

A Likelihoodist Appraisal of the Holy Spirit’s Influence — 09

Phil Stilwell
Independent Scholar
ORCID: 0009-0009-3725-8682
Revised August 2026

Central claim

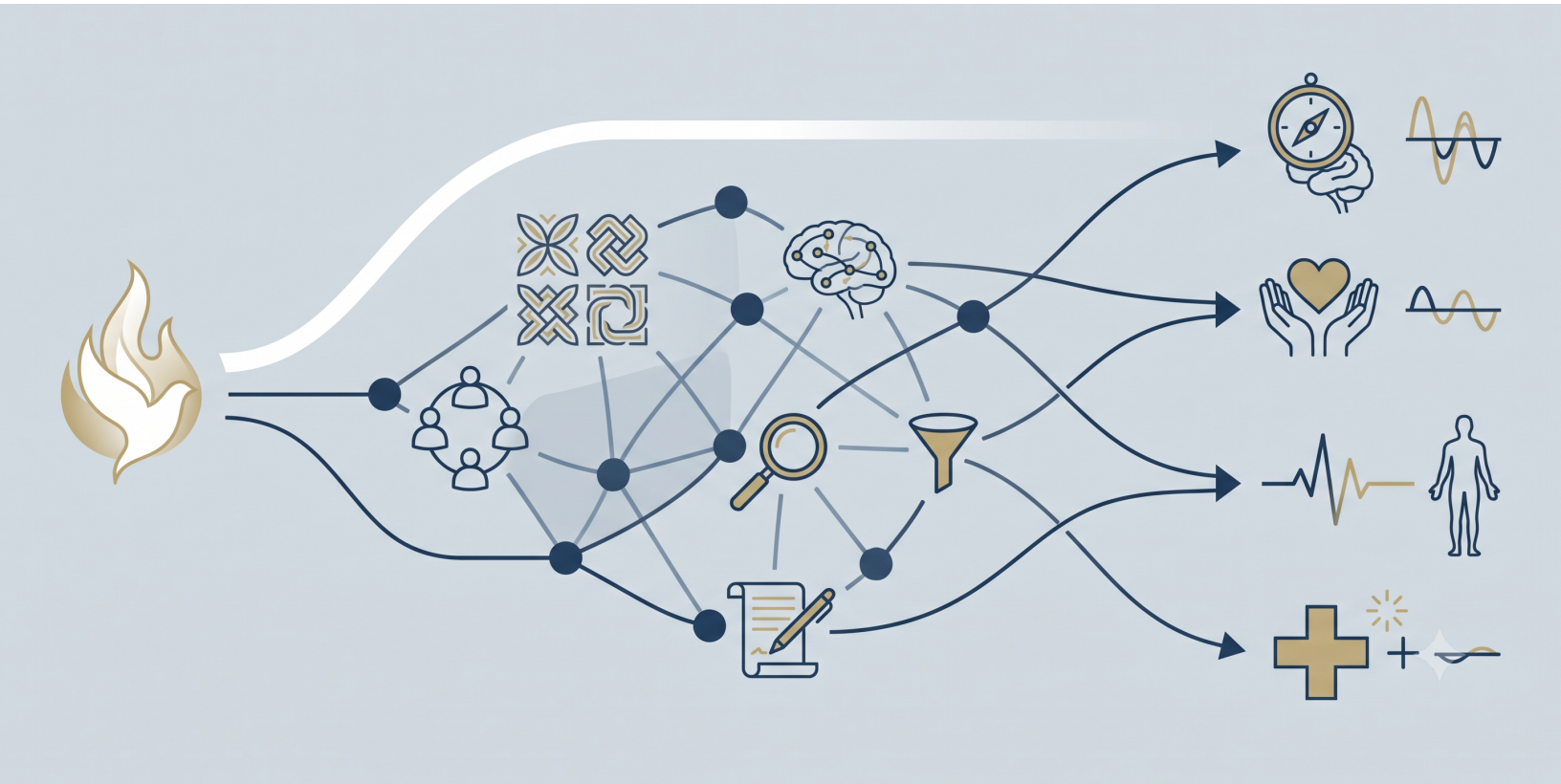
The current public evidence modestly disfavors a robust Christian Spirit-potency model—the claim that the Holy Spirit normally gives appropriately situated Christians widespread, distinctive, publicly detectable advantages. The small, mixed, proxy-laden record is more expected under a mediated-or-selective Spirit-action model and an ordinary-process benchmark, and it does little to distinguish those comparators from each other. This comparative result neither disproves theism nor excludes private, rare, or mediated Spirit action.

Preview:

Model under test	Expected difference	Observed record	Rational conclusion
Robust, widespread Spirit potency	Stable, widespread, distinctive public advantages	Small, mixed signals with little Spirit-specific measurement	Favor both comparators modestly; no clear winner between them

The comparison targets one declared public-effect model. Epistemic performance remains poorly measured, and the evidence neither chooses between mediated action and ordinary causes nor assigns a probability to any whole worldview.

Visual preview:



Four public evidence streams pass through ordinary mediators. Their small, mixed, overlapping signals pressure a distinctive public increment; they neither exclude mediated action nor confirm it over the ordinary-process benchmark.

