THE COLUMBIA
SCIENCE & TECHNOLOGY
LAW REVIEW

VOLUME 26 STLR.ORG NUMBER 2

ARTICLE

UNSTICKING LITIGATION SCIENCE

Edith Beerdsen*

Litigation science is increasingly out of step with academic, knowledge-producing science. Research practices in the social sciences have changed dramatically in the past fifteen years, in response to a knowledge crisis now popularly known as the "Replication Crisis." The Replication Crisis caused an evolution in scientists' understanding of what it takes to create reliable science and radically altered the way social science is conducted today. Central to these reforms is a focus on the elimination of "Analytical Flexibility"—the flexibility a researcher has to alter a research protocol along the way—in recognition of the fact that Analytical Flexibility has a propensity to lead to research results that are not only unreliable but also unreliable in undetectable ways.

The Replication Crisis represents a paradigm shift that has not yet been recognized by the legal community. In this symposium paper, I describe how litigation science and academic science are currently on divergent paths, and argue that it is critical for courts and litigators to start engaging with recent progress in research methodology. In the academic sciences, modern research practices such as preregistration are increasingly becoming routine and expected by journals, peer reviewers, and funders. Meanwhile, testifying experts retained in connection with litigation essentially proceed as they always have, and disclosure requirements have remained unchanged.

Litigation science is at risk of becoming a quaint, historically shaped discipline that bears scant relationship to its academic, knowledge-producing cousin. If we

* Assistant Professor of Law, Temple University, Beasley School of Law. I am grateful to the Columbia Law & Technology Law Review for the invitation to present this paper at its "Judging Science" symposium in November 2025. For insightful comments and observations, I thank Hon. Jed Rakoff, Edward K. Cheng, Orna Madigan, and participants of the symposium. Thanks also to Jinwan Cho, Ashley Pennington, and Anisha Amurthur, for expertly shepherding this article through the editing process.

do not reform how litigation science is produced and presented, it will increasingly be seen as incapable of producing information that can usefully inform relevant issues in litigation.

I. INTRODUCTION

In the second decade of the 21st century, a fundamental shift began in the way that research is conducted in the social sciences. The crisis that brought about this shift, now known as the Replication Crisis, raised alarm about research practices that undermine the reliability of research results in ways that are typically invisible and cannot be remedied or made visible after the fact.[1] The Replication Crisis caused an evolution in scientists' understanding of what it takes to create reliable science and radically altered the way social-science research is conducted today.[2] The crux of much of the recent reforms is a recognition of the perils of researchers' flexibility during the research process. It is this "Analytical Flexibility"—the flexibility to change or supplement research protocols *during* a study—that has

[1] *See* Leif D. Nelson, Joseph Simmons & Uri Simonsohn, *Psychology's Renaissance*, 69 ANN. REV. PSYCH. 511, 512 (2018) (methods of data collection and analysis that were in widespread use for decades made it "impossible to distinguish between findings that are true and replicable and those that are false and not replicable"); *see also* Andrew Gelman, *Essay: The Experiments Are Fascinating. But Nobody Can Repeat Them*, N.Y. TIMES (Nov. 19, 2018), https://www.nytimes.com/2018/11/19/science/science-research-fraud-reproducibility.html.

[2] *See, e.g.*, Megan L. Head, Luke Holman, Rob Lanfear, Andrew T. Kahn & Michael D. Jennions, *The Extent and Consequences of P-Hacking in Science*, PLOS BIOLOGY, Mar. 13, 2015, at 1, 1 (concluding that *p*-hacking is "widespread throughout science"); John P.A. Ioannidis, *Why Most Published Research Findings Are False*, 2 PLOS MED. 0696, 0697-98 (2005) (arguing that flexibility in design, number of tested relationships, and sample size all contribute to unreliable research outcomes); Marcus R. Munafò et al., *A Manifesto for Reproducible Science*, 1 NATURE HUM. BEHAV. 1, 1 (2017) (identifying apophenia, confirmation bias, and hindsight bias as contributors to flawed data analysis); Joseph P. Simmons, Leif D. Nelson & Uri Simonsohn, *False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant*, 22 PSYCH. SCI. 1359, 1359 (2011) (demonstrating how "unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis").

been proven to be dangerous even in a well-intentioned, honest researcher's hands.[3] And one of the most difficult aspects of Analytical Flexibility with which to come to terms is that it has a propensity to lead to research results that are not only unreliable, but unreliable *in undetectable ways*.[4]

The Replication Crisis represents a paradigm shift that has not yet been sufficiently recognized by the legal community. In this symposium contribution, I describe how litigation science and academic science are currently on divergent paths, and argue that it is critical for courts and litigants to start engaging with what has been going on in the academic sciences. Part II briefly reviews the history of the Replication Crisis and the main lessons that it taught the scientific community. Part III demonstrates that in many of the social sciences from which expert evidence is regularly drawn (psychology, sociology, criminology, etc.), modern, updated research practices are increasingly becoming routine. For example, researchers these days routinely preregister their studies and design their studies with conscious attention to the risks of Analytical Flexibility.[5] Researchers who fail to adopt these procedural safeguards are increasingly facing difficulties having their work published and accepted by the community of their peers.

Part IV surveys the extent to which the judiciary currently incorporates lessons from the Replication Crisis in making admissibility decisions regarding scientific evidence. That extent, thus far, has been almost nil. Testifying experts retained in connection with litigation essentially proceed as they always have, and disclosure requirements have remained unchanged. But as I have argued in earlier work, the Replication Crisis has made clear that current procedural practice, when it comes to the evaluation and admission of litigation science, is simply not designed to provide a court with the information it needs to make an accurate assessment of the reliability of scientific evidence.[6]

I argue (in Part V) that litigation science is being left behind and is increasingly out of step with academic, knowledge-producing science. If we do not try to reform the way that litigation science is produced, it will soon be a quaint, historically shaped discipline that will bear scant relationship to its academic, knowledge-

---

[3] Analytical Flexibility can take the form of adding extra datapoints, decisions about how to filter data, conducting additional, unplanned analyses, etc. Such methodological tinkering along the way can lead to a dramatic increase in the false-positives rate: the likelihood of finding a result that shows up as statistically significant, but is not a "true" result, in the sense that it is unlikely to be replicated if the experiment were to be repeated. Simmons et al., *supra* note 2, at 1359.

[4] *See* Nelson et al., *supra* note 1, at 512 (analytically flexible methods make it "impossible to distinguish between findings that are true and replicable and those that are false and not replicable").

[5] *See, e.g.*, Brian A. Nosek, Charles R. Ebersole, Alexander C. DeHaven & David T. Mellor, *The Preregistration Revolution*, 115 PNAS 2600, 2605 (2018) (reporting a "cultural shift" toward preregistration); Simine Vazire & Brian Nosek, *Introduction to Special Topic "Is Psychology Self-Correcting? Reflections on the Credibility Revolution in Social and Personality Psychology,"* SOC. PSYCH. BULL., Nov. 17, 2023, at 1, 1 (introducing a special volume "reflecting on the crisis and the ensuing 'credibility revolution'").

[6] Edith Beerdsen, *Litigation Science after the Knowledge Crisis*, 106 CORNELL L. REV. 529, 529 (2021).

producing cousin. It will also increasingly be seen as incapable of producing information that can usefully inform relevant issues in litigation.

II. THE REPLICATION CRISIS

The crisis of knowledge now known as the Replication Crisis started in the discipline of psychology, in the early 2010s.[7] A number of attention-grabbing studies, published in high-impact journals and catching fire in popular media, turned out not to be replicable.[8] Psychology researchers started to report on the widespread prevalence of "questionable research practices" that lead to unreliable (and therefore unreplicable) study results[9] and prominent psychologist Daniel Kahneman urged the field to engage in some serious self-reflection, to assess what the field could learn from these high-profile failures.[10] Several large-scale collaborative replication projects followed, involving hundreds of studies and labs, and they led to some startling conclusions.[11] Only about half of the studies selected for these replication attempts were successfully replicated.[12]

The field of psychology frantically tried to make sense of these failures and quickly homed in on some statistical practices that were widely used and that inadvertently injected unreliability into research results.[13] Researchers, under

---

[7] For a more detailed overview of the history, see Beerdsen, *supra* note 6, at 543-64.

[8] *See, e.g.*, Stuart J. Ritchie, Richard Wiseman & Christopher C. French, *Failing the Future: Three Unsuccessful Attempts to Replicate Bem's 'Retroactive Facilitation of Recall' Effect*, PLOS ONE, Mar. 14, 2012, at 1, 3-4 (reporting three failed attempts to replicate a finding that people can "[f]eel[] the [f]uture"); Stéphane Doyen, Olivier Klein, Cora-Lise Pichon & Axel Cleeremans, *Behavioral Priming: It's All in the Mind, but Whose Mind?*, PLOS ONE, Jan. 18, 2012, at 1, 5-6 (2012) (reporting a failed attempt to replicate a celebrated age-priming study).

[9] Leslie K. John, George Loewenstein & Drazen Prelec, *Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling*, 23 PSYCH.. SCI. 524, 524 (2012).

[10] *See Kahneman on the Storm of Doubts Surrounding Social Priming Research*, DECISION SCI. NEWS (Oct. 5, 2012), https://www.decisionsciencenews.com/?p=3634 [https://perma.cc/BV6Y-8LCB] (quoting Kahneman warning of "a train wreck looming" and proposing a systematic replication project).

[11] *See* Open Science Collaboration, *Estimating the Reproducibility of Psychological Science*, 349 SCIENCE aac4716-1, aac4716-2 (2015) (involving 100 research labs replicating a set of 100 studies); Richard A. Klein et al., *Many Labs 2: Investigating Variation in Replicability Across Samples and Settings*, 1 ADVANCES METHODS & PRACS. PSYCH. SCI. 443, 446 (2018) (involving 36 labs replicating a set of 28 studies).

[12] *See* Ed Yong, *Psychology's Replication Crisis Is Running out of Excuses*, ATLANTIC (Nov. 19, 2018), https://www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223 [https://perma.cc/8WZC-QL5P] ("Ironically enough, it seems that one of the most reliable findings in psychology is that only half of psychological studies can be successfully repeated.").

[13] *See, e.g.*, Simmons et al., *supra* note 2, at 1359 (demonstrating how common research tools can generate unreliable findings); Open Science Collaboration, *Maximizing the Reproducibility of Your Research*, *in* PSYCHOLOGICAL SCIENCE UNDER SCRUTINY 3, 3 (Scott O. Lilienfeld & Irwin D. Waldman eds., 2017) (suggesting best practices to improve reliability of research); Brian A. Nosek, Jeffrey R. Spies & Matt Motyl, *Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth over Publishability*, 7 PERSPS. ON PSYCH. SCI. 615, 616 (2012) (arguing for institutional change to improve reliability of social science research).

pressure to generate statistically significant findings, were using a variety of methods that were generally accepted and not exactly fraudulent, but that allowed a researcher to turn virtually any data set into statistically significant findings.[14] These practices typically shared a common feature: they exploited various kinds of Analytical Flexibility—a researcher's range of freedom to change a study protocol once data gathering has commenced, or even to make up the protocol along the way.[15] In the hunt for small *p*-values, a measure of statistical significance,[16] the presence of even relatively modest amounts of Analytical Flexibility could have large consequences for the reliability of the resulting findings.[17] The exploitation of Analytical Flexibility has come to be known as "*p*-hacking."[18] P-hacking can generate effects that look statistically significant, but that are not in fact real; that is, effects that are not replicable.[19] Practices that rely on Analytical Flexibility and that can reduce the reliability of research results include:

---

[14] *See, e.g.*, Nelson et al.*, supra* note 1, at 515 (asserting that *p*-hacking is "the only honest and practical way to consistently get underpowered studies to be statistically significant" (emphasis omitted)); Marjan Bakker, Annette van Dijk & Jelte M. Wicherts, *The Rules of the Game Called Psychological Science*, 7 PERSPS. ON PSYCH. SCI. 543, 547 (2012) (stating that approaches involving multiple small, underpowered studies are "the optimal strategy" for manufacturing statistically significant results where there is no true effect); John et al., *supra* note 9, at 524 ("[Q]uestionable practices may constitute the prevailing research norm.").

