

Answer Coverage vs Predictive Risk: Why “Answering More” Does Not Confirm More — 24

Phil Stilwell
Independent Scholar
ORCID: [0009-0009-3725-8682](https://orcid.org/0009-0009-3725-8682)
Revised August 2026

Central claim

Christians and apologists who treat a worldview’s many answers as evidence for its truth confuse an inventory of claims with confirmation. Compared with a live rival, an added answer strengthens the evidential case from the observed record only when it makes that record more expected; merely adding answers, or making every outcome fit, does not improve that comparison. Breadth can still earn strong support when one constrained account generates accurate, independently assessable expectations across domains. The target is breadth without evidential accountability, not comprehensiveness itself.

Preview:

Target inference

A worldview answers more big questions, therefore it is better confirmed.

Required bridge

An added answer must make relevant evidence more expected than a live alternative does.

Pressure

Answer count fixes no comparison; post-hoc flexibility can make every outcome fit.

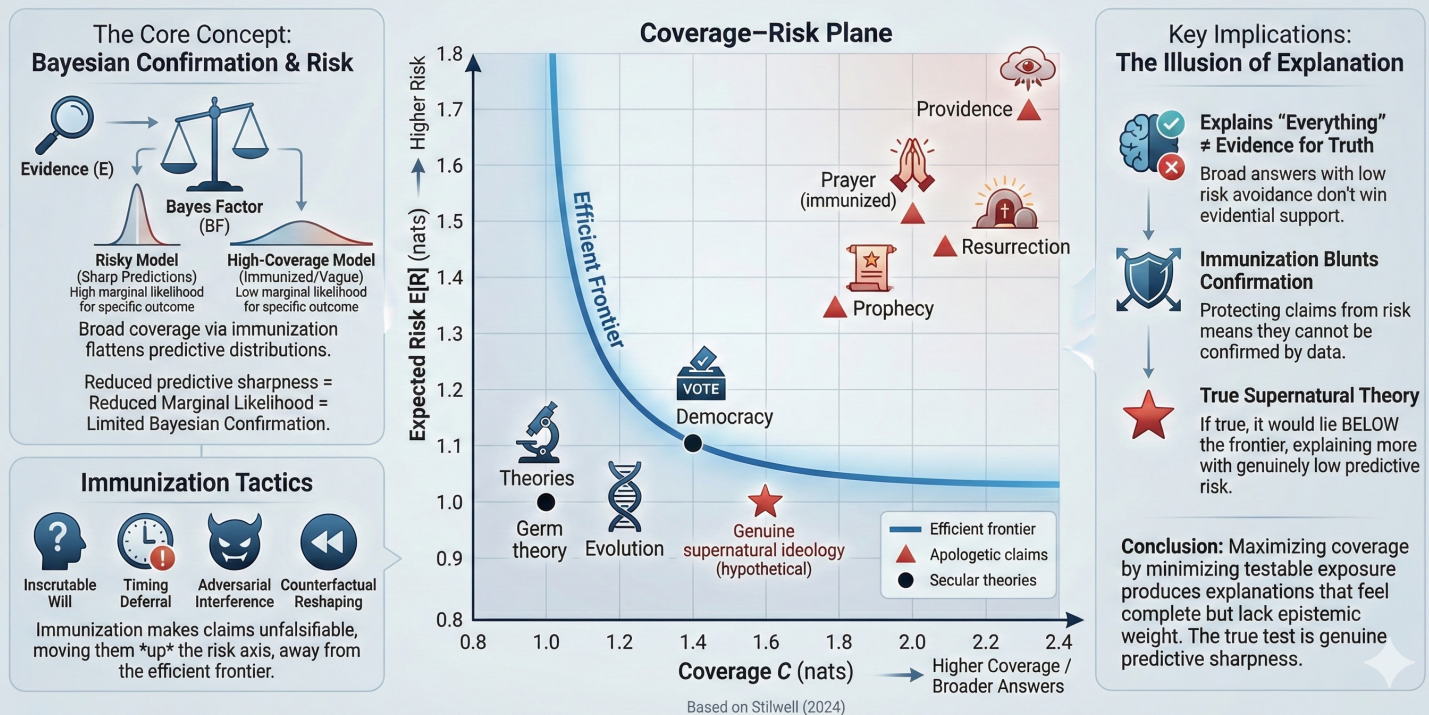
Rational conclusion

Treat coverage as an inventory, not evidence; credit breadth only when it discriminates.