Abstract

This paper asks whether current public evidence is better expected on a robust Christian Spirit-potency model—the claim that the Holy Spirit normally gives appropriately situated Christians widespread, distinctive, publicly detectable advantages—or on two comparators: a mediated-or-selective Spirit-action model and an ordinary-process benchmark. Its answer is that the evidence modestly disfavors the robust model relative to each comparator while giving little reason to choose between those comparators. This comparative conclusion neither disproves theism nor excludes private, rare, or mediated Spirit action. The method is likelihoodist: it compares how well each declared model expects the evidence without assigning a probability that any whole worldview is true. It treats epistemic performance, prosocial conduct, health and well-being, spiritual-presence experience, and extraordinary intervention as a dependence-aware cumulative case. No cited study directly measures exposure to the Spirit. Measures such as religiosity, attendance, and self-reported experience are therefore noisy in both directions: they can miss genuine Spirit action, and they can package ordinary causes such as selection, practice, identity, cultural learning, and social integration. Across the five domains, the recurring pattern is small, heterogeneous, and weakly source-specific. Directly observed prosocial effects are smaller than self-reported effects; favorable attendance associations coexist with weak within-person mental-health change; spiritual-presence reports covary with culture and absorption across traditions; controlled intercessory-prayer evidence has not produced a stable medical advantage; and Spirit-specific prospective evidence for epistemic performance remains scarce. Because these domains share causes and measurement problems, their contributions cannot simply be multiplied. Their convergence nevertheless makes the observed record more expected under the two comparators than under a robust model. Public appeals to Spirit-produced effects should therefore declare recipients, outcomes, time scales, ordinary mediators, and failure conditions before results are known.

Keywords: Holy Spirit; likelihoodism; religious epistemology; pneumatology; prosociality; well-being; intercessory prayer; causal inference

Argument in four moves

1. Specify	2. Observe	3. Compare	4. Limit
Declare a robust public-effect model.	Inspect five outcome domains.	Ask which model better expects the pattern.	Report only the supported contrast.
Not Christianity as a whole.	Religiosity and attendance are noisy proxies.	Compatibility is not confirmation.	No proof of naturalism or universal nonaction.

1 From “Influence” to Testable Contrasts

This paper addresses Christians, theologians, and philosophers who treat the Holy Spirit’s influence as a recurrent cause of publicly detectable transformation. Its governing thesis is straightforward: across five noisy evidence domains, the present record modestly disfavors a robust model that predicts widespread, distinctive advantages among appropriately situated Christians. The same small, mixed pattern is more expected under a mediated-or-selective Spirit-action model and an ordinary-process benchmark, but it does little to distinguish those comparators from each other. This is evidence against the robust model, not against theism as a whole. Readers who appeal to public Spirit effects should therefore require predictions that identify recipients, outcomes, time scales, ordinary mediators, and failure conditions before the results are known.

The claim that the Holy Spirit influences persons can mean many things. It may concern assurance of salvation, illumination while reading Scripture, sanctification over a lifetime, communal gifts, wise judgment, consolation, healing, or a rare miracle. Some of these claims are principally first-person or theological. Others imply recurring differences in outcomes accessible to an outside observer. An evidential appraisal becomes confused when it treats this entire range as one hypothesis.

Christian theology itself gives reason to expect influence to be experientially and practically recognizable. Paul’s language of empowerment, gifts, fruit, and communal formation does not reduce the Spirit to a hidden

metaphysical posit. Fee's exegetical synthesis emphasizes the Spirit as the empowering divine presence in Pauline communities; Macchia places baptism in the Spirit at the center of a global Pentecostal theology; and Zahl argues that accounts of the Spirit should remain connected to the practical recognizability of Christian experience (Fee, 1994; Macchia, 2006; Zahl, 2020). Yet none of these works entails a simple claim that Christians must outperform every comparison group on every secular measure. Theological traditions contain both public-facing expectations and qualifications concerning freedom, mediation, vocation, suffering, and goods that a standard instrument may fail to capture.

The proper question is therefore contrastive: which specified model makes the observed public record more expected? This formulation prevents two opposite errors. One error immunizes every claim by moving from a failed prediction to a less observable form of influence without acknowledging the evidential change. The other treats pressure on one strong model as evidence that all divine action is absent. A fair appraisal must preserve intermediate possibilities and expose the assumptions on which each evidential inference depends.

1.1 Three models, not a forced dichotomy

Let H_S denote a *robust Spirit-potency model*. It says that, among people and communities who satisfy ordinary Christian participation conditions, the Spirit normally produces recurrent effects that are appreciably more favorable and more distinctively patterned than those produced by ordinary religious and social participation alone. The relevant effects need not occur without exception, but they should be accessible often enough to support comparative detection across some combination of sound judgment, other-directed conduct, well-being, or extraordinary intervention.

Let H_M denote a *mediated or selective model*. On this model, the Spirit may work through ordinary psychological and social pathways, act rarely or for particular vocations, prioritize communion and holiness over externally scored success, or confer goods whose public measurement is difficult. It can predict genuine effects while expecting weak, context-sensitive, or sparse population signals.

Let H_O denote an *ordinary-process benchmark*. It says that the public patterns under examination are adequately generated by familiar causes such as social integration, norms, selection into communities, expectancy, absorption, reporting practices, and background health. It does not assert that theism is false, that no miracle has ever occurred, or that religious life has no value. It merely denies that a specifically Spirit-produced increment is needed to explain the selected public data.

The three models are not mutually exhaustive. A pluralist transcendent source, other proposed spiritual causes, or combinations of models remain logically available. Their purpose is to discipline a comparison that otherwise slides between an expansive Christian claim and metaphysical naturalism. Evidence favoring H_O over H_S within this contrast does not automatically favor H_O over every possible theology.