[15] Ioannidis, *supra* note 2, at 0696-98 ("The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true."); Simmons et al., *supra* note 2, at 1359 ("[F]lexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates."); John et al., *supra* note 9, at 524 (referring to methodologies that exploit Analytical Flexibility as "questionable research practices").

[16] A *p*-value represents the likelihood that, given the assumption that the "null hypothesis" (i.e., the assumption that there is no true effect) is true, the measured outcome came about purely by chance. Expressed as a number between 0 and 1, the smaller the *p*-value, the stronger the support the measurement lends to the tested hypothesis. For example, if the null hypothesis states that there is no correlation between eating pizza and an increase in IQ and a study finds a 3-point increase in test subjects' IQ upon eating pizza with a corresponding *p*-value of $p = 0.2$, that *p*-value expresses that the correlation is such that if there is no real correlation between eating pizza and an increase in IQ, we should expect to find a 3-point increase 20% of the time. I.e., if we ran the same test five times, by purely random chance we would find an increase in IQ by three points or more in (on average) one of those tests. While any cutoff is somewhat arbitrary, $p < 0.05$ has traditionally been used as the threshold for statistical significance. Nosek et al., *supra* note 13, at 617. For a thorough overview of *p*-values and their many common misinterpretations, see Sander Greenland et al., *Statistical Tests,* P-*Values, Confidence Intervals, and Power: A Guide to Misinterpretations*, 31 EUR. J. EPIDEMIOLOGY 337, 340-45 (2016).

[17] The mere ability to add ten data points to an initially collected data can raise the false-positive rate by 50 percent. Simmons*, supra* note 2, at 1361; s*ee also* Bakker et al., *supra* note 14, at 547 (demonstrating that commonly used "questionable research practices" can lead to false-positive findings up to 40 percent); Nelson et al.*, supra* note 1, at 516-17 ("In truth, it is not that hard to get a study's false-positive rate to be very close to 100%.").

[18] Munafò et al., *supra* note 2, at 3 (defining *p*-hacking as the intentional or unintentional use of "data-contingent analysis decisions," usually to obtain statistically significant results).

[19] Simmons*, supra* note 2, at 1361 (finding that the flexibility to add ten data points combined with a choice of variables and control variables virtually guarantees a researcher statistically significant support for a hypothesis, even if the hypothesis being tested is false).

- failure to commit to a hypothesis up front, instead gathering the data first and then retrofitting a hypothesis to suit the results;[20]
- failure to commit, prior to the start of data collection, to a study protocol that includes details about the manner in which data will be collected, selected, cleaned, and analyzed, instead leaving the researcher free to add additional data points, remove outlier data at will, vary approaches for handling missing data, slice and dice the data in any number of ways, and report only the methodology that produced the reported results;[21] and
- reporting only the findings that are interesting or desirable, instead of reporting all results obtained along with a full accounting of the methodology used and all analytical decisions made.[22]

A key insight from the Replication Crisis is that even well-intentioned researchers can easily be led astray when they use such methods. Even if they have no intention to deceive themselves or their audience, genuine curiosity coupled with cognitive biases and professional pressures make Analytical Flexibility hard to resist.[23] And the exploitation of Analytical Flexibility is usually impossible to detect, unless research results are reported along with extensive detail about how they were generated.[24] Prior to the Replication Crisis, journals did not typically

---

[20] This practice is known as "HARKing" (Hypothesizing After Results are Known). *See* Munafò et al., *supra* note 2, at 2.

[21] *See, e.g.*, Ioannidis, *supra* note 2, at 0697-98 (arguing that flexibility in study design, sample size, and number of correlations tested all contribute to unreliable research outcomes); Simmons et al., *supra* note 2, at 1359-60 (demonstrating how "unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis").

[22] This practice is known as "publication bias" or the "file-drawer problem." *See* Munafò et al., *supra* note 2, at 3.

[23] *See* Chris Allen & David M. A. Mehler, *Open Science Challenges, Benefits and Tips in Early Career and Beyond*, 17 PLOS BIOLOGY, May 1, 2019, at 1, 6 ("'[S]cience is an ongoing race between our inventing ways to fool ourselves, and our inventing ways to avoid fooling ourselves.'" (citation omitted)); Nelson et al., *supra* note 1, at 518 ("[*P*-hacking] is something that benevolent researchers engage in while trying to understand their otherwise imperfect results."); Munafò et al., *supra* note 2, at 2 (noting that scientists have a "natural tendency . . . to see patterns in noise," and identifying confirmation bias, and hindsight bias as additional contributors to flawed data analysis); Nelson et al., *supra* note 1, at 515 ("Researchers did not learn from experience to increase their sample sizes precisely because their underpowered studies *were not failing*. *P*-hacking allowed researchers to think, . . . 'most of my studies work; [critics of *p*-hacking] must be talking about other people.'" (emphasis in original)); Nosek et al., *supra* note 13, at 615 (institutional incentives "inflate the rate of false effects in published science"); *id.* at 626 ("[P]ublishing is a central, immediate, and concrete objective for [a researcher's] career success."); Munafò et al., *supra* note 2, at 2 (describing how perverse incentives undermine the scientific method); Nelson et al., *supra* note 1, at 512 (finding that researchers "rely[] on methods of data collection and analysis that make it too easy to publish false-positive, nonreplicable results"); Marc A. Edwards & Siddhartha Roy, *Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition*, 34 ENV'T. ENG'G. SCI. 51, 55 (2017) (describing the tendency of people to see themselves as more honest than their peers as the "Muhammad Ali effect"); Head et al., *supra* note 2, at 8 (concluding that "*p*-hacking is rife").

[24] *See* Nelson et al., *supra* note 1, at 512 (noting that methods of data collection and analysis that were in widespread use for decades made it "impossible to distinguish between findings that

require disclosure of this type of information,[25] and researchers did not sufficiently realize how important this type of disclosure was.[26]

The problems caused by Analytical Flexibility were not confined to the field of psychology. They have, thus far, received the most attention in psychology, however, in part because psychology was the epicenter of the Replication Crisis—the first discipline where celebrated studies turned out to be unreliable—and in part because it is a discipline whose studies tend to generate statistical data.[27] Analytical Flexibility poses a risk to the reliability of statistical data. That means that it does not threaten (or threatens to a much lesser extent) studies that are qualitative in nature.[28] Anthropologists and sociologists who engage in qualitative research or ethnography may not have much to learn from the Replication Crisis, but sociologists who do engage in statistical analysis, as well as many economists, medical researchers, biologists, pharmacologists, and others whose work tends to generate statistical data have been forced to take note and consider shoring up their methodology.[29]

---

are true and replicable and those that are false and not replicable"); Munafò et al., *supra* note 2, at 4 ("Improving the quality and transparency in the reporting of research is necessary to address" reproducibility problems.). For more detail, see *infra* Section III.A.

[25] *See* Andrew W. Brown, Kathryn A. Kaiser & David B. Allison, *Issues with Data and Analyses: Errors, Underlying Themes, and Potential Solutions*, 115 PNAS 2563, 2567 (2018) ("[A]uthors do not always report information specified in guidelines, nor do peer reviewers demand that the information be reported.").

[26] Nelson et al., *supra* note 1, at 518 ("P-hacking is a pervasive problem precisely because researchers usually do not realize that they are doing it or appreciate that what they are doing is consequential.").

[27] Katherine Nelson, *Quantitative and Qualitative Research in Psychological Science*, 10 BIOLOGICAL THEORY 263, 263 (2015) ("Certainly among the American social sciences psychology is most assertively the quantitative experimental laboratory science.").

[28] *See, e.g.*, Uri Simonsohn, *Preregistration Prevalence*, DATACOLADA (Nov. 13, 2023), https://datacolada.org/115 [https://perma.cc/48Z6-WFJ2] (explaining that review articles and observational studies are not "pre-registerable"). *But see* Tamarinde Haven, Fernando Rosenblatt, Rafael Pineiro & Florian G. Kern, *Qualitative Preregistration*, CTR. FOR OPEN RSCH. (Dec. 10, 2020), https://www.cos.io/blog/qualitative-preregistration [https://perma.cc/BN4K-CRG3] (suggesting preregistration of hypotheses, even when a study does not have a quantitative component).

[29] *See* Andrew C. Chang & Phillip Li, *Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not"* 1 (Fin. & Econ. Discussion Series, Working Paper No. 2015-083, 2015) (successfully replicating 29 of 59 economics studies); C. Glenn Begley & Lee M. Ellis, *Raise Standards for Preclinical Cancer Research*, 483 NATURE 531, 532 (2012) (confirming results for only 6 of 53 hematology and oncology studies); Munafò et al., *supra* note 2, at 1 ("85% of biomedical research efforts are wasted . . . ."); Open Science Collaboration, *supra* note 11, at aac4716-1 (describing two replication studies in molecular biology that had a replication rate of 11% and 25%, respectively); Joanna Diong, Annie A. Butler, Simon C. Gandevia & Martin E. Héroux, *Poor Statistical Reporting, Inadequate Data Presentation and Spin Persist Despite Editorial Advice*, PLOS ONE, Aug. 15, 2018, at 1, 2 (reporting wide use of inappropriate statistical techniques in pharmacology); Daniel Engber, *Cancer Research Is Broken*, SLATE (Apr. 19 2016), https://slate.com/technology/2016/04/biomedicine-facing-a-worse-replication-crisis-than-the-one-plaguing-psychology.html [https://perma.cc/UU2R-7P7V] ("[W]e face a replication crisis in the field of biomedicine, not unlike the one we've seen in psychology but with far more dire implications.").

III. SCIENTISTS' RESPONSE TO THE REPLICATION CRISIS

Psychology's "year of horrors" in 2011,[30] sparked a substantial investment in efforts to improve research practices and make research results more reliable, in psychology but also in other academic disciplines.[31] This meta-science effort focused on methodological questions: How should researchers design research procedures that reduce the risk of being led astray by Analytical Flexibility? And how should labs and other research communities foster a culture of rigor and discipline in the creation and evaluation of research results? This period of methodological self-reflection has been described as "psychology's renaissance,"[32] as it changed how research is performed in a profound way.[33]

In this Part, I first discuss at a high level (in Section III.A) what interventions have been recommended to reduce Analytical Flexibility and improve reliability. Next (in Section III.B), I discuss the strides that have been made—primarily in psychology but also in other academic disciplines—in implementing these interventions and the extent to which they are starting to become (or have already become) accepted norms.

