

This is a section of [doi:10.7551/mitpress/14922.001.0001](https://doi.org/10.7551/mitpress/14922.001.0001)

# Open Minded

## Searching for Truth about the Unconscious Mind

By: Ben R. Newell, David R. Shanks

### Citation:

*Open Minded: Searching for Truth about the Unconscious Mind*

By: Ben R. Newell, David R. Shanks

DOI: 10.7551/mitpress/14922.001.0001

ISBN (electronic): 9780262375375

Publisher: The MIT Press

Published: 2023

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

## 8 Feeling the Future; Precipitating a Crisis

In part I, we dove into the murky depths to search for evidence of a smart unconscious. We discovered that many high-profile examples of unconscious influences evaporate once scrutinized, or at least admit alternative and often more plausible explanations. In many ways, it seems that the very notion of unconscious thought is misguided. With this new perspective on the close connection between awareness and behavior established, we now turn to the equally puzzling question of how we, as a discipline, but also as a society, got to this point. How did we become hoodwinked into believing that our unconscious mind has a hold on our behavior?

### **Anatomy of a Train Wreck**

In early October 2012, Ed Yong, a staff writer for the scientific journal *Nature*, published a short article with a long-lasting impact.<sup>1</sup> The article discussed an email written by Daniel Kahneman, a Nobel laureate and perhaps the world's best-known psychologist, whom we read about in chapters 3 and 7. The email described Kahneman's growing concerns about an area of psychology known as "priming." As we saw in chapter 3, priming is the study of how subtle cues can apparently unconsciously influence our thoughts and behavior. For example, a person might walk more slowly down a corridor when thinking about elderly people or find cartoons funnier while "smiling" when holding a pen in her mouth.

In the email—intended for colleagues but leaked to *Nature*—Kahneman wrote provocatively of a "train wreck looming" because of doubts about the replicability of many priming effects. Kahneman's concern followed revelations of fraudulent conduct by some researchers. The pressure to publish

or perhaps the desire for fame had led some to simply invent data. Less dramatic but equally troubling, many researchers were unable to reproduce prominent, eye-catching findings and the bulging file drawers in which these “failures to replicate” had been hidden were finally spilling open.

The implication of these events for the “integrity of psychological research” played strongly on Kahneman’s mind. He advised colleagues to “collectively do something about this mess” and suggested methods for improving the replicability of research. In this brief chapter we provide the context behind the looming train wreck. We set the scene for chapters 9 and 10, which explore in more detail the nature of the research practices that precipitated the crisis in confidence and the scientific ecosystem that encourages behaviors that are at variance with the pure pursuit of truth.<sup>2</sup>

\* \* \*

It was the question Diederik Stapel had been dreading for years: “Diederik, I have to ask you: have you been faking your data?” Stapel, a professor of social psychology at the University of Tilburg at the time, was sitting in the living room of his friend, Maarten, the chair of the department. The question was direct and unexpected. Stapel’s first reaction was to deny the accusation flat out. Faking data was, as Stapel would write in his memoir *Derailment* a few years later, “professional suicide.”<sup>3</sup> An admission of guilt would end his extremely successful career. Instead he played it down. What evidence did Maarten have? Were there specific details, or was it just idle gossip at conferences?

Maarten had been taken aside after dinner at a recent conference by some young researchers who had collaborated with Stapel. What they told Maarten was deeply unsettling. The junior colleagues claimed that no one knew where Stapel was getting his data from. They feared that he’d simply been making up numbers. Stapel’s initial attempts to mollify Maarten appeared to work. The rest of the evening was genial—discussions about interesting presentations Maarten had seen at the conference—but Stapel knew that his “big, fat, outright lies” were now likely to become very public.

