

Daryl Bem Proved ESP Is Real. Which Means Science Is Broken.

By Daniel Engber

Which means science is broken.

May 17, 2017 Cover Story

It seemed obvious, at first, that Jade Wu was getting punked. In the fall of 2009, the Cornell University undergraduate had come across a posting for a job in the lab of one of the world's best-known social psychologists. A short while later, she found herself in a conference room, seated alongside several other undergraduate women. "Have you guys heard of extrasensory perception?" Daryl Bem asked the students. They shook their heads.

While most labs in the psych department were harshly lit with fluorescent ceiling bulbs, Bem's was set up for tranquility. A large tasseled tapestry stretched across one wall, and a cubicle partition was draped with soft, black fabric. It felt like the kind of place where one might stage a séance.

"Well, extrasensory perception, also called ESP, is when you can perceive things that are not immediately available in space or time," Bem said. "So, for example, when you can perceive something on the other side of the world, or in a different room, or something that hasn't happened yet."

It occurred to Wu that the flyer might have been a trick. What if she and the other women were themselves the subjects of Bem's experiment? What if he were testing whether they'd go along with total nonsense?

"I know this sounds kind of out there," Wu remembers Bem saying, "but there is evidence for ESP, and I really believe it. But I don't need you to believe it. In fact, it's better if you don't. It's better if I can say, 'Even my staff don't believe in this.'"

As Bem went on, Wu began to feel more at ease. He seemed genuine and kind, and he wasn't trying to convert her to his way of thinking. *OK, so maybe there's going to be a you-got-punked moment at the end of this*, she thought, *but at least this guy will pay me.*

In truth, Bem had no formal funding for his semisecret research program. For nearly a decade, he'd been paying undergraduates like Wu out of his own pocket, to help him demonstrate that we all possess some degree of precognition—a subtle sense of what will happen in the future. He rarely came into the lab himself, so he'd leave his lab assistants an envelope stuffed with bills. They dispensed \$5 from the kitty to each subject they ran through the experiment.

For the rest of that semester and into the one that followed, Wu and the other women tested hundreds of their fellow undergrads. Most of the subjects did as they were told, got their money, and departed happily. A few students—all of them white guys, Wu remembers—would hang around to ask about the research and to probe for flaws in its design. Wu still didn't believe in ESP, but she found herself defending the experiments

to these mansplaining guinea pigs. The methodology was sound, she told them—as sound as that of any other psychology experiment.

In the spring of 2010, not long after Wu signed on, Bem decided he'd done enough to prove his claim. In May, he wrote up the results of his 10-year study and sent them off to one of his field's most discerning peer-reviewed publications, the *Journal of Personality and Social Psychology*. (*JSPS* turns away some 85 percent of all submissions, making its acceptance rate comparable to that of the Cornell admissions office.) This was the same journal where Bem had published one of the first [papers](#) of his career, way back in 1965. Now he would return to *JSPS* with the most amazing research he'd ever done—that anyone had ever done, perhaps. It would be the capstone to what had already been a historic 50-year career.

Having served for a time as an associate editor of *JSPS*, Bem knew his methods would be up to snuff. With about 100 subjects in each experiment, his sample sizes were large. He'd used only the most conventional statistical analyses. He'd double- and triple-checked to make sure there were no glitches in the randomization of his stimuli.

Even with all that extra care, Bem would not have dared to send in such a controversial finding had he not been able to replicate the results in his lab, and replicate them again, and then replicate them five more times. His finished paper lists nine separate ministudies of ESP. Eight of those returned the same effect.

Bem's 10-year investigation, his nine experiments, his thousand subjects—all of it would have to be taken seriously. He'd shown, with more rigor than anyone ever had before, that it might be possible to see into the future. Bem knew his research would not convince the die-hard skeptics. But he also knew it couldn't be ignored.

When the study went public, about six months later, some of Bem's colleagues guessed it was a [hoax](#). Other scholars, those who believed in ESP—their is a small but fervent field of study—saw his paper as validation of their work and a chance for mainstream credibility.

“Reading it made me physically unwell.”—E.J. Wagenmakers on Daryl Bem's ESP paper

But for most observers, at least the mainstream ones, the paper posed a very difficult dilemma. It was both methodologically sound and logically insane. Daryl Bem had seemed to prove that time can flow in two directions—that ESP is real. If you bought into those results, you'd be admitting that much of what you understood about the universe was wrong. If you rejected them, you'd be admitting something almost as momentous: that the standard methods of psychology cannot be trusted, and that much of what gets published in the field—and thus, much of what we think we understand about the mind—could be total bunk.