A. Methodological Recommendations

As I have described in more detail in earlier work, much of the methodological improvement in the social sciences has come from reducing the presence of Analytical Flexibility.[34] It is likely not possible to eliminate a researcher's flexibility during the research process entirely, but the more this flexibility can be contained, the more likely the process is to result in reliable findings.[35] Reducing Analytical Flexibility begins at the planning stage and may require limiting a researcher's ability or desire to adjust a research protocol on the fly.[36] Central to

---

[30] *See, e.g.*, Eric-Jan Wagenmakers, *A Year of Horrors*, 27 DE PSYCHONOOM 12, 12 (2012) ("[T]he year 2011 can go in the books as a true annus horribilis."); *see also* Ottoline Leyser, Danny Kingsley & Jim Grange, *The Science 'Reproducibility Crisis'—and What Can Be Done About It*, THE CONVERSATION (Mar. 15, 2017), https://theconversation.com/the-science-reproducibility-crisis-and-what-can-be-done-about-it-74198 [https://perma.cc/2T27-CJT7] ("Murmurings of low reproducibility began in 2011—the 'year of horrors' for psychology . . . .").

[31] *See* Nelson et al., *supra* note 1, at 511 (observing that events from 2010 to 2012 "sparked a period of methodological reflection that we . . . call Psychology's Renaissance"); Munafò et al., *supra* note 2, at 1 ("The field of metascience . . . is flourishing.")

[32] Nelson et al., *supra* note 1, at 512.

[33] *Id.* ("If a team of research psychologists were to emerge today from a 7-year hibernation, they would not recognize their field.").

[34] *See* Beerdsen, *supra* note 6, at 555-565 (discussing measures proposed or adopted in response to the Replication Crisis, to reduce Analytical Flexibility).

[35] *See supra* note 15 and sources therein. *See also* Brian A. Nosek et al., *Replicability, Robustness, and Reproducibility in Psychological Sciences*, 73 ANN. REV. PSYCH. 719, 731 (2022) (describing "[i]nitial evidence" suggesting that precommitting to a research protocol increases replicability).

[36] *See* Simmons et al., *supra* note 2, at 1362-63 (stating that the more flexibility researchers can eliminate, the more reliable the resulting outcomes will be); Open Science Collaboration, *supra* note 13, at 5-6 ("The best defense against inflation of false positives is to reduce the degrees of freedom

many methodological reforms, therefore, is the desire for a researcher to commit up front to as many aspects of a research protocol as possible, leaving limited ability to tinker with the protocol along the way.[37] If a particular domain of tinkering cannot be eliminated, the next-best option is to make the tinkering visible and evaluable.[38]

One recommended measure is for a researcher to formulate a hypothesis before starting an experiment, rather than gathering data and then retrofitting a hypothesis to the results—a practice known as "HARKing" (Hypothesizing After Results are Known).[39] Generating a hypothesis to fit existing data reduces the reliability of research results, because the hypothesis can be influenced by psychological biases such as hindsight bias and apophenia (a desire to see patterns in data even when none exist).[40] It also allows for a fluke result to be passed off as a result that is true and even expected in light of the (custom-created) hypothesis.[41] A researcher can engage in exploratory work without a preformulated hypothesis, but to create research results that are worthy of reliance, such exploration needs to be followed by separate confirmatory work, with a preformulated hypothesis.[42]

A more demanding and likely more impactful measure than formulating a hypothesis up front calls on a researcher to design a study protocol with care and then commit to either executing that protocol precisely as planned, or documenting and disclosing any deviation from the original protocol.[43] A research process that is generated in an incremental manner while the study is underway provides for too much flexibility. When the researcher has the ability to decide midstream to collect additional data or run additional analyses, it may well be almost impossible not to take advantage of that opportunity to enhance their research results.[44] But it is this

available to the researcher by writing down, prior to analyzing the data, how the data will be analyzed.").

[37] *See id*.; s*ee also infra* notes 42-55 and accompanying text.

[38] *See* Nelson et al., *supra* note 1, at 512 (arguing that analytical flexibility makes it impossible to distinguish true findings from false ones).

[39] *See* Nosek et al., *supra* note 5, at 2600; *see also* Munafò et al., *supra* note 2, at 2 (explaining that HARKing is a threat to reproducibility).

[40] *See* Nosek et al., *supra* note 5, at 2600 (noting that hindsight bias allows a researcher to construct a narrative that imbues randomness with meaning); Munafò et al., *supra* note 2, at 1 (explaining how apophenia can lead to false conclusions).

[41] *See* Patrick E. Shrout & Joseph L. Rodgers, *Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis*, 69 ANN. REV. PSYCHOL. 487, 491 (2018) (noting that unexpected interactions are often attributable to randomness). For an example of how this plays out, see Beerdsen, *supra* note 6, at 551 nn.121-122, 552 nn.123-124, 558 n.168 and accompanying text.

[42] *See* Allen & Mehler, *supra* note 23, at 5; Morton Ann Gernsbacher, *Writing Empirical Articles: Transparency, Reproducibility, Clarity, and Memorability*, 1 ADVANCES METHODS & PRACS. PSYCHOL. SCI. 403, 405 (2018).

[43] *See* Open Science Collaboration, *supra* note 13, at 6 ("By committing to a pre-specified analysis plan, one can avoid common cognitive biases.").

[44] *See, e.g.*, Simmons et al., *supra* note 2, at 1361 (reporting that approximately 70% of surveyed researchers reported having continued data analysis based on an interim data analysis).

type of tinkering—"*p*-hacking"[45]—that injects unreliability into research results. A documented protocol created prior to the collection of data should include details such as the sample size[46] and an account of every data-collection, data-cleaning, and data-analysis step the researcher plans to take.[47] Every decision that can be laid out up front rather than left to be decided later (when the study is underway) serves to reduce the amount of Analytical Flexibility, so the greater the extent to which a research protocol can be turned into a "script" of predetermined steps, the greater the likelihood that the project will generate reliable results.[48] Early studies confirm that preregistration is indeed associated with improved replicability.[49]

To further eliminate Analytical Flexibility, the researcher could choose to submit the research protocol (either publicly or nonpublicly) to a registry.[50] This practice, known as "preregistration," has come to be considered as the gold standard for engaging in reliable research.[51] At the conclusion of the study, when publishing the study's results, the author either confirms that she followed the preregistered research protocol exactly as planned or identifies and explains any departures from the protocol.[52] Preregistration serves four purposes. First, it forces the preregistering researcher to think through her research protocol and document it, rather than allow the protocol to be created in an ad-hoc manner potentially driven

---

[45] *See supra* note 18.

[46] *See* Bakker et al., *supra* note 14, at 552.

[47] *See* Open Science Collaboration, *supra* note 13, at 6 ("By committing to a pre-specified analysis plan, one can avoid common cognitive biases); *id.* at 5, 11 (explaining that the plan should include all conditions, variables, covariates, as well as when data collection will stop and start); Nosek et al., *supra* note 5, at 2602 (emphasizing that an ideal research plan should include all steps to be taken).

[48] Simmons et al., *supra* note 2, at 1362-63; s*ee also* Marcus Munafò et al., *Scientific Rigor and the Art of Motorcycle Maintenance*, 32 NATURE BIOTECHNOLOGY 871, 872 (2014) (recommending checklists for the design and preregistration of studies); s*ee also* Nelson et al., *supra* note 1, at 516-17 (explaining that even levels of Analytical Flexibility that are generally thought to be acceptable can raise false-positive rates from 5 to 61 percent).

[49] Nosek et al., *supra* note 34, at 731 ("Initial evidence suggests that behavioral changes such as preregistering the study, using large samples, and sharing research materials are associated with high replicability.").

[50] *See* Open Science Collaboration, *supra* note 13, at 7; Nosek et al., *supra* note 5, at 2600.

[51] *See, e.g.*, Shrout & Rodgers*, supra* note 41, at 493 (noting that preregistration has been called "the most important outcome[] of the replication crisis"); Nelson et al.*, supra* note 1, at 519 (asserting that preregistration is "the only way for authors to convincingly demonstrate that their key analyses were not *p*-hacked"); Alexandra Sarafoglou et al., *A Survey on How Preregistration Affects the Research Workflow*, 9 ROYAL SOC. OPEN SCI. 2054-5703 (2022) (noting that preregistration is perceived to be the "gold standard").

[52] *See* Allen & Mehler, *supra* note 23, at 3-4 (suggesting that researchers may make changes to a preregistered plan but should justify and discuss each change at publication); Gernsbacher, *supra* note 42, at 404; Open Science Collaboration, *supra* note 13, at 17 ("[T]he investigator can later demonstrate that his or her published analysis matched his or her original plan—or, if any changes were necessarily, detail what was changed and why.").

by interim data.[53] Second, it serves as a commitment device, as a researcher will be less likely to depart from the protocol if each departure must be justified in an eventual publication.[54] Third, it helps make imperfect methodological practices visible and thereby provides the reading audience with the context necessary to evaluate the study's reliability.[55] Fourth, it helps make *all* studies discoverable, not only those studies that yielded publication-worthy results.[56]

Improving the reliability of research results also requires attention to the way results are reported, as certain forms of transparency in reporting can enhance the audience's ability to evaluate the reliability of the reported findings.[57] Ideally, when reporting results, authors should disclose not only all positive results and how they were obtained, but also all variables collected, all conditions applied, any data that were excluded from the analysis, any other analytical decisions made, and any analyses that failed, to allow the reading audience to evaluate the results in context.[58]

B. Methodological Reforms Becoming Normative

The field of psychology, where the Replication Crisis began, has changed dramatically in the past decade. As many of the recommendations described above have been taken up, many researchers have begun to conduct their studies using updated methods and to train new researchers accordingly.[59] Journals and grant-funding organizations are applying new standards to their review of submissions.[60] New organizations and networks have emerged with an aim to improve research standards and facilitate researchers' ability to engage in reliable science.[61] Not all proposed measures described above have been universally implemented in full, but even as early as 2018 observers remarked that "[i]f a team of research psychologists

---

[53] Eric Jan Wagenmakers & Gilles Dutilh, *Seven Selfish Reasons for Preregistration*, APS OBSERVER (Oct. 31, 2016), https://www.psychologicalscience.org/observer/seven-selfish-reasons-for-preregistration [https://perma.cc/HVD9-YED3].

[54] *See, e.g.*, *id.* (stating that preregistration helps prevent researchers "being taken hostage by [their] own data"); Allen & Mehler, *supra* note 23, at 2-3.

[55] *See* Nosek et al., *supra* note 5, at 2602 (stating that preregistration "does not eliminate the possibility of poor statistical practices, but it does make them detectable"). In some cases, it may even be possible to adjust reported measures of confidence based on known amounts of Analytical Flexibility. *Id.* at 2601 (noting that correcting reported p-values in this manner is rarely done but is sometimes possible).

[56] *See* Nosek et al., *supra* note 34, at 728.

[57] Munafò et al., supra note 2, at 4 ("Improving the quality and transparency in the reporting of research is necessary to address" reproducibility problems.).

[58] Open Science Collaboration, *supra* note 13, at 11 (noting that the journal *Psychological Science* requires full disclosure of sample-size selection, exclusion of observations, and all experimental conditions, including failed manipulations); Nelson et al., *supra* note 1, at 518-19 (noting that some peer reviewers now demand disclosure of all measures, conditions, exclusions, and other analytical decisions).

[59] *See infra* notes 102-111 and accompanying text.

[60] *See infra* notes 112-132 and accompanying text.

[61] *See infra* notes 63-98 and accompanying text.

were to emerge today from a 7-year hibernation, they would not recognize their field."[62] Progress has accelerated since then, and other disciplines have followed in psychology's footsteps.[63] This Section will describe, first, institutional and infrastructural efforts to improve scientific methodology and, second, the extent to which these have resulted in a cultural shift within the scientific community.