1.2 The robust model is a conditional family

H_S is not Christianity simpliciter, nor does minimal orthodox pneumatology entail it. It is a family of stronger public-effect models whose members become testable only when joined to auxiliary commitments. Table 1 states the evidential contract required before a favorable or unfavorable result can discriminate. These commitments should be elicited before the outcome is known; otherwise, an apparent prediction can be narrowed after the fact until almost any record fits.

Table 1: The evidential contract for a public-effect model

Commitment	Question to answer before data	If left open
Recipient and exposure	Who counts as an appropriately situated recipient, and what observable indicator tracks the relevant exposure?	Nulls may reflect misclassification; positive associations may lack Spirit specificity.
Intended effect	Which theological good has what direction, magnitude, and behavioral or experiential indicator?	Outcome switching and weak construct validity preserve almost any result.
Contrast and mediation	Which comparison group and ordinary pathways are expected, and what residual or patterned difference should remain?	The same mediated pattern may fit H_M and H_O equally well.
Time and failure rule	Over what interval should change appear, and what result would count against the declared public-effect member?	Delay, rarity, or resistance can be invoked without a pre-data limit.

1.3 What likelihoodism contributes

For an outcome profile D , the likelihood ratio comparing two models is

$$\Lambda_{S/O}(D) = \frac{p(D \mid H_S, A_S)}{p(D \mid H_O, A_O)}, \quad (1)$$

where A_S and A_O represent auxiliary assumptions linking each broad model to measurement. Values above one favor the robust Spirit-potency model over the ordinary benchmark; values below one favor the ordinary benchmark over that robust model. Likelihoodism treats this as a relation between evidence and specified hypotheses, not as a posterior probability of either hypothesis (Royall, 1997; Sober, 2008). Priors, alternative models, and further evidence are needed for a Bayesian posterior.

The auxiliaries do real work. A failure to observe an advantage matters only insofar as the robust model, together with plausible claims about recipients, time scale, measurement, and divine aims, made that advantage likely. Conversely, a favorable association matters only insofar as the ordinary benchmark did not already make it likely. Social support can be both an ordinary causal pathway and, on a mediated theology, an instrument of the Spirit. An observed benefit may therefore fit both models well.

Because H_S and H_M are composite families, no unique numerical likelihood exists until their auxiliaries and nuisance parameters are specified. The comparisons below are therefore ordinal and qualitative. Likelihoodism does not impose an automatic complexity penalty when a model is narrowed after seeing the data. The relevant evidential cost is instead a loss of contrast: a model flexible enough to predict success, failure, delay, rarity, or invisibility may remain true, but the observed outcome supplies it little confirmation over rivals (Royall, 1997; Sober, 2008).

The evidential bridge.

A powerful, purposive, population-wide cause should normally leave some stable comparative trace when its declared effects are repeatedly measured under suitable conditions. The bridge is defeasible: it weakens when action is selective, the intended good is nonpublic, the measure is poor, or the causal effect is routed through mechanisms equally predicted by the comparison model.

The expected-difference argument.

1. If the robust Spirit-potency model is true under declared recipient, outcome, timing, and measurement conditions, a stable and distinctive public trace should normally appear.
2. The examined record instead shows small, heterogeneous, proxy-laden, and largely nonspecific signals; the epistemic domain remains prospectively undermeasured.
3. That profile is more expected under mediated or selective action and under ordinary-process explanations than under the robust model.
4. Therefore the current evidence modestly disfavors the robust model relative to both comparators while weakly discriminating the comparators from each other.

2 Declaring the Evidence Before Interpreting It

A persuasive appraisal should specify outcomes and error controls before treating results as diagnostic. Otherwise, broad theological language can select whichever association happens to be favorable and redescribe unfavorable findings as irrelevant. Five domains are pertinent here.

- **Epistemic performance:** convergence on important judgments, resistance to demonstrable error, calibrated confidence, and decisions that can be checked independently of doctrinal agreement.
- **Prosocial conduct:** behavior toward ingroup and outgroup members, preferably measured directly or through records rather than by self-description alone.
- **Health and well-being:** longitudinal change in distress, flourishing, health behavior, and physical outcomes, with baseline status and selection into participation addressed.
- **Extraordinary intervention:** preregistered, independently verified outcomes for a declared target and population, while keeping the tested intervention narrower than miracle claims in general.
- **Spiritual-presence experience:** frequency and content of reported encounters, analyzed for cultural and dispositional predictors as well as for any source-specific accuracy.

Religious disagreement is relevant to the first domain, but disagreement alone is not a decisive refutation. Reflective peers may possess different evidence, different background commitments, or unequal access to a truth-conducive process. Still, persistent collision among sincere, informed religious agents creates pressure toward epistemic humility and toward independent error controls (Dormandy, 2023). A model predicting reliable guidance should say which judgments will converge, for whom, and how genuine guidance can be distinguished from confident mistake.

The evidence ledger in Table 2 reports what the central studies establish and, equally importantly, what they do not establish.

