instruction format (i.e., 'Based on the provided literature excerpts, evaluate how well the *Drake-S* equation satisfies the evidentiary standard of [e.g., logical architecture]. Use the four-point Likert scale defined below and provide a justification for your rating'). Each evaluation prompt included (a) the target standard and its operational definition, (b) the four-point Likert rating rubric (ranging from 'Very Poor Fit' to 'Very Good Fit'), and (c) selected, thematically relevant text segments from the literature set. This process was iterated independently for each standard.

(3) *Model configuration parameters.* The analyses were conducted in the LLMs' standard environments using the following parameters: (a) temperature = 0.7, (b) Top-p = 1.0, (c) Max tokens = adequate for full reasoning generation, and (d) System message = "You are an expert in epistemology and scientific methodology evaluating theoretical research on postmortem survival."

(4) *Human oversight and verification.* All outputs were independently reviewed by the two authors for accuracy, interpretive consistency, and fidelity to the rubric. While no AI responses were altered, discrepancies or ambiguities

were resolved via consensus. This ensured that the integration of AI remained transparent, methodologically sound, and aligned with ethical guidelines (cf. Hamilton et al., 2023; Resnik et al., 2025).

(5) *Reproducibility and documentation.* To facilitate replication and external evaluation, an example prompt–response transcript from the ChatGPT-4o evaluation—specifically, the analysis of the ‘logical architecture’ standard—is presented in the Appendix. The full process is formally summarized below as a sequential evaluation expression:

literature input → prompting strategy → evaluation output →  
human review → final scoring

## *Results and conclusions*

### ChatGPT-4o evaluation

Table 1 shows that the *Drake-S* equation scored ‘Good Fit’ to ‘Very Good Fit’ on six (or 75%) of Sudduth’s eight standards. The averaged AI rating across all eight standards was 2.88 (on a 1–4 scale), and the raw total was 23 out of a possible 32 points, or an overall 72% fit to the entire set of standards. These findings indicate that our equation framework has good conceptual clarity and engagement with objections. Accordingly, ChatGPT-4 concluded that the *Drake-S* approach was a reasonably robust and structured basis for abstracting or exploring the question of survival.

But two standards (i.e., 25%) received lower ratings, given potential issues with ‘inappropriate evidential criteria’ (i.e., application of legal standards) and ‘logical fallacies’ (i.e., lack of preregistration and scientific gaps do not imply survival evidence). Another standard (‘revisiting IBE’) received a ‘Good Fit’ rating but with the caveat that a Bayesian integration might “strengthen inferential support and address critiques regarding the limits of explanatory power in this context” (cf. Table 1). Part 2 therefore addresses these and other suggested areas for improvement that followed from the other AI analysis below.

### Copilot evaluation

GitHub Copilot, an AI model based on OpenAI’s Codex architecture and optimized for technical reasoning, independently evaluated the *Drake-S* equation framework using the same rubric and literature set provided to ChatGPT-4o (Table 2). Copilot rated the framework as ‘Good Fit’ to ‘Very Good Fit’ on seven of the eight standards (88%), yielding an average rating of 3.25 (on a 1–4 scale) and a total of 26 out of 32 points (81% overall fit). This high rating affirms the *Drake-S* equation’s structural soundness and conceptual defensibility.

Notably, ‘logical architecture’ received a top rating (4), with Copilot praising the framework’s transparent assumptions, defined variables, and stepwise derivations. ‘distinction of evidence and belief’ (4) was likewise commended for avoiding conflations between probabilistic outcomes and ontological claims.

TABLE 1.

*ChatGPT-4o ratings of the Drake-S equation approach compared with Sudduth's evidentiary standards*

Evidentiary standard	Rating*	Rating justification
Logical architecture	3	The <i>Drake-S</i> equation framework is structured with clear premises and logical progression. It defines the survival hypothesis and systematically accounts for KCs. However, it could improve by more explicitly delineating sub-arguments related to each confound and how they fit into the broader survival analysis, as noted in critiques around variable co-variation
Critical assessment of objections	3	The <i>Drake-S</i> equation addresses various objections, particularly skepticism around KCs and alternative explanations like LAP. The approach engages with critiques, but the authors admit that further work is needed to clarify distinctions between survival evidence and the potential effects of psi abilities of living individuals
Logical fallacies	2	There are issues with potentially fallacious reasoning, such as overstating the ability of confound exclusions to support the survival hypothesis and a lack of pre-registration in some studies, which commentators identified as a limitation that might introduce bias. This suggests a need for caution against assuming that gaps in confounds automatically point toward survival evidence
Distinction of evidence and belief	3	The approach largely distinguishes between what evidence indicates and what should be believed about survival, acknowledging that evidence from AEs could favor survival-related interpretations without compelling belief. However, some conclusions imply that residual variance unaccounted for by confounds supports belief in survival, which blurs this distinction slightly
Conceptual clarity	4	The framework is conceptually clear in defining survival-related phenomena and associated KCs, providing specific categories and examples (e.g., near-death experiences, reincarnation cases). The approach also includes clarifications around variables like LAP, contributing to a clear overall structure
Inappropriate evidential criteria	2	There is some reliance on legal standards, with references to the Daubert standard and its application to survival data. While meant to strengthen evidential quality, this may not be entirely appropriate in a scientific context. The legal analogy could be misleading, as scientific inquiry into survival differs fundamentally from legal proof

*Continued*

TABLE 1—continued.  
*ChatGPT-4o ratings of the Drake-S equation approach compared with Sudduth's evidentiary standards*

Evidentiary standard	Rating*	Rating justification
Revisiting IBE	3	The <i>Drake-S</i> equation approach uses probabilistic reasoning rather than pure IBE, which is advantageous. However, it could benefit from additional Bayesian integration to strengthen inferential support and address critiques regarding the limits of explanatory power in this context
Proper epistemic principles	3	The <i>Drake-S</i> equation generally applies appropriate epistemic principles, relying on probabilistic estimates and empirical data. Nevertheless, improvements could include tighter methodological controls and adjustments for co-variation among variables to further refine the survival probability estimates

\*Key: 1 = Very Poor Fit, 2 = Poor Fit, 3 = Good Fit, 4 = Very Good Fit.