If one had to choose a single moment that set off the “replication crisis” in psychology—an event that nudged the discipline into its present and [anarchic](#) state, where even [textbook findings](#) have been [cast in doubt](#)—this might be it: the publication, in early 2011, of Daryl Bem's experiments on second sight.

The replication crisis as it's understood today may yet prove to be a passing worry or else a mild problem calling for a soft corrective. It might also grow and spread in years to come, flaring from the [social sciences](#) into other disciplines, burning trails of cinder through [medicine](#), [neuroscience](#), and [chemistry](#). It's hard to see into the future. But here's one thing we can say about the past: The final research project of Bem's career landed like an ember in the underbrush and set his field ablaze.

Scientist With Tarot Cards

Lisa Larson-Walker

Daryl Bem has always had a knack for not fitting in. When he was still in kindergarten—a gentle Jewish kid from Denver who didn’t care for sports—he was [bullied so viciously](#) that his family was forced to move to a different neighborhood. At the age of 7, he grew interested in magic shows, and by the time he was a teenager, he’d become infatuated with mentalism. Bem would perform tricks of [mind-reading and clairvoyance](#) for friends and classmates and make it seem as though he were telepathic.

As a student, Bem was both mercurial and brash. He started graduate school in physics at MIT, then quickly changed his mind, transferring to the University of Michigan to study as a social psychologist. While at Michigan, still in his early 20s and not yet in possession of his Ph.D., Bem took aim at the leading figure in his field, Leon Festinger. For his dissertation, Bem proposed a [different explanation](#)—one based on the old and out-of-fashion writings of behaviorist B.F. Skinner—for the data that undergirded Festinger’s theory of [cognitive dissonance](#).

This would be Bem’s method throughout his career: He’d jab at established ways of thinking, rumble with important scholars, and champion some antique or half-forgotten body of research he felt had been ignored. Starting in the 1970s, he quarreled with famed personality psychologist [Walter Mischel](#) by proffering a [theory of personality](#) that dated to the 1930s. Later, Bem would argue against the biological theory of sexual orientation, favoring a developmental hypothesis that derived from “[theoretical and empirical building blocks ... already scattered about in the literature](#).”

As a young professor at Carnegie Mellon University, Bem liked to close out each semester by performing as a mentalist. After putting on his show, he’d tell his students that he didn’t really have ESP. In class, he also stressed how easily people can be fooled into believing they’ve witnessed paranormal phenomena.

Around that time, Bem met Robert McConnell, a biophysicist at the University of Pittsburgh and an [evangelist for ESP research](#). McConnell, the [founding president of the Parapsychological Association](#), told Bem the evidence for ESP was in fact quite strong. He invited Bem to join him for a meeting with Ted Serios, a man who could supposedly [project his thoughts onto Polaroid film](#). The magic was supposed to work best when Serios was inebriated. (The psychic called his booze “film juice.”) Bem spent some time with the drunken mind-photographer, but no pictures were produced. He was not impressed.

In his skepticism about ESP, Bem for once was not alone. The 1970s marked a [golden age for demystifying paranormal claims](#). James Randi, like Bem a trained stage magician, had made his name as a professional debunker by exposing the likes of [Uri Geller](#). Randi subsequently took aim at researchers who studied ESP in the lab, [sending a pair of stage performers](#) into a well-funded parapsychology lab at Washington University in 1979. The fake psychics convinced the lab their abilities were real, and Randi did not reveal the hoax until 1983.

As debunkers rose to prominence, the field of psychical research wallowed in its own early version of the replication crisis. The laboratory evidence for ESP had begun to shrivel under careful scrutiny and sometimes seemed to disappear entirely when others tried to reproduce the same experiments. In October 1983, the Parapsychology Foundation held a conference in San Antonio, Texas, to address the field’s “[repeatability problem](#).” What could be done to make ESP research more reliable, researchers asked, and more resilient to fraud?

