LITIGATION SCIENCE AFTER THE KNOWLEDGE CRISIS

Edith Beerdsen†

The knowledge crisis that has been causing turmoil in the social sciences for the past few years represents a fundamental shift in scientists' understanding of what it takes to create reliable science. This crisis, now known as the Replication Crisis, has thrown the spotlight on 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. There is ample reason to believe that the defects of method uncovered in the course of the Replication Crisis are at least as pervasive in litigation science as they are in academic science. Yet the legal profession has failed to recognize and address them.

This Article is the first to address the broad implications of the Replication Crisis for the production of scientific knowledge in a civil-litigation context. Drawing on insights from the Crisis, it argues that current procedural practice is simply incapable of providing a court with the information it needs to make an accurate assessment of the reliability of scientific evidence. The Article identifies a number of core principles, drawn from the response of academic science to the Replication Crisis, that can guide reforms to the treatment of scientific evidence in civil litigation. It argues that shoring up the courts' capacity to evaluate scientific evidence requires a rethinking of the entire chain of creation of scientific knowledge and a reframing of the role of the court in that chain.

---

† Acting Assistant Professor of Lawyering, New York University School of Law. I am grateful to Edward K. Cheng, Jason M. Chin, Vincent Daly, Jonah B. Gelbach, J. Benton Heath, Helen Hershkoff, Christopher B. Jaeger, Justin McCrary, Erin E. Murphy, Dale A. Nance, Alexander A. Reinert, Daniel C. Richman, John Sexton, David Simson, James Steiner-Dillon, Jacob Victor, Maggie Wittlin, and participants at the 2020 Evidence Summer Workshop and the NYU Lawyering Scholarship Colloquium for helpful discussions and suggestions. For diligent research assistance, I thank Yan (Céline) Wang. Thanks also to Lily A. Coad, Victor Flores, John R. Mucciolo, Lachanda R. Reid, Michelle J. Zhu, and the staff of the *Cornell Law Review* for their conscientious and thoughtful editing.

INTRODUCTION

Novelty catches attention. In 2011, psychologist Daryl J. Bem made the front page of the *New York Times* with a study showing that people can "feel the future"—that it is possible to improve one's performance on a test by studying *after* taking the test.[1] A year earlier, a team of psychologists had published a study demonstrating that adopting confident postures ("power poses") causes hormonal changes that reduce stress and increase "feelings of power."[2] Like Bem's paper on extrasensory perception, the power-pose study received extensive coverage by the popular press,[3] and even catapulted one of its coauthors to internet fame.[4] The two studies had more in com-

---

1 Daryl J. Bem, *Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect*, 100 J. PERSONALITY & SOC. PSYCHOL. 407, 407 (2011); Benedict Carey, *Journal's Paper on ESP Expected to Prompt Outrage*, N.Y. TIMES (Jan. 5, 2011), https://www.nytimes.com/2011/01/06/science/06esp.html [https://perma.cc/ZLW2-Y3XT].

2 Dana R. Carney, Amy J.C. Cuddy & Andy J. Yap, *Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance*, 21 PSYCHOL. SCI. 1363, 1366 (2010).

3 *See, e.g.*, Danielle Venton, *Power Postures Can Make You Feel More Powerful*, WIRED (May 15, 2012, 1:30 PM), https://www.wired.com/2012/05/st-cuddy [https://perma.cc/GHF5-FH6F] (interviewing Cuddy about power poses); Katy Waldman, *That Time I Tried to Be Wonder Woman*, SLATE (Dec. 12, 2012, 3:06 PM), https://slate.com/human-interest/2012/12/amy-cuddy-and-power-poses-can-body-language-affect-your-confidence.html [https://perma.cc/TGQ9-LJXL] (describing the author's experimentation with power poses).

4 In 2012, coauthor Amy Cuddy gave a presentation about power poses that, with more than fifty-five million views as of early 2020, is still the second-most watched TED talk in history. *The Most Popular Talks of All Time*, TED, https://

mon than their novelty and viral power. Both studies suffered from the same severe methodological flaws.[5] And in each case, the flaws were virtually undetectable in the published papers.[6]

Inside the field of psychology, the reception of the two studies was vastly different, at least initially. Bem's study was scrutinized and picked apart immediately.[7] The results were so hard to believe that many researchers assumed that the methodology that had produced them had to be flawed.[8] In contrast, psychology researchers cited the power-pose study largely without questioning its conclusions.[9] The findings, while novel, were not as bewildering as those found by Bem, and therefore did not attract the same immediate scrutiny. The power-pose paper would not be subjected to a searching inquiry until years after publication, when the field of psychology was in the midst of an epistemic crisis[10]—one that drew large swaths of research findings into question and that would radically alter the way research is done in the social sciences.[11] By 2017, both studies had been discredited, along with scores of

---

www.ted.com/playlists/171/the_most_popular_talks_of_all [https://perma.cc/3TTZ-99L7] (last visited Nov. 20, 2020).

5 *See* Eric-Jan Wagenmakers, Ruud Wetzels, Denny Borsboom & Han L. J. van der Maas, *Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011)*, 100 J. PERSONALITY & SOC. PSYCHOL. 426, 427 (2011) (arguing that Bem obtained his result through a "fishing expedition" using a variety of analyses and reporting only the successful ones); Letter from Dana R. Carney, Assoc. Professor, UC Berkeley Dep't of Psychology, https://faculty.haas.berkeley.edu/dana_carney/pdf_my%20position%20on%20power%20poses.pdf [https://perma.cc/5BFK-NXLL] (last visited Nov. 20, 2020) (conceding, in a letter by the power-pose study's co-author, that the results were "p-hacked," or generated by attempting numerous different analyses and reporting only "the ones that 'worked'").

6 *See, e.g.*, Wagenmakers, Wetzels, Borsboom & van der Maas, *supra* note 5, at 427 (When results are reported selectively, "we simply do not know how many other factors were taken into consideration only to come up short . . . [or] how many other hypotheses were . . . tested and discarded.").

7 For critiques immediately following Bem's study, see *infra* subpart II.A and notes 76–80.

8 *See, e.g.*, Carey, *supra* note 1 (noting that copies of Bem's article "have circulated widely among psychological researchers . . . and have generated a mixture of amusement and scorn").

9 *See, e.g.*, Li Huang, Adam D. Galinsky, Deborah H. Gruenfeld & Lucia E. Guillory, *Powerful Postures Versus Powerful Roles: Which Is the Proximate Correlate of Thought and Behavior?*, 22 PSYCHOL. SCI. 95, 96 (2011) (examining the effects of body posture on behavior and thought); Azim F. Shariff & Jessica L. Tracy, *What Are Emotion Expressions for?*, 20 CURRENT DIRECTIONS PSYCHOL. SCI. 395, 398 (2011) (examining nonverbal emotional expression).

10 *See* Larry Laudan, *Epistemic Crises and Justification Rules*, 29 PHIL. TOPICS 271, 273 (2001) ("[A]n epistemic crisis occurs when a group or community . . . finds itself with reasons to question the correctness of the rules and structures it has been using for fixing beliefs.").

11 *See infra* subpart II.A.

other psychology studies whose results researchers had long considered trustworthy and credible.[12]

This crisis of knowledge, now commonly known as the "Replication Crisis,"[13] was sparked by a series of developments in the early- to mid-2010s that signaled that much of published psychology research could not be trusted or replicated. It rapidly spread from psychology to fields as diverse as economics, pharmacology, and medicine,[14] throwing into doubt the reliability of research approaches that, until recently, were applied routinely and without attracting serious scrutiny.[15] A period of soul-searching led to a broad recognition among scientists that: (1) the reliability of published research depends crucially on methodological rigor in ways that had not been sufficiently appreciated; and (2) lack of methodological rigor can contaminate research results in a manner that is entirely undetectable to a peer examining those results.[16] The Replica-

---

12 *See* Eva Ranehill et al., *Assessing the Robustness of Power Posing: No Effect on Hormones and Risk Tolerance in a Large Sample of Men and Women*, 26 PSYCHOL. SCI. 653, 653 (2015) (attempting and largely failing to replicate Carney, Cuddy, and Yap's power-pose study); Joseph P. Simmons & Uri Simonsohn, *Power Posing: P-Curving the Evidence*, 28 PSYCHOL. SCI. 687, 690 (2017) (concluding based on meta-analysis that the power-pose hypotheses were "currently lacking in empirical support"); sources cited *infra* notes 84 and 89 (reporting failed attempts to replicate for large numbers of studies).

13 *See infra* note 69.

14 *See, e.g.*, Colin F. Camerer et al., *Evaluating Replicability of Laboratory Experiments in Economics*, 351 SCIENCE 1433, 1434 (2016) (reporting failed replication attempts in economics); Joanna Diong, Annie A. Butler, Simon C. Gandevia & Martin E. Héroux, *Poor Statistical Reporting, Inadequate Data Presentation and Spin Persist Despite Editorial Advice*, 13 PLOS ONE e0202121, at 2 (2018), https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0202121&type=printable [https://perma.cc/X85S-5E2K] (reporting wide use of inappropriate statistical techniques in pharmacology); Daniel Engber, *Cancer Research Is Broken*, SLATE (Apr. 19, 2016, 9:21 AM), https://slate.com/technology/2016/04/biomedicine-facing-a-worse-replication-crisis-than-the-one-plaguing-psychology.html [https://perma.cc/7WNG-BPFZ] ("[M]uch of cancer research in the lab . . . simply can't be trusted" due to "[s]loppy data analysis . . . and poor experimental design . . . ."); *see also infra* note 126 and source cited therein (explaining that certain common practices had been criticized earlier but have now come under renewed and wider-spread scrutiny).