To get a sense of the size of these lies, consider the following example of one of Stapel’s studies that turned out to have been rather economical with the truth. Imagine that you are walking down the street and are stopped by a person holding a clipboard. The interviewer asks you about your thoughts

and feelings toward particular minority groups in society (perhaps Muslims). The goal, although not stated explicitly, is to measure your tendency to stereotype people. After answering the questions, the interviewer offers you a payment of 5 euros but also asks whether you'd be willing to donate a portion of the money to a charity that helps immigrants and homeless people. You give back a few euros, you are thanked, and you head off down the street. This is a simple interaction, but is something more complex going on? Collecting data in the field like this is a common approach in many areas of psychology, especially those concerned with social interaction. There is a strong desire to get outside the confines of the sterile laboratory and into the messy "real world." In the study just described, it was literally a need for mess that pushed Stapel out on to the streets. He wanted to explore the broken windows theory—the idea that poor physical environments (dilapidated buildings, broken windows) beget poor social environments. His extension to this idea was that messy and rundown physical environments lead people to rely more on stereotypes and other forms of prejudice because stereotyping allows people to create structure in their mental world. In Stapel's words, "Stereotyping is a mental cleaning device that helps people to cope with physical chaos."<sup>4</sup>

What does this have to do with being stopped on the street and asked about your views on different groups in society? The key manipulation in the "coping with chaos" study was the presence or absence of messy street features. For one group of participants, the interviewer was standing in a spot with broken tiles on the pavement, a poorly parked car with its windows open, and an abandoned bicycle. For the other group, tested in exactly the same place on another occasion, none of these features were present: the street was neat and tidy. Stapel predicted that when people were asked about their thoughts and feelings in the presence of "chaos," they would show more stereotyping. That is, they would accentuate the differences between us (the Caucasian participants) and them (the various minority groups listed in the survey) to compensate for the disordered physical environment. Moreover, he predicted that those faced with chaos would donate more of their participation fee to the Money for Minorities charity than those on the ordered street.

This pattern of results is exactly what Stapel found. Except he didn't. Because none of it happened. No interviewer, no passersby, no broken

bicycles. No (real) data. And yet the paper on these results was reported in one of the world's most prestigious scientific outlets, the journal *Science*. How could this have happened?

Perhaps the simplest answer to this question is that Stapel had not wanted inconvenient data to get in the way of a good story. In fact, he had not wanted *any* data to get in the way. The idea that chaos might lead to stereotyping had come to Stapel several years before the faked street study. He had done some initial laboratory-based experiments in which he showed participants photos of disorderly and orderly scenes (walls with or without graffiti) and then had them rate their feelings toward different social groups. He had found small differences consistent with his hypothesis—the people shown the disorderly scenes tended to display more prejudice—but the differences were weak and transient. After a few more failed attempts to find the effect, he gave up. Eventually, though, the lure of a good idea and a good story to go with it overpowered Stapel's need for data.

For the kinds of experiments reported in the coping-with-chaos paper, faking data is relatively straightforward. Each (hypothetical) participant provides some numbers on a few rating scales indicating how much they tend to stereotype; these numbers can then be combined to produce average or mean ratings for the different conditions (chaotic versus ordered street scenes) and subjected to statistical analysis. Stapel describes the process in some detail in his memoir:

I would . . . make a careful list of all the results and effects I needed to create for the experiment I was doing. Neat tables with the results I expected based on extensive reading, theorizing, and thinking. Simple, elegant, comprehensible. Next I started to enter the data, column by column, row by row. I tried to imagine how the participants' answers to my questionnaire would look. What were some reasonable answers that might be expected? 3, 4, 6, 7, 8, 4, 5, 3, 5, 6, 7, 8, 5, 4, 3, 3, 2. When I'd input all the data, I ran some quick preliminary analyses. Often these didn't show what I was expecting, so I went back to the table of data to change a few things. 4, 6, 7, 5, 4, 7, 8, 2, 4, 4, 6, 5, 6, 7, 8, 5, 4. And so on, until the analyses provided the results I was looking for. That is, until the data showed what was logical, and therefore true.<sup>5</sup>

The brazenness of Stapel's approach is chilling. His ability to have gotten away with it for so long had emboldened him. If we are to believe his memoir, faking the data made him nauseous and terrified him but he was addicted—addicted to the success and fame that the next big (fake) discovery (and publication) would bring him. Stapel had been playing a dangerous game,

but he'd been playing it carefully. The results he created were simple and comprehensible—they made a good story—and they fit, more or less, with established theories. Thus, although some of the statistical results might have appeared too good to be true, the claims being made were not *too* extraordinary. We can make sense of the idea that messy surroundings are unsettling and that we might vent some frustration at the chaos by distancing ourselves from those we perceive as responsible.