TABLE 2.  
*Copilot ratings of the Drake-S equation approach compared with Sudduth's evidentiary standards*

Evidentiary standard	Rating*	Rating justification
Logical architecture	4	The <i>Drake-S</i> material systematically lays out assumptions, variables, and formula steps; only the commentary occasionally digresses among teams
Critical assessment of objections	3	The <i>Drake-S</i> approach acknowledges and engages major sceptic-and-advocate objections, with the multi-team piece offering the most thorough cross-examination
Logical fallacies	3	The core approach is coherent, but the <i>Drake-S</i> approach currently relies on an additive-independence assumption that overlooks covariation among error terms
Distinction of evidence and belief	4	The <i>Drake-S</i> approach consistently separates empirical effect-size data from belief commitments; the commentary especially labels data versus interpretive claims
Conceptual clarity	3	Key terms (PR, EN, Pp) and variables are defined but constructs like LAP and certain confounds sometimes remain imprecise
Inappropriate evidential criteria	4	The <i>Drake-S</i> approach draws solely on peer-reviewed meta-analyses, systematic reviews, and recognized statistical standards—no irrelevant analogies or legal proofs

*Continued*

TABLE 2—continued.

*Copilot ratings of the Drake-S equation approach compared with Sudduth's evidentiary standards*

Evidentiary standard	Rating*	Rating justification
Revisiting IBE	2	The Drake equation analogy frames a best-explanation approach, but it does not integrate formal Bayesian or likelihood-based IBE critique
Proper epistemic principles	3	The <i>Drake-S</i> approach invokes falsifiability and the Daubert standard; the commentary advances toward likelihood reasoning, though full Bayesian rigor is still absent

\*Key: 1 = Very Poor Fit, 2 = Poor Fit, 3 = Good Fit, 4 = Very Good Fit.

Both ‘conceptual clarity’ and ‘critical assessment of objections’ received a 3, reflecting mostly precise use of terminology and moderate engagement with opposing views. Copilot’s evaluation of ‘proper epistemic principles’ (3) acknowledged references to scientific admissibility norms (e.g., Popperian falsifiability and Daubert criteria) but flagged the absence of formal likelihood ratios or Bayesian methods. Interestingly, Copilot gave a top rating (4) for ‘inappropriate evidential criteria,’ showing tolerance for interdisciplinary analogies such as legal benchmarks, which it viewed as heuristically useful rather than epistemically flawed. The only lower rating (2) was for ‘revisiting IBE.’ Copilot noted that the framework employed IBE informally, lacking probabilistic scaffolding or structured model comparison techniques such as Bayesian updating or likelihood evaluation.

Taken together, Copilot’s ratings suggest a pragmatic and interdisciplinary stance. Compared with ChatGPT-4o, which emphasized formal rigor, Copilot favored heuristic plausibility and structural coherence. This divergence reinforces the utility of using multiple AI perspectives in evaluating complex or nuanced theoretical models.

#### Comparison of the AI rating sets

Comparing the ChatGPT-4o and Copilot evaluations revealed both convergence and divergence across the eight evidentiary standards. To assess the degree of alignment between these LLMs on ordinal data, two inter-rater reliability metrics were calculated over the eight paired ratings: Spearman’s (1904) rank-order correlation ( $\rho$ ) and weighted Cohen’s (1968) kappa ( $\kappa$ ). Spearman’s  $\rho$  (−0.25) indicated a weak and non-significant negative correlation ( $df = 6$ ,  $p = 0.55$ ), suggesting limited concordance in the relative ranking of standards between the models. Similarly, the weighted Cohen’s  $\kappa$ , which adjusts for chance agreement and accounts for the ordinal nature of Likert-scale ratings, was −0.22 (SE = 0.18, 95% CI = −0.57723 to 0.11570), indicating

systematic disagreement (Cohen, 1968; Landis & Koch, 1977) that potentially reflects consistent differences in the LLMs interpreted or applied the rating criteria. We address this issue in the next section.

Both models exhibited perfect agreement on only two standards: critical assessment of objections and proper epistemic principles (score = 3), indicating consensus regarding moderate responsiveness to criticism and partial alignment with established epistemic methodologies. However, discrepancies were evident on six standards. On 'logical architecture,' Copilot assigned the highest score (4), emphasizing structural clarity and formulaic derivation, whereas ChatGPT-4o rated it a 3, citing interruptions in narrative coherence. For 'logical fallacies,' Copilot's score of 3 reflected pragmatic tolerance for heuristic modelling, while ChatGPT-4o's lower score (2) indicated stricter adherence to formal scientific standards, including concerns regarding unregistered assumptions and causal oversimplification.

The models also diverged on distinction of evidence and belief: Copilot awarded a 4 for clear separation of descriptive claims and belief commitments; ChatGPT-4o assigned a 3, noting occasional advocacy language. In contrast, for conceptual clarity, ChatGPT-4o gave the highest score (4), highlighting well-defined concepts such as 'LAP,' whereas Copilot rated it a 3, citing minor ambiguities. Regarding 'inappropriate evidential criteria,' Copilot assigned a 4, accepting interdisciplinary analogies (e.g., the Daubert standard) as heuristically valuable. ChatGPT-4o, however, rated this a 2, critiquing such legal comparisons as epistemically unsound within scientific contexts—underscoring its stricter boundary between scientific and non-scientific standards. Finally, for 'revisiting IBE,' Copilot issued the lowest score (2), criticizing the absence of formal probabilistic integration, whereas ChatGPT-4o rated it a 3, acknowledging explanatory virtues despite lacking Bayesian scaffolding.

These findings illustrate Copilot's more pragmatic and interdisciplinary approach compared with ChatGPT-4o's formalist and conservative epistemology. The moderate statistical agreement underscores the value of AI cross-referencing for theoretical evaluations, where cross-model perspectives reveal subtle tensions in logic, inference, and evidentiary framing.

#### INTERPRETING RATING DISCREPANCIES BETWEEN LANGUAGE MODELS

The varying outputs between the two LLMs warrant closer scrutiny for the purpose of refining AI methodologies in future research. In particular, we suggest there are five interrelated issues that likely contributed to the divergence in their evaluations:

- (1) *Pragmatism versus formalism.* Copilot's evaluations tended to reflect a pragmatic acceptance of heuristic reasoning and interdisciplinary analogies. In contrast, ChatGPT adhered more closely to formal correctness and rigorous scientific standards. This suggests a

foundational difference in how each model operationalizes validity and relevance across disciplines.