A [raft of reforms](#) were proposed and implemented. Experimenters were advised to be wary of the classic test for “statistical significance,” for example, since it could often be misleading. They should avail themselves of larger groups of subjects, so they’d have sufficient power to detect a real effect. They should also attempt to replicate their work, ideally in adversarial collaborations with skeptics of the paranormal, and they should analyze the data from lots of different studies all at once, including those that had never gotten published. In short, the field of parapsychology decided to adopt the principles of solid scientific practice that had long been ignored by their mainstream academic peers.

As part of this bid to be taken seriously by the scientific establishment, a noted ESP researcher named Chuck Honorton asked Bem to visit his lab in Princeton, New Jersey. He thought he’d found [strong evidence in favor of telepathy](#), and he wanted Bem to tell him why he might be wrong.

Bem didn’t have an answer. In 1983, the scientist and stage performer made a careful audit of the Honorton experiments. To his surprise, they appeared to be airtight. By then, Bem had already started to reconsider his doubts about the field, but this was something different. Daryl Bem had found his faith in ESP.

Not long after she was hired, Jade Wu found herself staring at a bunch of retro pornography: naked men with poofy mullets and naked girls with feathered hair. “I’m gay, so I don’t know what’s sexy for heterosexuals,” Bem had said, in asking for her thoughts. Wu didn’t want to say out loud that the professor’s porno pictures weren’t hot, so she lied: *Yeah, sure, they’re erotic.*

These would be the stimuli for the first of Bem’s experiments on ESP (or at least the first one to be reported in his published paper). Research subjects—all of them Cornell undergraduates—saw an image of a pair of curtains on a computer monitor. They were then prompted to guess which of the curtains concealed a hidden image. The trick was that the correct answer would be randomly determined only *after* the student made her choice. If she managed to perform better than chance, it would be evidence that she’d intuited the future.

Bem had a reason for selecting porn: He figured that if people did have ESP, then it would have to be an adaptive trait—a sixth sense that developed over millions of years of evolution. If our sixth sense really had such ancient origins, he guessed it would likely be attuned to our most ancient needs and drives. In keeping with this theory, he set up the experiment so that a subset of the hidden images would be arousing to the students. Would the premonition of a pornographic image encourage them to look behind the correct curtain?

The data seemed to bear out Bem’s hypothesis. In the trials where he’d used erotic pictures, students selected their location 53 percent of the time. That marked a small but significant improvement over random guessing.

For another experiment, Bem designed a simple test of verbal memory. Students were given several minutes to examine a set of words, then were allotted extra time to practice typing out a subset of those words. When they were asked to list as many of the words as possible, they did much better on the ones they’d seen a second time. That much was straightforward: Practice can improve your recall. But when it was time to run the study, Bem flipped the tasks around. Now the students had to list the words just *before* the extra practice phase instead of after it. Still, he found signs of an effect: Students were better at remembering the words they would type out later. It seemed as though the practice session had benefits that extended backward through time.

Similar experiments, with the sequence of the tasks and stimuli reversed, showed

students could have their emotions primed by words they hadn't seen, that they would recoil from scary pictures that hadn't yet appeared, and that they would get habituated to unpleasant imagery to which they would later be exposed. Almost every study worked as Bem expected. When he looked at all his findings together, he concluded that the chances of this being a statistical artifact—that is to say, the product of dumb luck—were infinitesimal.

This did not surprise him. By the time he'd begun this research, around the turn of the millennium, he already believed ESP was real. He'd delved into the published work on telepathy and clairvoyance and concluded that Robert McDonnell was right: The evidence in favor of such phenomena, known to connoisseurs as "[psi](#)" processes, was compelling.

Indeed, a belief in ESP fit into Bem's way of thinking—it tempted his contrarianism. As with his attacks on cognitive dissonance and personality theory, Bem could draw his arguments from a well-developed research literature—this one [dating to the 1930s](#)—which had been, he thought, unfairly rejected and ignored.

Together with Chuck Honorton, the paranormal researcher in Princeton, Bem set out to summarize this research for his mainstream colleagues in psychology. In the early 1990s, they put together a review of all the work on ESP that had been done using Honorton's approach and sent it to Bem's associate Robert Sternberg, then the editor of *Psychological Bulletin*. "We believe that the replication rates and effect sizes achieved ... are now sufficient to warrant bringing this body of data to the attention of the wider psychological community," he and Honorton wrote in a paper titled "[Does Psi Exist?](#)" Sternberg made the article the lead of the January 1994 issue.