Achieving this high-wire act of balancing plausible yet sufficiently surprising findings had opened the door to a slew of high impact publications for Stapel. However, the Levelt Report—an investigation into Stapel's fraudulent activity—concluded that in addition to his own dishonesty, there had also been serious shortcomings in the publication process and a widespread failure of scientific criticism when it came to his work.<sup>6</sup> The report established that the peer-review process (where other researchers get to evaluate work before it can be published) had often been “strongly in favor of telling an interesting, elegant, concise and compelling story, possibly at the expense of the necessary scientific diligence.” There was, it seems, an uncritical acceptance of findings that “felt right” despite a lack of clear evidence, let alone reliable replication.

Although Stapel had created a raft of false-positive results—or simply false results—he had been careful not to make his claims too extraordinary. He'd prided himself on keeping the stories of his faked data simple, elegant, and within the bounds of existing theory. But sometimes the pendulum swings the other way. The data may be genuine but the theoretical claims made on their behalf are simply off the map. And when the claims are extraordinary, as the astronomer Carl Sagan was fond of saying, the evidence needs to be extraordinary too. Precognition anyone?

### Time-Traveling Porn

“Science has finally discovered time-traveling porn,” declared Stephen Colbert on his hugely popular US TV show *The Colbert Report* in January 2011. The topic: the claim by Cornell social psychologist Daryl Bem to have found evidence that participants in his experiments could “feel the future”—specifically that they knew, in advance, what picture was about to be displayed on a computer screen in front of them. But not any old picture; it only worked for erotic images. In discussing his results, Bem quoted from

the celebrated quantum physicist Richard Feynman: “Do not keep saying to yourself . . . ‘But how can it be like that?’ because you will get . . . into a blind alley from which nobody has yet escaped. Nobody knows how it can be like that.” Quite.<sup>7</sup>

The experiment itself was very simple.<sup>8</sup> Cornell University undergraduates sat in front of a computer monitor displaying images of two curtains side by side. They were told that behind one curtain was a blank wall and behind the other was a picture. All they needed to do was “click on the curtain that you feel has the picture behind it.” Once clicked, the curtain was drawn back to reveal either the picture or the wall. The procedure was repeated for thirty-six trials. The position of the picture, relative to the wall, was randomized on each trial, as was the type of picture displayed. On some trials, the pictures were erotic (“couples engaged in nonviolent but explicit consensual acts”), on some they were negative images like snakes and spiders, and on others they were positive and romantic but not erotic, such as a bride and groom kissing at their wedding. A final type was simply neutral (landscapes).

With a fifty-fifty chance of picking the curtain covering the image, one would expect people to be 50 percent accurate—that is, at chance—across the thirty-six trials of the experiment. The extraordinary finding was that the hit rate (choosing the correct curtain) for the erotic pictures was 53.1 percent. This might not seem like much above the expected 50 percent, but it was statistically significant (an issue to which we will return), which is to say meaningfully greater than 50 percent. Moreover, as Bem pointed out to Stephen Colbert, although it seems small, it is in fact similar to the margin by which Obama defeated McCain in the 2008 presidential election; it is also about the same margin the house has over the punter in casinos. Hit rates for the nonerotic pictures—whether neutral, positive, negative, or romantic—all fell very close to 50 percent and did not pass the statistical test for being a “real” result. So people can feel the future, but only if it is porn! Stephen Colbert, along with large chunks of the media, had a field day with this finding.

Within academic circles, the reaction was one of puzzlement but also deep concern. Here was an article published in one of the top journals in social psychology reporting apparent evidence undermining a fundamental belief in the direction of causation. We all know that a cause must precede an effect: you cannot hear the bell ringing until someone strikes the bell. But Bem wanted us to believe that the content of an unseen image can have

a retroactive influence on our choice: “The participant is, in fact, accessing information yet to be determined in the future, implying that the direction of the causal arrow has been reversed.”<sup>9</sup> The yet-to-be-determined aspect is a subtle but crucial feature of the experimental design. In the experiments, participants made a selection *and then* the computer used a special algorithm to randomly determine which picture would be displayed behind the selected curtain. Thus, the participants were not displaying clairvoyance; they were showing *precognition*. This is knowledge of a future event that could not otherwise be anticipated through any known inferential process.