- (2) *Interdisciplinary benchmarks.* Copilot's high rating for the criterion of 'inappropriate evidential criteria' indicates an openness to legal-judicial standards—such as Daubert criteria—as useful proxies for evidentiary relevance. This reflects a broader willingness to integrate cross-disciplinary norms. However, ChatGPT-4o's more conservative stance underscores the necessity of epistemic justification when importing such frameworks into scientific contexts (cf. Faigman et al., 2014).
- (3) *Calibration of rubrics.* The six-point spread observed across eight evaluation criteria highlights the importance of narrowly defined rubrics and joint calibration exercises to align the raters' underlying assumptions. This is particularly salient for criteria involving 'fallacy' identification and 'evidential appropriateness.' Future evaluations would benefit from explicitly declaring the philosophical commitments informing each model's reasoning and from specifying acceptable levels of model simplification (Bechtel & Richardson, 2010).
- (4) *Sensitivity to advocacy language.* ChatGPT-4o showed a heightened sensitivity to rhetorical cues, as evidenced by its lower rating on the criterion 'distinction of evidence and belief' due to mild advocacy phrasing. This suggests a stricter demand for neutral and descriptive reporting. Copilot, by contrast, appeared to place greater emphasis on structural clarity than on rhetorical precision.
- (5) *Standards of explanatory scope.* ChatGPT-4o rewarded discussions that acknowledged explanatory virtues such as simplicity and scope, particularly in the context of 'revisiting IBE.' This reflects a broader valuation of narrative breadth in explanatory reasoning. Conversely, Copilot emphasized formal rigor and penalized broader interpretative moves lacking tight evidential justification.

Overall, these discrepancies illuminate how underlying epistemological assumptions and disciplinary allegiances shape AI-driven evaluations. Addressing such differences through clearer rubric design, interdisciplinary dialog, and philosophical transparency will be essential for enhancing methodological robustness in future applications.

## PART 2: EXPLORING POTENTIAL REFINEMENTS TO THE DRAKE-S EQUATION

Building on the comparative assessments generated by AI platforms, we now turn to targeted refinements of the *Drake-S* equation itself. These evaluations revealed areas of agreement—as well as divergence—in how our framework is interpreted and applied, particularly concerning evidential criteria and statistical robustness. While the equation has gained traction as a novel tool to quantify competing explanations for survival-type phenomena, its credibility and utility hinge on methodological clarity, logical coherence, and adaptability

to emerging data. In this section, we outline two key domains of enhancement: (a) mitigating risks related to inappropriate evidential criteria and logical fallacies and (b) integrating Bayesian reasoning to improve probabilistic sensitivity and interpretability. Together, these proposals aim to reinforce the equation's scientific rigor while accommodating new theoretical and empirical insights.

### *Inappropriate evidential criteria and logical fallacies*

A review of our prior works (Houran et al., 2023; Laythe & Houran, 2022; Rock et al., 2023) shows that the *Drake-S* equation needs little substantive revision with respect to evidential standards and logical reasoning. The AI analyses affirmed the importance of clearly contextualizing the rationale and conclusions of the model. We acknowledge, however, that Laythe and Houran's (2022, pp. 146–147) discussion of the *Drake-S* equation relative to the Daubert standard<sup>3</sup> in legal settings could be overstated or misapplied. Importantly, this analogy does not compromise the validity of the model's core premise. To address this concern, we endorse Sudduth's (2024, p. 517) recommendation to disentangle presentations of the *Drake-S* framework from legal standards such as 'beyond a reasonable doubt.'

Further, the *Drake-S* equation does not interpret knowledge gaps as direct or inherent evidence of postmortem survival. That approach would exemplify an inference argument, i.e., drawing conclusions without adequate support, and thus risks perpetuating biased or inaccurate beliefs (Govier, 2013). Instead, our probabilistic framework characterizes certain experiences—interpreted by some as survival evidence—as legitimate scientific anomalies requiring deeper investigation. Both materialist and LAP-oriented explanations remain insufficient, thereby justifying continued study of survival-type reports as essential for advancing scientific understanding of perception and consciousness. Additionally, we propose two methodological upgrades to improve evidential reliability and formal coherence:

- (1) *Model covariation among error terms.* The traditional assumption that error sources are additive and independent oversimplifies their interaction. A more robust approach involves estimating partial correlations among key error factors—such as identification,

---

3. The Daubert standard is a legal guideline used by US courts to decide whether scientific or expert testimony can be presented as evidence during a trial. It helps judges determine if an expert's methods are reliable and the evidence is relevant to the case. This standard comes from the 1993 Supreme Court case *Daubert v. Merrell Dow Pharmaceuticals, Inc.* To meet the Daubert standard, the testimony must generally satisfy these key criteria: (a) testability—the methods used by the expert must be testable and have been tested; (b) peer review—the work or methods should have been reviewed by other experts in the field; (c) error rate—there should be a known or potential error rate for the methods; (d) standards and controls—the methods should follow established standards; and (e) acceptance—the methods must be generally accepted in the relevant scientific community. This standard empowers judges to act as 'gatekeepers,' ensuring that unreliable or irrelevant scientific evidence is not presented to juries.



expectancy, environmental, fraud, measurement, and susceptibility. Even basic pairwise covariances can yield a more accurate 'net error' term and restore internal probabilistic consistency to the model.

- (2) *Pre-register, calibrate, and document rubrics.* Before deploying the *Drake-S* framework, researchers should publish a detailed rating protocol. This should define each error category, explain procedures for estimating covariances, and include exemplar calculations to ensure transparency and reproducibility. Moreover, conducting joint calibration exercises—whether among human experts or AI raters—helps to standardize interpretations of what qualifies as a 'fallacy' or 'inappropriate evidential criterion,' narrowing subjective leeway and enhancing inter-rater reliability.

### *Bayesian integration*

Integrating Bayesian methodology could transform the *Drake-S* equation into a more evidence-sensitive and -adaptive tool, capable of continuous probabilistic updates. Bayesian analysis applies probability to represent uncertainty, combining prior knowledge with new data to revise hypotheses dynamically (McGrayne, 2011). Unlike static models, Bayesian reasoning allows beliefs to be recalibrated as new findings emerge—an approach especially well suited to fields like medicine, machine learning, and finance (Gelman et al., 2013). Applied to the *Drake-S* equation, Bayesian refinements might include:

- (1) *Incorporating prior probabilities.* Assigning prior probabilities for each survival-related variable (e.g., AEs and KCs, including LAP) grounds the framework in established data. For instance, Rock et al. (2023) derived a 30.3% prior probability for the survival hypothesis, though a more conservative starting point—e.g.,  $\leq 10\%$ —could be used before adjusting for KCs.
- (2) *Modelling likelihood ratios.* Comparing the likelihood of survival versus confound hypotheses allows for quantitative adjustment of survival probability. Existing effect sizes (e.g., expectancy-suggestion effects = 0.097 or environmental factors = 0.078) offer benchmarks to model these ratios, while LAP explanations might use their upper variance limit of 8.3% as a confound likelihood.
- (3) *Sequential updates with new evidence.* Bayesian analysis supports iterative revision, recalculating the *Drake-S* equation as new studies (e.g., Frisaldi et al., 2023; Stein et al., 2025) update known effect sizes. This ensures the survival hypothesis remains robust and evidence-aligned over time.
- (4) *Assigning evidence weights.* Evidence quality varies across studies. Bayesian techniques can assign differential weights to AEs and KCs based on sample size, methodology, and effect magnitude, producing a more refined probability model (cf. Wahbeh et al., 2023).