15 *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*, 13 PLOS BIOLOGY e1002106, at 8 (2015), https://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.1002106&type=printable [https://perma.cc/B9VS-BP2K] (concluding that "p-hacking is rife"); Leif D. Nelson, Joseph Simmons & Uri Simonsohn, *Psychology's Renaissance*, 69 ANN. REV. PSYCHOL. 511, 514–15 & n.4 (2018) (reporting evidence of widespread use of *p*-hacking and reluctance by researchers to consider it "a problem worth worrying about").

16 *See, e.g.*, Nelson, Simmons & Simonsohn, *supra* note 15, at 512 (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").

tion Crisis represents a fundamental shift in scientists' understanding of what it takes to create reliable scientific knowledge.[17] Its focus transcends individual techniques or tools; rather, it centers on methodology itself—on how the process by which research protocols are designed affects the reliability of research outcomes, and on the role that institutional structures, culture, and incentives play in methodological design.[18]

The Replication Crisis, this Article argues, holds important and as-yet-unappreciated lessons for the treatment of scientific evidence by the courts. While the conversation is still ongoing, a consensus is developing in the social sciences around a number of institutional reforms aimed at improving the reliability of research.[19] These reforms are largely procedural in nature, and include stricter regimes for planning a research project, restrictions on modifications while the project is underway, and radically increased openness of data.[20] The common thread in much of these reforms is the need to constrain researchers' flexibility during the research process. It is this "analytical flexibility"—the flexibility to adapt, supplement, or even completely make up research protocols when the study is already underway—that has a propensity to lead to research results that are unreliable in undetectable ways.[21] Unless the types and degrees of analytical flexibility are either exposed or eliminated, there is typically no way to assess the extent to which their presence has contaminated reported research results.[22]

There is ample reason to believe that the defects of method uncovered in the course of the Replication Crisis are at least as pervasive in litigation science[23] as they are in academic sci-

---

17 *See, e.g.*, *id.* ("Researchers now understand that the old ways of collecting and analyzing data produce results that are not diagnostic of truth and that a new, more enlightened approach is needed. Thousands of [researchers] have embraced this notion.").

18 *See infra* subpart II.A.

19 *See infra* subpart II.B. While the problems uncovered in the Replication Crisis first came up in the context of projects involving statistical analysis, they are not limited to this setting. Any scientific project that involves multiple steps has the potential to have its reliability undermined by analytical flexibility.

20 *See id.*

21 *See* Nelson, Simmons & Simonsohn, *supra* note 15, at 512 (arguing that analytical flexibility makes it impossible to distinguish true findings from false ones).

22 *See id.*

23 In this Article, I will use the term "litigation science" for the mode of science produced in connection with civil litigation by an expert witness retained by one or

ence.[24] *First*, in both modes of science, researchers face powerful incentives associated with unreliable research results: cultural and institutional norms and expectations demand that academic researchers produce publishable results, while researchers in a litigation setting are under pressure to produce results that can be presented to a jury. *Second*, analytical flexibility does not distinguish between the two modes of science: when present, it impairs the reliability of scientific findings regardless of the purpose for which the findings were generated. *Third*, since both modes of science have traditionally allowed experimenters to present research results selectively, the undetectable nature of this impairment can be expected to be endemic in both.

This Article is the first to address the broad implications of the Replication Crisis for the production of scientific knowledge in a civil-litigation context. Few scholars have made a connection between the Crisis and evidence law.[25] No scholarship has focused on the lessons the Crisis holds for judges and practitioners of litigation science. Moreover, a fundamental point has been left unaddressed thus far: that, in light of the Replication Crisis, current procedural practice is simply incapable of providing a court with the information it needs to make an accurate assessment of the reliability of scientific evidence.

---

more parties and the term "academic science" for science conducted for the general purpose of expanding knowledge.

24 *See infra* subpart III.A.

25 *See* Kevin D. Hill, *The Crisis in Scientific Publishing and Its Effect on the Admissibility of Technical and Scientific Evidence*, 49 J. MARSHALL L. REV. 727 (2016) (questioning the utility of peer review as a marker of reliability in the wake of the Replication Crisis and focusing largely on scientific fraud rather than structural methodological problems); Krin Irvine, David A. Hoffman & Tess Wilkinson-Ryan, *Law and Psychology Grows Up, Goes Online, and Replicates*, 15 J. EMPIRICAL LEGAL STUD. 320 (2018) (assessing the implications of the Replication Crisis for empirical legal research; *see also* Jason M. Chin, Rory McFadden & Gary Edmond, *Forensic Science Needs Registered Reports*, 2 FORENSIC SCI. INT'L: SYNERGY 41 (urging academic forensic science to adopt practices of the "open science" movement); Jason M. Chin, Gianni Ribeiro & Alicia Rairden, *Open Forensic Science*, 6 J.L. & BIOSCIENCES 255 (2019) (same). The most in-depth treatments of the Crisis's import for litigation are Jason M. Chin, *Psychological Science's Replicability Crisis and What It Means for Science in the Courtroom*, 20 PSYCHOL. PUB. POL'Y & L. 225 (2014) [hereinafter Chin, *Psychological Science's Replicability Crisis*] (identifying weaknesses in the *Daubert* admissibility standard for the evaluation of "framework evidence"—expert testimony that relies on existing scientific literature) and Jason M. Chin, Bethany Growns & David T. Mellor, *Improving Expert Evidence: The Role of Open Science and Transparency*, 50 OTTAWA L. REV. 365 (2019) (proposing reforms for the treatment of forensic evidence in Canadian courts in light of the Replication Crisis), but even these articles do not address the most critical implications for the production and assessment of scientific knowledge in the context of civil litigation.

It is critical that courts assess and shore up their capacity to evaluate scientific evidence. The Replication Crisis has demonstrated that institutional structures and incentives can greatly enhance or undermine the reliability of scientific-knowledge production.[26] Containing or rendering visible the "invisible contamination" of analytical flexibility will therefore require procedural and institutional reforms. I argue that these reforms can be implemented within existing legal frameworks,[27] but doing so requires a careful rethinking of the entire chain of creation of scientific evidence and a reframing of the role of the court in that chain.

This Article identifies a number of core principles—drawn from the response of academic science to the Replication Crisis—that can guide reforms to the treatment of scientific evidence in civil actions.[28] Accounting for the role analytical flexibility plays in the creation of scientific evidence will require much more rigorous planning of studies than is currently the norm; commitment to predetermined study protocols, with or without involvement of the court; and a wholesale rethinking of the timing and nature of information exchanges between litigants and the court.

These insights also have broader theoretical implications for the intersection between law and science. In particular, they disrupt both the traditional view that science offers judges a set of objective principles for the evaluation of scientific evidence[29] and the more critical view that the quality of scientific evidence can only be evaluated with reference to both process and results.[30] The Replication Crisis undermines the traditional view that judges can meet their gatekeeping obligations and prevent insufficiently reliable evidence from reaching the finder of fact by "thinking like a scientist."[31] By the time they examine proposed scientific evidence—after it has been fully

---

26 *See infra* Part II.

27 *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 592–94 (1993); Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

28 *See infra* subpart III.B.

29 *See Daubert*, 509 U.S. at 593 (assigning judges a gatekeeping role with respect to scientific evidence, "confident that federal judges possess the capacity to undertake this review" based on a flexible set of factors).

30 *See* Sheila Jasanoff, *Representation and Re-Presentation in Litigation Science*, 116 ENVTL. HEALTH PERSP. 123, 125 (2008) (arguing that *Daubert* rests on a flawed assumption that criteria external to the legal process can guide judges in their screening of scientific evidence).

31 *Id.* at 128.

compiled and summarized—it is too late to detect or resolve the kind of unreliability that is caused by analytical flexibility.[32]

The critical view, however, also requires modification. Sheila Jasanoff, a leading voice in science and technology studies, has argued that in evaluating scientific evidence, courts should focus less on the end result and more on the process of its creation as an indicator of reliability.[33] While assigning courts a monitoring role would be a step in the right direction, it is insufficient to prevent or solve the problems that have been exposed by the Replication Crisis.[34] Reliable, reproducible litigation science requires courts not merely to assume a supervisory role with respect to its production but to create the settings, conditions, and culture necessary to enable party experts to produce it. Instead of waiting by the gate for the evidence to arrive in its full-fledged form, courts should ride out and meet evidence creators as soon as they appear on the horizon, engaging with the creation of the scientific evidence at an early stage and guarding the process of its creation.

More broadly, this Article aims to spark a discussion about methodological standards in litigation science and how these standards interact with prevailing evidentiary standards. The stakes are high. Evidentiary standards "determine important issues such as who gets to trial in the first place [and] which verdicts will be allowed to stand."[35] Scientific evidence plays a large role in many civil actions, and a large number of important cases—particularly those involving toxic torts and environmental damage—cannot be brought at all without the support of scientific studies conducted by expert witnesses.[36] If courts and litigants fail to absorb the lessons of the Replication Crisis, practices that are now understood to inject unreliability into the creation of scientific data will continue to contaminate litigation science in significant and largely undetectable ways. The just outcome of civil actions that depend on the accuracy and reliability of scientific evidence demands an understanding of the misconceptions and practices that engen-

---

32 *See infra* subparts II.A and III.B.