The argument rejects a shortcut from completeness to confirmation. “Prediction” includes historical and other independently specified evidential expectations; it is not limited to laboratory forecasts.

Visual preview:

The Coverage–Risk Tradeoff in Apologetics Why Answering More Does Not Confirm More



The preserved conceptual infographic illustrates one strand of the wider argument: sharp predictions versus immunized breadth. Its frontier and plotted ideologies are illustrative, not measured data. The red star does not mean that supernatural truth alone guarantees predictive superiority; it marks only the conditional possibility of outperforming rivals through accurate, discriminating public predictions.

Abstract

Answer coverage is not confirmation: a worldview's ability to answer many questions does not by itself make that worldview better supported. The apologetic inference from breadth to truth fails because answer count, predictive sharpness, exposure to refutation, and predictive accuracy are different properties. The governing bridge is contrastive: an added answer strengthens the case from relevant evidence only when it makes that evidence more expected under the worldview than under a live rival. Answers that introduce no independently specified evidential difference leave that comparison unchanged, while post-hoc auxiliaries that accommodate every result may preserve logical compatibility without earning evidential support. Likelihood comparison, proper scoring rules, and minimum-description-length methods converge on this same discipline by asking whether a claim was constrained before assessment and whether it outperformed alternatives afterward. A bounded intercessory-prayer case illustrates the practical difference between a pre-specified intervention claim and a God-concept—a specified conception of a deity—that treats improvement, no change, and deterioration as equally compatible with benevolent action. Here “prediction” includes historical and other evidential expectations fixed independently of the result, not only future laboratory forecasts. The conclusion is not that broad theories are epistemically defective. A unified theory can gain strong support across many domains when its answers generate successful, independently assessable differences. Nor does the argument disprove theism or any particular God-concept. It blocks one shortcut: explanatory completeness, narrative satisfaction, or an answer for every outcome cannot substitute for evidence that discriminates among alternatives. Readers should ask not “How much does this worldview answer?” but “What evidential difference does this answer make, relative to what rival?”

Keywords: apologetics; Bayesian confirmation; answer coverage; predictive risk; proper scoring rules; model selection

1. The Inference Under Pressure

Christian apologetics often organizes worldview comparison around answers to “big questions”: why anything exists, why persons matter, why moral obligation feels binding, why suffering occurs, and what death means. A prominent worldview-catalogue approach, for example, compares positions through a common set of basic questions (Sire, 2020). That practice is legitimate: a catalogue can reveal commitments, scope, and tensions. Explanations organize experience, and a unified account can be intellectually valuable. Inference to the best explanation explicitly treats explanatory considerations as guides to belief, while psychological research shows that simplicity and causal structure shape explanatory preference (Lipton, 2004; Lombrozo, 2007). People can also feel that they understand a system more deeply than they can actually explain it, an effect documented as the illusion of explanatory depth (Rozenblit & Keil, 2002).

The mistake begins only when an apologetic presentation moves from “this worldview has an answer” to “therefore the evidence confirms this worldview.” An answer may be meaningful, coherent, comforting, or even true without the fact that it is an answer supplying evidence for its truth. A detective can invent a story that assigns a motive to every suspect and a cause to every clue. The story's coverage does not show that it identifies the culprit. Coverage inventories claims; confirmation compares how well those claims account for the record relative to competitors.

This paper is addressed especially to Christians and apologists who use answer breadth as public evidence. The claim under pressure is not Christianity as a whole, and it is not theism in general. A *God-concept* is a specified conception of a deity. Different God-concepts assign different intentions, capacities, and expected relations to public evidence. The narrower target is an *answer-breadth inference*: because a Christian God-concept and its associated worldview offer answers across more existential and empirical domains than a secular alternative, the Christian account is thereby better confirmed.

The governing thesis is direct: that inference is invalid unless a bridge is supplied between the additional answers and a comparative evidential advantage. If the answers make no relevant evidence more expected under the Christian account than under live alternatives, counting them does not change the comparison. If every possible result is reinterpreted as consistent with the account, logical survival has been purchased at the cost

of discrimination. A broad account may nevertheless be well confirmed when a unified structure generates successful evidential expectations across domains. The objection is therefore not to breadth. It is to breadth being counted as confirmation before it has done evidential work.