- (5) *Credible intervals for survival probability.* Unlike traditional confidence intervals, Bayesian credible intervals offer direct probability estimates. This clarifies the interpretation of survival likelihood, adding statistical depth to *Drake-S* conclusions.
- (6) *Sensitivity analysis.* Bayesian frameworks facilitate testing how different priors or assumptions affect results. Sensitivity testing reveals which components most influence survival probability and bolsters the model's transparency.

Together, these Bayesian enhancements should yield a more nuanced, interpretable, and adaptable *Drake-S* framework, enabling a clearer comparison between survival interpretations and alternative hypotheses. This structure is also compatible with adversarial collaborations (cf. Corcoran et al., 2023). Nonetheless, setting a prior probability for survival—such as the 30.3% cited by Rock et al. (2023)—may invite debate. Critics might argue this estimate is overly generous. For context, Houran et al. (2023) compared the survival hypothesis with the simulation argument—i.e., that perceived consciousness could result from a technological construct or program as depicted in the popular sci-fi motion picture franchise *The Matrix*. Bostrom (2003, p. 255; cf. Bostrom & Kulczycki, 2011) estimated a 33% likelihood for simulated reality, while Chalmers (2022, p. 101) endorsed a 25% chance with qualifications. More recently, Vopson (2023) advanced a physics-based rationale via his theoretical law of ‘infodynamics.’ While such analogies are intellectually provocative, we do not claim that these Bayesian modifications offer the definitive solution. Rather, prior probabilities and model calibration require ongoing refinement as addressed in the next section.

## GENERAL DISCUSSION

We reiterate that the *Drake-S* equation (Houran et al., 2023; Laythe & Houran, 2022; Rock et al., 2023) was never positioned as either direct or exclusive evidence for the survival hypothesis. Rather, our probability framework simply offers empirical justification for the assertion that survival-type experiences currently elude comprehensive scientific explanations rooted in either materialism or putative psi. This conclusion—and the methodological foundation supporting it—is not mere conjecture but rather anchored to systematic data collection and analysis. In this light, some percipients’ accounts arguably represent legitimate anomalies that warrant continued, rigorous, and dispassionate investigation by consciousness researchers (for discussions on the concept of scientific anomalies, see e.g., Brewer & Chinn, 1994; Foorthuis, 2021; Lightman & Gingerich, 1992). It is also important to recognize that our mathematical approach to death-related phenomena does not stand alone. For instance, Himeoka et al. (2024) modelled cell death as a controllability problem in biochemical systems. By defining cell death as a state from which a cell cannot return to life, the study used a method called ‘stoichiometric rays’ to

map the boundary between living and dead states through enzymatic reactions, providing a quantitative tool to analyze cellular mortality. Their study similarly exemplifies how empirical modelling—rather than speculative theorizing—can illuminate complex phenomena at the edge of current understanding.

Integrated information theory (IIT) likewise is a mathematical approach for understanding and measuring the concept of consciousness, more broadly. It asserts that consciousness depends on the amount of integrated information generated by a system. This is captured by the value  $\Phi$ , which quantifies how much a system's parts work together in a unified, irreducible way (Oizumi et al., 2014; Tononi, 2004). A high  $\Phi$  coefficient means the system has strong internal connections that cannot be broken down into independent components, indicating a high degree of consciousness. If  $\Phi$  is zero, the system lacks integration and is not conscious. IIT is based on five core properties of conscious experience—such as its unity and specificity—which are translated into physical and mathematical terms. It then analyses the causal relationships within any system (biological or artificial) to calculate  $\Phi$ . This allows IIT to be applied beyond the brain, offering a general framework to evaluate consciousness in any physical system (Tononi et al., 2016).

But irrespective of the ultimate practicability of the survival hypothesis, Sudduth's (2024) eight guidelines can effectively steer future argument and research in this domain. In turn, the results of any subsequent analyses should better help to (a) explore and bridge the various focus areas and associated findings in consciousness studies as broached in the Introduction, and correspondingly (b) refine, refute, or rewrite major cognitive theories across the social and biomedical sciences. However, our study has several important limitations. LLMs are convenient and efficient tools, but this technology arguably lacks at this time a highly nuanced or sophisticated understanding of textual data or its context (Batubara et al., 2024). Using other LLMs might also produce different results. Indeed, by comparing ratings from both ChatGPT-4o and Copilot we uncovered systematic differences in formal versus pragmatic evaluation styles—underscoring the need to cross-validate AI assessments. Some readers also may question the ethics of AI-generated analyses or data (Resnik et al., 2025) or the epistemic value of our four-point Likert system that was filtered through different evidential criteria with their own nuances or complexities.

Finally, and relatedly, although the *Drake-S* equation is one mathematical way to contextualize the survival question, it currently says nothing about the *quality* of evidence used in its calculation. Evidence quality is perhaps *the* critical issue that complicates the interpretation and significance of research results in this domain (Augustine, 2022; Sudduth, 2024; Tressoldi et al., 2022). Future iterations of the equation could account for this confound, similar to its incorporation within some systematic literature reviews or meta-analyses (Luchini et al., 2021). Future studies also might supplement an AI-driven

analysis with expert evaluators (e.g., Hamilton et al., 2023; Houran et al., 2025; O’Keeffe et al., 2024). Notably, using two to three independent reviewers is arguably the standard across the social and biomedical sciences (e.g., Fernandez-Llimos et al., 2020). Securing reports from at least two experts helps to ensure that a range of voices is presented and promotes a fuller and fairer evaluation. Moreover, AI programs often lack deep, domain-specific expertise. This means that they can identify premises and conclusions but may miss nuanced sub-arguments or embedded assumptions without extensive fine-tuning (Gupta et al., 2024). Incorrect evaluations (‘hallucinations’) or import biases from the AI’s training data can also occur (Batubara et al., 2024). This can lead to overconfident ratings of logical coherence where gaps actually exist. Without transparent chains of reasoning, it can be difficult to verify why a particular ‘Good Fit’ versus ‘Poor Fit’ was assigned (Kaur et al., 2022). This undermines trust unless outputs are carefully audited. All that said, any refinement(s) of the current *Drake-S* equation only serves to calculate a more precise measurement of the amount of unexplained variance in the reported prevalence rates of survival-type experiences after controlling for the potential influence (per aggregated effect sizes) of KCs.