Psychology has a long tradition of interest in psi phenomena: anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. This fascination seems in part to be driven by the widespread acceptance among the public that something like psi exists. For example, according to a US CBS Newspann, 57 percent of Americans believe in extrasensory perception, and other surveys show that over 40 percent of people believe in telepathy.<sup>10</sup> *If so many people believe in it, then perhaps there is something worth investigating* appears to be the logic that Bem and his fellow parapsychologists relied on. But the old saying that the plural of anecdotes is not anecdota has long been recognized as relevant here. Joseph Jastrow, one of the founders of modern psychology, wrote a book, *Fact and Fable in Psychology*, way back in 1901 that was largely concerned with separating what he saw as the legitimate inquiry of the then relatively new discipline of psychology (the facts) from the vagaries of “psychical” research (the fables). He suggested:

If the problems of psychical research, or that portion of the problems in which investigation seems profitable, are ever to be illuminated and exhibited in an intelligible form, it will only come about when they are investigated by the same methods and in the same spirit as are other psychological problems.<sup>11</sup>

Arguably, by bringing phenomena like precognition into the laboratory and performing simple and potentially easily replicable experiments, Bem was following this advice. He went beyond anecdotal reports and attempted to use statistical methods to demonstrate reliable—and yet inexplicable—effects in his data. But again, some prescient words from Jastrow are worth heeding here:

Data cannot claim serious attention before they are strong in their validity, and extensive in their scope and consistently significant in their structure; then, and not before, are they ready for the crucible of scientific logic, from which they may



or may not emerge as standard metal, to be stamped and circulated as accepted coin of the realm.<sup>12</sup>

As Bem was about to discover, many in the discipline were not quite ready to accept his findings as legal tender. E.-J. Wagenmakers and colleagues from the University of Amsterdam were quick to critique Bem's research.<sup>13</sup> They urged fellow psychologists to reconsider the ways in which they analyze data and to be clear about the differences between *exploratory* and *confirmatory* research. Exploration is fine—it is often the way we make new discoveries; noticing the anomalous pattern in the data (or the mold in the petri dish) can lead to significant breakthroughs. But novel hypotheses discovered on an exploratory trawl through the data must be tested in new, confirmatory experiments. A single surprising finding is just that—a one-off that might have occurred simply by chance. To have confidence that a finding should be newly minted and circulated to the community, it needs to be replicated. And the more surprising the finding, the more urgent the need for replications.

Wagenmakers and colleagues pointed to several instances of what they perceived to be exploratory practices in Bem's feeling-the-future paper. Why, for example, was precognition only found for erotic and not neutral, negative, or positive pictures? Was this an a priori prediction or something discovered by slicing and dicing the data after collection? Why in some experiments were anomalous influences found only for women and not men? According to Bem, the psi literature does not show systematic sex differences in psi ability, so why test for gender unless to squeeze out something—anything—interesting? The problem with this exploration is that each time the data are examined in a different way, the possibility of finding a false-positive result—a result that looks genuine, but is in fact spurious—increases. Researchers know these as type 1 errors.

The principal guard against such errors is to use statistics—essentially methods for inferring whether observed patterns in data are real or occurred by chance. The most commonly used technique in psychology (and many other disciplines) is null hypothesis significance testing (NHST). This is what Bem (and Stapel) used. The basic idea is to test whether the data provide enough evidence to reject the hypothesis that there is no impact of a manipulation on behavior—the so-called *null hypothesis*. Rejecting the null when in reality there is no evidence for a difference in the data is the definition of a type 1 error. So why didn't the statistics that Bem used control for these kinds of errors? Why could he not be confident in rejecting the null

hypothesis that the 53.1 percent hit rate for selection of the erotic images was a chance result? Wagenmakers and colleagues illustrate the problem of relying solely on NHST with the following example.

Imagine you have won \$10 million in the state lottery—an extremely happy event, but one that might lead to jealousy in some friends and acquaintances. Perhaps this jealousy leads one acquaintance to accuse you of cheating: the probability of winning the lottery is very low so you must have cheated. We immediately see that this argument is not logical: the low probability by itself cannot be taken as evidence for cheating. The evidence becomes useful only if we compare it to an alternative hypothesis—with an even lower probability—that you were somehow able to obtain advanced knowledge of the winning numbers (perhaps you are a “precog”?).