The expected-difference bridge

If added answers confirm a worldview, they must change how expected the observed record is under that worldview relative to a live rival. If adding those answers leaves the comparative likelihood of the evidence unchanged, the evidence's confirmatory force is unchanged. Answer count alone therefore cannot carry the inference from breadth to truth.

This bridge preserves the paper's polemical point without pretending that one formula settles every question about explanation. It asks an advocate to identify the missing step: what observed difference follows from the added answer, compared with what alternative, and why should that difference be more likely if the answer is correct?

Throughout the paper, "prediction" is shorthand for an expectation fixed independently of the result being assessed. The expectation may concern future observations, traces of a past event, coherence relations, or another public evidential pattern. The thesis does not privilege laboratories; it requires a difference that is not manufactured after the evidence is known.

2. Four Ideas That Must Not Be Conflated

The phrase "coverage versus risk" can hide several different ideas. The distinctions matter to one governing point: coverage records how much an account says, while confirmation depends on whether what it says discriminates among alternatives. Ordinary language complicates matters by using *risk* for both a theory's exposure to failure and the error suffered by a bad forecast. Those quantities can move in opposite directions. A sharp forecast takes a greater chance of being badly wrong, yet if it is accurate it earns a lower predictive loss. A vague forecast avoids assigning tiny probability to many outcomes, yet it may perform poorly because it never assigns high probability to the outcome that occurs.

Four distinctions are therefore essential.

1. **Answer coverage** is the range of questions for which an account supplies some answer. This is semantic and thematic breadth: origin, purpose, morality, suffering, destiny, and so forth.
2. **Predictive sharpness** is how concentrated a probability forecast is over mutually exclusive outcomes. A sharp forecast assigns much more probability to some outcomes than to others.
3. **Test exposure** is whether some admissible result would count against a claim. A claim is exposed when it rules out or strongly disfavors something that could occur.
4. **Predictive accuracy** is how well the probabilities issued before observation match what subsequently happens.

There is no automatic equation between these dimensions. A worldview can answer many questions yet remain vague about evidential differences. A narrow hypothesis can make one sharp prediction. A broad scientific theory can make many precise predictions because one mechanism unifies several domains. And a sharp prediction can fail. Answer coverage is not Shannon entropy; test exposure is not realized logarithmic loss; truth is not guaranteed by either sharpness or breadth. The point of separating the terms is to prevent answer count from borrowing evidential credit earned, if at all, by discrimination and accuracy.

Table 1 records how these forms differ. It is an argument ledger, not a scoring system.

Table 1: Four forms often collapsed into “answer coverage.” The table separates their evidential roles and does not rank worldviews by itself.

Form	What it supplies	What evidence can do	Main limit
Narrative breadth	Answers across many human questions	May organize experience or suggest hypotheses	Answer count alone fixes no comparative likelihood
Risky, discriminating answer	A difference fixed independently of the evidence assessed	Can confirm or disconfirm relative to a rival	Clear expectations can be wrong
Immunized answer	A reinterpretation for every possible result	Preserves compatibility with the record	Compatibility is not confirmation
Broad calibrated theory	One mechanism with successful predictions across domains	Can receive cumulative, cross-domain support	Breadth earns credit only through performance

The distinction matters because a defender can otherwise switch meanings mid-argument. Before a test, “risk” may mean that the claim boldly predicts healing. After a null result, “risk” may mean that no finite creature should expect to understand the deity’s reasons. The first formulation seeks public confirmation; the second protects logical possibility. Both may be coherent positions, but they are not the same evidential claim. The table therefore serves the central thesis: coverage is one property, whereas discriminating support must be earned by others.