By 2001, Bem had mostly set aside his mainstream work and turned to writing [commentaries and book reviews](#) on psi phenomena. He'd also quietly embarked upon a major scientific quest, to find what he called "the holy grail" of parapsychology research: a fully reproducible experiment on ESP that any lab could replicate. His most important tool, as a scientist and rhetorician, would be simplicity. He'd work with well-established protocols, using nothing more than basic tests of verbal memory, priming, and habituation. He'd show that his studies weren't underpowered, that his procedures weren't overcomplicated, and that his statistics weren't convoluted. He'd make his methods bland and unremarkable.

In 2003, 2004, 2005, and 2008, Bem presented pilot data to the annual meeting of the Parapsychological Association. Finally, in 2010, after about a decade's worth of calibration and refinement, he figured he'd done enough. A thousand subjects, nine experiments, eight significant results. This would be his solid, mainstream proof of ESP—a set of tasks that could be transferred to any other lab.

On May 12, 2010, he sent a manuscript to the *Journal of Personality and Social Psychology*. He called it "Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect."

Scientist With Crystal Ball

Lisa Larson-Walker

The first time E.J. Wagenmakers read Bem's ESP paper, he was having lunch at a neuroscience conference in Berlin. "I had to put it away several times," he recalls. "Reading it made me physically unwell."

Wagenmakers, a research methodologist from the University of Amsterdam, believed

the paper had at least one glaring problem: There was no clear dividing line between the exploratory and confirmatory phases of Bem's research. He noticed, for example, that there were lots of different ways Bem might have analyzed the data in his erotic pictures study. He could have looked for ESP on neutral pictures instead of just erotic ones, or exclusively on happy pictures, or on nonerotic pictures that happened to be romantic. If you give yourself a dozen different ways to slice and dice your data, you're at much greater risk of finding patterns in a set of random blips. That's not so bad at the start of your research, when you're working out the best approach for your experiments, but later on it can be disastrous. If Bem hadn't decided well ahead of time exactly how he planned to crunch his numbers, all his findings would be suspect.

Had Bem made those choices in advance? The wording of his paper suggested that he had. But that's the way papers in his field are written—or at least it's how they were written back in 2010: People would act as if they'd preplanned everything, even when they'd bushwhacked their way through a thicket of results, ignoring all the dead ends they came across along the way. Statistician Andrew Gelman refers to this as “the garden of forking paths”: If you don't specify your route before you start, any place you end up can be made to seem like a meaningful destination. (Gelman [ascribed this problem to Bem's research](#) in a 2013 piece for *Slate*.)

Bem's paper did acknowledge that he'd done pilot testing and even cites three sets of findings that he did not include in his final analysis. “As all research psychologists know, many procedures get tried and discarded,” he wrote. He did not, however, mention the other forks he traversed throughout the research process. Wu, for one, remembers Bem making lots of tweaks to his experiments. He would adjust the numbers of trials and the timing of the stimuli, she says.

That's to be expected: Lab research is a messy process, and it's not always clear in retrospect—even to the author of a study—when and how every decision got made. “I would start one [experiment], and if it just wasn't going anywhere, I would abandon it and restart it with changes,” Bem told me recently. Some of these changes were reported in the article; others weren't. “I didn't keep very close track of which ones I had discarded and which ones I hadn't,” he said. Given that the studies spanned a decade, Bem can't remember all the details of the early work. “I was probably very sloppy at the beginning,” he said. “I think probably some of the criticism could well be valid. I was never dishonest, but on the other hand, the critics were correct.”

There's now a movement in psychology to “pre-register” your research, so you commit yourself in advance to a plan for running your experiment and analyzing the data. Even now, the wisdom of this practice is [contested](#), and it's certainly the case that journal editors never would have expected Bem to pre-register anything circa the early 2000s, when he started on his ESP research.

“Clearly by the normal rules that we [used] in evaluating research, we would accept this paper,” said Lee Ross, a noted social psychologist at Stanford who served as one of Bem's peer reviewers. “The level of proof here was ordinary. I mean that positively as well as negatively. I mean it was exactly the kind of conventional psychology analysis that [one often sees], with the same failings and concerns that most research has.”