1. Institutional and Infrastructural Efforts

Newly created organizations and resources have played a cardinal role in driving cultural change. The Center for Open Science ("COS") was founded in 2013 to "start, scale, and sustain open research practices that enhance accountability to research integrity, facilitate the self-corrective process of science, expand transparency and sharing of all research content, and improve research rigor and reproducibility."[64] COS was conceived as a "non-profit technology and culture-change organization."[65] It has served as a central hub for a large variety of efforts aimed at improving the quality of research studies, and as a repository of related resources for researchers. COS has been so influential that in 2015 (just two years after its founding) co-founder Brian Nosek was named one of *Nature*'s "10 people who mattered this year."[66]

Membership of COS provides researchers with tools and trainings in research methodologies that promote reproducibility and rigor.[67] COS maintains the Open Science Framework ("OSF"), created in 2012 "with a mission to increase openness,

---

[62] Nelson et al., *supra* note 1, at 512.

[63] *See, e.g.,* Diong et al., *supra* note 28, at 7 (listing awareness and transparency initiatives across disciplines and noting that "[t]here is considerable momentum throughout science"); Munafò et al., *supra* note 47, at 872 (describing preregistration checklists used in medicine, animal studies, observational epidemiology, and other fields); Leigh W. Simmons, Editorial, *Guidelines for Transparency and Openness (TOP)*, 28 BEHAV. ECOLOGY 347, 347 (2017) (describing a symposium focused on transparency initiatives in behavioral ecology); Christophe Bernard, Editorial, *Improving the Way Science Is Done, Evaluated, and Published*, ENEURO, Nov./Dec. 2017, at 1, 2 (noting that ENEURO "nearly guarantees publication" of preregistered studies); Agnès Dechartres, Philippe Ravaud, Ignacio Atal, Carolina Riveros & Isabelle Boutron, *Association Between Trial Registration and Treatment Effect Estimates: A Meta Epidemiological Study*, BMC MED., July 4, 2016, at 1, 1 (observing that preregistration has become the norm rather than the exception in medicine); s*ee also* Paul Bloom, *Psychology's Replication Crisis Has a Silver Lining*, ATLANTIC (Feb. 19, 2016), https://www.theatlantic.com/science/archive/2016/02/psychology-studies-replicate/468537 (arguing that the "Replication Crisis" is an opportunity for the field of psychology to "lead the way").

[64] *About*, CTR. FOR OPEN SCI., https://www.cos.io/about [https://perma.cc/Z7ZE-L6Z7] (last visited Mar. 14, 2025).

[65] *Strengthening Transparency or Silencing Science? The Future of Science in EPA Rulemaking: Hearing Before the H. Comm. on Sci., Space, and Tech.*, 116th Cong. (2019) (statement of Brian A. Nosek, co-founder and Executive Director of COS) [hereinafter Nosek, *Strengthening Transparency* statement].

[66] Nature, *Nature's 10*, 528 NATURE 459, 459 (2015).

[67] *See OSF for Institutions*, OPEN SCI. FRAMEWORK, https://www.cos.io/products/osf-institutions [https://perma.cc/4L8Q-2DJM] (last visited Feb. 2, 2025).

integrity, and reproducibility of scientific research."[68] Through OSF, researchers can preregister their studies in the form of "time-stamped, archived, read-only versions of a study plan." [69] Journals and funders that would like to encourage or demand preregistration of the studies they publish or fund can rely on OSF for the preregistration process. Since 2014, OSF has also provided the infrastructure for "registered reports," a form of preregistration whereby (some of) the peer-review process takes place at the registration stage, before results are known.[70]

In addition to maintaining OSF, COS brought together representatives from journals, funders, and societies to promulgate a series of Transparency and Openness Promotion ("TOP") Guidelines, which were published in SCIENCE[71] and aim to help journals develop standards that improve the quality of published research.[72] Already as of 2020, more than 5,000 journals had adopted a version of these guidelines.[73] As part of its meta-science efforts, COS spearheaded multiple reproducibility projects aimed at assessing the replicability (and, therefore, reliability) of existing studies in psychology, to investigate further what makes a study replicable or unreplicable.[74] Among many other projects, COS collaborates with the National Science Foundation to train researchers in rigorous and reliable

---

[68] *About*, OPEN SCI. FRAMEWORK, https://osf.io/4znzp/wiki/home (last visited May 8, 2025).

[69] Nici Pfeiffer & Mark Call, *Surpassing 100,000 Registrations on OSF: Strides in Adoption of Open and Reproducible Research*, CTR. FOR OPEN SCI. (Sept. 7, 2022), https://www.cos.io/blog/surpassing-100000-registrations-on-osf [https://perma.cc/C884-H5GS]. OSF is not the only preregistration platform. *See, e.g.*, AS PREDICTED, https://aspredicted.org [https://perma.cc/EU67-Y5GS] (last visited Feb. 2, 2025) (hosted by the University of Pennsylvania); ZENODO, https://zenodo.org [https://perma.cc/G2EM-RU4Y] (last visited Feb. 2, 2025) (developed by a European partnership of 50 organization and operated by the European Organization of Nuclear Research (CERN)); *see also Pre-Registration Information*, STAN. GRADUATE SCH. OF BUS., https://www.gsb.stanford.edu/faculty-research/behavioral-lab/conduct-research/pre-registration-information [https://perma.cc/J3RE-CSGF] (last visited Feb. 2, 2025) (listing specialized preregistration platforms for disciplines including economics, international development, and medicine).

[70] *Registered Reports: Peer Review Before Results Are Known to Align Scientific Values and Practices*, CTR. FOR OPEN SCI., https://www.cos.io/initiatives/registered-reports [https://perma.cc/XW5S-BE7Q] (last visited Feb. 2, 2025).

[71] Brian A. Nosek et al., *Promoting an Open Research Culture*, 348 SCIENCE 1422 (2015).

[72] Center for Open Science, *Guidelines for Transparency and Openness Promotion (TOP) in Journal Policies and Practices*, OPEN SCI. FOUNDATION, https://osf.io/ud578 [https://perma.cc/UTW2-UHDR] (last visited Feb. 2, 2025). An alternative set of principles are the "FAIR" principles for reporting data, developed by a group of academics, industry representatives, publishers, and government agencies. Mark D. Wilkinson et al., *The FAIR Guiding Principles for Scientific Data Management and Stewardship*, SCI. DATA, Mar. 15, 2016, at 1.

[73] *See The TOP Guidelines Are a Recognized Standard in Publishing and Funding*, CTR. FOR OPEN SCI., https://www.cos.io/initiatives/top-guidelines [https://perma.cc/GW7P-9LC9] (last visited Feb. 2, 2025) (click "Signatories") (listing over 5,000 signatories, including journals published by Elsevier, Springer, the American Psychological Association, BioMed Central, Oxford University Press, and Cambridge University Press, and indicating that as of 2020, any additional signatories are no longer being added to the list).

[74] *E.g.*, Open Science Collaboration, *supra* note 11.

research practices,[75] and with the Robert Wood Johnson Foundation to create a "SCORE" program for assessing the reliability of already-existing research results.[76]

The Society for the Improvement of Psychological Science ("SIPS") was created in 2016, also in direct response to the Replication Crisis, and aims to "improve methods and practices in psychological science."[77] It organizes an annual conference, publishes a journal, and regularly publishes resources for authors seeking to engage in reliable science.[78] Other newly formed institutions and centers aiming to improve the production of reliable research results include the Meta-Research Innovation Center at Stanford ("METRICS"), housed within the Stanford School of Medicine ("dedicated to finding ways to improve the validity and transparency of scientific research"),[79] and several similar organizations outside of the United States.[80]

The booming meta-science movement[81] that came out of the Replication Crisis also included the creation of new outlets to publish the meta-research it was generating, including journals such as *Meta-Psychology*[82] and *Advances in Methods*

---

[75] *The Center for Open Science Partners on New National Science Foundation Large-Scale Research Infrastructure for Education,* CTR. FOR OPEN SCI. (Apr, 30, 2024), https://www.cos.io/about/news/cos-partners-on-new-national-science-foundation-large-scale-research-infrastructure-for-education [https://perma.cc/4GNQ-RQVM].

[76] *Center for Open Science Expands Systematizing Confidence in Open Research and Evidence (SCORE) Program Efforts* CTR. FOR OPEN SCI. (Nov. 14, 2023), https://www.cos.io/about/news/cos-expands-score-program-efforts [https://perma.cc/Z7E3-LE89].

[77] *Mission Statement*, SOC'Y FOR THE IMPROVEMENT OF PSYCH. SCI., https://improvingpsych.org/mission [https://perma.cc/EZ54-8237] (last visited Feb. 2, 2025).

[78] *SIPS Products*, SOC'Y FOR THE IMPROVEMENT OF PSYCH. SCI., https://improvingpsych.org/sipsinaction/sips-products [https://perma.cc/W7AM-S4M9] (last visited Feb. 2, 2025).

[79] META-RESEARCH INNOVATION CTR. AT STAN. (METRICS), https://metrics.stanford.edu [https://perma.cc/FE2R-N8MR] (last visited Feb. 2, 2025).

[80] *See e.g.*, Meta-Research Center at Tilburg University, META-RESEARCH CTR., https://metaresearch.nl [https://perma.cc/5WHF-7MV9] (last visited Feb. 2, 2025) ("We study the scientific system in psychology and in general to find its flaws and empirically test potential solutions"); the Research On Research Institute in London, UK, *About*, RSCH. ON RSCH. INST., https://researchonresearch.org/about [https://perma.cc/MC6X-87FW] (last visited Feb. 2, 2025) ("ensuring that we have the evidence we need to realise the full potential of research"); the Meta-research and Evidence Synthesis Unit at the George Institute in New Delhi, India, *Meta-research and Evidence Synthesis Unit*, GEORGE INST. FOR GLOB. HEALTH https://www.georgeinstitute.org.in/units/meta-research-and-evidence-synthesis-unit [https://perma.cc/Q6WY-9GEV] (last visited Feb. 2, 2025); the Australia-based Association for Interdisciplinary Meta-Research and Open Science (AIMOS), AIMOS, https://aimos.community [https://perma.cc/4Q9P-AVWB] (last visited Feb. 2, 2025) ("to improve the quality of scientific research"); and the QUEST Center for Responsible Research in Berlin, *QUEST Center for Responsible Research*, BERLIN INST. OF HEALTH, https://www.bihealth.org/en/translation/innovation-enabler/quest-center/mission-approaches [https://perma.cc/B4GN-G396] (last visited Feb. 2, 2025) (aiming to make biomedical research "accessible, reproducible and generalizable and thus have a high value").

[81] Munafò et al., *supra* note 2, at 93 ("The field of metascience . . . is flourishing.").