33 Jasanoff, *supra* note 30, at 129.

34 *See infra* subparts III.B and IV.A.

35 Michael S. Pardo, *The Nature and Purpose of Evidence Theory*, 66 VAND. L. REV. 547, 554 (2013).

36 *See, e.g.*, Margaret A. Berger & Aaron D. Twerski, *Uncertainty and Informed Choice: Unmasking* Daubert, 104 MICH. L. REV. 257, 267 (2005) (arguing that as standards of admissibility of scientific evidence have become more exacting, it has become "nigh impossible" in certain toxic torts cases for plaintiffs to maintain a civil action).

dered the Replication Crisis and a reconceptualization of the court's role in the creation and evaluation of litigation science.

The remainder of this Article is organized as follows. Part I reviews the frameworks for the creation and evaluation of litigation science as they are currently applied in federal courts and the courts of most states, as well as the implicit assumptions embedded in them. Part II describes the Replication Crisis that swept through the social sciences and the reforms and changed norms that are garnering growing acceptance in those scientific communities. Part III argues that the problems uncovered in social science research are likely to be at least as prevalent in litigation science and draws on social-science responses to the Crisis to propose reforms for the creation and presentation of scientific evidence in civil litigation. Part IV considers the implications of the proposed measures on the creation of litigation science, including how they reframe the role of judges and advocates in the expert-discovery phase of a litigation and improve judges' ability to fulfill their gatekeeping obligation to keep unreliable evidence from reaching the finder of fact.

I
SCIENTIFIC EVIDENCE IN CIVIL LITIGATION

Testifying expert witnesses play a central role in the U.S. litigation system. They are often thought of as interpreters of evidence: they present and explain technical, scientific, or other specialized matter that the finder of fact might find difficult to interpret without an expert's guidance.[37] But experts very frequently are also *creators* of evidence. They routinely create evidence in support of or against a party's claims or defenses, by conducting experiments and analyses tailor-made for the case at hand.[38] The studies that experts can conduct

---

37 *See, e.g.*, FED. R. EVID. 702(a) (An expert can testify if "the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue."); CAL. EVID. CODE § 801(a) (An expert can testify on topics "sufficiently beyond common experience that the opinion of an expert would assist the trier of fact.").

38 *See, e.g.*, Anthony Champagne, Daniel Shuman & Elizabeth Whitaker, *An Empirical Examination of the Use of Expert Witnesses in American Courts*, 31 JURIMETRICS J. 375, 381 (1991) (reporting results of a three-month study of civil cases in Dallas, Texas, and finding that experts, including experimental psychologists, economists, biochemists, and engineers, were used in more than half of cases); Samuel R. Gross, *Expert Evidence*, 1991 WIS. L. REV. 1113, 1119 (analyzing data on 529 civil jury trials in California State Superior Courts from 1985 and 1986 and finding that experts testified in eighty-six percent of those trials); Andrew W. Jurs, *Expert Prevalence, Persuasion, and Price: What Trial Participants*

are too numerous to list. Experts can examine soil or water samples in connection with environmental tort cases, analyze financial data relevant to securities litigation, dissect music samples in connection with copyright disputes—the list is endless.[39] In some areas of litigation, it is not unusual for a plaintiff's claims to rely heavily, or even primarily, on litigation science.[40]

Expert evidence is not automatically admissible in state or federal court. To promote just determinations, evidentiary rules put limits on the types and scope of expert testimony that can be heard.[41] Before an expert can provide testimony in a civil case, the party proffering the expert must justify why this outsider—an individual with no inherent connection to the case and no independent knowledge of the facts at issue—should be heard as part of the proceeding.[42] In determining whether to admit expert testimony, courts apply jurisdiction-dependent admissibility criteria to an expert's credentials and proposed testimony. *First*, as with any other type of evidence, expert testimony must be relevant to be admissible.[43] If it is not relevant to the claims or defenses in the case, it cannot be admitted.[44] *Second*, for testimony to be admissible, it must be

---

Really Think About Experts, 91 IND. L.J. 353, 355 (2016) (stating that expert witnesses appeared in 86% of civil trials in an urban Iowa county, in line with results in earlier studies in different locales); Robert J. Shaughnessy, *Dirty Little Secrets of Expert Testimony*, 33 LITIG. 47, 52 (2007) ("In large cases, it is not unheard of for one side to designate 20 or more experts . . . .").

39 *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, 1022 (C.D. Cal. 2013) (involving analysis of mortgage loans in support of securities litigation); Swirsky v. Carey, 376 F.3d 841, 845 (9th Cir. 2004) (involving analysis of lyrics and melodies in support of copyright infringement action); 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).

40 *See, e.g.*, Berger & Twerski, *supra* note 36 (asserting that certain environmental tort cases for plaintiffs cannot be brought without expert evidence).

41 *See, e.g.*, FED. R. EVID. 102 (Federal Rules of Evidence "should be construed . . . to the end of ascertaining the truth and securing a just determination.").

42 *See, e.g.*, FED. R. EVID. 702 (providing requirements for testimony by expert witness in federal court).

43 *See, e.g.*, FED. R. EVID. 402 (providing that relevant evidence is admissible and irrelevant evidence is not admissible); CAL. EVID. CODE § 350 ("No evidence is admissible except relevant evidence.").

44 *See* FED. R. EVID. 402. For views on what it means for evidence to be relevant, see, for example, Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1025–26 (1977) (arguing that evidence is relevant if it tends to make a fact more or less probable than it would be without the evidence); Pardo, *supra* note 35, at 558 (arguing that evidence is relevant if it supports a theory of the

sufficiently reliable: (a) the expert proffering the testimony must be qualified to testify regarding the subject matter,[45] and (b) the analysis or methodology that forms the basis for the testimony must meet certain jurisdiction-dependent standards of reliability.[46] A typical third requirement is that the testimony be helpful to the trier of fact.[47] The focus of this Article is on the reliability determination.[48]

The most widely applied standard of reliability of scientific or technical evidence is the one articulated by the U.S. Supreme Court in *Daubert v. Merrill Dow Pharmaceuticals, Inc.* and since codified in a revision of Federal Rule of Evidence 702.[49] When evaluating the admissibility of expert testimony

---

case, even if it does not make that theory more likely than a competing theory); Michael S. Pardo & Ronald J. Allen, *Juridical Proof and the Best Explanation*, 27 L. & PHIL. 223, 224 & n.2 (2008) (presenting probability-based conceptions of relevance); *cf.* FED. R. EVID. 401(a) (providing that evidence is relevant if "it has any tendency to make a fact [of consequence in determining the action] more or less probable than it would be without the evidence").

45 *See, e.g.*, FED. R. EVID. 702 (providing that experts are witnesses who are "qualified as an expert by knowledge, skill, experience, training, or education"); DEL. R. EVID. 702 (same); ILL. R. EVID. 702 (same); N.J. R. EVID. 702 (same); *see also* CAL. EVID. CODE § 720(a) ("A person is qualified to testify as an expert if he has special knowledge, skill, experience, training, or education sufficient to qualify him as an expert on the subject to which his testimony relates.").

46 *See, e.g.*, FED. R. EVID. 702 (providing that expert testimony must be "based on sufficient facts or data," and "the product of reliable principles and methods" that have been "reliably applied . . . to the facts of the case"); DEL. R. EVID. 702 (same); *see also* CAL. EVID. CODE § 801(b) (providing an expert may testify "[b]ased on matter . . . that is of a type that reasonably may be relied upon by an expert in forming an opinion upon the subject to which his testimony relates").

47 *See, e.g.*, FED. R. EVID. 702(a) (providing that an expert may testify if "the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue"); DEL. R. EVID. 702(a) (same); ILL. R. EVID. 702 (same); N.J. R. EVID. 702 (same); *see also* CAL. EVID. CODE § 801(a) (providing that expert testimony is limited to opinions "[r]elated to a subject that is sufficiently beyond common experience that the opinion of an expert would assist the trier of fact").

48 The reliability of scientific evidence introduced in criminal cases is often attacked on the basis that the methodology that produced it has not been validated. *See, e.g.*, COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY., NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES 87 (2009), https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf [https://perma.cc/Y7V8-KB93] (offering recommendations to improve the reliability of forensic evidence); PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS 44 n.94 (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf [https://perma.cc/6FES-UMXT] (same). The scope of this Article is limited to civil proceedings, where such arguments feature much less prominently.