3. Confirmation Is Contrastive

For the comparative claim at issue, evidence favors one hypothesis over another when it is more expected under the first. Let H be the target hypothesis, K a live comparator, and E the evidence. Their Bayes factor is

$$\text{BF}_{H:K}(E) = \frac{P(E | H)}{P(E | K)}. \quad (1)$$

When the ratio is greater than one, E favors H over K ; when it is less than one, it favors K over H ; when it is one, the observation does not discriminate between them. This likelihood comparison is a standard core of Bayesian model comparison and likelihood reasoning (Jaynes, 2003; Kass & Raftery, 1995; Sober, 2008). Prior probabilities still matter to final belief, and real models may include parameters and uncertainty, but the central point survives those complications: the evidential contribution of E depends on a comparison, not on the number of sentences a worldview can produce (Gelman et al., 2013).

Suppose two worldviews make exactly the same predictions about a body of public evidence. One also supplies a detailed account of cosmic purpose. If that account produces no new evidential expectation, the Bayes factor for the existing evidence does not change. The added answer may matter existentially. It may become testable later. It may cohere with other commitments. But the mere existence of the answer has not increased the likelihood of the evidence. If an advocate instead says that the answer raises the worldview’s prior plausibility through unification or simplicity, that is a different argument and must be defended as such. The work would then be done by independently supported unity or constraint, not by the number of answered questions.

This exposes the answer-count fallacy. An advocate may list ten questions for which Christianity offers an answer and a secular rival allegedly offers none. The list does not yet contain ten pieces of evidence. It contains ten proposed claims. Each claim must be connected to some observation and compared with alternatives. An untested proposition cannot confirm itself by occupying an empty cell in a worldview chart.

Nor does a precise-sounding answer automatically help. Specificity in prose is not predictive specificity. “Suffering serves soul-making, displays justice, preserves freedom, deepens dependence, or advances purposes beyond human understanding” contains many conceptual details while accommodating sharply different worlds.

The relevant question is not how many clauses the account contains. It is how their truth would alter expectations before the outcome is known.

The same discipline applies to nonreligious theories. A secular ideology does not gain confirmation merely by attaching a naturalistic label to every event. “Evolution,” “social construction,” “emergence,” or “brain process” can function as placeholders when no mechanism or risky implication is supplied. The standard is symmetrical: a catalogue of labels remains a catalogue, while explanation earns evidential credit through comparative success, not through vocabulary or tribal ownership.

4. The Every-Outcome Dilemma

The negative core of the central thesis is the every-outcome problem: if an answer can absorb every result without evidential cost, its coverage cannot also be counted as confirmation. Test exposure has long been associated with serious empirical inquiry. Popper’s account of falsifiability is not the last word on confirmation, but it correctly identifies the asymmetry between a claim that excludes possible observations and one that is protected from every result (Popper, 1959). Bayesian reasoning refines rather than erases that insight. A claim need not be strictly falsifiable in one step, yet it must assign different probabilities from its rivals somewhere if evidence is to discriminate.

The central dilemma can be stated without technical language.

The every-outcome dilemma

1. Either an answer makes some public outcomes more expected than others, or it does not.
2. If it does, the answer is exposed to evidence: favored outcomes can support it and disfavored outcomes can count against it.
3. If it does not, the answer may remain compatible with every outcome, but no favorable outcome can be counted as confirmation merely because the answer can absorb it.
4. Therefore an advocate cannot claim evidential credit from a risky reading before the result and then retreat to an every-outcome reading only after the risk materializes without reducing the original claim’s confirmatory force.

This is not the crude claim that every auxiliary assumption is illegitimate. All serious theories rely on background conditions, measurement assumptions, and supporting hypotheses. An unexpected result may reveal a faulty instrument rather than a false central theory. The problem is selective flexibility: adding or invoking auxiliaries only to prevent any result from lowering confidence, while continuing to treat favorable results as confirmation.

Consider a petitionary-prayer claim. A specified interventionist God-concept may predict that, under stated conditions, prayed-for patients will recover at a higher rate than comparable patients who are not assigned intercessory prayer. That claim risks a difference and permits a test. A different God-concept may hold that the deity’s reasons are opaque, that any outcome can express benevolence, that benefits may occur outside the measured window, or that the experiment itself blocks intervention. This second account may preserve theism. What it cannot do is use an observed recovery as straightforward public confirmation while declaring a null or adverse pattern evidentially irrelevant. The evidential rule cannot be “heads confirms; tails is mystery.”

