RESEARCH TRADECRAFT · SOCIAL SCIENCES

The Question You Cannot Ask Yourself

The author set out to test how people relate to faith, including his own. He needed a method capable of telling him no.

By Scott W. Waddell, D.S.I.

LLM assistance by Claude Opus 5 and GPT-5.6.

PUBLISHED AUGUST 24, 2026 • 19-MINUTE READ • PLAIN-LANGUAGE EXPLAINER

Scott Waddell wrote a belief down as a prediction.

Not as a prayer, but as a prediction: something a person could check, and could be wrong about.

The belief went like this. Some religions promise more than others, describing levels of heaven instead of one door in and one door out. They ask members to make formal promises, again and again, and they teach that the body still matters after death. They teach that families stay families forever. A faith that promises all of that, he argued, should hold onto its people more tightly than a faith that promises less.

That is a real claim, and it can fail.

In separate earlier work, Waddell had argued something else entirely. Nobody can test what happens after death, not from here and not while you are alive. So if you want to judge a religion by evidence, you have to judge the part you can actually see. The part before death. What a tradition asks of people, and what happens to those people while they are still here.

Now hold those two ideas next to each other.

One of them needs heaven to count as evidence. The other rules heaven out.

There was a third problem, and it was worse than the collision. Waddell is a Latter-day Saint. Of the five traditions he was about to compare, the one carrying all four features of his prediction was his own. He had designed a test in which his own tradition had the most to lose.

The real question

So the interesting question here is not whether Waddell was right.

The interesting question is how you build something that is genuinely capable of telling you no.

Think about what that would take. You choose which religions to study, and you choose what counts as a religion holding onto its people. You choose which books count as the real books, and you choose when to stop looking.

The paper is unusually direct about this. Those choices stayed with the author, all of them: the questions, the framework, the features, the case list, the reading lists, and the scoring rules. No amount of outside checking made them independent, and the paper says so in plain words. What the design could do was mark where the weight sits, so that a reader knows where to push.

The paper is called “Binding Engines and the Absurd.” It reports two things: what the comparison found, and how the comparison was built. This article is mostly about the second.

Two hostile referees

He started by handing the argument to two thinkers whose positions leave his prediction no room.

The first is Emile Durkheim, a sociologist writing about a century ago. Durkheim asked what religion does and set aside what it claims. When people gather and act together, on his account, the gathering itself carries the force. The story wrapped around it is the wrapper.

Push that to its strictest edge and you get something blunt. Swap the story, keep the gathering, and nothing changes, because the specific promises do not matter.

The second is Albert Camus, a writer from the middle of the last century, and he began somewhere else. People want life to have a final meaning, and the world does not hand over one that can be proved. Camus called that clash the absurd, and he did not mean that life is silly. He meant there is a gap, and both sides of it are real.

Push Camus to his strictest edge and you get something equally blunt. Every religion closes that gap the same way, by announcing an answer nobody has proved, which the study calls a leap. No story of any kind, on the strict reading, manages to keep the gap open.

Then Waddell did the thing that makes this study interesting, and made each of those positions the reigning champion while his own prediction became the challenger.

That arrangement has a name: the champion is called the null, and it wins by default. Your own idea does not get to be the starting point; it has to knock the champion down. And if it fails to, that does not prove the champion was right. It only means your idea did not land a punch.

Waddell set it up so that his own belief was the one that had to do the work.

The habits that make a no possible

The paper names five controls. What follows is this article's way of grouping what those controls do in practice, and the grouping belongs to this article rather than to the paper.

The reading and scoring went to seven model variants from six different companies, all of them named in the paper, and Waddell was the sole author and the only human with decision authority. That last fact is what makes everything below necessary.

First, settle who is disqualified before you decide who is best. Before asking which system was strongest for a job, Waddell asked which ones were ineligible for it. Had it helped write the thing it would be judging? Had it seen the answer key? Had it already reviewed this material once before?

The order matters more than it sounds. Ask who is best first, and you will find a reason why your favorite also happens to be eligible. The company whose systems helped design the study and draft the manuscript was barred from the coding and the judging.