In his submission to the editors of *JPSP*, Bem had recommended Ross as a reviewer. The two men are very close: Each is a godparent to the other's children. Although Ross did not believe in psi phenomena, he was knowledgeable about the field, having analyzed data from parapsychology experiments as a graduate student in the 1960s. Indeed, Ross didn't trust the data in the paper—he still does not believe in ESP—but he also knew there was no chance whatsoever that his friend had been deceptive or incompetent.

Bem had shown that even a rigorous scientist could cart himself to crazy-land, just by following the rules of the road.

Ross made his name as a psychologist, in part, by demonstrating that people often cling to their beliefs in the face of any challenge, [no matter how profound](#). Bem's paper struck him as an interesting challenge for the field. "You have a belief, and here's some data that contradict it," he said. "I thought it was time for a discussion [of] how we deal with surprising results in psychology."

Meanwhile, at the conference in Berlin, Wagenmakers finally managed to get through Bem's paper. "I was shocked," he says. "The paper made it clear that just by doing things the regular way, you could find just about anything."

On the train back to Amsterdam, Wagenmakers drafted a rebuttal, to be published in *JPSP* alongside the original research. The problems he saw in Bem's paper were not particular to paranormal research. "Something is deeply wrong with the way experimental psychologists design their studies and report their statistical results," Wagenmakers wrote. "We hope the Bem article will become a signpost for change, a writing on the wall: Psychologists must change the way they analyze their data."

The final version of Bem's paper was scheduled to appear in the March 2011 issue of *JPSP*. In advance of its release, the Cornell communications office put out a [story on the work](#), which it called the cap to Bem's career. The work "gladdened the hearts of psi researchers," it said, "but stumped doubting social psychologists, who [cannot fault Bem's mainstream and widely accepted methodology](#)."

By the beginning of January, the ESP study had become a media phenomenon. The reaction was intense and at times derisive. A front-page story in the *New York Times* quoted one psi-skeptic as saying, "[It's craziness, pure craziness](#). I can't believe a major journal is allowing this work in."

["Only time will tell if the data holds up,"](#) wrote Jonah Lehrer at *Wired*.

At first, Bem was thrilled by the attention. He called a lab meeting at Cornell to thank his undergraduate assistants, then invited them to New York City to watch him [appear on an episode](#) of *The Colbert Report*.

Shortly thereafter, Bem posted his [response](#) to the Wagenmakers piece, as well as to some [other critiques](#) of his methodology. The negative reviews had not deterred him. Maybe there had been some sloppiness in the earliest experiments, he thought, but it wasn't as if the skeptics were ever going to believe his results. The real test would come through replication. That was the whole point of this exercise—the holy grail of ESP research. If he could get his mainstream colleagues to run the studies for themselves, and if they could find the same results, he'd be vindicated once and for all.

To help get this project underway, Bem had granted researchers full access to his data and provided a detailed how-to guide for redoing his experiments—a level of transparency that was pretty much unheard of at the time. Meanwhile, in 2010, a former student of Bem's had passed along an early version of the ESP paper to a young University of California–Berkeley business school professor named Leif Nelson. Within a few weeks, Nelson and another professor, Carnegie Mellon's Jeff Galak, had coded up an online version of Bem's word-recall study, the one in which people practiced for a test after having taken it. Within a couple of days, they had [results from more than a hundred people](#). On Oct. 14, 2010, Galak sent Nelson an email with the subject line, "There is no such thing as ESP."

Bem would later argue that you cannot do this kind of work with online samples. He also says the word-recall test may not work as well for ESP as the erotic-picture task or any of the others in his paper. He's come to think that it relies too much on what Nobel Prize winner Daniel Kahneman calls the mind's "[slow mode](#)" of thinking. Slow thinking might be less conducive to producing psi-phenomena, Bem argues.

There were other replication failures, too. But then, there were also some successes. Bem has since put out a [meta-analysis](#) that includes 23 exact replications of his original experiments, going back to 2003. When he pooled all those studies with his own, creating a pool of more than 2,000 subjects, he found a positive effect. In his view, the data showed ESP was real.

Others have disputed this assessment. Wagenmakers notes that if Bem restricted his analysis to those studies that came out after his—that is to say, if he'd looked at the efforts of mainstream researchers and skipped the ones by fellow travelers who'd heard about his work at meetings of the Parapsychological Association—the positive effect would [disappear](#).

In any case, those replications soon became a footnote. Within a month or two, the fallout from Bem's initial paper had broadened into something bigger than a referendum on precognition. It had become a referendum on evidence itself.