49 *See* FED. R. EVID. 702 & advisory committee note, *reprinted in* 28 U.S.C. at 398–401; Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 592–94 (1993).

under the *Daubert* standard, a court is to assess whether "the reasoning or methodology underlying the testimony is scientifically valid," as well as "whether that reasoning or methodology properly can be applied to the facts in issue."[50] Construing the reliability inquiry to be undertaken as "a flexible one," the *Daubert* Court steered clear of providing a "definitive checklist or test" for performing this evaluation, but offered guidance in the form of four considerations: (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."[51]

The *Daubert* decision thrust the court into the role of "gatekeeper" tasked with keeping out expert testimony that does not meet these requirements.[52] Courts exercise their gatekeeper function when ruling on so-called "*Daubert* motions" and determine whether proffered expert testimony meets the standard of reliability.[53] Although the Federal Rules of Evidence and *Daubert* do not prescribe the point in time at which these hearings are to take place, they typically take the form of an in limine motion after expert discovery has been completed (i.e., shortly before or in conjunction with a motion for summary judgment, in the lead-up to trial) or even during trial, before the challenged expert is called to the stand to testify.[54] The *Daubert* standard is observed in federal courts as well as in the courts of the more than forty states that have adopted it.[55] Even in states that have not adopted the *Daubert* standard, pretrial proceedings on the admissibility of expert testimony have become commonplace.[56]

---

50 *Daubert*, 509 U.S. at 600. Since 2000, these requirements are codified in FED. R. EVID. 702(b)–(d).

51 *Daubert*, 509 U.S. at 593–94.

52 *See id.* at 589, 590–91 n.9.

53 *See generally* JAMES C. COOPER, TIMING AND DISPOSITION OF DAUBERT MOTIONS IN FEDERAL DISTRICT COURTS: AN EMPIRICAL EXAMINATION 1 (2015), https://masonlec.org/site/rte_uploads/files/Daubert%20Report%5B1%5D.pdf [https://perma.cc/YXK6-LSEP] (analyzing the effect of *Daubert* on litigation practice).

54 *See id.* at 9, 10 tbl.6 (reporting the median time from case commencement to *Daubert* motion of 647 days in federal courts, with 74% of *Daubert* motions filed prior to or with a summary judgment motion and 26% after summary judgment and before or during trial).

55 *See, e.g.*, THOMAS A. MAUET & WARREN D. WOLFSON, TRIAL EVIDENCE 275 (7th ed. 2020) ("Thus far, more than 40 states have adopted *Daubert*.").

56 *See, e.g.*, Margaret A. Berger, *What Has a Decade of* Daubert *Wrought?*, 95 AM. J. PUB. HEALTH S59, S61 (2005) ("Even in jurisdictions that purportedly follow

The admission or exclusion of expert evidence can alter the course of a case dramatically. When a plaintiff's claims lean heavily on the outcome of an expert's analysis, it is not hyperbole to say that the admissibility or inadmissibility of that opinion can make or break a case. When an element of the plaintiff's claim rests entirely on expert evidence, the case crumbles when the expert is barred from testifying or when the court substantially limits the scope of the expert's testimony.[57] A ruling of inadmissibility can alter the outcome of a trial or determine whether the case will survive summary judgment.[58] Further, having expert evidence in play also affects the parties' settlement power: the greater a party's confidence that its expert will be permitted to testify at trial, the stronger a position it can take in settlement negotiations.

The procedural mechanics of expert discovery differ from jurisdiction to jurisdiction,[59] and default requirements are often supplemented by agreements between the parties or through case-management orders issued by the court.[60] While the details thus vary from court to court and from case to case,

---

the [older] *Frye* 'general acceptance' test, judges are citing and analyzing *Daubert* and its progeny . . . ."); David E. Bernstein, Frye, Frye, *Again: The Past, Present, and Future of the General Acceptance Test*, 41 JURIMETRICS 385, 388 (2001) ("[C]ase law under *Frye* is slowly converging with *Daubert* jurisprudence."). While standards of relevance differ from jurisdiction to jurisdiction, I am not aware of any court of general jurisdiction within the United States where expert testimony is admitted without regards to reliability.

57 *See* Weisgram v. Marley Co., 528 U.S. 440, 455–56 (2000) (holding that when an expert is excluded shortly before trial, plaintiffs do not have a right to salvage their case by substituting a new expert).

58 *See id.*; Pardo, *supra* note 35 ("[E]videntiary rules and standards . . . determine important issues such as who gets to trial in the first place, which verdicts will be allowed to stand, and which convictions will be overturned."); *see also* Berger, *supra* note 56, at S64 (arguing that *Daubert* has "made it more difficult for plaintiffs to litigate successfully"); Nat'l Ctr. for State Courts, *The Changing Role of Judges in the Admissibility of Expert Evidence*, 5 CIV. ACTION 1, 3 (2006) (concluding that the *Daubert* standard "appears to . . . reduc[e] the number of cases that 'survive' *Daubert* challenges and result in a summary judgment and encourag[e] the defense to settle if their challenge is not successful").

59 *See, e.g.*, FED. R. CIV. P. 26(a)(2) (requiring disclosure of the expert's identity and expert's written report at a specified time); CAL. CODE CIV. P. § 2034.260 (requiring exchange of rudimentary information about an expert and the expert's expected testimony, but not an exchange of expert reports); TEX. R. CIV. P. 166(i), 192.3(e), 195 (providing detailed description of disclosure requirements relating to expert discovery).

60 *See, e.g.*, FED. R. CIV. P. 16(b)(3)(B) (providing that a scheduling order may modify the timing and nature of disclosures and the scope of discovery); *Sample Expert Discovery Stipulation*, DEL. CTS., https://courts.delaware.gov/chancery/docs/Sample_Expert_Discovery_Stipulation1.pdf [https://perma.cc/R9Y4-VPV7] (sample expert-discovery stipulation modifying statutory disclosure requirements).

the road to presentation of expert-generated scientific evidence at trial typically proceeds as follows.[61] First, an expert is retained by a party, and starts work on an analysis or project. At some point during the discovery period, the party's retention of the expert and its intention to call the expert as a witness at trial are disclosed to opposing parties.[62] After the expert has completed her analysis, an expert report is served on opposing parties.[63] Expert reports generally contain all of the opinions the expert intends to present at trial, along with all the bases for those opinions.[64] If any opinions are based on a study or analysis conducted by the expert, the report typically includes those findings that the expert considers part of the basis for her opinions, along with enough information about the expert's project to allow a reader to understand the steps that were taken to arrive at the reported results. Next, opposing parties have an opportunity to depose the expert. Finally, if a party chooses to challenge the admissibility of (some of) the expert's opinions into evidence, it files a motion to exclude or limit the expert's testimony. Following briefing on the motion and, typically, an evidentiary hearing, the court makes a determination on the expert's qualifications, as well as the reliability, relevance, and helpfulness of the expert's proposed testimony, and the testimony is either admitted, limited, or excluded.

Daubert and other standards of reliability rely on an implicit assumption that, when the time comes to assess the reliability of scientific evidence, the information necessary to make that assessment will be available to the court—i.e., an assump-

---

61 The following characterization of the expert-discovery process is based on the author's experience in eight years of litigation practice.

62 Expert disclosures can happen by rule-based or court-imposed deadlines, by deadlines agreed upon between the parties in discovery stipulations, or in response to interrogatories.

63 Some jurisdictions do not require the exchange of expert reports. *See, e.g.*, Beck v. Hirchag, No. G041955, 2011 Cal. App. Unpub. LEXIS 2649, at *15 (Cal. Ct. App. Apr. 11, 2011) (acknowledging that California procedure does not require an expert to prepare an expert report and "practice guides recommend attorneys 'instruct their experts *not* to prepare a formal report'"). It is not uncommon in these jurisdictions for parties to agree to exchange expert reports regardless. *See supra* note 60.

64 *See, e.g.*, FED. R. CIV. P. 26(a)(2)(B)(i)–(ii) (providing that an expert report must contain "a complete statement of all opinions the [expert] witness will express and the basis and reasons for them"); N.Y. C.P.L.R. 3101(a)(4)(d)(1) (McKinney 2014) (requiring the proffering party to disclose "the substance of the facts and opinions on which each expert is expected to testify . . . and a summary of the grounds for each expert's opinion"); TEX. R. CIV. P. 166(i), 192.3(e), 194.2(f), 195.1 (requiring disclosure of an expert's opinions, the expert's bases for those opinions, and the facts that "relate to or form the basis of" those opinions and limiting expert discovery beyond the substance explicitly permitted or required by rule).

tion that the necessary information can be obtained from reports submitted by the expert or through deposition or hearing testimony. As a discussion of the Replication Crisis and its implications for litigation science will make clear, this assumption is incorrect.[65] Experts typically do not disclose their hypotheses and methodology until they submit a report in which they present the results of their completed study, generally toward the end of the expert-discovery period. And while the typical report describes the steps that were taken to arrive at the reported results, it usually does not describe at what point in time the expert settled on a set of hypotheses, when exactly the research protocol that led to the findings was designed, and to what extent the protocol was modified or supplemented along the way.[66] Ordinarily, the report omits experimental results that do not support the expert's opinions, as well as research steps that did not end up leading to the results that are being reported.

As will be discussed in Parts II and III of this Article, the information included in the typical expert report does not permit the court to assess whether the expert's findings meet the required standard of reliability. There is a dimension of unreliability that is missed when reliability is assessed based exclusively on expert reports and testimony, and by the time the reliability of the expert's work is up for adjudication, it is too late to recover this dimension.

II
THE KNOWLEDGE CRISIS IN THE SOCIAL SCIENCES

In 2018, psychologists Nelson, Simmons, and Simonsohn noted that "[i]f a team of research psychologists were to emerge today from a 7-year hibernation, they would not recognize their field."[67] In less than a decade, the field had (a) learned that a majority of its research results could not and should not be relied upon; (b) figured out that this unreliability could be blamed on a number of commonly used research practices; (c) devised practices and standards to improve the quality and reproducibility of research results; and (d) had these updated practices and standards adopted by a large and growing group

---

65 *See infra* subparts II.A and III.A.