[82] LINNAEUS UNIV.: META-PSYCHOLOGY, https://open.lnu.se/index.php/metapsychology [https://perma.cc/K6YH-J6VE] (last visited Feb. 2, 2025)

& Practices in Psychological Science;[83] methodological training platforms such as the Framework for Open and Reproducible Research Training ("FORRT"),[84] awareness campaigns,[85] motivational campaigns,[86] symposiums,[87] and at least two task forces.[88]

Attention to methodological rigor and transparency has not been confined to the realm of researchers and journals. Organizations that provide funding and coordination of research efforts have also joined the replicability train. The National Science Foundation has funded replication-focused research proposals, as have the Defense Advanced Research Projects Agency ("DARPA"), the German Research Foundation ("DFG"), the Netherlands Organization for Scientific Research ("NWO"), the Bill & Melinda Gates Foundation, and the European Commission.[89]

Various influential governmental entities have taken note as well. The White House's Office of Science and Technology Policy ("OSTP") convened a Task Force in 2021, which formulated a number of principles promoting transparency

---

[83] Daniel J. Simons, *Editorial, Introducing Advances in Methods and Practices in Psychological Science*, 1 ADVANCES METHODS & PRACTICES PSYCH. SCI. 3, 3 (2018) (inaugurating new journal); *see also* PLOS BIOLOGY *and* PLOS ONE *Meta-Research Collection* (Nov. 16, 2021) https://collections.plos.org/collection/plos-biology-and-plos-one-meta-research-collection [https://perma.cc/8ZEX-FVWC] (gathering "research on research" articles published in PLOS BIOLOGY and PLOS ONE).

[84] FRAMEWORK FOR OPEN & REPRODUCIBLE RSCH. TRAINING (FORRT), https://forrt.org [https://perma.cc/G8ED-83ER] (last visited Feb. 2, 2025).

[85] *See, e.g.*, Wagenmakers & Dutilh, *supra* note 53 (explaining to researchers why it is in their interest to preregister their studies); Kai Kupferschmidt, *More and More Scientists Are Preregistering Their Studies; Should You?*, SCIENCE (Sept. 21, 2018), https://www.science.org/content/article/more-and-more-scientists-are-preregistering-their-studies-should-you; DIV. OF RSCH., COUNCIL FOR EXCEPTIONAL CHILDREN, OPEN SCIENCE BRIEF: REGISTERED REPORTS (Feb. 2023), https://cecdr.org/sites/default/files/2023-02/06_open_science_brief_6._registered_reports_draft34.pdf [https://perma.cc/2FWP-NNNS].

[86] *See, e.g.*, Brian A. Nosek, *We Are In It Together*, 32 PSYCH. INQUIRY 45, 47 (2021) (making the case for collective responsibility for a discipline's research output, arguing that "[f]ields need not be embarrassed by discovering errors, but they should worry about failing to discover them," and that "[s]cience's strength and credibility does not come from being right, it comes from a relentless commitment to getting it right").

[87] *E.g.*, METASCIENCE 2019 SYMPOSIUM, https://www.metascience2019.org [https://perma.cc/X4BC-TYFJ] (last visited Feb. 2, 2025).

[88] David C. Funder et al., *Improving the Dependability of Research in Personality and Social Psychology: Recommendations for Research and Educational Practice*, 18 PERSONALITY & SOC. PSYCH. REV. 3 (2014) (reporting recommendations of the Society for Personality and Social Psychology's Presidential Task Force on Publication and Research); *Scientific Integrity Task Force*, WHITE HOUSE: OFFICE OF SCI. & TECH., https://www.whitehouse.gov/ostp/ostps-teams/nstc/scientific-integrity-task-force [https://perma.cc/PSU5-ALUK] (last visited Feb. 2, 2025).

[89] Nosek et al., *supra* note 34, at 737; NAT. ACADS. OF SCI., MED., & ENG'G, OPEN SCIENCE BY DESIGN: REALIZING A VISION FOR 21ST CENTURY RESEARCH 3-5 (2018); *see also* Nosek, *Strengthening Transparency* statement, *supra* note 64, at 3 ("[M]any federal agencies have taken steps toward improving policies supporting transparency and reproducibility of research.").

and accountability,[90] a move that COS's leadership has called "probably the most important event for open science in the United States to date."[91] The Environmental Protection Agency ("EPA") adopted a rule that requires the agency to consider transparency and reproducibility of data to be taken into account when it makes policy decisions.[92] DARPA initiated a program to investigate whether machine algorithms might be used to assess the credibility of research data.[93] The National Academies of Science, Medicine, and Engineering published a "Vision for 21st Century Research" that discussed concerns about replicability and addresses the need for preregistration, transparent reporting, and culture change in research settings.[94] The United Nations Educational, Scientific and Cultural Organization ("UNESCO") adopted a Recommendation on Open Science in 2021, [95] and convened no fewer than five working groups to implement key aspects of the Recommendation, plus a Steering Committee and an Intersectoral Task Team.[96] The White House named 2023 the Year of Open Science.[97] The National Aeronautics and Space Administration ("NASA") participated in the Year of Open Science as part of its own Transform to Open Science ("TOPS") initiative[98] and enlisted the support of numerous influential organizations, including the Federation

---

[90] Memorandum from Alondra Nelson, Deputy Director Office of Science and Technology Policy for the Heads of Executive Departments and Agencies (Aug. 25, 2022), https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf [https://web.archive.org/web/20250102041624/https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf] (providing "policy guidance to federal agencies with research and development expenditures," which includes a focus on fraud and corruption, but also on transparent reporting of scientific data in other contexts).

[91] *A Win for Open Science: White House OSTP's Updated Guidance Advances Open Access and Data Sharing Across Federal Agencies*, CTR. FOR OPEN SCI. (Sept. 1, 2022), https://www.cos.io/blog/white-house-ostp-guidance-advances-open-access-and-data-sharing [https://perma.cc/Z6Q8-S2LV]. Note that the "Open Science" movement encompasses more than just replicability-focused efforts. Some Open Science efforts focus on transparency as a means to increase access to research data, fight corruption, or improve the equitable allocation of research funding. However, these objectives frequently overlap, and many Open Science initiatives are aimed at increasing transparency as an effort to generate more reliable science.

[92] Nosek, *Strengthening Transparency* statement, *supra* note 65, at 3.

[93] *Id*., at 4.

[94] National Academies, *supra* note 89, at 5-7.

[95] *Recommendation on Open Science*, UNESCO, https://www.unesco.org/en/open-science/about?hub=686 [https://perma.cc/B3Z3-QXKT] (last updated Sept. 21, 2023).

[96] *Implementation of the UNESCO Recommendation on Open Science*, UNESCO, https://www.unesco.org/en/open-science/implementation?hub=686#open-science-working-groups [https://perma.cc/369F-25DM] (last updated Mar. 13, 2025).

[97] *FACT SHEET: Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research*, WHITE HOUSE: OFFICE OF SCI. & TECH. (Jan. 11, 2023), https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research [https://perma.cc/D4YY-XCEK].

[98] *NASA Boosts Open Science through Innovative Training*, NASA (Apr. 11, 2023), https://www.nasa.gov/centers-and-facilities/marshall/nasa-boosts-open-science-through-innovative-training/ (describing NASA's Open Science efforts).

of American Scientists, the Health Research Alliance, the Life Sciences Editors Foundation, and many others.[99] Suffice it to say, attention to the reliability of scientific research has transcended the narrow confines of psychology labs and journals, and major players on the world stage are paying attention and are taking action.

2. Cultural Change

As discussed above, energetic institutional efforts have rapidly created an infrastructure that allows researchers to change the way they perform their studies and report the resulting data. But have these efforts actually resulted in a change in community norms and a shift toward more reliable research methodology? It appears that they have, though more work remains to be done.

As metascience researchers have pointed out, eliminating technical barriers may not be sufficient if the main barriers are social or cultural.[100] But as one group of authors put it, "[t]he rules of statistical inference have no empathy for how hard it is to acquire the data."[101] "Make it Normative" is therefore one of COS's operating principles with respect to reliable methodology, as is "Make it Required."[102]

The use of preregistration has been growing exponentially since the introduction of preregistration platforms in the early 2010s.[103] As early as 2018, a widely cited article reported a "cultural shift" toward preregistration.[104] During the period of 2017-2018, preregistration was described as not yet the norm, but quickly becoming the norm in psychology research.[105] Whereas in 2011, 8% of surveyed psychologists reported having preregistered a study, by 2017, this number had risen to 44%.[106] By 2018, the use of the recommended practices such as preregistration had become "orders of magnitude more common" than before the Replication

---

[99] *Center for Open Science Coordinates NASA-Funded Initiative to Support the Year of Open Science*, CTR. FOR OPEN SCI. (Sept. 12, 2023), https://www.cos.io/about/news/cos-coordinates-nasa-funded-initiative [https://perma.cc/EU2W-GKPN].

[100] Nosek et al., *supra* note 13, at 626.

[101] Nosek et al., *supra* note 5, at 2604.

[102] *A Decade of Impact: 2023 Impact Report*, CTR. FOR OPEN SCI., https://www.cos.io/impact [https://web.archive.org/web/20240824004904/https://www.cos.io/impact] (last visited Aug. 24, 2024) [hereinafter *COS 2023 Impact Report*].

[103] Nosek et al., *supra* note 35, at 736.

[104] Nosek et al., *supra* note 5, at 2605.

[105] Nelson et al., *supra* note 1, at 520 (observing in 2018 that preregistration should become the norm in the next three to five years); Brian A. Nosek & D. Stephen Lindsay, *Preregistration Becoming the Norm in Psychological Science*, APS OBSERVER (Feb. 28, 2018), https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science (preregistration is becoming the norm, much of it driven by APS journals).

[106] Nosek et al., *supra* note 35, at 736 (citing Garrett Christensen et al., *Open Science Practices are on the Rise: The State of Social Science (3S) Survey* (Cnt. for Effective Glob. Action, Working Paper No. 106, 2020)).

Crisis.[107] A 2019 survey among sixty-nine psychology departments in the United States, 42%-57% of participants in social, quantitative, and cognitive psychology had active OSF accounts (ownership of OSF accounts was much lower in clinical and education psychology, but higher in more quantitatively oriented fields).[108] Many of the world's most highly regarded universities, including Harvard, MIT, and Oxford, became members of OSF.[109] By mid-2022, more than 100,000 studies had been preregistered through the OSF platform.[110] The platform received more than 10 million unique visitors in 2021.[111] As of 2023, it has almost 700,000 registered users.[112]

As for scientific journals, change has been slower but is underfoot as well. Few journals *require* preregistration, but many encourage the practice. For example, Nature Portfolio journals (which include leading scientific journal *Nature* itself as well as several dozen more specialized journals) do not require preregistration, but do require authors to indicate the preregistration status of their studies when submitting their work.[113] The highly influential PLOS (Public Library of Science) journals do not require preregistration, but label the practice "highly encouraged."[114] The American Psychological Association ("APA"), which is "psychology's largest publisher,"[115] likewise does not require preregistration but "recommend[s] that researchers try out preregistration, and if a design isn't preregistered, be transparent in the reporting of research results, asking authors to distinguish between planned and unplanned analyses."[116] The APA recognizes that expectations vary by subdiscipline, but many of its editors encourage

---

[107] Nelson et al., *supra* note 1, at 529.

[108] Nosek et al., *supra* note 35, at 736.

[109] *OSF Institutions*, OPEN SCI. FRAMEWORK, https://osf.io/institutions [https://perma.cc/4GAW-Q9K8] (last visited Mar. 12, 2025).

[110] *Surpassing 100,000 Registrations on OSF*, CTR. FOR OPEN SCI., https://www.cos.io/blog/surpassing-100000-registrations-on-osf [https://perma.cc/7K2M-6SAE] (last visited Mar. 12, 2025).

[111] Pfeiffer & Call, *supra* note 69.

[112] *COS 2023 Impact Report*, *supra* note 102.