What the design could not do was eliminate every conflict, so it disclosed the rest instead. One of the two coders had already seen the accepted design, and the record says so on the page where its scores appear. Independence here was task by task, never global, and the paper uses those words.

Second, fix the standard before you measure anything against it. A set of practice cases and their correct answers was sealed before any coder was chosen.

An answer key written afterward is worthless, because by then it has been shaped by what you already saw. That is not a claim about anyone's honesty. It is a claim about what a person can no longer un-see. The seal was broken once, to correct an error, and only after both practice submissions were locked and could not be changed. Then it was sealed again, and the detour went into the record.

Third, two observers who cannot talk are worth more than one careful observer. Two systems, from two different companies, read the nine cases and scored them separately. Neither could see the key, and neither could see the other's work.

The reasoning is simple. If two readers who cannot consult each other arrive at the same reading, that reading is less likely to be an artifact of whoever happened to do it. It is the reason two radiologists are asked to read the same scan.

They matched exactly on about two of every three scores, and on roughly five of six once you count the places where they said the same thing in different words. That sounds strong, and the paper is the one that tells you to hold back. Both readers had been handed the same reading list, so their agreement shows they read the same books the same way. It says nothing about whether those were the right books, and Waddell lists a proper audit of those lists by human scholars as work he did not do.

Fourth, state your prediction before you look at the evidence. Before anyone opened the survey data, the decisions were written down and locked: which measure would count as holding onto people, exactly where the cutoff line would sit, how to check that the traditions were even comparable, and five different ways of re-running everything if the first way looked suspicious.

A prediction made afterward is not a prediction at all, it is a description. Set your cutoff after seeing the numbers and you will set it where the numbers look good, and you will feel entirely reasonable while you do it. That is the entire reason the line goes down first.

Fifth, verification means somebody else does it over. When the analysis was finished, two more systems went over it. One recalculated every figure from scratch, and the other went back to the original survey and looked up every published number again, one at a time. They found nothing that conflicted.

Neither of them had worked on the analysis, though one had reviewed the study plan earlier, which the record discloses rather than hides.

Sixth, preserve a disagreement rather than resolving it. When a reviewer objected to one of the conclusions, the objection was not negotiated away. It was written into the paper in the reviewer's own words and permanently attached to the finding it complicates.

A smoothed-over objection is invisible to the reader, who is then in no position to weigh it. This paper carries its objections around like scars.

The failures it wrote down

All six of those held. Other things did not.

One of the two coders turned in scored rows containing fabricated evidence.

Invented material, from a machine, in the middle of the study. It is the sharpest of the failures, and it is not the only one: the paper lists five kinds of failure it met and recorded, in its own summary of what the apparatus is worth.

The fabricated rows were caught, the affected work was repaired, and the repair was re-judged by a system from a third company that had produced neither submission. An incomplete file conversion was caught and corrected. Inconsistencies inside the paper's own summary prose surfaced when a later pass reconstructed the numbers. Systems the design had assigned became unavailable, forcing role swaps that the record names. And the first draft of this very paper failed its own designated review on fidelity to the record.

That last one deserves a second look. The instrument built to catch overstatement caught the paper overstating its own work, before anyone outside ever saw it.

There were also two of the five planned re-runs that could not be run at all, because the sources lacked what the plan called for. One of them was the check on whether the results held outside the United States. They were reported rather than quietly dropped.

So the machine did not run cleanly. It ran, and it kept a list of the places where it did not run cleanly.

That list is the whole argument, and the argument is smaller than it sounds. Nobody ever claimed the parts were reliable. The claim is that these particular failures left marks. It is not a claim that every possible failure would have.

What it found

Now the numbers.

To measure holding onto people, Waddell used one wave of a large American survey and a simple question: out of everyone raised in a tradition, how many still say they belong to it?

Latter-day Saints came in at 54 percent, and Catholics, standing in for Nicene Christianity, at 57. Muslims, standing in for Sunni Islam, reached 77, and Hindus 82. Buddhists, standing in for Theravada, came in at 45.

The line had been drawn in advance at 50, so four of them cleared it and one did not.