66 *See, e.g.*, FED. R. CIV. P. 26(a)(2)(B)(i)–(ii) (providing that an expert report must include "all opinions the witness will express and the basis and reasons for them" and "the facts or data considered by the witness in forming them").

67 Nelson, Simmons & Simonsohn, *supra* note 15, at 512.

of researchers, research centers, journals, and organizations.[68] The crisis of confidence that shook the field in the early 2010s and that precipitated this overhaul of research practices is commonly referred to as the "Replication Crisis."[69]

While the Replication Crisis originated in psychology research, other fields soon followed. Economics, biology, pharmacology, and other fields each in turn discovered (or in some cases rediscovered) that their research methodologies were prone to the same type of unreliability that had triggered alarm bells in the field of psychology.[70] In recognition of psychology's frontrunner position in uncovering and addressing unreliable research practices, psychologists have started referring to this period of reconstruction as "Psychology's Renaissance,"[71] but in the past few years, each of these fields has engaged in efforts to shore up the rigor of its research practices, and the number of organizations adopting or recommending updated standards is still growing.[72]

This Part describes the events that heralded the Replication Crisis, the research practices that stood at its center, and the lessons that the social sciences have drawn from it. Subpart II.A presents a brief chronology of the Crisis, how its magnitude gradually came into focus, and how it spread like a wildfire from field to field. Subpart II.B describes the "Renaissance" period that followed the early Crisis years. It reviews the causes of unreliability that were uncovered during the Crisis

---

68 *See id.*

69 *See, e.g.*, *id.* ("Many have been referring to this period as psychology's 'replication crisis.'"); 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 [https://perma.cc/X6LW-CKBV] (arguing that the "Replication Crisis" is an opportunity for the field of psychology to "lead the way"); 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 [https://perma.cc/R9M8-AA7D] ("Science is mired in a 'replication' crisis."); 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/J75C-HPJ6] ("[I]t has become painfully clear that psychology is facing a 'reproducibility crisis.'").

70 *See infra* notes 93–97; *see also infra* note 125 (noting that certain practices that are now the subject of renewed interest had been acknowledged previously in economics, psychology, and other fields).

71 *See, e.g.*, Nelson, Simmons & Simonsohn, *supra* note 15, at 512 ("Many have been referring to this period as psychology's 'replication crisis.' This makes no sense. We do not call the rain that follows a long drought a water crisis. . . . This is psychology's *renaissance*."); ELI J. FINKEL & ROY F. BAUMEISTER, *Social Psychology: Crisis and Renaissance*, *in* ADVANCED SOCIAL PSYCHOLOGY: THE STATE OF THE SCIENCE 1, 1 (Eli J. Finkel & Roy F. Baumeister eds., 2d ed. 2019).

72 *See infra* notes 93–97.

period and some of the reforms that have been proposed and implemented in response.

A. Replication Crisis

Daryl J. Bem, the social psychologist who demonstrated that people can "[f]eel[ ] the [f]uture,"[73] was a well-respected researcher with a decades-long career of research and publication.[74] The journal in which he published his 2011 study is a peer-reviewed, high-impact journal in his field.[75] As one might expect to happen with a study purporting to show that the future can change the past, Bem's study was met with skepticism. Researchers delved into the data underlying the reported results and almost immediately identified a number of flaws, such as defects in Bem's calculation of statistical significance,[76] improper modifications to experimental procedures,[77] and incorrect experimental design.[78] Researchers who reanalyzed the data[79] or attempted to replicate Bem's work[80] found no support for his findings. Around the same time, another high-profile psychology study failed to replicate,[81] and in the same year, a team of psychologists published a paper demonstrating that large numbers of psychology researchers were

---

73 *See* Bem, *supra* note 1, at 407.

74 *See* James E. Alcock, *Back from the Future: Parapsychology and the Bem Affair*, 35 SKEPTICAL INQUIRER 31, 31 (2011) (describing Bem's study as particularly newsworthy because of the "academic stature of its author"); DARYL J. BEM (Dec. 25, 2014), https://dbem.org [https://perma.cc/M7GE-CN68].

75 *See* Nelson, Simmons & Simonsohn, *supra* note 15, at 513 (*The Journal of Personality and Social Psychology* is "[s]ocial psychology's most prestigious journal.").

76 *See* Wagenmakers, Wetzels, Borsboom & van der Maas, *supra* note 5, at 428.

77 *See* Alcock, *supra* note 74, at 33.

78 *See* Ulrich Schimmack, *The Ironic Effect of Significant Results on the Credibility of Multiple-Study Articles*, 17 PSYCHOL. METHODS 551, 556–59 (2012).

79 *See* Jeffrey N. Rouder & Richard D. Morey, *A Bayes Factor Meta-Analysis of Bem's ESP Claim*, 18 PSYCHONOMIC BULL. REV. 682, 683 (2011).

80 *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*, 7 PLOS ONE e33423, at 3 (2012) (reporting three failed replication attempts).

81 *See* Stéphane Doyen, Olivier Klein, Cora-Lise Pichon & Axel Cleeremans, *Behavioral Priming: It's All in the Mind, but Whose Mind?*, 7 PLOS ONE e29081, at 5–6 (2012) (reporting a failed attempt to replicate a celebrated age-priming study); Nelson, Simmons & Simonsohn, *supra* note 15, at 513–14 (describing the Doyen, Klein, Pichon, and Cleeremans study as one of the events that precipitated the Replication Crisis).

in the habit of using what they termed "questionable research practices" that undermine the reliability of their findings.[82]

Prominent researchers issued calls for large-scale replication efforts, to assess how widespread the use of unreliable methodology was and how consequential its use.[83] In one large-scale project, researchers selected 100 papers that had been widely cited and relied upon in the field of psychology and coordinated systematic replication efforts by 100 different labs.[84] Teams attempting to replicate the original studies aimed to reproduce the original testing conditions as closely as possible. To this end, they requested (and in almost all cases received) the original materials from the original authors,[85] and in most cases obtained the original authors' endorsement of the replication attempt.[86] Only 39 of the 100 studies replicated successfully.[87] To rule out the possibility that the low replication rate had been caused by inartfully conducted replication attempts,[88] a "Many Labs 2" project was designed in a manner

---

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

83 *See Kahneman on the Storm of Doubts Surrounding Social Priming Research*, DECISION SCI. NEWS (Oct. 5, 2012), http://www.decisionsciencenews.com/2012/10/05/kahneman-on-the-storm-of-doubts-surrounding-social-priming-research/ [https://perma.cc/M3GJ-8YHB] (quoting email by Daniel Kahneman, dated September 26, 2012, in which he warns of "a train wreck looming" and proposes a systematic replication project).

84 Open Sci. Collaboration, *Estimating the Reproducibility of Psychological Science*, 349 SCIENCE aac4716-1, aac4716-1 (2015).

85 *See* Open Sci. Collaboration, *Supplementary Material for* Estimating the Reproducibility of Psychological Science 4, SCIENCE (Aug. 28, 2015), https://science.sciencemag.org/content/sci/suppl/2015/08/26/349.6251.aac4716.DC1/Aarts-SM.pdf [https://perma.cc/89GD-CGGP] (reporting that 89 teams were able to obtain original materials).

86 *See id.* (reporting that 69 original teams endorsed the replication attempts).

87 *See id.* at 17. Note that it was not expected that all 100 studies would yield consistent results on replication. Given the variability reported in the original studies, the a priori expectation was that approximately 89 studies would replicate. *See id.* at 4.

88 For criticism of early replication projects, see, for example, Daniel T. Gilbert, Gary King, Stephen Pettigrew & Timothy D. Wilson, *Comment on "Estimating the Reproducibility of Psychological Science"*, 351 SCIENCE 1037-b, 1037-b (2016) (criticizing sample size, experimental setup, and analysis employed). For responses to these critiques, see *Comment on "Estimating the Reproducibility of Psychological Science,"* PUBPEER, https://pubpeer.com/publications/120FE2AC75B4C873787611246291A8 [https://perma.cc/BK4G-4B92] (last visited May 27, 2020); Daniël Lakens, *The Statistical Conclusions in Gilbert et al. (2016) Are Completely Invalid*, 20% STATISTICIAN (Mar. 6, 2016), https://daniellakens.blogspot.com/2016/03/the-statistical-conclusions-in-gilbert.html [https://perma.cc/B94V-B62V].

intended to address early criticisms.[89] As part of this project, 36 labs in 36 countries collaborated to rerun a set of 28 well-known studies.[90] Only 15 of the 28 selected studies were rated to have replicated.[91] "Ironically enough," one commentator noted, "it seems that one of the most reliable findings in psychology is that only half of psychological studies can be successfully repeated."[92]

Similar failures to replicate were found in economics,[93] biology and biomedicine,[94] pharmacology,[95] neuro-imaging,[96] and other fields.[97] A sizable methodological meta-field sprang up that aimed to make sense of these dismal replication rates, focusing both on the research practices that evidently risked producing irreproducible results and the cultural and institutional incentives that led to their use.[98] The emphasis on pub-

---

89 *See* Richard A. Klein et al., *Many Labs 2: Investigating Variation in Replicability Across Samples and Settings*, 1 ADVANCES METHODS & PRACTICES PSYCHOL. SCI. 443, 446 (2018).