The preceding helps us to relate the *Drake-S* equation to Sudduth’s (2024) suggestion that only three kinds of generic claims should be the focus of inquiry about whether any facts provide evidential support for the survival hypothesis, and assuming so, in what way and to what degree those facts support the survival hypothesis. *First*, Sudduth asked if the observational data in question favor the survival hypothesis over alternative hypotheses. The *Drake-S* approach and calculation do not implicitly or necessarily endorse the survival hypothesis, but its results certainly suggest that either (a) alternative hypotheses or (b) substantial refinements of existing hypotheses are needed to account for the apparent prevalence rate of certain survival-type experiences. *Second*, Sudduth asked if the observational data in question are evidence that the survival hypothesis is true. The *Drake-S* approach, at best, perhaps buttresses the viability of the survival hypothesis. That is, the equation’s current outcomes are clearly consistent with a survival interpretation and would be expected if, in fact, survival was the correct explanation for some of the AEs considered here. And *third*, Sudduth asked if the observational data in question show that the survival hypothesis is probably true. The *Drake-S* equation does not necessarily establish the definitive feasibility of the survival hypothesis for explaining any AE category. These three conclusions limit the value of the present approach, but the *Drake-S* framework can nonetheless help to filter or vet categories of AEs that are most likely to yield chance scientific discoveries or challenges to them. Wuestman et al. (2020) explained that chance-type discoveries are driven by new observations or evidence that could agree with existing literature or not, whereas challenge-type discoveries are motivated by new or existing evidence that rebuts the existing literature.

Consciousness is one of the most elusive problems in science (e.g., Browning & Veit, 2021; Havlik et al., 2023; Melloni et al., 2021), with currently no consensus about its neural correlates, measurement methods, or underlying mechanisms. That said, some altered-exceptional experiences seemingly undermine a purely materialist model (Kofman & Levin, 2025). Cunningham's (2024) useful review of competing consciousness models likewise does not explicitly address the survival hypothesis, but one of its basic conclusions is certainly relevant here. In particular, we strongly agree with his assertion that "contemporary debates about the 'hard problem' of consciousness (i.e., how does 'mindless' matter produce 'matterless' mind?) cannot be resolved through philosophical analysis alone and need to be anchored to a comprehensive empirical foundation that includes psychophysiological research of psychosomatic phenomena and exceptional human experience" (p. 28). But it remains to be seen whether pro-survival researchers will develop new study designs or more robust arguments that purposely engage all of Sudduth's (2024) evidentiary standards.

This is probably easier said than done, and we have no specific proposals to offer at this time. But testable models involving objective measurements have been put forward for some survival-related phenomena (e.g., Persinger & St-Pierre, 2011) and consciousness theories more broadly (e.g., Hu & Wu, 2013). Moreover, some authors have forged adversarial collaborations to confront the daunting challenge of the mind–brain problem (e.g., Escolà-Gascón et al., 2023; Gomez-Marin & Seth, 2025; Kekecs et al., 2023; LeBel et al., 2022; Melloni et al., 2023; Mossbridge et al., 2025; Negro, 2024; Parnia et al., 2022; Schlitz et al., 2006; Sheldrake & Shermer, 2016; Wiseman & Schlitz, 1997). In our view, future efforts should avoid dogmatic and ideologically rooted campaigns that specifically seek either to establish or refute the ontological reality of survival. We instead think that research will advance more rapidly via the principles of curiosity, open-mindedness, and intellectual humility to various evidence-based perspectives—including those with low prior probabilities. This way, investigators can more effectively navigate the long and winding road to understanding the ultimate nature and limits of human consciousness. The ontological reality of postmortem survival could well be unlikely, but the *Drake-S* equation suggests that this provocative hypothesis should not be discounted just yet.

**Corresponding author**

JAMES HOURAN

*Integrated Knowledge Systems*

*1052 Summerplace Lane*

*Southlake, TX 76092 USA*

*Email: jim@drjimhouran.com*

*ORCID ID: <https://orcid.org/0000-0003-1725-582X>*

## ACKNOWLEDGMENTS

We thank the Society for Psychical Research Survival Research Committee (now The Yew-Kwang Ng (黃有光) Fund for Survival Research) for its support of this research. Appreciation also goes to Adam J. Rock, Patrizio Tressoldi, and Michael Sudduth for earlier consultations that helped to shape certain aspects of this paper.

## REFERENCES

- Altenmüller, M. S., Nuding, S., & Gollwitzer, M. (2021). No harm in being self-corrective: Self-criticism and reform intentions increase researchers' epistemic trustworthiness and credibility in the eyes of the public. *Public Understanding of Science, 30*, 962–976. <https://doi.org/10.1177/09636625211022181>
- Augustine, K. (2022). How *not* to do survival research: Reflections on the Bigelow Institute essay competition. *Journal of Scientific Exploration, 36*, 366–398. <https://doi.org/10.31275/20222581>
- Bak, W., Wójtowicz, B., & Kutnik, J. (2022). Intellectual humility: An old problem in a new psychological perspective. *Current Issues in Personality Psychology, 10*, 85–97. <https://doi.org/10.5114/cipp.2021.106999>
- Barušs, I. (2023). *Death as an altered state of consciousness: A scientific approach*. American Psychological Association. <https://doi.org/10.1037/0000361-000>
- Batubara, M. H., Nasution, A. K. P., Nurmalina, & Rizha, F. (2024). ChatGPT in communication: A systematic literature review. *Applied Computer Science, 20*, 96–115. <https://doi.org/10.35784/acs-2024-31>
- Beaty, W. (2024). Journey to credibility: 13 rules for maverick scientists [Guest editorial]. *Journal of Scientific Exploration, 38*, 1–10. <https://doi.org/10.31275/20243353>
- Beischel, J., Rock, A. J., Pekala, R. J., & Boccuzzi, M. (2021). Survival psi and somatic psi: Exploratory quantitative phenomenological analyses of blinded mediums' experiences of communication with the deceased and psychic readings for the living. *Journal of Near-Death Studies, 39*, 61–102. <https://doi.org/10.17514/JNDS-2021-39-2-p61-102>
- Benford, J. (2021). A Drake equation for alien artifacts. *Astrobiology, 21*, 757–763. <https://doi.org/10.1089/ast.2020.23>
- Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research* (2nd ed.). MIT Press. <https://doi.org/10.7551/mitpress/8328.001.0001>
- Boccuzzi, M. (2024, September 15). *Exploring the fit of quantum models with psi phenomena, survival research, and traditional science: A comparative analysis*. Psi Hacking Blog. <https://www.psihacking.com/posts/exploring-the-fit-of-quantum-models-with-psi-phenomena-afterlife-research-and-traditional-science-a-comparative-analysis/>
- Bostrom, N. (2003). Are we living in a computer simulation? *Philosophical Quarterly, 53*, 243–255. <https://doi.org/10.1111/1467-9213.00309>
- Bostrom, N., & Kulczycki, M. (2011). A patch for the simulation argument. *Analysis, 71*, 54–61. <https://doi.org/10.1093/analys/anq107>
- Brewer, W. F., & Chinn, C. A. (1994). Scientists' responses to anomalous data: Evidence from psychology, history, and philosophy of science. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1994*, 304–313. <https://doi.org/10.1086/psaprocbienmeetp.1994.1.193035>