Table 2: A dependence-aware evidence ledger

Domain	Observed profile	Comparative force and limit
Epistemic performance	Deep religious disagreement is visible, but Spirit-specific prospective forecasting and calibration studies are scarce.	Weak pressure on claims of readily identifiable guidance; insufficient for a quantified comparison.
Prosocial conduct	A 2024 meta-analysis finds a small positive association, substantial heterogeneity, and a smaller effect for directly measured behavior than for self-report.	Resists a blanket-null claim, yet fits robust transformation less well than a large, distinctive behavioral advantage would.
Health and well-being	Prospective attendance studies find several favorable associations; a nine-year within-person study finds almost no beneficial lagged effects.	Mixed and strongly exposed to selection, social integration, construct choice, and time scale.
Extraordinary intervention	A major cardiac-bypass prayer trial did not show improved complication-free recovery; a Cochrane review judged the wider trial record equivocal.	Counts against broad guaranteed effects in the tested setting, not against every petition, miracle, or form of Spirit action.
Spiritual presence	Cross-cultural reports covary with models of mind and individual absorption across different faiths.	Supports ordinary mediation and weakens source specificity; does not show that every experience is unreal or nondivine.

What these studies actually test.

Religiosity, attendance, and self-reported experience are constructed indicators, not direct observations of the Holy Spirit. Their measurement error and mixed content attenuate inference in both directions: a null result may miss genuine action, while a positive result may be produced by selection, practice, identity, or social integration. The present studies test whether a broad public-effect model leaves a stable comparative trace through noisy proxies; they do not directly test Spirit causation (VanderWeele, 2022).

3 What the Current Record Shows

3.1 Epistemic performance: an open evidential cell

The epistemic domain has the clearest conceptual criterion and the thinnest direct record. Widespread religious disagreement is compatible with human error, unequal evidence, and freedom; it is not itself a controlled observation of the Spirit's absence. More importantly, the studies reviewed below do not prospectively identify recipients who claim Spirit guidance, record their confidence before outcomes are known, and score independently resolvable judgments. The present appraisal therefore assigns no negative domain-specific likelihood contribution merely from disagreement.

The lacuna is still evidentially important. Claims of illumination or guidance often imply some capacity to distinguish truth-conducive prompting from sincere projection. Without prospective calibration, error denominators, and a declared comparison, public appeals to discernment remain difficult to assess. This domain should be treated as a research priority rather than quietly converted into either supporting testimony or disconfirming absence.

3.2 Conduct: positive, small, and measurement-sensitive

The most comprehensive recent synthesis of religiosity and prosociality analyzed 701 effects from 237 samples comprising 811,663 participants. It found a small overall association, $r = .13$, with substantial heterogeneity. The association was stronger for self-reported prosociality, $r = .15$, than for directly measured behavior, $r = .06$ (Kelly et al., 2024). The difference permits more than one interpretation: direct laboratory measures may omit

real-world conduct, self-reports may contain self-enhancement, or both measurement families may be imperfect.

This is not evidence that religion produces no prosocial benefit. It is positive evidence of an average association, and some contexts or practices may generate larger effects. Yet it is not the profile one would most naturally predict from an unqualified H_S under which widespread indwelling yields a large and readily discriminable behavioral transformation. Small effects, large between-study variation, and sensitivity to self-report are more expected under H_O , which already predicts norm learning, reputation management, group selection, and differences between reported identity and observed action. They are also compatible with H_M , particularly if sanctification is gradual, uneven, mediated by community, or poorly represented by brief behavioral tasks.

The comparison therefore cuts both ways. A strict no-benefit version of H_O would fare poorly, but that is not the benchmark used here. An ordinary-process model can expect modest benefits from stable community, moral rehearsal, mutual monitoring, and costly commitments. The evidential question is not whether religiosity correlates with any good outcome; it is whether the magnitude and pattern are more expected if a distinctive Spirit increment is present. On the present evidence, that increment is difficult to isolate.

3.3 Well-being: favorable associations and weak within-person change

Chen, Kim, and VanderWeele examined religious-service attendance across three large prospective U.S. cohorts. At least weekly attendance was associated with lower subsequent mortality, heavy drinking, smoking, and depression, along with higher social integration, purpose, and several well-being outcomes. The authors adjusted for numerous measured confounders and baseline outcomes while acknowledging that observational data do not definitively establish causation (Chen et al., 2020). These findings deserve genuine evidential weight. Any appraisal that says religious participation has no measurable benefits is empirically inadequate.

The measures, however, chiefly concern service attendance rather than Christian orthodoxy, reception of a spiritual gift, or an independently identified episode of Spirit action. Attendance itself packages social integration, routine, behavioral norms, purpose, and selection by prior health. Those mechanisms are predicted under H_O and can also serve as media under H_M . Favorable attendance associations therefore have limited source specificity.

Prati's nine-year panel analysis sharpens the causal question by separating stable between-person differences from lagged within-person change. It found almost no evidence that changes in religious variables produced later improvements in mental health; the few significant associations did not yield a simple beneficial pattern (Prati, 2024). A later analysis of 29,298 participants across eighteen years of the British Household Panel Survey likewise found mostly nonsignificant within-person associations between religious-service attendance and mental health; the few significant paths pointed in both beneficial and detrimental directions (Prati, 2025). These designs have limitations, including measurement intervals, attrition, and the possibility that the relevant spiritual goods do not fluctuate on the modeled schedule. Still, the divergence between favorable between-person associations and weak within-person effects is precisely what selection and stable background differences can produce.

The combined record is thus mixed rather than null. H_S can accommodate it by narrowing the relevant recipients, extending the time horizon, or redefining the intended good. Each move may be theologically motivated. But the more the model expects no stable difference on observable well-being measures, the less those measures can confirm its claimed public potency. H_M predicts this ambiguity more readily because it treats social and psychological pathways as integral rather than as noise around a large added effect.