90 *See id.*

91 *See id.* at 477.

92 Yong, *supra* note 69.

93 *See* Camerer et al., *supra* note 14 (successfully replicating 11 of 18 economics studies); 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), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2669564 [https://perma.cc/UN2F-9PWW] (successfully replicating 29 of 59 economics studies).

94 *See, e.g.*, C. Glenn Begley & Lee M. Ellis, *Raise Standards for Preclinical Cancer Research*, 483 NATURE 531, 532 (2012) (reporting that the Amgen team succeeded in confirming results for only 6 of 53 hematology and oncology studies); *see also* Marcus R. Munafò et al., *A Manifesto for Reproducible Science*, 1 NATURE HUM. BEHAV. 1, 1 (2017) ("85% of biomedical research efforts are wasted . . . ."); Open Sci. Collaboration, *supra* note 84 (describing two molecular biology replication studies that had a replication rate of 11% and 25%, respectively).

95 *See* Christopher H. George et al., *Updating the Guidelines for Data Transparency in the British Journal of Pharmacology—Data Sharing and the Use of Scatter Plots Instead of Bar Charts*, 174 BRIT. J. PHARMACOLOGY 2801, 2801 (2017) (reporting mixed results of a series of replication attempts in pharmacology).

96 *See* Russell A. Poldrack et al., *Scanning the Horizon: Towards Transparent and Reproducible Neuroimaging Research*, 18 NATURE REVIEWS: NEUROSCIENCE 115, 115 (2017) (reporting that neuroimaging has replication problems similar to those found in other scientific fields).

97 *See, e.g.*, Head, Holman, Lanfear, Kahn & Jennions, *supra* note 15 (examining published study results across disciplines including cognitive sciences, information and computing sciences, and medical and health sciences to conclude that reproducibility concerns are "rife").

98 *See, e.g.*, Joseph P. Simmons, Leif D. Nelson & Uri Simonsohn, *False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant*, 22 PSYCHOL. SCI. 1359, 1359 (2011) (demonstrating how common research tools can generate unreliable findings); *see also* Open Sci. Collaboration, *Maximizing the Reproducibility of Your Research*, *in* PSYCHOLOGICAL SCIENCE UNDER SCRUTINY 3, 3 (Scott O. Lilienfeld & Irwin D. Waldman

lication of statistically significant findings was identified as a major culprit.[99]

Statistical significance has long been the holy grail for researchers aiming to publish their results. Traditionally, many scientific journals would not publish results that did not meet threshold levels of statistical significance.[100] This put pressure on researchers to generate results that would meet those requirements.[101] As early as 2011, Joseph P. Simmons and coauthors demonstrated how "unacceptably easy" it is to find statistically significant evidence for any hypothesis, whether true or false, through the use of everyday analytical methods.[102] Subsequent work confirmed that these common practices, employed in pursuit of statistical significance, lay at the heart of the Replication Crisis.[103] Passed on from generation to generation in social science labs and the labs of adjacent disciplines, they had endowed researchers with the flexibility they needed to squeeze maximum results from limited data, generate statistically significant conclusions where there should not have been any, and thereby, wittingly or unwittingly, achieve and publish results that could not be relied upon.[104]

These practices included:[105]

---

ed., 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 PERSP. ON PSYCHOL. SCI. 615, 616 (2012) (arguing for institutional change to improve reliability of social science research).

99 *See, e.g.*, Nosek, Spies & Motyl, *supra* note 98, at 617.

100 *See id.* ("[T]he nominal false-positive rate of a $p < 0.05$ has become a de facto criterion for publishing.").

101 *See id.*

102 *See* Simmons, Nelson & Simonsohn, *supra* note 98, at 1359–60 (using standard research methodology to prove that listening to the Beatles' *When I'm Sixty-Four* shaved a year and a half off participants' age).

103 *See, e.g.*, John, Loewenstein & Prelec, *supra* note 82 ("[Q]uestionable practices may constitute the prevailing research norm.").

104 *See id.*; Marjan Bakker, Annette van Dijk & Jelte M. Wicherts, *The Rules of the Game Called Psychological Science*, 7 PERSP. ON PSYCHOL. SCI. 543, 547 (2012) (approaches involving multiple small, underpowered studies are "the optimal strategy" for manufacturing statistically significant results where there is no true effect); Nelson, Simmons & Simonsohn, *supra* note 15, at 515 (arguing that *p*-hacking is "the only honest and practical way to consistently get underpowered studies to be statistically significant" (emphasis omitted)).

105 *See generally* Head, Holman, Lanfear, Kahn & Jennions, *supra* note 15, at 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); Munafò et al., *supra* note 94 (identifying apophenia, confirmation bias, and hindsight bias as contributors to flawed data analysis); Simmons, Nelson & Simonsohn, *supra* note 98 (demonstrating how "unacceptably easy it is to accumulate (and report) statis-

1. failing to commit to a hypothesis before the start of the study, instead leaving the researcher free to reverse-engineer a hypothesis after collecting and analyzing the data;
2. failing to commit to an overall research protocol before the start of the study, instead leaving the researcher free to choose the exact sequence of steps sometime during or after data collection;
3. failing to commit to a data-gathering protocol before the start of the study, instead leaving the researcher free to examine and take into account incoming data in deciding whether to stop data collection, collect additional data, or make changes to the way data are collected;
4. failing to commit to a protocol for the cleaning up of data—including criteria for the removal of outlier data and for handling missing or corrupted data—leaving the researcher free to try different approaches until identifying one that results in a desirable outcome;
5. tweaking the methodology used to analyze the data—a potentially infinite number of times—until it results in a desirable outcome;
6. performing regressions and other correlative analyses with a potentially infinite number of variables and variations, until hitting on a desirable outcome; and
7. testing a wide variety of variables and reporting only the ones that show a desirable outcome.

Practice one is referred to as "hypothesizing after the results are known" or "**HARKing**."[106] Practices two through seven are forms of a practice now known as "***p*-hacking**,"[107] because they aim to create a *p*-value—the most commonly used indicator of statistical significance—in a range that will allow the research results to be published.[108] Practice seven

---

tically significant evidence for a false hypothesis"). The list provided here is not intended to be exhaustive. Other practices criticized for impairing reliability include the use of small samples and plain fraud. *See, e.g.*, Bakker, van Dijk & Wicherts, *supra* note 104; Gelman, *supra* note 69.

106 *See* Munafò et al., *supra* note 94, at 2.

107 *P*-hacking is the intentional or unintentional use of "data-contingent analysis decisions," typically aimed at manufacturing statistically significant results. *Id.* at 3.

108 A *p*-value expresses as a number between 0 and 1 the likelihood of finding the observed outcome (or an outcome farther removed from the null hypothesis) in the event that the null hypothesis is true. In other words, it expresses the likelihood that there is no effect and that the measured outcome came about by pure chance. The lower the *p*-value, the stronger support the measurement lends to the tested hypothesis. *P*-values are frequently misunderstood, even by scientists who use them. For an overview of twenty-five common misconceptions of *p*-values and other measures of confidence, 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).

has been termed "publication bias" or, when it causes entire research projects to remain unpublished and stored away in metaphorical file drawers, the "**file drawer problem**."[109]

What these practices have in common is that they impart an undue level of flexibility on a researcher.[110] In this Article I use the term "Analytical Flexibility" to refer to a researcher's range of freedom to adapt or even completely make up a research protocol while the research is already underway.[111] Even small measures of Analytical Flexibility, such as the ability to add a few more data points or the use of different variables and control variables, can have large consequences[112] by raising significantly 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.[113] When used opportunistically, Analytical Flexibility virtually guarantees a researcher statistically significant (and therefore publishable) evidence for a false hypothesis.[114]

It is worth noting that this type of flexibility has been shown to be harmful even in the hands of an honest, well-intentioned researcher.[115] To understand why this is the case, imagine such a researcher setting to work analyzing a set of data, choosing an initial approach that she, in good faith, believes to be the most suitable approach. If the researcher's first choice does not result in a publishable result, even a well-

---

109 *See* Munafò et al., *supra* note 94, at 3.

110 *See generally* Head, Holman, Lanfear, Kahn & Jennions, *supra* note 15, at 2, 8 (arguing that *p*-hacking is widespread and harmful); John, Lowenstein & Prelec, *supra* note 82 (terming methodologies capitalizing on analytical flexibility "questionable research practices"); Munafò et al., *supra* note 94 (arguing that research efforts are "wasted" when they do not yield reliable results); Ioannidis, *supra* note 105, 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, Nelson & Simonsohn, *supra* note 98 ("[F]lexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates.").

111 This flexibility has also been termed "researcher degrees of freedom." *See, e.g.*, Simmons, Nelson & Simonsohn, *supra* note 98 (emphasis omitted).

112 For example, the mere flexibility to choose to add ten data points to an initially collected data set raises the false-positive rate by 50 percent. *Id.* at 1361.

113 *See id.* at 1359 ("Perhaps the most costly error is a *false positive* . . . .").

114 *See id.* at 1361 (finding that mere 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).