- Browning, H., & Veit, W. (2021). The measurement problem of consciousness. *Philosophical Topics*, 48, 85–108. <https://doi.org/10.5840/philtopics20204815>
- Cardena, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist*, 73, 663–677. <https://doi.org/10.1037/amp0000236>
- Chalmers, D. J. (2022). *Reality +: Virtual worlds and the problems of philosophy*. Norton.
- Chen, M., Tworek, J., Jun, H., Yuan, Q., de Oliveira Pinto, H. P., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., ... Zaremba, W. (2021). *Evaluating large language models trained on code*. arXiv. <https://arxiv.org/abs/2107.03374>
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213–220. <https://doi.org/10.1037/h0026256>
- Corcoran, A. W., Hohwy, J., & Friston, K. J. (2023). Accelerating scientific progress through Bayesian adversarial collaboration. *Neuron*, 111, 3505–3516. <https://doi.org/10.1016/j.neuron.2023.08.027>
- Cunningham, P. F. (2024). The real ‘hard problem’ of consciousness: Where do thoughts come from, if not the brain? *Journal of Consciousness Studies*, 31, 28–54. <https://doi.org/10.53765/20512201.31.7.028>
- Drake, F. D. (1965). The radio search for intelligent extraterrestrial life. In G. Mamikunian & M. H. Briggs (Eds), *Current aspects of exobiology* (pp. 323–345). Oxford University Press.
- Duffy, B., May, G., Voas, D., Wright, J., Harsanyi, R., Hewlett, K., & Stoneman, P. (2023). *Lost faith? The UK's changing attitudes to religion*. The Policy Institute at King's College London. <https://doi.org/10.18742/pub01-134>
- Escollà-Gascón, A., Houran, J., Dagnall, N., Drinkwater, K., & Denovan, A. (2023). Follow-up on the U.S. Central Intelligence Agency's (CIA) remote viewing experiments. *Brain and Behavior*, 13, Article e3026. <https://doi.org/10.1002/brb3.3026>
- Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J. (2014). *Modern scientific evidence: The law and science of expert testimony* (vols 1–5). Thomson Reuters/West.
- Fernandez-Llimos, F., Salgado, T. M., Tonin, F. S., & Pharmacy Practice 2019 peer reviewers (2020). How many manuscripts should I peer review per year? *Pharmacy Practice*, 18, Article 1804. <https://doi.org/10.18549/PharmPract.2020.1.1804>
- Foorthuis, R. (2021). On the nature and types of anomalies: A review of deviations in data. *International Journal of Data Science and Analytics*, 12, 297–331. <https://doi.org/10.1007/s41060-021-00265-1>
- Frisaldi, E., Shaibani, A., Benedetti, F., & Pagnini, F. (2023). Placebo and nocebo effects and mechanisms associated with pharmacological interventions: An umbrella review. *BMJ Open*, 13, Article e077243. <https://doi.org/10.1136/bmjopen-2023-077243>
- Gallup (2023). *Gallup poll social series: Values and beliefs*. <https://news.gallup.com/poll/508886/belief-five-spiritual-entities-edges-down-new-lows.aspx>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>
- Gomez-Marin, A., & Seth, A. K. (2025). A science of consciousness beyond pseudo-science and pseudo-consciousness. *Nature Neuroscience*, 28, 703–706. <https://doi.org/10.1038/s41593-025-01913-6>
- Govier, T. (2013). *A practical study of argument* (7th ed.). Cengage Learning.
- Gupta, P., Ding, B., Guan, C., & Ding, D. (2024). Generative AI: A systematic review using topic modelling techniques. *Data and Information Management*, 8, Article 100066. <https://dx.doi.org/10.1016/j.dim.2024.100066>

- Hamilton, L., Elliott, D., Quick, A., Smith, S., & Choplin, V. (2023). Exploring the use of AI in qualitative analysis: A comparative study of guaranteed income data. *International Journal of Qualitative Methods*, 22, 1–13. <https://doi.org/10.1177/16094069231201504>
- Hashemi, A., Oroojan, A., Rassouli, M., & Ashrafizadeh, H. (2023). Explanation of near-death experiences: A systematic analysis of case reports and qualitative research. *Frontiers in Psychology*, 14, Article 1048929. <https://doi.org/10.3389/fpsyg.2023.1048929>.
- Havlík, M., Hlinka, J., Klírová, M., Adámek, P., & Horáček, J. (2023). Towards causal mechanisms of consciousness through focused transcranial brain stimulation. *Neuroscience of Consciousness*, 2023(1), Article niad008. <https://doi.org/10.1093/nc/niad008>
- Himeoka, Y., Horiguchi, S. A., & Kobayashi, T. J. (2024). *A theoretical basis for cell deaths*. arXiv. <https://doi.org/10.48550/arXiv.2403.02169>
- Houran, J. (2024). Reader advisory for Sudduth's BICS treatise: Preface to long essay section. *Journal of Scientific Exploration*, 38, 466–467. <https://doi.org/10.31275/20243333>
- Houran, J., Katz, D. L., Williamson, J., Koren, S. A., Wahbeh, H., & Woollacott, M. H. (2025). Magic flights or mind's eye? Further quali-quantitative explorations of dimensional-slip narratives. *Journal of Anomalistics*, 25, 61–118. <https://doi.org/10.23793/zfa.2025.061>
- Houran, J., Rock, A. J., Laythe, B., & Tressoldi, P. E. (2023). Dead reckoning: A multiteam system approach to commentaries on the *Drake-S* equation for survival. *International Journal of Transpersonal Studies*, 42, 80–105. <https://doi.org/10.24972/ijts.2023.42.1.80>
- Hu, H., & Wu, M. (2013). Groundbreaking new results in consciousness, quantum brain & nonlocality research. *Journal of Consciousness Exploration & Research*, 4, 69–74. <https://jcer.com/index.php/jcj/article/viewFile/282/310>
- Kar, S., Roy, C., Das, M., Mullick, S., & Saha, R. (2023). AI horizons: Unveiling the future of generative intelligence. *International Journal of Advanced Research in Science, Communication and Technology*, 3, 387–391. <https://doi.org/10.48175/ijarsct-12969>
- Kaur, D., Uslu, S., Rittichier, K., & Duresi, A. (2022). Trustworthy artificial intelligence: A review. *ACM Computing Surveys*, 55, 1–38. <https://doi.org/10.1145/3491209>
- Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, 73, 1–2. <https://doi.org/10.1037/amp0000263>
- Kekecs, Z., Palfi, B., Szaszi, B., Szecsi, P., Zrubka, M., Kovacs, M., Bakos, B. E., Cousineau, D., Tressoldi, P., Schmidt, K., Grassi, M., Evans, T. R., Yamada, Y., Miller, J. K., Liu, H., Yonemitsu, F., Dubrov, D., Röer, J. P., Becker, M., ... Aczel, B. (2023). Raising the value of research studies in psychological science by increasing the credibility of research reports: The transparent psi project. *Royal Society Open Science*, 10, Article 191375. <https://doi.org/10.1098/rsos.191375>
- Kelleher, C., & Bigelow, R. (2022). The 2021 Bigelow Institute for Consciousness Studies (BICS) essay contest. *Journal of Scientific Exploration*, 36, 350–365. <https://doi.org/10.31275/20222693>
- Kofman, K., & Levin, M. (2025). Cases of unconventional information flow across the mind–body interface. *Mind and Matter*, 23, 13–69. <https://doi.org/10.5376/mm2025.13>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. <https://doi.org/10.2307/2529310>
- Laythe, B., & Houran, J. (2022). Adversarial collaboration on a *Drake-S* equation for the survival question. *Journal of Scientific Exploration*, 36, 130–160. <https://doi.org/10.31275/20222357>
- LeBel, E., Augustine, K., & Rock, A. (2022). Beyond the BICS essays: Envisioning a more rigorous preregistered survival study. *Journal of Scientific Exploration*, 36, 436–447. <https://doi.org/10.31275/20222691>