In 2005, while Bem was still working on his ESP experiments, medical doctor and statistician John Ioannidis published a short but often-cited essay arguing that "[most published research findings are false](#)." Among the major sources of this problem, according to Ioannidis, was that researchers gave themselves too much flexibility in designing and analyzing experiments—that is, they might be trying lots of different methods and reporting only the "best" results.

Bem's colleagues in psychology had, for their part, been engaged in methodological debates for decades, with many pointing out that sample sizes were [far too small](#), that treatments of statistics could be [quite misleading](#), and that researchers often conjured their hypotheses [after collecting all their data](#). And every once in a while, someone would bemoan the [lack of replications](#) in the research literature. (Notice how these concerns mirror those the parapsychological community alighted upon in the early 1980s.)

Even by the mid-2000s, the darker implications of these warnings hadn't really broken through. Certain papers might be sloppy or even spurious, but major swaths of published work? Only Chicken Little types would go that far.

"You felt so alone. You knew something was wrong, but nobody was listening," says Uli Schimmack, a psychologist at the University of Toronto Mississauga and something of a Chicken Little. "I felt very depressed until the Bem paper came out."

At his university, there was to be a discussion of the newly published ESP research. "I thought we would all just go and trash it," Schimmack says, but he was shocked to find that his colleagues seemed impressed by the study's rigorous design. "There was a group of people who said that we should keep an open mind, since there's all this evidence. ... I'm like, 'Look, I don't have to believe any of these results because they're clearly fudged.' ... And someone said, 'You don't want to end up on the wrong side of history.'"

Frustrated, Schimmack set out to write his own rebuttal to the Bem paper and to the approach to science that it represented. Bem had reported running nine experiments, eight of which yielded significant results. This repetition seemed to show that ESP might be a real, robust effect. Schimmack, though, argued that such consistency was too good

to be true. Bem would have needed tremendous luck to score so many hits with his experiments, given the relatively small size of the effect. Ironically, Schimmack argued, the success of all those extra studies made Bem's finding [less believable](#).

Other skeptics—not of ESP, but of the field of social psychology more broadly—felt similarly emboldened by Bem's research. “It couldn't be true, and yet here was this collection of evidence that was thoroughly presented, seemed generally compelling, and was in one of our leading journals,” says Nelson. “In some ways this was just the perfect exemplar of, ‘Oh, if we play by the rules that we've all agreed upon, then we can end up with something like this. And that was a crystallizing moment.’”

In November 2010, while he was still working on a replication of Bem's verbal memory experiment, Nelson met up with a pair of fellow researchers, Joe Simmons and Uri Simonsohn. Over dinner, they talked about all the bogus findings in their field. It started as a game: What's the most ridiculous paper you've ever read? But pretty soon, their conversation turned to deeper questions: How could such silliness make its way to print? And, more importantly, why were so many clever, well-trained researchers turning out illegitimate results?

In the weeks to come, Simmons, Nelson, and Simonsohn continued their discussion over email. First, they made a list of ways that research could go wrong. There were lots of options to consider. Instead of deciding on a sample size ahead of time, psychologists might analyze the data from their studies as they went along, adding new subjects until they found results they liked. Or they might do lots of different tests, based on lots of different variables, then pick out the ones that delivered clean results. They might report unexpected findings as if they'd been predicted. They might neglect to mention all their failed experiments.

These dodgy methods were clearly rife in academic science. A 2011 survey of more than 2,000 university psychologists had found that [more than half of those researchers admitted using them](#). But how badly could they really screw things up? By running 15,000 simulations, Simmons, Nelson, and Simonsohn showed that a researcher could almost double her false-positive rate (often [treated](#) as if it were 5 percent) with just a single, seemingly innocuous manipulation. And if a researcher combined several questionable (but common) research practices—fiddling with the sample size and choosing among dependent variables after the fact, for instance—the false-positive rate might soar to more than 60 percent.

“It wasn't until we ran those simulations that we understood how much these things mattered,” Nelson said. “You could have an entire field that is trying to be noble, actively generating false findings.”

To underline their point, Nelson and the others ran their own dummy experiment to show how easy it could be to gin up a totally impossible result. The trio had a bunch of undergraduates listen to the Beatles' “When I'm Sixty-Four,” then used statistical shenanigans to make it seem as though the music had made the students several years younger than they were before the song started playing.