115 *See* Nelson, Simmons & Simonsohn, *supra* note 15, 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." (citation omitted)).

intentioned researcher is likely to try a number of alternative statistical methods.[116] If any of these methods renders a result that is *close* to statistically significant, she might be tempted to run the analysis a second time after gathering a few more data points, to see if these points confirm the earlier results.[117] All of these practices have traditionally been exceedingly common in the social sciences.[118] And all of these practices contribute to unreliability (i.e., irreproducibility) of research results.[119] While a research journey can be unpredictable, and adjustments are sometimes necessary, it is now understood that constraining Analytical Flexibility to the extent possible is essential to the creation of reliable, reproducible research.[120]

The practices listed above have another important feature in common: unless a researcher reporting research results is held to stringent recording and disclosure requirements, these practices tend to be invisible. When Analytical Flexibility that was present in the study is neither explicitly eliminated nor documented, there is typically no way to assess the extent to which its presence may have contaminated or even invalidated the research results. Consider as an example two scenarios involving a researcher studying the link between acne and jelly bean consumption.[121] In Scenario A, the researcher has a hunch that green jelly beans cause acne. He performs a study using only green jelly beans and finds a statistically significant positive correlation ($p<0.05$) between consumption of green jelly beans and incidence of acne.[122] In Scenario B, the researcher performs 20 separate studies, each examining the relationship between acne and consumption of a specific color

---

116 *See id.* (*P*-hacking is "something that benevolent researchers engage in while trying to understand their otherwise imperfect results.").

117 *See* Simmons, Nelson & Simonsohn, *supra* note 98, at 1362.

118 *See* Head, Holman, Lanfear, Kahn & Jennions, *supra* note 15 (analyzing published study results to conclude that "*p*-hacking is rife").

119 *See* Nelson, Simmons & Simonsohn, *supra* note 15, at 512 (finding that researchers "rely[ ] on methods of data collection and analysis that make it too easy to publish false-positive, nonreplicable results"); Simmons, Nelson & Simonsohn, *supra* note 98, at 1362; Ioannidis, *supra* note 105, at 0699 ("Most research findings are false for most research designs and for most fields.") (capitalization omitted).

120 *See* Munafò et al., *supra* note 94, at 3; Simmons, Nelson & Simonsohn, *supra* note 98, at 1362–63.

121 This example is adapted from Randall Munroe's xkcd web comic. *See* Randall Munroe, *Significant*, XKCD (Apr. 6, 2011), https://xkcd.com/882 [https://perma.cc/PY7N-3QTU].

122 A *p*-value below 0.05 indicates a probability of less than 5 percent that the result came about by pure chance. By convention in many fields of science, results with associated *p*-values smaller than 0.05 are considered statistically significant, while results with larger *p*-values are not.

of jelly bean. Even if there is no link between jelly bean consumption and acne, the researcher is likely to find a statistically significant result ($p<0.05$) for one of the 20 colors (say, green), purely by chance.[123] In both scenarios, the researcher can report with a 95% confidence level that green jelly beans cause acne. In Scenario A, this confidence level provides a true measure of the variability of the finding: there is only a 5% likelihood that the result was obtained purely by chance, and if the researcher repeated Scenario A, he would likely obtain the same result. If the study in Scenario B were to be repeated, conversely, it is unlikely that the same result would be obtained; in Scenario B, the reported confidence level overstates the reliability of the result. Unless the researcher publishes full details of the research protocol he used, a peer reviewing the findings is unable to tell by looking at the reported result whether the reported finding is the result of Scenario A (a single experiment yielding a statistically significant result) or Scenario B (a selectively reported result from a series of experiments that collectively yielded a statistically significant result just by chance) and unable to predict whether a repeat experiment would be likely to reproduce the finding (as in Scenario A) or not (as in Scenario B).[124]

The methodological problems raised by the Replication Crisis were not novel.[125] But many researchers did not take the threat of unreliable research methodology seriously until they were confronted with the reality that the very foundations of their discipline might be compromised.[126] The replication

---

123 A *p*-value of 0.05 indicates a 1 in 20 probability that a result came about by pure chance, assuming the null hypothesis is true. This means that if 20 alternative false hypotheses are tested, one would expect (on average) one of the results to have an associated *p*-value below 0.05. In fact, the probability of *not* obtaining at least one result with $p < 0.05$ (here: that *none* of the 20 types of jelly beans are found with statistical significance to cause acne) is only 36%. In other words, even if the null hypothesis is true (jelly beans, regardless of their color, do not cause acne), a researcher who tests 20 alternative hypotheses is more likely than not to find at least one result that shows up as statistically significant.

124 Reported measures of confidence can sometimes be corrected to account for the presence of flexibility, but only if the presence and nature of the flexibility is reported in detail. *See infra* note 180.

125 *See, e.g.*, Ioannidis, *supra* note 105, at 0696 (raising concern in 2005 that "most current published research findings are false"); Justin McCrary, Garret Christensen & Daniele Fanelli, *Conservative Tests Under Satisficing Models of Publication Bias*, 11 PLOS ONE e0149590, at 1 (2016) (collecting sources indicating that publication bias, to various extents, had been recognized in psychology, economics, medicine, and other fields more than three decades before the start of the Replication Crisis).

126 *See* Nelson, Simmons & Simonsohn, *supra* note 15, at 514 & n.4 ("Methodologists in other fields had brought up the problem we now know as *p*-hacking.

projects, however, left little room for doubt: some of the most commonly applied research methods were simply not as robust as many researchers had assumed them to be.

It has been suggested that methodological lapses and harmful shortcuts were long allowed to proliferate unchecked among researchers because incentives were aligned in dangerous ways.[127] At all levels of experience, a researcher's career depends on being able to publish novel findings.[128] Hiring, promotion, tenure, research funding, and the ability to recruit students and collaborators all depend on a steady stream of publications.[129] Peer-reviewed journals typically publish only results that meet acceptable levels of statistical significance.[130]

The Replication Crisis made clear: those levels of statistical significance are easier to achieve by researchers who (willfully or blunderingly) take flexible approaches to research.[131] While a carefully planned study is more likely to yield results of a high degree of reliability and replicability, a scattershot approach involving a large set of analyses run on the same dataset in hopes that one will lead to a statistically significant result is more likely to yield results that meet publication criteria.[132] Experiments on small samples, moreover, are a dream come true for a researcher pressed for time: smaller samples are both more likely than larger samples to yield a statistically significant result when there is no "true" result *and* typically involve less work for the researcher.[133] In short, researchers have am-

---

However, perhaps because they did not demonstrate that this was a problem worth worrying about or because they did not propose concrete and practical solutions to prevent it . . . their concerns did not have perceivable consequences on how research was conducted and reported in their fields." (citation omitted)).

127 *See, e.g.*, Marcus Munafò, *Reproducibility Blues*, 543 NATURE 619, 619 (2017) (describing how perverse incentives undermine the scientific method); Nosek, Spies & Motyl, *supra* note 98, at 615 (arguing that institutional incentives "inflate the rate of false effects in published science").

128 *See* Nosek, Spies & Motyl, *supra* note 98, at 626 ("[P]ublishing is a central, immediate, and concrete objective for [a researcher's] career success."); Ottoline Leyser, Danny Kingsley & Jim Grange, *The Science 'Reproducibility Crisis'—and What Can Be Done About It*, THE CONVERSATION (Mar. 15, 2017, 5:49 AM), https://theconversation.com/the-science-reproducibility-crisis-and-what-can-be-done-about-it-74198 [https://perma.cc/DN4B-9ZFT] (arguing that the Replication Crisis "is actually a publication bias crisis").

129 *See* Nosek, Spies & Motyl, *supra* note 98.

130 *See id.* at 617 ("[T]he nominal false-positive rate of a $p < 0.05$ has become a de facto criterion for publishing.").

131 *See* Nelson, Simmons & Simonsohn, *supra* note 15, at 515 ("*P*-hacking is the only honest and practical way to *consistently*" obtain statistically significant results.).

132 *See* Simmons, Nelson & Simonsohn, *supra* note 98, at 1361.

133 *See* Bakker, van Dijk & Wicherts, *supra* note 104 (stating that approaches involving multiple small, underpowered studies described as "the optimal strat-

ple incentive to gravitate toward low-effort, high-reward, but low-reliability methodologies.[134]

In addition to publication pressure, even honest scientists face a risk of being carried away by enthusiasm. Enthusiastic scientists motivated by a search for knowledge have been said to have a "natural tendency . . . to see patterns in noise"[135]—a tendency known as "apophenia."[136] Indeed, science has been described as "an ongoing race between our inventing ways to fool ourselves, and our inventing ways to avoid fooling ourselves."[137]

Prior to the Replication Crisis, there were few impediments to the—conscious or unconscious—use of shortcuts and creative exploitation of data. In a research culture that was lacking awareness and understanding of the way in which methodological decision making might contaminate research results, authors of papers submitted for publication would not typically include the kind of information that would help a reader assess the extent to which Analytical Flexibility may have contaminated the reported research results.[138] Indeed, by now it is clear that *p*-hacking was rife in the social sciences.[139] Journals were not requiring disclosures of this information and peer

---

egy" for manufacturing statistically significant results where there is no true effect); Poldrack et al., *supra* note 96, at 115–16 (explaining that low power reduces the likelihood of finding a true result and raises the likelihood of finding a false-positive result).

134 *See generally* Ioannidis, *supra* note 105 ("The smaller the studies conducted in a scientific field, the less likely the research findings are to be true. . . . The smaller the effect sizes in a scientific field, the less likely the research findings are to be true. . . . The greater the number of and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true. . . . 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. . . . The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true. . . . The hotter a scientific field . . . , the less likely the research findings are to be true.").