- Lee, L., Dabirian, A., McCarthy, I., & Kietzmann, J. (2020). Making sense of text: Artificial intelligence-enabled content analysis. *European Journal of Marketing*, 54, 615–644. <https://doi.org/10.1108/ejm-02-2019-0219>
- Lightman, A., & Gingerich, O. (1992). When do anomalies begin? *Science*, 255, 690–695. <https://doi.org/10.1126/science.255.5045.690>
- Luchini, C., Veronese, N., Nottegar, A., Shin, J. I., Gentile, G., Granzio, U., Soysal, P., Alexinschi, O., Smith, L., & Solmi, M. (2021). Assessing the quality of studies in meta-research: Review/guidelines on the most important quality assessment tools. *Pharmaceutical Statistics*, 20, 185–195. <https://doi.org/10.1002/pst.2068>
- Mann, T. (1937). *Freud, Goethe, Wagner*. Alfred A. Knopf. (Original work published as *Freud und die Zukunft (Vortrag gehalten in Wien am 8 Mai 1936 zur Feier von Sigmund Freuds 80 Geburtstag)*, Bormann-Fischer, 1936)
- Matlock, J. G., & Hilton, B. (2025). After-death communications from non-human animals: Suggestions of post-mortem survival. *Journal of Scientific Exploration*, 39, 63–84. <https://doi.org/10.31275/20253321>
- Matlock, J. G., Hilton, B., Sheldrake, R., Smart, P., & Nahm, M. (2024). After-death communications (ADCs) from non-human animals: Parallels with human ADCs. *Journal of Scientific Exploration*, 38, 61–78. <https://doi.org/10.31275/20243087>
- McGrayne, S. B. (2011). *The theory that would not die: How Bayes' rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. Yale University Press.
- Melloni, L., Mudrik, L., Pitts, M., Bendtz, K., Ferrante, O., Gorska, U., Hirschhorn, R., Khalaf, A., Kozma, C., Lepauvre, A., Liu, L., Mazumder, D., Richter, D., Zhou, H., Blumenfeld, H., Boly, M., Chalmers, D. J., Devore, S., Fallon, F., ... Tononi, G. (2023) An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. *PLoS One*, 18, Article e0268577. <https://doi.org/10.1371/journal.pone.0268577>
- Melloni, L., Mudrik, L., Pitts, M., & Koch, C. (2021). Making the hard problem of consciousness easier. *Science*, 372, 911–912. <https://doi.org/10.1126/science.abj3259>
- Microsoft (2025). *GitHub Copilot* [AI code generation tool]. <https://github.com/features/copilot>
- Mossbridge, J., Green, D., French, C. C., Pickering, A., & Abraham, D. (2025). Future dreams of electric sheep: Case study of a possibly precognitive lucid dreamer with AI scoring. *International Journal of Dream Research*, 18, 151–168. <https://journals.ub.uni-heidelberg.de/index.php/IJoDR/article/view/108750/109002>
- Negro, N. (2024). (Dis)confirming theories of consciousness and their predictions: Towards a Lakatosian consciousness science. *Neuroscience of Consciousness*, 2024, Article niae012. <https://doi.org/10.1093/nc/niae012>
- Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology*, 10, Article e1003588. <https://doi.org/10.1371/journal.pcbi.1003588>
- O'Keeffe, C., Houran, J., Caputo, G. B., Katz, D. L., & Williamson, J. (2024). Magic flights or mind's eye? Competing content analyses of dimensional-slip narratives. *Journal of Transpersonal Psychology*, 56, 147–186.
- OpenAI (2024). *GPT-4 technical report and software*. <https://openai.com/research/gpt-4>
- O'Sullivan, L., Ma, L., & Doran, P. (2021). An overview of post-publication peer review. *Scholarly Assessment Reports*, 3, Article 6. <https://doi.org/10.29024/sar.26>
- Overgaard, M. (2017). The status and future of consciousness research. *Frontiers in Psychology*, 8, Article 1719. <https://doi.org/10.3389/fpsyg.2017.01719>