[113] *Editorial Policies: Reporting Standards and Availability of Data, Materials, Code and Protocols*, NATURE PORTFOLIO, https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards [https://perma.cc/F5BW-LLQS] (last visited Mar. 12, 2025) ("Nature Portfolio journals support study pre-registration (including clinical trials) and pre-registration of analysis plans in public repositories; details of pre-registration should be provided with submission.") [hereinafter NATURE PORTFOLIO policies].

[114] *Research Integrity and Publication Ethics*, PLOS, https://plos.org/research-integrity-and-ethics [https://perma.cc/J5D6-SV4X] (last visited Apr. 12, 2025) ; *see also How to Report Statistics*, PLOS, https://plos.org/resource/how-to-report-statistics [https://perma.cc/2VSM-5HYX] (last visited Mar. 12, 2025) [hereinafter PLOS *How to Report Statistics*] (recommending preregistration for any study reporting statistical data).

[115] Nosek et al., *supra* note 35, at 737.

[116] *Preregistration*, AM. PSYCH. ASS'N, https://www.apa.org/pubs/journals/resources/preregistration [https://perma.cc/VXT6-FW7E] (last visited Mar. 12, 2025) [hereinafter APA *Preregistration*].

preregistration.[117] APA journals offer badges to recognize preregistered research.[118] The APA has also partnered with the British Psychological Society, the German Psychological Society, the Leibniz Institute for Psychology and Center for Open Science, and the Center for Open Science to create a set of Preregistration Standards for Quantitative Research in Psychology.[119]

Going a step beyond simple preregistration, as of early 2025, over 300 journals offer a publication track that involves Registered Reports.[120] These journals include some of Nature's specialty journals and several APA journals.[121] In a Registered Reports track, a proposed study protocol is subjected to peer review.[122] If the protocol is accepted, the final manuscript reporting the results is presumptively accepted for publication, and reviewed only for adherence to the planned protocol.[123]

Even though many journals do not (yet) require preregistration as a condition for publication, preregistration can still be in authors' own interest. Preregistration can bolster an author's reputation and (as noted above) at some journals it virtually guarantees acceptance.[124] A study of papers published in 2022 in four major psychology journals found that 43% of papers included preregistered studies.[125] While this may seem lower than expected in light of the above, some research published in these journals could not have been preregistered because it is not quantitative in nature (e.g., review articles or observational studies), so the preregistration rate for reported statistical data is higher.[126]

While few journals make preregistration mandatory, large numbers of journals have made transparency of research methodology and data a priority. *Nature* and its related journals require that authors make available all data, code, and protocols,

---

[117] *APA Supports Openness and Rigor in Psychological Science*, AM. PSYCH. ASS'N, https://www.apa.org/pubs/journals/resources/open-science [https://perma.cc/5DS7-GLKK] (last visited Mar. 13, 2025) [hereinafter *APA Open Science*].

[118] APA *Preregistration*, *supra* note 117.

[119] *International Open Science: A Preregistration Template for Quantitative Research in Psychology,* LEIBNIZ INSTITUT FÜR PSYCHOLOGIE (Jan. 27, 2021), https://leibniz-psychology.org/en/news/detail/internationale-zusammenarbeit-prae-registrierungsvorlage-fuer-die-quantitative-forschung-in-der-psych-1 [https://perma.cc/X38V-ZASX].

[120] To view participating journals, see *Registered Reports*, CTR. FOR OPEN SCI., https://www.cos.io/initiatives/registered-reports [https://perma.cc/J75R-QBR7] (last visited Mar. 12, 2025) (click "Participating Journals").

[121] NATURE PORTFOLIO policies, *supra* note 113 ("Registered Reports, a research article format intended to reduce publication bias and increase methodological rigour is available at Nature Human Behaviour, Communication Psychology and for select disciplines at Nature Communications, Nature Methods, and Scientific Reports. Registered Reports involves a two-stage peer review approach where methods and analysis are pre-registered and peer reviewed before carrying out the intended research.").

[122] *APA Open Science*, *supra* note 118.

[123] *Id*.

[124] Wagenmakers & Dutilh, *supra* note 53.

[125] Simonsohn, *supra* note 28.

[126] *Id.*

as a precondition for publication.[127] So do many other journals.[128] APA journals do not employ a unified set of guidelines across the board but "support" open science efforts and encourage editors to set requirements as appropriate to the relevant subdiscipline.[129] PLOS journals use checklists and detailed guidelines for the reporting of statistical data, citing the need for "rigorously reported" methodology in light of the "flexible analysis" that led to the Replication Crisis.[130] Peer reviewers increasingly demand disclosure of such information as well.[131] As of 2023, more than 400 journals have achieved a "TOP" rating of at least six, with points awarded for every recommended research-publication practice.[132]

While research culture has not yet changed so dramatically that publishing of research results has become impossible without preregistration or data transparency, these "Open Science" procedures are increasingly expected. As one psychologist remarked, "[i]t is pretty incredible to be living in a period in which we get to observe the field undergo a historical transformation."[133]

IV. THE LEGAL PROFESSION'S RESPONSE TO THE REPLICATION CRISIS

This Part of the paper will be short. It summarizes the legal world's response to the Replication Crisis but there is not much to summarize. Law's response to the Replication Crisis has been almost nonexistent. Although scientific research is presented in courtrooms in this country every day,[134] the legal profession has barely taken note of the problems raised by the Replication Crisis.

---

[127] NATURE PORTFOLIO policies, *supra* note 113.

[128] *See, e.g.*, Tom E. Hardwicke & Simine Vazire, *Transparency Is Now the Default at* Psychological Science, 35 PSYCH. SCI. 708, 708-09 (explaining that transparency is now the default at PSYCHOLOGICAL SCIENCE); Bernard, *supra* note 63, at 1 (noting that since 2017, ENEURO requires inclusion of computer code); Simmons et al., *supra* note 63, at 347 (stating that BEHAVIORAL ECOLOGY will only publish papers documenting all design decisions and analytical decisions); Wolf Vanpaemel et al., *Are We Wasting a Good Crisis? The Availability of Psychological Research Data After the Storm*, COLLABRA, Oct. 9, 2015, at 1, 1 (2015) (explaining that Science and Nature journals require openness of data); Mercè Crosas et al., *Data Policies of Highly-Ranked Social Science Journals*, SOCARXIV 8 (Mar. 30, 2018), https://osf.io/preprints/socarxiv/9h7ay [https://perma.cc/9Q96-5SUY] (noting that as of January 2018, 155 of 291 social sciences journals surveyed have a data-disclosure policy).

[129] *APA Open Science*, *supra* note 118.

[130] PLOS *How to Report Statistics*, *supra* note 114.

[131] Nelson et al., *supra* note 1, at 518-19.

[132] *Anniversary Timeline*, CTR. FOR OPEN. SCI., https://www.cos.io/timeline [https://perma.cc/Q3ZA-U3XG] (last visited Apr. 13, 2025).

[133] Vazire & Nosek, *supra* note 5, at 1 ("Their comment reflects a sentiment that many share—much has changed in how psychology is done and communicated during the last decade, perhaps more than in any other.").

[134] Studies measuring the prevalence of expert testimony have consistently found that experts are used in more than half of civil cases that reach trial. *See, e.g.*, Andrew W. Jurs, *Expert Prevalence, Persuasion, and Price: What Trial Participants Really Think About Experts*, 91 IND. L.J. 353, 355 (2016) (reporting that expert witnesses appeared in 86% of civil trials in an urban Iowa county); Anthony Champagne, Daniel Shuman & Elizabeth Whitaker, *An Empirical Examination*

A search for "Replication Crisis" across all legal opinions in the Lexis Plus database (state and federal, trial-level and appellate, without date restrictions) brings up just three opinions.[135] In two of these opinions, the judge references the Replication Crisis in holding that a proposition supported by a single study should not be considered conclusively established, because science is fallible.[136] In the third opinion, the judge references the Replication Crisis even more broadly, in support of the point that science is not perfect and mistakes are to be expected.[137] There are no reported opinions in which research results presented by experts, witnesses, or other party representatives are subjected to methodological scrutiny with reference to any of the lessons that have come out of the Replication Crisis.[138]

Even the term "*p*-hacking"—the criticized practice central to the Replication Crisis—[139] turns up in only two relevant opinions.[140] In one opinion, in ruling on a summary-judgment motion, the Eastern District of Pennsylvania notes that the parties disagree as to appropriateness of an expert's methodology, happens to cite an article that includes the term "*p*-hacking" in its title, and concludes without much discussion that the parties' disagreement is fodder for cross-examination rather than

---

of the Use of Expert Witnesses in American Courts, 31 JURIMETRICS J. 375, 381 (1991) (reporting that experts testified in more than half of civil cases in Dallas); Samuel R. Gross, *Expert Evidence*, 1991 WIS. L. REV. 1113, 1119 (reporting that expert testified in 86% of civil trials in California state court).

[135] Lehmann v. Ohr Pharm., No. 18 Civ. 1284, 2019 U.S. Dist. LEXIS 161164 (S.D.N.Y. Sept. 20, 2019) *aff'd and remanded,* 830 F. App'x 349 (2d Cir. 2020); Federoff v. Geisinger Clinic, 571 F. Supp. 3d 376 (M.D. Pa. 2023); *Chiles v. Salazar*, 116 F.4th 1178 (10th Cir. Sept. 12, 2024), *cert. granted,* No. 24-539, 2025 U.S. LEXIS 1025 (U.S. Mar. 10, 2025) (Hartz J., dissenting). The search was performed on Sept. 13, 2024. A confirmatory search in the Westlaw Edge database turns up the same three opinions.

[136] *Chiles*, 116 F.4th at 1238 (Hartz J., dissenting) ("Not every study published in a peer-reviewed journal can be relied on."); *Federoff*, 571 F. Supp. 3d at 389 ("Over the past decade, the medical sciences (among others) have struggled through a replication crisis—essentially, that the results of studies are not reproducible, making them unreliable. . . . That doesn't mean we can't trust science; it means that we can't just take any single study off the shelf that supports our view and assume its findings will hold.").

[137] *Lehmann*, 2019 U.S. Dist. LEXIS 161164, at *15 (rejecting shareholder suit complaining that defendant pharmaceutical company should not have proceeded to phase III medical trials and stating, "This Court will not adopt a rule that discourages free scientific inquiry in the name of shielding investors from risks of failure. Science is risky. . . . With science suffering from a replication crisis, this Court is happy to report that the law does not abide attempts at using the judiciary to stifle the risk-taking that undergirds scientific advancement and human progress. The answer to bad science is more science." (citations omitted)).

[138] "Replicability Crisis," an alternative term for the crisis used by some scientists, turns up zero search hits in Lexis Plus. "Preregist!" (which should capture instances of "preregistration, "preregister," "preregistered," etc.) does turn up about 1,000 search hits, but these tend to relate to unrelated topics, such as preregistration of copyright claims and preregistration to vote or to attend an event. A search for "preregist!" in the same sentence as either "study" or "protocol" turns up zero relevant results. A search for "p-hacking" turns up two opinions, discussed *supra* note 141.

[139] *See supra* note 18.