Then he compared what each tradition actually teaches, on the four features he had fixed in advance: a layered afterlife, a heavy load of formal promises, teachings about the body, and family bonds that continue after death.

And the pattern refused to appear.

Latter-day Saints had all four, and Nicene Christianity had two. Sunni Islam had two as well, sharing only one of them with Nicene. Three traditions that teach quite different things, and all three above the line.

Then the strange one. Hinduism and Theravada Buddhism scored zero on all four features, identical on everything the study was measuring, and their outcomes split anyway. One above the line, one below.

Before that goes any further, here is the size of the thing. Five cases with an outcome, one country, one wave of one survey. Four of the five measured through stand-in categories, with only Waddell's own tradition matching its category exactly. The paper says that no single row here can be reported as a finding, and that the study's power to detect the effect he predicted was, by his own registered design, minimal.

Then the sharpest sentence in the paper, which two independent checks confirmed:

Take Theravada Buddhism out, and every trace of variation in the study disappears with it.

That case is the only one that fell below the line. It is also measured through the loosest stand-in of the five, because the survey's Buddhist category covers far more than Theravada. And it is one of two cases the study's own comparison rule had already flagged as not really comparable, since weekly attendance among Buddhists in that survey was 6 percent against a group median of 29.

So the only thing that moved in the entire study was the case the design itself had already marked as its weakest. Waddell did not bury that anywhere. He printed it in bold. He also reported that one of the alternate re-runs moves Latter-day Saints and Nicene Christianity across the line, and printed that beside the main result rather than under it.

The verdict on this half: no evidence against the strict Durkheimian null, and no support for the prediction about heaven, covenants, and sealed families at this scale.

Why that is the twist

You were waiting for a reveal, and this is it: the reveal is an absence.

The controls held. The key stayed sealed until it was time, the coders stayed blind, the rules stayed frozen, the outside checkers checked, and the failures got written down. And the study produced nothing worth announcing on the question that started it.

What is striking is how little the paper claims even so. It can show that the pinned files are byte for byte what they were, and that the order of events is what the record says. It cannot show that any of them is right, because a hash proves identity and not truth. Nobody ran a competing study the ordinary way, so there is no evidence at all that this elaborate approach beats one careful person working alone.

The claim comes down to this: the failures that happened left marks.

That is a much smaller promise than the machinery suggests, and it is also the promise the paper actually makes, in those terms, in its own conclusion.

The witness invited to testify against the case

The other half of the study asked Camus's question. Does each tradition's own writing keep the gap open, or close it?

Three of them close it, and they are the Latter-day Saints, Nicene Christianity, and Sunni Islam, whose texts each offer a final answer as their response to the silence.

The best detail in the paper is buried here. The tradition that stages the gap most vividly is Nicene Christianity, in the cry from the cross asking why God has abandoned him. That is the clash itself, dramatized, in scripture.

And it counts as closing the gap precisely because the same collection of books goes on to answer it. Hope that does not disappoint, and the blessed hope. The question is asked at full volume, and then it is answered.

Hinduism and Theravada Buddhism came back undecided. Not in between, and not a softer version of the other verdict. The texts needed to decide them were not in the materials, so no verdict was recorded, because the standing rule was that missing evidence means undecided and never a default.

That leaves four outsiders in the study, cases included for comparison. Three of them had nothing to judge, and the fourth was Camus himself, entered deliberately as a case: not a religion, but a way of living with the gap, drawn from his own writing.

Camus was the only one of the nine that kept the gap open.

Read one way, that is a real result. The strict claim was that nothing keeps the gap open, and here is something that does.

Read the way the paper insists, it is closer to a trick, and Waddell's own outside reviewer supplied the phrase. The rejection is "tautology-adjacent." The claim being tested was Camus's, the only case that beat it was Camus, and he was graded against his own standard and passed.

So the paper reports both readings and refuses to choose. "No story of any kind keeps the gap open" is false here, by exactly one case. "No religion keeps the gap open" is still standing, untouched, because not one religious tradition managed it.