The conclusion is proportionate. An inscrutability reply may be a theological reason not to expect a controlled effect. If so, the public predictive claim should be withdrawn or narrowed. A retreat from prediction can preserve logical possibility, but it also gives up the evidence that depended on that prediction. The answer may still cover the outcome; its coverage is no longer support for the risky claim.

5. Three Disciplines for Evidential Accountability

Bayes factors, proper scoring rules, and minimum description length are not three proofs of the same theorem, and none directly scores the number of questions an account answers. They ask related but different questions about comparison, advance constraint, and performance. Together they expose why breadth without evidential accountability is cheap.

5.1 Likelihood comparison: did the answer discriminate?

Equation 1 asks whether the observed result was more expected under the target than under a comparator. This is the most direct response to an apologetic breadth claim. If a worldview predicts the same evidence as its rival, the evidence does not choose between them. If it predicts the evidence less well, additional narrative coverage does not reverse that local evidential loss unless some other, independently supported evidence enters the comparison.

Bayes factors do not provide a mechanical answer when hypotheses are vague, priors are contested, data are selected after the fact, or dependencies are ignored. Those are reasons to state assumptions and test robustness, not reasons to replace comparison with answer counting. A verbal likelihood judgment can be more honest than fabricated numerical precision: “this result is at least as expected under the rival” is sometimes the strongest defensible conclusion. This discipline supplies the paper’s central bridge directly: an answer matters evidentially only through a defensible comparative difference.

5.2 Proper scoring rules: did the forecast perform?

A probabilistic forecast should be judged by the probabilities it issued before the outcome occurred. The logarithmic loss for forecast distribution q and realized outcome e is

$$S_{\log}(q, e) = -\log q(e). \quad (2)$$

Assigning high probability to what occurs yields a lower loss; assigning tiny probability yields a high loss. Strictly proper scoring rules make honest probability reporting optimal in expectation and permit comparative forecast evaluation (Gneiting & Raftery, 2007). The prequential approach—judging a model through its sequence of probabilistic forecasts and outcomes—similarly keeps attention on performance rather than on hidden parameters or retrospective storytelling (Dawid, 1984).

Sharpness is valuable only subject to calibration. A forecaster who always predicts one unlikely outcome with certainty is sharp and terrible. A forecaster who assigns equal probability to everything may be safe from catastrophic surprise but uninformative. Good predictive performance combines warranted concentration with accuracy over repeated or suitably structured cases.

This corrects a common slogan. Flattening a prediction does not always reduce marginal likelihood, and a broad forecast is not automatically unconfirmable. Its performance depends on the outcome-generating process and on the comparator. What immunization reliably does is reduce the content of the pre-result commitment. If the model is adjusted until every outcome receives an acceptable interpretation, the adjustment cannot be counted as though it were a sharp forecast issued in advance. Proper scoring therefore evaluates one possible route from answers to support—forecast performance—rather than rewarding answer count itself.

5.3 Minimum description length: did flexibility buy only a redescription?

Minimum description length (MDL) approaches model selection through compression: a good account balances the cost of describing the model with the cost of describing the data given that model (Grünwald & Roos, 2019; Grünwald, 2007). A flexible account can fit many observations, but its extra clauses, exception rules, and

parameter choices carry a cost. The point is not the crude maxim that simple theories are always true. It is that fit achieved by adding enough freedom to reproduce known data may not generalize.

This is highly relevant to worldview comparisons. “The deity healed when healing occurred, withheld healing for a greater purpose when it did not, acted through ordinary medicine when treatment worked, and preserved spiritual goods when treatment failed” compresses little unless the clauses generate constraints beyond the observations they redescribe. The account covers every result in ordinary language, but the coverage may be purchased one exception at a time.

A unified broad theory can do the opposite. Germ theory explains many disease patterns because a common mechanism yields novel expectations about transmission, sterilization, intervention, and recurrence. Its breadth is not an inventory of post-hoc answers. It is a consequence of a model that repeatedly succeeds under constraint. That is exactly the kind of breadth a worldview advocate should seek.

The three disciplines thus converge on one sequence: identify a live comparator, specify the evidential difference without tailoring it to the result, and evaluate performance against the evidence. They do not show that breadth is bad. They show where breadth must enter the argument if it is to become evidence rather than remain an inventory of claims.