[140] *In re* Roundup Prods. Liab. Litig., 390 F. Supp. 3d 1102 (N.D. Cal. 2018), *aff'd sub nom.* Hardeman v. Monsanto Co., 997 F.3d 941 (9th Cir. 2021); Allegheny Cnty. Emps. Ret. Sys. v. Energy Transfer LP, 744 F. Supp. 3d 350 (E.D. Pa. 2024).

grounds for summary judgment.[141] In other words, the court holds that the alleged *p*-hacking goes to weight rather than admissibility. The other opinion, issued in the *In re Roundup* multidistrict litigation against Monsanto, contains a slightly more in-depth discussion. Monsanto moved to exclude the testimony of an expert witness, among other reasons because the expert had engaged in "*p*-hacking, manipulation of data to obtain statistically significant results."[142] Specifically, Monsanto alleged that the expert had allowed his approach to data analysis to evolve over time.[143] In denying the motion, the court appears somewhat sympathetic to this critique, encouraging Monsanto to emphasize these discrepancies to the finder of fact, but it nevertheless declines to exclude the analyses at issue in light of the fact that the expert was relying on multiple sources of data and not solely on the allegedly *p*-hacked data.[144]

There is some recognition among courts that significant departures from a research protocol (when it exists) should cast doubt on a study's reliability. A Florida district court excluded a study after departures from the study protocol "[called] its reliability into question."[145] A West Virginia district court rejected a study where the researcher (among other alleged misconduct) "fail[ed] to abide by his own protocol and plans for conducting the study."[146] A South Carolina district judge excluded an analysis performed by an expert who did not prespecify a statistical model, instead trying "at least five different statistical models."[147] A Massachusetts district judge excluded expert testimony based on a "measure that was not the primary outcome identified before the trial was conducted."[148] But to the best of my knowledge, no court has affirmatively required an expert witness to preformulate a study protocol as a matter of course.

Litigants have used the term "Replication Crisis" more frequently than judges, but still not often. A search on Lexis Plus turns up twenty-seven documents that are categorized as "pleadings," "motions," or "briefs," and that reference the term. Many of these are meaningless hits,[149] but some are not. In a handful of cases,

---

[141] *Allegheny*, 2024 U.S. Dist. LEXIS 140849, at 363 & n.14.

[142] *In re Roundup*, 390 F. Supp. 3d. at 1137.

[143] *Id.*

[144] *Id.*

[145] *In re* Denture Cream Prod. Liab. Litig., No. 09-2051-MD, 2015 WL 392021, at *11 (S.D. Fla. Jan. 28, 2015), *aff'd sub nom.* Jones v. SmithKline Beecham, 652 F. App'x 848 (11th Cir. 2016).

[146] Black v. Rhone-Poulenc, Inc., 19 F. Supp. 2d 592, 598-604 (S.D.W. Va. 1998).

[147] *In re* Lipitor (Atorvastatin Calcium) Mtkg., Sales Pracs. & Prods. Liab. Litig., 145 F. Supp. 3d 573, 588 (D.S.C. 2015).

[148] *In re* Neurontin Mktg. & Sales Practices Litig., No. 04-CV-10739-PBS, 2011 WL 3852254, at *110 (D. Mass. Aug. 31, 2011), *aff'd*, 712 F.3d 21 (1st Cir. 2013).

[149] Meaningless hits include duplicates of the three opinions mentioned above, *supra* note 135, and exhibits to a court filing that happen to cite an article that includes the term "Replication Crisis" in the title but that are filed for an unrelated reason. *E.g.*, Brief of Brandon Butler as Amicus Curiae Supporting Defendant, Apple v. Corellium, No. 21-12835, 2023 U.S. App. LEXIS 11225 (11th Cir. May 8, 2023), 2022 U.S. 11th Cir. Briefs LEXIS 3093; Declaration of Peter Herzog in Support of Defendant's Opposition to Class Certification, Ex. 2, Ngethpharat v. State Farm, 339 F.R.D. 154 (W.D. Wash. 2021) (No. 20-CV-00454-MJP), 2021 U.S. Dist. Ct. Motions LEXIS 673651.

parties have raised lessons from the Replication Crisis to attack the validity of an opinion offered by an expert witness. For example, in *In re Zofran*, a multidistrict litigation case about anti-nausea drug Zofran, defendant GlaxoSmithKline ("GSK") moved to exclude the testimony of an expert on the basis that the expert had engaged in "significant results-driven deviations from the study protocol."[150] It noted that "Plaintiffs do not dispute that developing a study plan in advance—before data is gathered—and then strictly adhering to that plan is a bedrock of scientific reliability,"[151] and that even the expert at issue himself had written that "if the analytical strategy and process is not identified before analyzing the data, the reported p-values are not interpretable."[152] GSK also relied on a statement by the American Statistical Association to argue that the expert should have reported "all hypotheses explored, all data collection decisions, all statistical analyses conducted and all *p*-values computed," which he had not, and that "[d]eviation from this practice is unaffectionately labeled as cherry picking, fishing, p-hacking, or data dredging."[153] The court granted summary judgment for the defendant on other grounds, without explicitly ruling on this motion.[154]

In *United States v. Godsey*, the defendant moved for a sentencing variance, relying on social scientific studies finding a limited deterrent effect when punishment is increased.[155] In opposing the motion, the prosecution referenced the Replication Crisis, pointing out that the studies cited by the defendant included "no detail about methodology, or even more importantly, evidence or replication by later studies employing similar methodology."[156] Phrases such as "studies show" and "research suggests" lead to "bad thinking," the prosecution argued, in particular "given the extent to which so many studies aren't or can't be replicated."[157] The court denied the defendant's motion in a form order.[158]

In *Moitoso v. FMR LLC*, an ERISA case, the plaintiff moved to limit the testimony of an expert economist.[159] Here, too, the moving party relied on guidance

---

[150] GlaxoSmithKline LLC's Reply Memorandum in Support of Its Motion in Limine to Exclude the Zambelli-Weiner Study and to Preclude Plaintiffs' Experts from Relying upon It at 2, *In re:* Zofran (Ondasetron) Prod. Liab. Litig., 541 F. Supp. 3d 164 (D. Mass. 2021) (MDL No. 15-md-2657-FDS), ECF No. 1915.

[151] *Id.* at 10.

[152] *Id.* at 12.

[153] *Id.* at 15 (internal quotation marks omitted).

[154] *In re:* Zofran (Ondasetron) Prod. Liab. Litig., 541 F. Supp. 3d 164, 167 (D. Mass. 2021).

[155] Defendant's Sentencing Memorandum and Motion for Variance Sentence at 10-11, United States v. Godsey, No. 3:23-cr-058 (E.D. Va. Dec. 20, 2023) (minute order), 2023 U.S. Dist. Ct. Briefs LEXIS 22986.

[156] Government's Response to Defendant's Motion for Variance Sentence at 2, United States v. Godsey, No. 3:23-cr-058 (E.D. Va. Dec. 20, 2023) (minute order), 2023 U.S. Dist. Ct. Motions LEXIS 272595.

[157] *Id.*

[158] United States v. Godsey, 3:23-cr-058 (E.D. Va. Dec. 20, 2023) (minute order).

[159] Memorandum of Law in Support of Plaintiff's Motion to Exclude in Part the Expert Report and Testimony of Bruce Strombom at 14, Moitoso v. FMR LLC, No. 1:18-cv-12122 (D. Mass. Jun. 18, 2020) (minute order), ECF No. 124.

from the American Statistical Association, this time to argue that conclusions should not be based "solely on whether an association or effect was found to be statistically significant," because the "widespread use of 'statistical significance' (generally interpreted as '$p \leq 0.05$') as a license for making a claim of a scientific finding . . . leads to considerable distortion of the scientific process."[160] Many months later, the motion ended up being denied as moot in light of a series of stipulations.[161]

Finally, in a pair of cases challenging Idaho and Tennessee state statutes that criminalize gender-affirming care, a number of states repeatedly filed amicus briefs in support of the statutory bans.[162] The states accuse American medical organizations who argue against such bans of engaging in "performative" conduct and of supporting their advocacy with illegitimate data.[163] The Replication Crisis, they argue, has highlighted "the ability of researchers to use statistical tricks to make their preferred findings appear significant."[164] The Idaho case is awaiting a decision by the Ninth Circuit;[165] the Tennessee case is currently before the Supreme Court.[166]

In all, conversations about *p*-hacking, departures from study protocols, and untransparent reporting of data occasionally happen in the courts of this country, but they are far from common. Courts have not had many opportunities to rule on issues relating to Analytical Flexibility, and even when they have had a chance, they have been hesitant to use alleged *p*-hacking as a reason to exclude expert testimony, impose requirements, or otherwise engage with these methodological issues. Furthermore, the institutional efforts described in Part III find no counterpart in the legal profession. There are no ABA committees, no task forces, and even resources such as the *Reference Manual on Scientific Evidence* and the *Science Bench Book for Judges* (each of which include a basic chapter on statistics) do not delve into the topics at issue.[167] Indeed, the *Reference Manual on Scientific Evidence* has not been updated since 2011.[168]

---

[160] *Id.* at 14.

[161] Moitoso v. FMR LLC, No. 1:18-cv-12122 (D. Mass. Jun. 18, 2020) (minute order).

[162] Brief of Alabama, Arkansas, and 21 Other States as Amici Curiae Supporting Appellants and Reversal, Poe v. Labrador, No. 24-142 (9th Cir. filed Feb. 13, 2024), 2024 U.S. 9th Cir. Briefs Lexis 1543; Brief of Alabama as Amicus Curiae Supporting Respondents, United States v. Skrmetti, Nos. 23-466, 23-477, 23-492 (U.S. filed Feb. 2, 2024).

[163] Brief of Alabama as Amicus Curiae Supporting Respondents, *supra* note 163, at 24.

[164] *See id.*

[165] *See* Poe v. Labrador, 144 S. Ct. 921, 921 (2024) (lower-court's universal injunction stayed pending appeal in the Ninth Circuit).

[166] United States v. Skrmetti, No. 23-477 (U.S. June 24, 2024) (granting certiorari), https://www.supremecourt.gov/docket/docketfiles/html/public/23-477.html.

[167] *See* David H. Kaye & David A. Freeman, *Reference Guide On Statistics*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 211 (3d ed. 2011); Eryn Blagg & Alicia Carriquiry, *Introduction to Statistical Thinking for Judges*, *in* SCIENCE BENCH BOOK FOR JUDGES 131 (2d. ed. 2020).

[168] REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (3d ed. 2011); s*ee also Science for Judges - Development of the Reference Manual on Scientific Evidence, 4th Edition*, NAT. ACADS. OF SCI.. ENG'G, & MED., https://www.nationalacademies.org/our-work/science-for-judges-development-of-

V. DIVERGENT PATHS

In the 2010s, the field of psychology underwent a "renaissance."[169] Other fields that have a statistical bent have followed suit and recognized the importance of precommitment to research protocols and transparent reporting of results.[170] The law, meanwhile, appears to be stuck in the Dark Ages.

It is important to consider the lessons of the Replication Crisis in the legal context. Litigants frequently make use of experts.[171] Expert testimony on psychological topics is extremely common in American courts.[172] Expert testimony relying on other statistics from other disciplines, such as economics or medicine, are very common as well.[173] Whenever statistical data are being presented, it is important to consider how they were obtained and how they are being reported, and whether there is any risk that the data have been infected by Analytical Flexibility. Many of these experts testify about statistical data, either based on existing academic research or based on their own statistical analyses.[174] When experts rely on "off the shelf" studies and thereby bring the results of academic studies into the courtroom,[175] there is the possibility that these studies have been tainted by Analytical Flexibility. When experts conduct their own study, the expert herself may engage in analytically flexible practices that render her results unreliable.

---

the-reference-manual-on-scientific-evidence-4th-edition (last visited Feb. 11, 2025) (discussing ongoing work on a fourth edition).