3.4 Experience: reality without source specificity

Reports of divine presence are data even when their ultimate cause is disputed. Luhrmann and colleagues studied spiritual-presence events across faiths and cultures and found that cultural models of a more "porous" mind and the disposition of absorption predicted such events across multiple studies (Luhrmann et al., 2021). These findings show that reported encounters are systematically related to learned models and modes of attention. They do not establish that no reported event is divinely caused. A divine cause could operate through culturally learned attention, and the study was not designed to adjudicate Christian pneumatology.

The findings nonetheless matter for likelihoods. If phenomenology and content vary with prior formation and

absorption across incompatible traditions, the occurrence of vivid presence by itself has low specificity for H_S . The same accessible feature is expected under ordinary cognitive-cultural pathways. A defender may answer that the Spirit graciously accommodates human culture, that other experiences have different sources, or that fruits over time identify the genuine cases. Those are coherent replies, but each calls for an independent classifier: a measure that identifies which culturally scaffolded experiences are Spirit-produced without simply presupposing the preferred interpretation.

3.5 Extraordinary intervention: a narrow negative result

The STEP trial randomly assigned 1,802 coronary-artery-bypass patients to conditions involving uncertainty or certainty about receiving study-arranged intercessory prayer. Receiving prayer did not improve complication-free recovery, and patients certain that they would receive it had more complications than those uncertain about receiving it (Benson et al., 2006). A Cochrane review of intercessory-prayer trials judged the evidence equivocal and did not find a basis for a confident health-effect conclusion (Roberts et al., 2009).

These results are easy to overstate. The intervention was distant intercessory prayer for a specified medical population, not the totality of Christian prayer, healing, miracles, sanctification, or the Spirit's influence. Background prayer could not be eliminated; divine freedom may resist experimental control; and a theology need not predict that a requested medical outcome will be granted. Philosophical assessment of miracle testimony also extends well beyond randomized clinical trials (Anderson, 2023).

The legitimate inference is narrower. A broad claim that organized Christian intercession reliably produces measurable recovery advantages in comparable surgical settings received no support from the strongest single trial, and the trial literature did not yield a stable effect. This makes the observed profile less likely under versions of H_S that actually predict such an advantage. It says little against models that make no such prediction. The discipline is symmetrical: one may not use an undefined miracle claim to predict success before a trial and then retreat to divine freedom only after failure without lowering the claim's empirical content.

4 Why the Domains Cannot Be Naively Multiplied

If the five domains were conditionally independent under each model, the joint likelihood ratio could be written as a product of domain-specific ratios. That assumption is implausible. The relevant joint comparison is instead

$$\Lambda_{S/O}(D_1, \dots, D_5) = \frac{p(D_1, \dots, D_5 \mid H_S, A_S)}{p(D_1, \dots, D_5 \mid H_O, A_O)}, \quad (2)$$

where the D_j denote the declared evidence domains. Religious participation, self-selection, community norms, self-report styles, socioeconomic background, and cultural learning influence multiple domains at once. A person selected into weekly attendance may acquire social support, lower substance use, stronger purpose, a vocabulary for spiritual presence, and more favorable self-description through overlapping pathways. Counting each association as an independent confirmation would double-count shared causes.

Dependence also blocks a simple cumulative case from negative findings. Measurement insensitivity, recipient misclassification, or a mismatch between theological goods and secular instruments could depress several domains together. The proper synthesis is qualitative unless a defensible joint model is supplied. Three features nevertheless recur across otherwise different designs: observed benefits are often modest or heterogeneous; many are strongly tied to ordinary mediators or self-report; and distinctive, prospectively declared Spirit-specific advantages are rarely measured. Their recurrence makes the total record more expected under H_O and H_M than under an unqualified H_S , but it does not justify a numerical ratio assembled from invented inputs.

The comparison between H_M and H_O is much less decisive. Both expect ordinary mediation, contextual variation, and weak public specificity. If the Spirit acts mainly through mechanisms that the ordinary model independently predicts, observational equivalence is not surprising. The theological model may remain true, but these data cannot do much to select it. Compatibility is not confirmation.

5 Strong Objections and Their Evidential Costs

The following objections are not ad hoc merely because they reduce public expectations; several arise from serious theological commitments. Their “cost” is not a generic penalty for doctrinal complexity. It is the specific loss of evidential contrast that occurs when the revised model makes approximately the same public record likely as H_O .

5.1 Freedom, resistance, and sanctification

Christian agency is not mechanical. People can resist grace, misinterpret guidance, or remain morally inconsistent. Sanctification may be slow, and suffering rather than comfort may accompany faithfulness. These points undermine any expectation of universal or immediate superiority. They do not eliminate every population expectation, however. A model can allow substantial noncompliance while still predicting that an efficacious and widespread influence shifts the distribution among suitable recipients. If no comparison population, time scale, or outcome could reasonably reveal such a shift, the public-effect component has been relinquished rather than defended.

5.2 Communion and holiness are not secular success

The Spirit's primary goods may be communion with the deity specified by Christian theology, Christlike love, fidelity, or ecclesial unity rather than health, wealth, or cognitive performance. This objection rightly rejects crude prosperity measures. It also improves research design: constructs should be derived from the theology being tested, evaluated by believing and nonbelieving experts, and linked to observable error criteria where possible. Yet love of neighbor, honesty, reconciliation, and wise judgment have behavioral manifestations. A model that predicts these goods should permit at least some risky expectations about longitudinal conduct rather than relying entirely on private description.