6. A Bounded Test Case: Intercessory Prayer

Intercessory prayer is a bounded illustration of the every-outcome problem, not the sole evidence for the paper’s thesis. It displays the difference between a public causal claim and a protected theological interpretation. It must be handled carefully. A clinical trial cannot test all prayer, all divine purposes, or theism. It can test a pre-specified claim about whether a defined prayer intervention changes a defined outcome under stated conditions.

The STEP trial randomized 1,802 coronary artery bypass patients to conditions involving receipt or nonreceipt of distant intercessory prayer and uncertainty or certainty about receiving it. Among patients uncertain about assignment, complications occurred in 52 percent of those assigned prayer and 51 percent of those not assigned prayer. Complications occurred in 59 percent of patients certain that they would receive prayer, compared with 52 percent of prayed-for patients who were uncertain (Benson et al., 2006). The trial did not establish why the certainty condition had more complications. Nor does the study show that no deity exists, that no prayer has value, or that every interventionist God-concept is false. It addresses one protocol and one 30-day outcome window.

Its relevance to answer coverage is methodological. Before a test, the claim “intercessory prayer improves recovery” supplies a risky answer: it predicts a group difference. After a null result, an advocate has several legitimate options.

- Reject or reduce confidence in that causal formulation.
- Argue, with independent reasons, that the protocol failed to represent the relevant practice.
- Adopt a God-concept that predicts no controlled difference because divine action is not available for this form of public testing.
- Treat prayer chiefly as a devotional, communal, or psychological practice rather than as a reliable remote clinical intervention.

Each option has a different evidential consequence. The third and fourth may be coherent theological positions, but they cannot retain the original clinical prediction as public confirmation. If improvement, no change, deterioration, delayed benefit, hidden benefit, and refusal to cooperate are all equally admissible, then the outcome space has been covered by interpretation rather than predicted.

The strongest reply is that divine action could be personal and context-sensitive rather than governed by a stable population effect. That is possible. Yet context sensitivity becomes evidentially useful only when the relevant context can be identified independently and used to improve prediction. Otherwise it functions as an

unrestricted auxiliary: whatever happens reveals that the unobserved context called for exactly that result. Again, possibility survives while public discrimination disappears.

This standard is not anti-theistic. A genuine interventionist claim could outperform its rivals. It would need a clear outcome, a defined comparison, pre-specified conditions, safeguards against selective reporting, and replication. The paper does not exclude that success. The case simply makes the governing distinction visible: a claim can cover an outcome through interpretation, but it confirms only through a comparison it was constrained enough to win.

7. Strong Objections and Replies

7.1 “Explanatory power is evidence even without prediction”

Explanation can guide inquiry and rational belief. Lipton’s account of inference to the best explanation is not reducible to one likelihood ratio, and virtues such as unification, mechanism, simplicity, and fit can matter (Lipton, 2004). But explanatory virtues must connect to truth somehow. A merely lovely story is not automatically a likely story. Contrastive explanation asks why this event rather than another occurred, and that contrast already points toward differing expectations.

The present thesis does not require every meaningful claim to issue a numerical or future forecast. Historical explanations, for example, may be assessed through source dependence, causal detail, counterfactual consequences, and comparative fit. An advocate may also argue that unification raises prior plausibility. The requirement is more modest: if an explanation is offered as evidence for truth, the advocate must show why its unity is truth-conducive or why the relevant record is better explained by it than by serious alternatives. “It has an answer” is not that showing.

7.2 “A worldview is confirmed holistically, not claim by claim”

Holism can strengthen the case when one structure generates mutually supporting expectations across domains. It can also hide failure when evidence from different domains is counted without dependence or when every problem is answered by a new auxiliary. Cumulative support is not the sum of answer boxes. It depends on whether the strands are genuinely diverse, whether they were predicted independently, and whether the same flexible assumption is doing the work repeatedly.

The holistic reply therefore reinforces the demand for an evidence ledger. For each strand, state the target claim, the comparator, the expected difference, the observed record, and the scope of the conclusion. Then ask whether the strands converge or merely repeat the same interpretive permission. A worldview chart inventories breadth; only this second ledger can display cumulative confirmation.