[169] Nelson, *supra* note 1, at 512. ("[R]eferring to this period as psychology's 'replication crisis' . . . makes no sense. We do not call the rain that follows a long drought a water crisis. . . . This is psychology's *renaissance*.").

[170] *See supra* notes 28 & 62 and sources cited therein.

[171] *See supra* note 134 and sources cited therein.

[172] *See, e.g.*, David L. Faigman & John Monahan, *Psychological Evidence at the Dawn of the Law's Scientific Age*, 56 ANN. REV. PSYCH. 631, 651 (2005) (discussing a variety of contexts in which psychology experts testify in court); Christopher B. Jaeger, Daniel Levin & Evan Porter, *Justice is (Change) Blind: Applying Research on Visual Metacognition in Legal Settings*, 23 PSYCH. PUB. POL'Y & L. 259, 273 (2017) (discussing the role of expert testimony on the topic of visual metacognition).

[173] *See, e.g.*, Champagne et al., *supra* note 135, at 381 (finding that experts, including experimental psychologists, economists, biochemists, and engineers, were used in more than half of civil cases during a three-month period in Dallas, Texas).

[174] *See, e.g.*, Jones v. United States, No. 2:16-CV-00435-JRS-DLP, 2019 U.S. Dist. LEXIS 14382, at *7 (S.D. Ind. Jan. 30, 2019) (involving an analysis of water samples in defense of environmental tort claims); *In re* Countrywide Fin. Corp. Mortg.-Backed Sec. Litig., 984 F. Supp. 2d 1021, 1025 (C.D. Cal. 2013) (involving analysis of mortgage loans in support of securities litigation); Chin v. Port Auth. of N.Y. & N.J., 685 F.3d 135, 143-44 (2d Cir. 2012) (involving a statistical study in support of race-discrimination action).

[175] These studies might be introduced in the context of what has been termed "framework evidence." *See* David L. Faigman, John Monahan & Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. 417, 424 (2014) ("We refer to the scientist who testifies to the empirical framework of which a particular case might be an instance as a 'framework expert.'").

Thirty years post-*Daubert v. Merrell Dow Pharms., Inc.*,[176] courts are well aware of their role in ensuring that expert testimony on scientific and technical topics meets certain thresholds of reliability. Whether an expert testifies about another researcher's study or about her own, the court is tasked with evaluating her testimony for reliability.[177] The problem is that Analytical Flexibility adds a dimension of unreliability that is typically left unaddressed when courts perform this evaluation—a dimension that is undetectable unless the study protocol was preregistered and any departures from the protocol are disclosed in detail.[178]

In federal court and in the courts of most states, an expert's proffered testimony is evaluated under the standard of Federal Rule of Evidence 702, which codifies principles first articulated in *Daubert*.[179] The court is required to consider whether the "methodology underlying the testimony is scientifically valid."[180] In the language of Rule 702, this inquiry requires the court to consider (among other criteria) whether the testimony is "the product of reliable principles and methods."[181] The *Daubert* court shied away from articulating a "definitive checklist" for making this assessment, but offered a number of relevant factors: (1) whether the theory or technique on which the expert relies "can be (and has been) tested"; (2) whether it "has been subjected to peer review and publication"; (3) "the known or potential rate of error" associated with the theory or technique; and (4) whether the theory or technique has been "generally accept[ed]" in the "relevant scientific community."[182] The *Daubert* standard could easily accommodate an inquiry into the presence of Analytical Flexibility and the transparency of reporting. Whether a study was preregistered and whether it is being reported along with the types of detailed disclosures that allow the audience to determine the extent to which Analytical Flexibility was present could be considered dimensions of "the known or potential rate of error" and whether the methodology is "generally accept[ed]."[183] If a court believed that this would stretch the *Daubert* factors too far, it could also add preregistration and detailed disclosures as additional factors; after all, in promulgating the *Daubert* standard, the Supreme Court did "not presume to set out a definitive checklist or test."[184] The *Daubert* standard predates the Replication Crisis by two decades, and replication issues would not have been on the Justices' minds. They should be today.

---

[176] Daubert v. Merrell Dow Pharms. Inc., 509 U.S. 579 (1993).

[177] *See, e.g.*, FED. R. EVID. 702.

[178] *See supra* Section III.A.

[179] FED. R. EVID. 702; *Daubert*, 509 U.S. at 593-595 (1993) (laying out a non-exhaustive list of factors courts may consider when evaluating expert testimony); THOMAS A. MAUET & WARREN D. WOLFSON, TRIAL EVIDENCE 273 (8th ed. 2024) ("Thus far, more than 40 states have adopted *Daubert*. . . .").

[180] *Daubert*, 509 U.S. at 593.

[181] FED. R. EVID. 702(c).

[182] *Daubert*, 509 U.S. at 593-94.

[183] *See* Beerdsen, *supra* note 6, at 572-74.

[184] *Daubert*, 509 U.S. at 593.

A few states still follow the older *Frye* standard for the admissibility of scientific and technical evidence, which limits the inquiry to only the last *Daubert* factor: whether the expert's methodology has gained "general acceptance" in its field.[185] As noted above, this factor could be construed to require an inquiry into whether an expert conducted her work in a manner sufficiently conscious of the risk of Analytical Flexibility.[186] The *Frye* standard is a hundred years old, so it is unsurprising that it was not designed with replicability issues in mind. That does not mean, however, that they should not be on judges' and litigants' minds today.

Not all measures implemented in the academic sciences are equally applicable in a legal context. To some extent, each community that deals with scientific research will have to find mechanisms that are suitable to its particular ecosystem, and the solutions that have been implemented in an academic setting—preregistration and extensive disclosures—may not translate directly to a courtroom setting. Much as these measures contribute to the reliability (and therefore presumed accuracy) of results, in court, accuracy is not the only consideration. Litigation should also be efficient, affordable, and completed in a reasonable amount of time.[187] This means that some tradeoffs may be called for. Openness of data is likely to run into different types of pushback in a litigation setting compared to in an academic setting. Litigants are typically allowed some room for strategy, including in the form of selective disclosures.[188] And, in a criminal context, demanding full disclosures on the defendant's side may well run afoul, in certain instances, of the privilege against self-incrimination.[189] Thus far, though, this is a discussion we, as a legal profession, have not even begun to have. As for preregistration, I still believe that demanding preregistration of expert-conducted studies, coupled with increased transparency, as I have proposed previously,[190] is an option, but it is by no means the only option, and even if courts were to start experimenting with preregistration, they would have many possible options for going about it.[191]

---

[185] Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

[186] *See, e.g.*, Jason M. Chin, *Psychological Science's Replicability Crisis and What It Means for Science in the Courtroom*, 20 PSYCH. PUB. POL'Y & L. 225 (2014). Arguably, the Consensus framework recently proposed by Edward Cheng may be able to accommodate an inquiry into replicability issues as well. *See* Edward K. Cheng, *The Consensus Rule: A New Approach to Scientific Evidence*, 75 VAND. L. REV. 407 (2022). The central type of question in that framework, "Does the scientific community believe a particular fact?," *id.*, could invite testimony about the quality of the studies on the basis of which the relevant scientific community would be expected to make that determination.

[187] *See, e.g.*, FED. R. CIV. PROC. 1.

[188] *See* Edith Beerdsen, *Strategy for Strategy's Sake*, 103 N.C. L. REV. 733, 787 (2025) (describing forms and motivations for strategic concealment).

[189] *See* U.S. CONST. amend. V ("No person . . . shall be compelled in any criminal case to be a witness against himself . . . .").

[190] *See* Beerdsen, *supra* note 6, at 572-82 (proposing early admissibility procedures that could mirror preregistration practices, but could be implemented in a variety of ways).

[191] *See id.*

The first priority is to recognize that Analytical Flexibility is a source of unreliability that litigants and courts should start thinking about. It should be a consideration in the evaluation of expert evidence (including potentially in the selection of expert witnesses) and in the design of procedures that lead to its creation. Analytical Flexibility is a somewhat technical concept and asking judges and lawyers with no technical training to understand it may pose some challenges. But just as "[t]he rules of statistical inference have no empathy for how hard it is to acquire the data,"[192] justice has no empathy for how hard it is to acquire the necessary understanding.[193] Only if judges and lawyers know to be on the lookout for Analytical Flexibility can they start trying to avoid it and detect it. And only then can they begin demanding processes that allow for the creation of scientific evidence whose reliability can be properly scrutinized.

Measures that could be taken to limit the influence of Analytical Flexibility on results that come into court through expert testimony include applying a presumption that older studies (performed before the Replication Crisis) contain exaggerated results,[194] subjecting testimony to additional scrutiny if it relies on a single study,[195] supplementing *Daubert* or *Frye* with additional criteria,[196] a redesign of the expert discovery process to allow for early involvement of the court,[197] various types of preregistration of proposed research protocols by expert witnesses,[198] engaging specialists to help courts assess the reliability of expert evidence,[199] extensive disclosures of research procedures including (potentially) failed analyses and negative results,[200] and amendments to the rules of discovery to require more transparent reporting.[201] Other mechanisms could be devised and many options are possible. The point is: we should be thinking about how Analytical Flexibility affects legal proceedings and looking for ways to address it.

---

[192] Nosek *et al.*, *supra* note 5, at 2604.

[193] *See also* Allen & Mehler, *supra* note 23, at 3 (A reduction in the efficiency of continuous learning "may be the price of unbiased science.").

[194] *See, e.g.*, Jason M. Chin, Alex O. Holcombe, Kathryn Zeiler, Patrick S. Forscher & Ann Guo, *Metaresearch, Psychology, and Law: A Case Study on Implicit Bias*, 56 CONN. L. REV. 225, 294 (2023).

[195] *See* Maggie Wittlin, *Binding Hercules: A Proposal for Bench Trials*, 76 VAND. L. REV. 1735, 1753-54 n.109 (2023) ("[A]fter the 'Replication Crisis,' it may be prudent to avoid putting too much stock in any individual study before closely evaluating its methodology.").

[196] *See, e.g.*, Jarkko Jalava, Stephanie Griffiths & Rasmus Rosenberg Larsen, *How to Keep Unreproducible Neuroimaging Evidence Out of Court*, 29 PSYCH. PUB. POL'Y & L. 1, 1 (2023).

[197] *See* Beerdsen, *supra* note 6, at 572-82.

[198] *See* Jason M. Chin, Bethany Growns & David T. Mellor, *Improving Expert Evidence: The Role of Open Science and Transparency*, 50 OTTAWA L. REV. 365, 390 (2019) (proposing voluntary preregistration by expert witnesses); Beerdsen, *supra* note 6, at 572-82 (discussing a variety of preregistration models that could be adopted).

[199] *See* Chin et al., *supra* note 195, at 294.

[200] *See* Beerdsen, *supra* note 6, at 582-83 (proposing updated disclosure requirements).

[201] *See* Kaitlin McCormick-Huhn, *Preventing a (Replication) Crisis in the Courtroom*, 23 NEV. L. J. 643, 660 (2023) (proposing amendments to FED. R. CIV. P. 26(a)(2)(B) to require disclosure of a "replicability synopsis").

In the field of evidence, it is well known that some rules are based on outdated understandings of human behavior and psychology, and yet they are still being applied.[202] We should prevent the treatment of scientific and technical evidence from becoming another area that entrenches outdated practices in our legal system.

[202] *See, e.g.*, Richard A. Posner, *On Hearsay*, 84 FORDHAM L. REV. 1465, 1471 (2016) ("The hearsay exception for dying declarations, as for present sense impressions and excited utterances, is a fossil.").