- Parker, A., & Brusewitz, G. (2003). A compendium of evidence for psi. *European Journal of Parapsychology* 18, 33–52.
- Parnia, S., Post, S. G., Lee, M. T., Lyubomirsky, S., Aufderheide, T. P., Deakin, C. D., Greyson, B., Long, J., Gonzales, A. M., Huppert, E. L., Dickinson, A., Mayer, S., Locicero, B., Levin, J., Bossis, A., Worthington, E., Fenwick, P., & Keshavarz Shirazi, T. (2022). Guidelines and standards for the study of death and recalled experiences of death—A multidisciplinary consensus statement and proposed future directions. *Annals of the New York Academy of Sciences*, 1511, 5–21. <https://doi.org/10.1111/nyas.14740>
- Pérez-Núñez, A. (2023). Exploring the potential of generative AI (ChatGPT) for foreign language instruction: Applications and challenges. *Hispania*, 106, 355–362. <https://doi.org/10.1353/hpn.2023.a906568>.
- Persinger, M. A., & St-Pierre, L.S. (2011). The biophysics at death: Three hypotheses with potential application to paranormal phenomena. *NeuroQuantology*, 9, 36–40. <https://doi.org/10.14704/nq.2011.9.1.388>
- Platt, M., & Platt, D. (2023). Effectiveness of generative artificial intelligence for scientific content analysis. *2023 IEEE 17th International Conference on Application of Information and Communication Technologies (AICT)*, 1–4. <https://doi.org/10.1109/AICT59525.2023.10313167>.
- Resnik, D. B., Hosseini, M., Kim, J. J., Epiphanou, G., & Maple, C. (2025). GenAI synthetic data create ethical challenges for scientists. Here’s how to address them. *Proceedings of the National Academy of Sciences of the United States of America*, 122, Article e2409182122. <https://doi.org/10.1073/pnas.2409182122>
- Rock, A. J., Houran, J., Tressoldi, P. E., & Laythe, B. (2023). Is biological death final? Recomputing the *Drake-S* equation for postmortem survival. *International Journal of Transpersonal Studies*, 42, 9–26. <https://doi.org/10.24972/ijts.2023.42.1.9>
- Royall, R. (1997). *Statistical evidence: A likelihood paradigm*. Chapman & Hall/CRC.
- Schlitz, M., Wiseman, R., Watt, C., & Radin, D. (2006). Of two minds: Sceptic–proponent collaboration within parapsychology. *British Journal of Psychology*, 97, 313–322. <https://doi.org/10.1348/000712605X80704>
- Schurger, A., & Graziano, M. (2022). Consciousness explained or described? *Neuroscience of Consciousness*, 2022, Article niac001. <https://doi.org/10.1093/nc/niac001>
- Sheldrake, R., & Shermer, M. (2016). *Arguing science: A dialogue on the future of science and spirit*. Monkfish.
- Shushan, G. (2024). Diversity and similarity of near-death experiences across cultures and history: Implications for the survival hypothesis. *International Review of Psychiatry*, 37, 95–101. <https://doi.org/10.1080/09540261.2024.2402429>
- Sorini, D., Peacock, J. A., & Lombriser, L. (2024). The impact of the cosmological constant on past and future star formation. *Monthly Notices of the Royal Astronomical Society*, 535, 1449–1474. <https://doi.org/10.1093/mnras/stae2236>
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15, 72–101. <https://doi.org/10.2307/1412159>
- Stein, M. V., Heller, M., Hughes, N., Marr, D., Brake, B., Chapman, S., Rubin, G. J., & Terhune, D. B. (2025). Moderators of placebo effects in controlled experiments: A multi-level meta-analysis. *Neuroscience and Biobehavioral Reviews*, 172, Article 106042. <https://doi.org/10.1016/j.neubiorev.2025.106042>
- Sudduth, M. (2009). *The reformed objection to natural theology*. Ashgate. <https://doi.org/10.4324/9781315553580>

- Sudduth, M. (2016). *A philosophical critique of empirical arguments for postmortem survival*. Palgrave Macmillan. <https://doi.org/10.1057/9781137440945>
- Sudduth, M. (2024). The Augustine–Braude Bigelow survival debate: A postmortem and prospects for future directions. *Journal of Scientific Exploration*, 38, 468–531. <https://doi.org/10.31275/20243309>.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5, Article 42. <https://doi.org/10.1186/1471-2202-5-42>
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17, 450–461. <https://doi.org/10.1038/nrn.2016.44>
- Tressoldi, P., Rock, A. J., Pederzoli, L., & Houran, J. (2022). The case for postmortem survival from the winners of the Bigelow Institute for Consciousness Studies essay contest: A level of evidence analysis. *Australian Journal of Parapsychology*, 22, 7–29.
- Vernon, D. (2021). *Dark cognition: Evidence for psi and its implications for consciousness*. Routledge. <https://doi.org/10.4324/9780429447488>
- Vopson, M. M. (2023). The second law of infodynamics and its implications for the simulated universe hypothesis. *AIP Advances*, 13, Article 105308. <https://doi.org/10.1063/5.0173278>
- Wahbeh, H., Delorme, A., & Radin, D. (2023). Rating the persuasiveness of empirical evidence for the survival of consciousness after bodily death: A cross-sectional study. *Journal of Anomalous Experience and Cognition*, 3, 78–109. <https://doi.org/10.31156/jaex.24125>
- Wahbeh, H., Radin, D., Cannard, C., & Delorme, A. (2022). What if consciousness is not an emergent property of the brain? Observational and empirical challenges to materialistic models. *Frontiers in Psychology*, 13, Article 955594. <https://doi.org/10.3389/fpsyg.2022.955594>
- Wiseman, R., & Schlitz, M. (1997). Experimenter effects and the remote detection of staring. *Journal of Parapsychology*, 61, 197–208.
- Wuestman, M., Hoekman, J., & Frenken, K. (2020). A typology of scientific breakthroughs. *Quantitative Science Studies*, 1, 1203–1222. [https://doi.org/10.1162/qss\\_a\\_00079](https://doi.org/10.1162/qss_a_00079)

## APPENDIX

### Example prompt and output from ChatGPT-4o evaluation

EVIDENTIARY STANDARD: LOGICAL ARCHITECTURE

#### **Prompt input**

You are an expert in philosophy of science and research methods. Based on the following excerpt from Laythe & Houran (2022), evaluate how well the Drake-S equation satisfies the evidentiary standard of 'logical architecture.' Use the 4-point rubric provided below along with the detailed scoring parameters already provided, and explain your reasoning.

*Rubric:*

- 1 = Very Poor Fit
- 2 = Poor Fit
- 3 = Good Fit
- 4 = Very Good Fit

*Definition of logical architecture:*

Do not pile up facts without clearly structured arguments. Arguments should present premises, conclusions, and sub-arguments in standard forms to avoid confusion.

*Excerpt:*

Laythe and Houran (2022) developed a Drake-Survival equation as mathematical evidence that absolutist claims (pro or con) about survival-related data are presently insufficient and unwarranted. They sourced hundreds of peer-reviewed studies for a probabilistic exercise that found the representative effect sizes of several known confounds (e.g., expectancy effects, mental illness, fraud) explained only 61.4% of the variance in survival-type experiences such as NDEs and reincarnation cases. Rock et al. (2023) added psi effects, raising explanatory power to 69.7%, yet a notable portion remained unexplained.

**ChatGPT-4o response**

*Rating: 3 (Good Fit)*

The *Drake-S* equation approach exhibits a reasonably clear logical structure. The argument begins with a well-defined problem—insufficient explanatory power of known confounds—and proceeds to a probabilistic modelling solution. The premises (e.g., prevalence of AEs, limited explanatory scope of KCs and LAP) are logically connected to the conclusion that further explanations (including survival) warrant consideration. However, while the overarching logic is coherent, the presentation could benefit from more explicit sub-argument structuring (e.g., showing how each KC quantitatively contributes to variance explained). Thus, the argument is well-formed but could be sharpened for greater clarity.