Simmons, Nelson, and Simonsohn submitted their “When I'm Sixty-Four” paper for publication in early March 2011, two months after Bem's ESP results had landed on the front page of the *New York Times*.

“I saw an advance copy of that [“When I'm Sixty-Four”] paper, and I was like, holy shit,” says Simine Vazire, a personality psychologist at University of California–Davis and one of the founders of the [Society for the Improvement of Psychological Science](#). “I realized this is a big deal. This is a problem.”

The paper turned out to be far more influential than its authors guessed. “We believed that it was unlikely to be published, less likely to be read, virtually un-citable, and generally best of service as a three-authored effort at catharsis,” they remembered in a recent [essay](#) on the work. Their work has now been cited nearly 1,000 times and in 380 different journals.

Bem had shown that even a smart and rigorous scientist could cart himself to crazyland, just by following the rules of the road. But Simmons, Nelson, and Simonsohn revealed that Bem’s ESP paper was not a matter of poor judgment—or not merely that—but one of flawed mechanics. They’d popped the hood on the bullshit-mobile of science and pointed to its busted engine. They’d shown that anyone could be a Daryl Bem, and any study could end up as a smoking pile of debris.

Nelson says his “When I’m Sixty-Four” paper wasn’t meant as a rejoinder to Bem’s ESP study. Still, the ESP results—and Bem’s open invitation to try to replicate them—came at the start of a seismic 18 months for the study of psychology. Simmons, Nelson, and Simonsohn submitted their paper a few months after Bem’s went public. A few months later came the revelation that a classic finding in the field of social priming had [failed to replicate](#). Soon after that, it was revealed that the prominent social psychologist Diederik Stapel had engaged in [rampant fraud](#). Further [replication failures](#) and [new examples of research fraud](#) continued to accumulate through the following year. Finally, in September 2012, Daniel Kahneman sent a dire warning to his senior colleagues, one that would be repeated often in the years to come: “[I see a train wreck looming](#).”

In retrospect, it looks as though the Bem results helped release a store of pent-up energy. “That paper had dynamite written all over it,” Wagenmakers says. “I had some pre-existing concerns, but the Bem paper really brought them out. It inspired me to look more closely at the problem.”

Wagenmakers would later write that psi researchers such as Bem deserve “substantial credit” for the current state of introspection in psychology, as well as for the crisis of confidence that is now spreading into other areas of study. “It is their work that has helped convince other researchers that the academic system is broken,” he said, “for if our standard scientific methods allow one to prove the impossible, than these methods are surely up for revision.”

Scientist With Crystal Ball

Lisa Larson-Walker

Even now, Jade Wu wonders whether Bem planned this out from the very start. Wu is now a doctoral student in clinical psychology, so she’s seen first-hand how research practice has been changing in her field. “I still think it’s possible that Daryl Bem did all of this as a way to make plain the problems of statistical methods in psychology,” she says. Other academics I spoke to shared similar suspicions. One well-known psychologist, whom I knew Bem from when he went to grad school at Cornell, said at first he thought the ESP results might have been a version of the [Sokal hoax](#).

But for Bem’s fellow members of the Parapsychological Association, the publication marked a great success. “He brought a lot of attention to the possibility that this research can be done, and that it can be done in a mainstream establishment,” says Marilyn Schlitz, a sociolinguist who studies psi phenomena and has an appointment at the Institute of Noetic Sciences in Petaluma, California.

It’s plain to see that Bem achieved his major goal: to promote replications of the research into ESP. (“Everything rests on replication,” he told me.) As far as he and other

parapsychologists are concerned, these replications have so far been equivocal. “I think the jury is still out,” says Jonathan Schooler, a psychologist at University of California–Santa Barbara who was one of the original peer reviewers for Bem’s paper. Schooler, who is very open to the evidence for ESP, admits it’s possible that Bem’s results are nothing more than artifacts of flawed experimental design. But then he says that precognition could be real. In fact, “there’s no reason why you can’t entertain both of these possibilities at once.”

That’s more or less Bem’s position. “The critics said that I put psychologists in an uncomfortable position and that they’d have to revise their views of the physical world or their views on research practice,” he told me. “I think both are true. I still believe in psi, but I also think that methods in the field need to be cleaned up.”

Bem meant to satisfy the skeptics, but in the end he energized their doubts.