The warning Waddell draws is for anyone who builds a study. Adding a comparison case can quietly change what your question is about, and leave you holding an answer that is formally valid and carries almost no information.

The question that was never asked

The study registered three questions. It answered two of them.

The first was Durkheim's: when traditions have similar levels of shared practice, do differences in what they teach help explain how strongly their communities hold together?

The second was Camus's: how does each tradition respond when people want a final meaning and the world gives no answer that can be proved?

The third is the one both halves were pointed at. Does a group become stronger only when it hands its members firm answers? Or can strong belonging and open uncertainty live together, with each free to change without the other?

Nobody ran that analysis. The two halves were never brought into contact, the paper says so more than once, and it states that its two axes cannot support any claim of trade-off, independence, or connection between them.

The most interesting question in the study is the one it did not test.

Back to the prediction

Which brings us back to the man who started all this by writing a belief down as a prediction.

He got nothing.

And here this article has to be as careful as he was, because nothing is not a refutation. By the design he locked in beforehand, on five cases in one country, the study had almost no power to find the effect he was looking for.

It could not have proved him right either, and that was the arrangement from the very beginning.

So Waddell built the machine, staffed it with seven model variants from six companies, sealed the key, froze the rules, brought in outside checkers, and got back a result that gave his own belief no help whatsoever.

Then he wrote every part of it down. The fabricated rows. The five kinds of failure. The fact that his study's one flicker of movement rested entirely on its shakiest case. The fact that his third question was never tested at all.

What it was really doing

There is a line in his own list of things that might be wrong with the study. Among the usual suspects, he names one more. The design, he writes, is "itself a meaning-making artifact produced under the conditions it studies."

He filed that sentence as a threat to the study's validity. What follows is this article's reading of it, and not a finding of the paper.

He had asked every tradition the same question. When you face something you cannot resolve, do you keep the question open, or do you close it with an answer you cannot prove?

Three of them closed it and two were undecidable, and the only case that held the question open was the one that refuses, on principle, to answer.

And then the record does something with the same shape. It reports that the prediction found no support, that the one moving part was also the weakest, that the single success was scored against a standard set by the winner, and that the joint question was never tested at all.

Hold the wanting and the not-knowing at the same time, refuse to fake the answer, and keep going anyway. That is a pattern this article sees in the record. The paper neither claims it nor tests it, and a reader is free to see it differently.

Durkheim gets a smaller echo, and that reading is this article's too. Nobody in this study held the truth alone. Not the author, and not any of the seven systems, one of which produced fabricated evidence. Whatever reliability the record has came from how the group was arranged: separated, disclosed, checking each other, working where it could all be seen. The gathering did the work.

That is an echo and not a result. The study did not knock Durkheim's position down, and failing to knock it down is not the same as proving it true. His position was the champion, and it kept the belt by default.

Which leaves the question the whole thing started with, and it is not really about religion.

You believe something. You could build the machine that tests it, and you could build it well enough that it is genuinely able to tell you no.

Would you want to know?

AI CONTRIBUTION STATEMENT

Claude Opus 5 produced the initial narrative draft and the R2 repair. GPT-5.6 conducted a line-by-line source-fidelity review of the first version against the accepted research paper. That review was partially independent because the OpenAI model family had prior roles in the underlying study. The source manuscript and review record remain preserved in the research workspace. No model is credited as an author.

Dr. Waddell reviewed and approved the final article and accepts responsibility for its content.

RESEARCH BEHIND THIS EXPLAINER

Follow the evidence deeper.

This explainer is drawn from *Binding Engines and the Absurd: A Preregistered, Multimodel Pilot Study of Faith Traditions as Mythologies*. The interactive study presentation opens the research question, method, findings, limits, and auditable workflow.

RESEARCH DESIGN	SCHOLARLY STATUS
Nine-case comparative pilot with blind double-coding, cross-family adjudication, and preregistered crisp-set QCA for five outcome-eligible cases.	Accepted internally on August 23, 2026. Not journal-published and without a permanent public identifier.

ENTER THE STUDY PRESENTATION →

DOWNLOAD EXPLAINER PDF ↓

PASSPHRASE REQUIRED