7.3 “Ultimate questions cannot be tested like clinical forecasts”

Correct: not every philosophical or existential question has a clinical endpoint. Theism, moral realism, personal identity, and meaning include conceptual, phenomenological, and metaphysical considerations. The paper does not impose laboratory measurement on all rational inquiry.

The point is conditional. When an advocate claims that answer breadth itself is evidence, the advocate owes an account of how those answers bear on truth. The bridge may involve coherence, explanatory necessity, phenomenology, historical evidence, or moral experience rather than a randomized trial. But it must still discriminate. If incompatible worldviews can generate equally complete answers from the same record, completeness alone cannot choose among them. This is evidential accountability without laboratory reductionism.

7.4 “Secular accounts also use escape hatches”

They do, and the same criticism applies. A naturalistic label does not excuse absent mechanism, failed prediction, or endless auxiliary repair. Relevant safeguards include preregistration, held-out prediction, replication, model comparison, and public criticism. Imperfect as they are, each helps separate accommodation after the fact from prediction before the fact.

Symmetry is not a concession here; it is the standard. Any worldview that counts successes and absorbs failures without cost is evading comparison. Apologetics should not ask for an exemption from the evidential discipline it rightly applies to rival religions, conspiracy theories, or secular ideologies.

8. What Broad Confirmation Would Require

The paper’s conclusion is not merely negative. A worldview can turn breadth into evidence when its answers form a constrained, successful body of inquiry. The following five features operationalize the central bridge; they demand comparison and independent constraint, not laboratory prediction in every domain.

1. **A live comparator.** State what serious alternative is being assessed rather than contrasting a full worldview with a blank space.
2. **An expected difference.** Identify what should be observed more often, more clearly, or in a different pattern—including in historical traces or coherence relations—if the target answer is correct.
3. **Independent specification.** Fix the relevant expectation and permissible auxiliaries without tailoring them to the evidence under assessment; for future forecasts, do so before learning the result.
4. **Independent success.** Prefer evidence not used to construct the answer and do not count dependent reports as repeated confirmation.
5. **Cross-domain compression.** Reward a common mechanism that predicts several results with fewer independent assumptions, not a separate rescue clause for every problem.

These features reveal why broad scientific theories can be impressive. Their scope is not evidence merely because it is large. The same constrained mechanism survives many opportunities to fail and often predicts previously unknown relations. The scope magnifies the test rather than replacing it.

A Christian worldview could in principle meet the same standard. A specified God-concept might predict a distinctive, replicable public pattern; a historical claim might fit independent sources much better than alternatives; a moral or metaphysical theory might explain data that competitors leave fragmented while also generating costly commitments elsewhere. Nothing in Bayesian comparison, proper scoring, or MDL excludes such success. What these disciplines exclude is free confirmation: the claim that occupying more explanatory territory earns evidential credit even when no observation distinguishes the occupation from invention.

The practical reader-facing question is therefore simple: *What evidential difference does this answer make relative to a live rival?* Often that becomes: *What would look different if this answer were false?* If no answer can be given, the claim may still express hope, identity, or metaphysical possibility. It should not be advertised as publicly confirmed by the very outcomes it was designed to absorb.

9. Conclusion

Answer coverage inventories claims; confirmation compares expectations. Answering more therefore does not confirm more unless the added answers improve a real evidential comparison with alternatives. Predictive sharpness, test exposure, and predictive accuracy are different properties from coverage. Collapsing them allows a worldview to claim credit twice: first for supplying an answer and again for fitting an outcome that the answer did not risk.

The cumulative case is straightforward. Likelihood comparison requires the evidence to be more expected under the target than under a rival. The every-outcome dilemma shows that a claim cannot gain public support from favorable results while treating every unfavorable result as irrelevant. Proper scoring rules reward calibrated prediction rather than retrospective flexibility. Minimum description length distinguishes unification from a growing list of rescue clauses. The intercessory-prayer case illustrates the resulting choice between a bounded causal claim that can face evidence and a protected interpretation that preserves possibility while surrendering public confirmation.

The principal limit is equally important. Breadth is not an epistemic defect. A broad theory can be strongly confirmed when one constrained account makes successful, independently assessable predictions or other evidential expectations across many domains. The argument neither disproves theism nor settles any particular God-concept. It removes one shortcut. Christians, apologists, and their critics should stop treating the number of answers as though it were the weight of evidence. The rational change is to ask, for every advertised answer: compared with what rival, what evidential difference should it make, and what result would have reduced confidence in it?