5.3 Ordinary mediation is not theological defeat

Finding that social support, ritual attention, expectancy, or moral practice mediates an outcome does not by itself refute divine action. Under H_M , the Spirit may work precisely through those causes rather than compete with them. The causal route can therefore be shared while the proposed ontologies differ: H_M treats the ordinary mechanism as divinely used, whereas H_O treats it as explanatorily sufficient.

The consequence is underdetermination, not a victory by definition for either side. A mediation analysis can illuminate how an effect occurs without identifying its ultimate source. To favor H_M over H_O , researchers need some additional conditional pattern that the mediated theology expects more strongly—for example, a preregistered recipient-by-practice interaction, a source-specific accuracy gain, or an out-of-sample pattern not captured by ordinary predictors. Without such a divergence, the data can remain compatible with divine mediation while supplying little public confirmation of it.

5.4 Selective gifts and rare interventions

The Spirit may act only in particular persons or moments. Rare-event models can survive population averages that would disfavor a widespread-effect model. But rarity generates a verification burden. Case selection must be prospective or independently audited; base rates and multiple opportunities must be included; outcomes must be specified before they occur; and failures must remain in the denominator. Without those controls, a large population supplies an inexhaustible pool of striking coincidences.

5.5 Divine agency cannot be treated as a laboratory reagent

A personal agent is not a reagent, and a morally serious theology may deny that the deity under examination is obligated to respond to experimental prompts. That blocks some experimental interpretations, especially a demand that a requested miracle occur on command. It does not block all evidence. Personal agents regularly

generate probabilistic expectations from stable purposes without being controllable. If a model says the Spirit generally forms communities toward love, truth, or discernment, longitudinal observational predictions remain possible. Divine freedom cautions against deterministic tests; it does not turn every observed profile into equally strong evidence.

5.6 Private warrant need not be public evidence

A believer may have first-person grounds, a properly functioning faculty, or a cumulative life history unavailable to public study. This possibility should be preserved. It also changes the conclusion being defended. A claim of private warrant does not show that population data discriminate the Spirit's operation, and weak public discriminability does not refute all private warrant. The two questions should not borrow authority from one another.

6 A More Discriminating Research Program

Future work can compare specified God-concepts—specified conceptions of a deity—without pretending that one experiment could decide theism as a whole. The first step is registered model elicitation: theologians and practitioners should state recipient criteria, intended goods, ordinary mediators, time scales, and failure conditions before investigators inspect outcomes. The empirical protocol should then separate confirmatory tests from exploratory refinement. Outcome-wide longitudinal design offers a useful backbone because it emphasizes pre-exposure outcomes, confounder assessment, transparent reporting across outcomes, and sensitivity to multiple testing rather than selection of a favorable endpoint after the fact (VanderWeele et al., 2020).

First, researchers could preregister a multi-tradition longitudinal study of discernment. Christian theologians and practitioners would specify recipient conditions, practices, time horizons, and outcomes central to Spirit guidance. Comparison traditions would do the same for their claimed sources. Participants would make dated, confidence-rated judgments about independently resolvable questions; scoring would include accuracy, calibration, revision after feedback, and false-positive rates. The study would distinguish doctrinal agreement from checkable performance and retain unanswered or failed predictions in the denominator.

Second, prosocial research should combine administrative or peer-verified behavior with ecologically valid field measures, separate ingroup from outgroup beneficiaries, and model prior disposition and community selection. Theology-derived constructs such as reconciliation, costly love, fidelity, or communal unity should not be reduced to a single self-report scale: they require multiple indicators, behavioral validation, and tests that the measures function comparably across traditions and time (VanderWeele, 2022). If H_S predicts distinctive transformation, its defenders should state which practices and recipient conditions generate the difference. If H_M predicts ordinary mediation, analyses should test the proposed mediators rather than statistically remove them as though they were necessarily rivals.

Third, studies of experience should collect reports in real time, blind coders to the claimant's tradition where possible, and register criteria for source-specific success before examining narratives. Culture, absorption, expectancy, trauma, sleep, and communal reinforcement should be measured. A genuine Christian signal would need to improve prediction beyond those variables and beyond generic spiritual salience.

Fourth, extraordinary-event research should use prospective registries rather than retrospective collections alone. A registry would preserve failures, denominators, medical records, timing, and alternative diagnoses. Rare events could then be evaluated against a declared background rate without assuming that statistical rarity alone establishes divine agency. Such designs would not coerce divine action; they would test the human claim that a publicly recognizable pattern is already occurring.

Across all four programs, the analytic plan should include baseline outcomes, measured confounders, negative-control exposures or outcomes where feasible, attrition diagnostics, sensitivity analysis for unmeasured confounding, and multiplicity control. Exploratory interactions can generate revised models, but the claimed discriminator should then survive a separate confirmatory sample, independent replication, and out-of-sample prediction.

These controls do not convert an imperfect proxy into direct detection of the Spirit; they reveal whether a declared public-effect model produces a stable pattern beyond ordinary predictors.

The research program would produce informative results even if no model decisively wins. A failure of a preregistered robust prediction would rationally shift attention toward a mediated, selective, or private model. A stable advantage that survived blinding, baseline adjustment, negative controls, independent replication, and out-of-sample prediction would force the ordinary benchmark to become more complex. Either outcome is epistemically preferable to flexible interpretation after the fact.