The key point here is that the strength of evidence for a particular hypothesis—precognition exists—needs to be evaluated against a specific alternative hypothesis—precognition does not exist—rather than just a “nothing” or null hypothesis. Wagenmakers and colleagues argue that the best way to conduct such comparative hypothesis testing is by using a different statistical approach based on Bayesian methods. The details need not concern us, but the results should. When reanalyzed using these arguably more appropriate techniques, the evidence for precognition in all nine of Bem’s experiments all but evaporated and indeed in three of the nine substantial evidence in favor of the nonexistence of precognition emerged. These reanalyses should give us considerable pause for thought. But even if the statistical arguments are rather abstract, there are other, more fundamental reasons for questioning Bem’s conclusions.<sup>14</sup>

Klaus Fiedler, a social psychologist from the University of Heidelberg in Germany and long-term observer of the field, also found himself troubled by the publication of Bem’s work. Writing with colleague Joachim Kruger from Brown University, they took Bem to task on the absence of any theoretical explanation of his findings and the confusion between the explanans (the argument used for explanation) and the explanandum (the event to be explained).<sup>15</sup> In essence, Fiedler and Kruger argued that Bem’s article lacked any solid theoretical base. Although there were some allusions to quantum physics in the discussion of the results (and the appeal to Feynman), the bottom line was no real theory and no real explanation. Bem admitted as much to Colbert in his interview; when Colbert asked, “How is this working?” he replied, “We have no idea.”

More than that, the explanation that Bem appeared to be seeking was of the wrong phenomenon. According to Fiedler and Kruger the explanandum and the explanans had been reversed. In Bem's experiments, the crucial finding was that participants chose the erotic pictures at a rate above chance. If we hang on to the (pretty-well-established) idea that antecedent conditions explain consequent events, then the event to be explained is the computer's bias to produce responses congruent with participants' choices. Or as Fiedler and Kruger write, if we think that the results are valid (which is of course questionable), then they "must be interpreted as owing to metaphysics in the computer's chips rather than precognition in the human brain."<sup>16</sup>

What this boils down to is a debate about whether random number generators used to determine the location of pictures was truly random or somehow correlated with the participants' preceding responses. While this might seem unlikely, the key question is whether it is less likely than a reversal in the laws of causal inference. Winning versus cheating on the lottery, random number errors versus psychic retroactive anomalous influences: it is all in the balance of probabilities. And as Sherlock Holmes famously advised, "When you have eliminated the impossible, whatever remains, however improbable, must be the truth."<sup>17</sup>

The Stapel and Bem cases and others like them have been very painful for the behavioral sciences, but they have also been useful. They have forced the discipline to introspect on its research practices and review the role of psychological theory in our explanations of mental life. It is hard to exaggerate the transformation in research methods and rigor that has occurred in the decade or so since Bem's research was published and Stapel's fraud was revealed, and our contention is that much of the evidence for unconscious influences on human behavior falls victim to these improved methods. In the next two chapters, we delve more deeply into the features of our research practices and the scientific ecosystem that got us into this mess (as Daniel Kahneman put it) and see yet again that weak claims about the powers of the unconscious play a crucial role. In the final chapter, we point to how a focus on transparency and strong theory can help get us out of the mess and back on track to reclaiming the science of the mind.

© 2023 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC-BY-NC-ND license.  
Subject to such license, all rights are reserved.



The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Newell, Benjamin R., 1972– author. | Shanks, David R.

Title: Open minded : searching for truth about the unconscious mind /  
Ben R. Newell and David R. Shanks.

Description: Cambridge, Massachusetts : The MIT Press, [2023] | Includes  
bibliographical references and index.

Identifiers: LCCN 2022038725 (print) | LCCN 2022038726 (ebook) |  
ISBN 9780262546195 (paperback) | ISBN 9780262375368 (epub) |  
ISBN 9780262375375 (pdf)

Subjects: LCSH: Subconsciousness. | Thought and thinking. | Self-consciousness  
(Awareness)

Classification: LCC BF315 .N479 2023 (print) | LCC BF315 (ebook) |  
DDC 154.2—dc23/eng/20230316

LC record available at <https://lcn.loc.gov/2022038725>

LC ebook record available at <https://lcn.loc.gov/2022038726>