A. Formal Reconstruction of the Central Thesis

A.1 Plain-language reconstruction

- P1.** Evidence confirms a target hypothesis over a live comparator only if the evidence is more expected under the target than under the comparator.
- P2.** The fact that a target supplies answers to more questions does not, by itself, establish that any observed evidence is more expected under the target than under the comparator.
- P3.** Adding answers that leave the comparative likelihood of the evidence unchanged leaves the evidence's confirmatory force unchanged.
- P4.** A broad target can nevertheless be confirmed when its answers generate an independently supported comparative likelihood advantage.

Conclusion. Answer breadth alone does not confirm the target over its comparator. Breadth can contribute to confirmation only when the added answers generate an independently supported likelihood advantage; this conclusion does not show that the target is false or that comprehensive theories cannot be well confirmed.

A.2 Symbols

H : the target hypothesis or worldview.

K : a live comparator to H .

E : admissible evidence used in the comparison.

$A(H)$: H supplies answers across many questions.

$C_E(H, K)$: E confirms H over K .

$P(E | H)$: the likelihood of E if H is true; $P(E | K)$ is defined analogously.

A.3 Symbolic form

The contrastive definition of confirmation is

$$C_E(H, K) \leftrightarrow P(E | H) > P(E | K). \quad (3)$$

Answer breadth alone does not entail the required likelihood inequality:

$$A(H) \not\Rightarrow [P(E | H) > P(E | K)]. \quad (4)$$

Therefore,

$$A(H) \not\Rightarrow C_E(H, K). \quad (5)$$

The scope boundary can be shown in the same notation:

$$[A(H) \wedge P(E | H) > P(E | K)] \Rightarrow C_E(H, K). \quad (6)$$

Here $\not\Rightarrow$ marks non-entailment, not a material conditional. The inference from Equations 3 and 4 to Equation 5 is valid: if confirmation requires a likelihood advantage and answer breadth alone supplies no such advantage, breadth alone does not entail confirmation. Equation 6 makes the principal limit explicit: breadth does not block confirmation when an independently supported likelihood advantage is present. Premise 1 is a stipulated contrastive Bayesian criterion; Premise 2 is a conceptual claim about what answer count determines and is supported by counterexamples in which $A(H)$ changes while $P(E | H)$ and $P(E | K)$ do not. The empirical or interpretive work begins only when a concrete advocate specifies H , K , and E and defends the relevant likelihood comparison. The reconstruction thus blocks a general inference; it does not deductively disprove H .

References

- Benson, H., Dusek, J. A., Sherwood, J. B., Lam, P., Bethea, C. F., Carpenter, W., Levitsky, S., Hill, P. C., Clem, D. W., Jain, M. K., Drumel, D., Kopecky, S. L., Mueller, P. S., Marek, D., Rollins, S., & Hibberd, P. L. (2006). Study of the therapeutic effects of intercessory prayer (STEP) in cardiac bypass patients: A multicenter randomized trial of uncertainty and certainty of receiving intercessory prayer. *American Heart Journal*, 151(4), 934–942. <https://doi.org/10.1016/j.ahj.2005.05.028>
- Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society: Series A (General)*, 147(2), 278–290. <https://doi.org/10.2307/2981683>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/016214506000001437>
- Grünwald, P., & Roos, T. (2019). Minimum description length revisited. *International Journal of Mathematics for Industry*, 11(1), 1930001. <https://doi.org/10.1142/S2661335219300018>
- Grünwald, P. D. (2007). *The minimum description length principle*. MIT Press.
- Jaynes, E. T. (2003). *Probability theory: The logic of science* (G. L. Bretthorst, Ed.). Cambridge University Press.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). Routledge.
- Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, 55(3), 232–257. <https://doi.org/10.1016/j.cogpsych.2006.09.006>
- Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson.
- Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26(5), 521–562. https://doi.org/10.1207/S15516709COG2605_1
- Sire, J. W. (2020). *The universe next door: A basic worldview catalog* (6th ed.). IVP Academic.
- Sober, E. (2008). *Evidence and evolution: The logic behind the science*. Cambridge University Press.