7 Conclusion

Main finding. The evidence reviewed here counts modestly against one strong claim: that the Holy Spirit normally gives appropriately situated Christians widespread and publicly recognizable advantages. The evidence fits two other possibilities better—that Spirit action is selective or works through ordinary means, and that ordinary social and psychological causes are enough to explain the public pattern. The evidence does not clearly favor either of those two possibilities over the other. This is a conclusion about one strong public claim, not about Christianity or theism as a whole.

Why. Religious participation is associated with some real benefits, including modest prosociality and several favorable health and well-being outcomes. But the overall pattern is mixed. Observed prosocial behavior shows a smaller effect than self-reports. Long-term studies find little stable improvement in a person's mental health as attendance changes. Reports of spiritual presence vary with culture and a person's tendency toward absorption. Controlled studies of intercessory prayer have not found a stable medical advantage. Direct, forward-looking tests of specifically Spirit-guided judgment remain scarce. Taken together, these findings do not show the broad, distinctive public pattern that the strong claim leads us to expect.

Limits. The available measures are imperfect. Attendance, religiosity, and self-reported experience can miss genuine Spirit action; they can also reflect ordinary influences such as community, practice, expectation, and self-selection. The argument therefore does not show that religious communities do no good, that all spiritual experiences have ordinary causes, that miracles are impossible, or that naturalism is true. Private, rare, selective, or mediated Spirit action remains possible. But if such action produces the same public pattern that ordinary causes predict, that pattern cannot by itself confirm Spirit action over those causes. A possibility can fit the evidence without being supported by it.

What follows. Anyone who appeals to public effects of the Spirit should say in advance who is expected to show them, what difference should appear, when it should appear, which ordinary causes are allowed for, and what outcome would count against the claim. Failures must remain part of the record. A claim becomes publicly testable only when some possible results can count against it.

A Formal Reconstruction of the Central Thesis

The central conclusion is comparative and defeasible. It concerns the public evidence available for a robust Spirit-potency model, not a deductive disproof of theism or of every account of Spirit action. The reconstruction below first states the reasoning in ordinary language and then shows exactly what the notation adds.

A.1 Plain-language reconstruction

1. **Premise 1.** If a robust Spirit-potency model is true under declared recipient, outcome, timing, and measurement conditions, appropriately situated Christians should normally display some stable, widespread, and distinctive public trace across the relevant domains.
2. **Premise 2.** The current record instead shows small, heterogeneous, proxy-laden, and largely nonspecific signals across conduct, well-being, experience, and extraordinary intervention, while Spirit-specific epistemic performance remains prospectively undermeasured.

3. **Premise 3.** That evidence profile is more expected under a mediated-or-selective model and under an ordinary-process benchmark than under the robust model, although it weakly distinguishes the two comparators from each other.
4. **Premise 4.** On a likelihoodist account, evidence favors one specified model over another when that evidence is more expected under the first model than under the second.
5. **Conclusion.** Therefore the current public evidence modestly disfavors the robust Spirit-potency model relative to a mediated-or-selective model and the ordinary-process benchmark; it neither disproves theism nor publicly confirms the weaker theological model over ordinary causes.

A.2 Symbols

The model symbols, likelihood expressions, and logical operators used below mean:

- H_S : the robust Spirit-potency model, including its declared assumptions about recipients, outcomes, timing, and measurement.
- H_M : the mediated-or-selective Spirit-action model, including its corresponding declared assumptions.
- H_O : the ordinary-process benchmark, including its corresponding declared assumptions.
- E : the current dependence-aware public evidence profile across the five domains.
- p : the conditional likelihood function used to represent how strongly an assumed model expects the evidence.
- $p(E | H_S)$, $p(E | H_M)$, and $p(E | H_O)$: the likelihoods of observing E when each corresponding model is assumed for purposes of comparison. They are not posterior probabilities that the models are true.
- $>$: “is more expected than.” Thus $p(E | H_M) > p(E | H_S)$ says that E fits the mediated or selective model better than it fits the robust model.
- $\wedge$: both pairwise comparisons hold together.
- $\implies$: the stated likelihood ordering warrants the comparative evidential conclusion by Premise 4; it does not turn the empirical premises into certainties.

A.3 Symbolic form

The bounded likelihoodist inference has two parts. First, each comparator fits the observed evidence better than the robust model:

$$\begin{aligned}
 & [p(E | H_M) > p(E | H_S)] \wedge [p(E | H_O) > p(E | H_S)] \\
 & \implies \\
 & E \text{ modestly favors } H_M \text{ over } H_S, \\
 & E \text{ modestly favors } H_O \text{ over } H_S.
 \end{aligned}$$

Second, the same evidence supplies little basis for ranking the two comparators against each other:

$$\begin{aligned}
 & \text{Current evidence establishes neither } p(E | H_M) > p(E | H_O) \\
 & \text{nor } p(E | H_O) > p(E | H_M) \\
 & \implies \\
 & E \text{ does not clearly distinguish } H_M \text{ from } H_O.
 \end{aligned}$$

A.4 Annotated reading

1. **The first two lines are separate comparisons.** They do not combine H_M and H_O into one disjunctive hypothesis. The evidence is compared with the robust model twice: once against mediated or selective action and once against the ordinary benchmark.
2. **Greater likelihood is not greater posterior probability.** The sign $>$ says that the observed profile is more expected under one assumed model than another. It does not state the probability that either model, Christianity, theism, or naturalism is true.
3. **The arrow marks the likelihoodist bridge.** Premise 4 licenses the move from greater expectedness to pairwise evidential favoring. The word “modestly” records the uncertainty and qualitative character of the premises; it is not a hidden numerical value in the formula.
4. **The no-ordering line records weak discrimination.** It does not claim that two calculated likelihoods are exactly equal. It says that the present evidence establishes no stable, material ordering between the mediated-or-selective model and the ordinary benchmark.
5. **The formula is deliberately silent about stronger claims.** It does not entail that H_S is false, that H_O is true, or that H_M is publicly confirmed. It expresses only the pairwise, comparative evidential result defended in the body.