Both Schooler and Bem now propose that replications might be more likely to succeed when they’re performed by believers rather than by skeptics. Such “experimenter effects” have been [well-documented in the psychology literature](#) since the 1960s, and they’re often seen as arising from scientists’ hidden bias. But psi researchers proffer a different interpretation: Maybe this has less to do with the researcher’s expectation than with his ability as a psychic medium. “If it’s possible that consciousness influences reality and is sensitive to reality in ways that we don’t currently understand, then this might be part of the scientific process itself,” says Schooler. “Parapsychological factors may play out in the science of doing this research.”

In order to [test this proposition](#), and to help resolve the problem of the dueling replications, Bem joined up with Marilyn Schlitz and French neuroscientist Arnaud Delorme to have another go at finding precognition in the lab. With [funding](#) from the owner of a Portuguese pharmaceutical firm who happens to believe in ESP, they tried to replicate one of Bem’s original experiments on “retrocausal priming.” The idea was that people might be quicker to react to pleasant photographs (of, say, a polar bear cub) when they’re primed, after the fact, with a pleasant word (e.g. *love*).

To distinguish this replication from earlier attempts, Bem, Schlitz, and Delorme took extra steps to rule out any possibility of bias. They planned to run the same battery of tests at a dozen different laboratories, and to [publish the design of the experiment](#) and its planned analysis ahead of time, so there could be no quibbling over the “garden of forking paths.”

They presented their results last summer, at the most recent annual meeting of the Parapsychological Association. According to their pre-registered analysis, there was no evidence at all for ESP, nor was there any correlation between the attitudes of the experimenters—whether they were believers or skeptics when it came to psi—and the outcomes of the study. In summary, their large-scale, multisite, pre-registered replication ended in a failure.

In their conference abstract, though, Bem and his co-authors found a way to wring some droplets of confirmation from the data. After adding in a set of new statistical tests, *ex post facto*, they concluded that the evidence for ESP was indeed “highly significant.” Since then they’ve pre-registered a pair of follow-up experiments to test this new approach. Both of those efforts are in progress; meanwhile, the original attempt has not yet been published in a journal.

I asked Bem if he’d ever budge on his belief in ESP. What if, for example, pre-registered replications like the one he’d done with Schlitz and Delorme continued to turn up negative results? “If things continue to fail on that one, I’m always willing to update my

beliefs,” he said. “But it just seems unlikely. There’s too much literature on all of these experiments ... so I doubt you could get me to totally switch religions.”

“I’m all for rigor,” he continued, “but I prefer other people do it. I see its importance—it’s fun for some people—but I don’t have the patience for it.” It’s been hard for him, he said, to move into a field where the data count for so much. “If you looked at all my past experiments, they were always rhetorical devices. I gathered data to show how my point would be made. I used data as a point of persuasion, and I never really worried about, ‘Will this replicate or will this not?’ ”

When Bem started investigating ESP, he realized the details of his research methods would be scrutinized with far more care than they had been before. In the years since his work was published, those higher standards have increasingly applied to a broad range of research, not just studies of the paranormal. “I get more credit for having started the revolution in questioning mainstream psychological methods than I deserve,” Bem told me. “I was in the right place at the right time. The groundwork was already pre-prepared, and I just made it all startlingly clear.”

Looking back, however, his research offered something more than a vivid illustration of problems in the field of psychology. It opened up a platform for discussion. Bem hadn’t simply published a set of inconceivable findings; he’d done so in a way that explicitly invited introspection. In his paper proving ESP is real, Bem used the word *replication* 33 times. Even as he made the claim for precognition, he pleaded for its review.

“Credit to Daryl Bem himself,” Leif Nelson told me. “He’s such a smart, interesting man. ... In that paper, he actively encouraged replication in a way that no one ever does. He said, ‘This is an extraordinary claim, so we need to be open with our procedures.’ ... It was a prompt for skepticism and action.”

Bem meant to satisfy the skeptics, but in the end he did the opposite: He energized their doubts and helped incite a dawning revolution. Yet again, one of the world’s leading social psychologists had made a lasting contribution and influenced his peers. “I’m sort of proud of that,” Bem conceded at the end of our conversation. “But I’d rather they started to believe in psi as well. I’d rather they remember my work for the ideas.”

Published May 17, 2017, 9:00 PM ET