Validity. Given Premise 3's pairwise likelihood ordering and Premise 4's likelihoodist rule, the two conclusions favoring H_M and H_O over H_S follow by direct application of that rule. Premises 1 and 2 do not deductively generate the likelihood ordering by themselves; they supply the substantive, defeasible grounds for Premise 3. The second displayed line is likewise a calibrated interpretation of weak comparative separation, not a theorem of pure logic.

Premise support. Premise 1 is a defeasible conceptual bridge from a powerful, purposive, widespread cause to an expected public difference. Premise 2 is an inductive summary of the cited record, limited by noisy proxies, shared causal pathways, and an undermeasured epistemic domain. Premise 3 is a qualitative comparative judgment based on that bridge and record, not a measured likelihood ratio. Premise 4 states the likelihoodist evidential rule. These different kinds of support explain why the conclusion is modest and contrastive rather than a deductive disproof.

Scope. The formal result pressures only the robust public-effect model under its declared conditions. It leaves rare, private, selective, mediated, and other God-concepts logically open while emphasizing that logical compatibility with E is not public confirmation over H_O .

References

- Anderson, C. (2023). Rationality and miracles. In J. Fuqua, J. Greco, & T. McNabb (Eds.), *The cambridge handbook of religious epistemology* (pp. 56–69). Cambridge University Press.
<https://doi.org/10.1017/9781009047180.006>
- Benson, H., Dusek, J. A., Sherwood, J. B., Lam, P., Bethea, C. F., Carpenter, W., Levitsky, S., Hill, P. C., Clem, J., Donald 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>
- Chen, Y., Kim, E. S., & VanderWeele, T. J. (2020). Religious-service attendance and subsequent health and well-being throughout adulthood: Evidence from three prospective cohorts. *International Journal of Epidemiology*, 49(6), 2030–2040. <https://doi.org/10.1093/ije/dyaa120>
- Dormandy, K. (2023). Religious disagreement. In J. Fuqua, J. Greco, & T. McNabb (Eds.), *The cambridge handbook of religious epistemology* (pp. 208–223). Cambridge University Press.
<https://doi.org/10.1017/9781009047180.018>
- Fee, G. D. (1994). *God's empowering presence: The holy spirit in the letters of paul*. Hendrickson.
- Kelly, J. M., Kramer, S. R., & Shariff, A. F. (2024). Religiosity predicts prosociality, especially when measured by self-report: A meta-analysis of almost 60 years of research. *Psychological Bulletin*, 150(3), 284–318.
<https://doi.org/10.1037/bul0000413>
- Luhrmann, T. M., Weisman, K., Aulino, F., Brahinsky, J. D., Dulin, J. C., Dzokoto, V. A., Legare, C. H., Lifshitz, M., Ng, E., Ross-Zehnder, N., & Smith, R. E. (2021). Sensing the presence of gods and spirits across cultures and faiths. *Proceedings of the National Academy of Sciences*, 118(5), Article e2016649118. <https://doi.org/10.1073/pnas.2016649118>
- Macchia, F. D. (2006). *Baptized in the spirit: A global pentecostal theology*. Zondervan.
- Prati, G. (2024). Is religion beneficial for mental health? a 9-year longitudinal study. *International Journal of Clinical and Health Psychology*, 24(3), Article 100491. <https://doi.org/10.1016/j.ijchp.2024.100491>
- Prati, G. (2025). Does religious-service attendance increase mental health? a random-intercept cross-lagged panel analysis across 18 years. *Psychological Science*, 36(3), 157–167.
<https://doi.org/10.1177/09567976251325449>
- Roberts, L., Ahmed, I., Hall, S., & Davison, A. (2009). Intercessory prayer for the alleviation of ill health. *Cochrane Database of Systematic Reviews*, (2), Article CD000368.
<https://doi.org/10.1002/14651858.CD000368.pub3>
- Royall, R. M. (1997). *Statistical evidence: A likelihood paradigm*. Chapman & Hall.
- Sober, E. (2008). *Evidence and evolution: The logic behind the science*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511806285>
- VanderWeele, T. J. (2022). Constructed measures and causal inference: Towards a new model of measurement for psychosocial constructs. *Epidemiology*, 33(1), 141–151.
<https://doi.org/10.1097/EDE.0000000000001434>
- VanderWeele, T. J., Mathur, M. B., & Chen, Y. (2020). Outcome-wide longitudinal designs for causal inference: A new template for empirical studies. *Statistical Science*, 35(3), 437–466.
<https://doi.org/10.1214/19-STS728>
- Zahl, S. (2020). *The holy spirit and christian experience*. Oxford University Press.
<https://doi.org/10.1093/oso/9780198827788.001.0001>