135 Munafò et al., *supra* note 94, at 2.

136 *See id.* at 1 (defining apophenia as "the tendency to see patterns in random data").

137 Chris Allen & David M. A. Mehler, *Open Science Challenges, Benefits and Tips in Early Career and Beyond*, 17 PLOS BIOLOGY e3000246, at 6 (2019); *see also* Munafò et al., *supra* note 94, at 7 (quoting Richard Feynman as saying, "The first principle is that you must not fool yourself—and you are the easiest person to fool").

138 *See* Munafò et al., *supra* note 94, at 4 ("Improving the quality and transparency in the reporting of research is necessary to address" reproducibility problems.).

139 *See* Head, Holman, Lanfear, Kahn & Jennions, *supra* note 15; *see also* Nelson, Simmons & Simonsohn, *supra* note 15, at 517 (arguing that failed replication attempts and the prevalence of successful, underpowered studies in the scientific press are evidence of widespread *p*-hacking).

reviewers were unlikely to ask for the information.[140] Even researchers who were generally aware of the dangers of Analytical Flexibility were not necessarily aware that they were engaging in research practices that had been discredited and could undermine the reliability of their outcomes.[141] Students were absorbing bad research practices as they learned from their more experienced supervisors and in many cases became successful *because of* rather than *in spite of* improper research methodology.[142] Indeed, some have quipped that there is a "natural selection of bad science," as unrigorous practices were passed on from generation to generation of researcher.[143]

B. Replication Renaissance

By the mid- to late-2010s, many researchers in the field of psychology had come to the realization that the field was experiencing a crisis.[144] The year 2011 was being described as the "year of horrors" for psychology;[145] there was a growing awareness of failed replication attempts[146] and a growing concern

---

140 *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) (even when there are reporting guidelines, "authors do not always report information specified in guidelines, nor do peer reviewers demand that the information be reported").

141 *See, e.g.*, Nelson, Simmons & Simonsohn, *supra* note 15, at 515 ("*P*-hacking allowed researchers to think, . . . 'most of my studies work; [critics of *p*-hacking] must be talking about other people.'"); *see also* Marc A. Edwards & Siddhartha Roy, *Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition*, 34 ENVTL. ENGINEERING SCI. 51, 55 (2017) (describing the tendency of people to see themselves as more honest than their peers as the "Muhammad Ali effect").

142 *See* Nelson, Simmons & Simonsohn, *supra* note 15, at 515 ("Researchers did not learn from experience to increase their sample sizes precisely because their underpowered studies *were not failing*."); Paul E. Smaldino & Richard McElreath, *The Natural Selection of Bad Science*, 3 ROYAL SOC'Y OPEN SCI. 160384, at 2 (2016) (arguing that science is a cultural activity that is evolutionary in nature and that powerful incentives propagate poor research methods).

143 Smaldino & McElreath, *supra* note 142.

144 *See, e.g.*, Sophia Crüwell et al., *7 Easy Steps to Open Science: An Annotated Reading List* 3, PSYARXIV PREPRINTS (Nov. 16, 2018), https://psyarxiv.com/cfzyx [https://perma.cc/5QV4-3S35] ("Most scientists agree that there is a reproducibility crisis, at least to some extent." (citation omitted)); Munafò et al., *supra* note 94, at 1 ("90% of respondents to a recent survey in *Nature* agreed that there is a 'reproducibility crisis.'").

145 *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* Leyser, Kingsley & Grange, *supra* note 128 ("Murmurings of low reproducibility began in 2011—the 'year of horrors' for psychology . . . .").

146 *See* Munafò et al., *supra* note 94 (stating that "90% of respondents to a [2017] survey in *Nature* agreed that there is a 'reproducibility crisis'").

that published research findings could not be trusted.[147] The research practices that were facing criticism were described as "the biggest threat to the integrity of [the] discipline,"[148] as they "enabled researchers to achieve the otherwise mathematically impossible feat of getting most of their underpowered studies to be significant."[149]

The Crisis sparked a period of methodological reflection that is still ongoing today. It prompted a flurry of papers, conference presentations, and policy proposals[150] and centers, societies, and journals dedicated to improvement of methods and practices.[151] By 2018, practices aimed at increasing the integrity of the discipline of psychology were "orders of magnitude more common."[152] Optimistic researchers spoke of "psychology's renaissance,"[153] noting that psychology was becoming a leader in developing reliable research methodology,[154] and that discussions about research practices have gone mainstream

---

147 *See, e.g.*, Nelson, Simmons & Simonsohn, *supra* note 15, at 515 ("[J]ournals are filled with the 5%" of studies that show false-positive results, "while the file drawers . . . are filled with the [other] 95% . . . .").

148 *Id.* at 517.

149 *Id.*

150 *See, e.g.*, Morton Ann Gernsbacher, *Writing Empirical Articles: Transparency, Reproducibility, Clarity, and Memorability*, 1 ADVANCES METHODS & PRACTICES PSYCHOL. SCI. 403, 403 (2018) (asserting that psychology is now "confront[ing] . . . questionable research practices"); Munafò et al., *supra* note 93 ("The field of metascience . . . is flourishing."); Nelson, Simmons & Simonsohn, *supra* note 14, at 511 (stating that events from 2010 to 2012 "sparked a period of methodological reflection that we . . . call Psychology's Renaissance"); Patrick E. Shrout & Joseph L. Rodgers, *Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis*, 69 ANN. REV. PSYCHOL. 487, 491–93 (2018) (reviewing recent methodological literature).

151 *See, e.g.*, Crüwell et al., *supra* note 144 (explaining the Society for the Improvement of Psychological Science was founded "to further promote improved methods and practices in the psychological research field"); Daniel J. Simons, Editorial, *Introducing Advances in Methods and Practices in Psychological Science*, 1 ADVANCES METHODS & PRACTICES PSYCHOL. SCI. 3, 3 (2018) (inaugurating new *Advances in Methods and Practices in Psychological Sciences* journal, whose mission it is to foster discussions about research); Anthony N. Washburn et al., *Why Do Some Psychology Researchers Resist Adopting Proposed Reforms to Research Practices? A Description of Researchers' Rationales*, 1 ADVANCES METHODS & PRACTICES PSYCHOL. SCI. 166, 166 (2018) (noting in 2016, the Society for Personality and Social Psychology set up a Presidential Task Force on Publication and Research); *About*, OPEN SCI. FOUND., https://osf.io/4znzp/wiki/home/ [https://perma.cc/TD3Q-WQSS] (explaining that the Center for Open Science was founded in 2013, "with a mission to increase openness, integrity, and reproducibility of scientific research").

152 Nelson, Simmons & Simonsohn, *supra* note 14, at 529.

153 *See id.* 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*.").

154 *See* Bloom, *supra* note 69 (arguing that the Replication Crisis is an opportunity for psychology to "lead the way").

and upgrades to research and publishing practices are happening with unprecedented speed.[155] Similar advances were made in pharmacology,[156] behavioral ecology,[157] neuroscience,[158] medicine,[159] and other fields.[160]

It is widely acknowledged that many scientific fields still have a lot of work to do to improve the reliability of the research they produce,[161] but a growing consensus has been forming around a number of proposals to modify accepted research practices. Some of these modifications are true innovations—methodological or procedural devices that were not in use before.[162] Others merely represent increased attention to or insistence on existing best practices.[163] While perfection is likely unattainable, there is much that can—and should—be done to reduce the deleterious effects of the research practices that led to the Crisis.[164]

The remainder of this subpart reviews some of the measures that have been proposed in response to the Replication Crisis. They are by no means the only proposals for reform, but they are among the ones that have garnered the most support

---

155 Simons, *supra* note 150.

156 *See* Diong, Butler, Gandevia & Héroux, *supra* note 14 (noting that pharmacology journals published a series of editorials aimed at improving standards of data analysis and reporting).

157 *See* Leigh W. Simmons, Editorial, *Guidelines for Transparency and Openness (TOP)*, 28 BEHAV. ECOLOGY 347, 347 (2017) (stating that symposium convened to discuss transparency initiatives in behavioral ecology).

158 *See* Shrout & Rodgers, *supra* note 150, at 502 (stating that replication attempts by original researchers are now common in neuroscience).

159 *See* Poldrack et al., *supra* note 96 (stating that U.K. Academy of Medical Sciences convened a meeting about reproducibility).

160 *See* Diong, Butler, Gandevia & Héroux, *supra* note 14, at 7 (listing initiatives from different fields to address awareness and bad reporting and noting that "[t]here is considerable momentum throughout science").

161 *See, e.g.*, Crüwell et al., *supra* note 144 (noting that there is still a lot of confusion and misinformation and identifying a need for a guide); Leyser, Kingsley & Grange, *supra* note 128 (quoting psychologist Jim Grange opining that psychology is leading the way but is not yet out of the woods); Nelson, Simmons & Simonsohn, *supra* note 15, at 529 ("[T]he Middle Ages are behind us, and the Enlightenment is just around the corner.").

162 *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).

163 *See* Nosek, Spies & Motyl, *supra* note 98, at 626 (arguing that the main barriers to change may not be technical or financial, but rather social or cultural within the research community); *see also supra* note 125 and sources therein (indicating earlier, less widely spread knowledge of the importance of certain aspects of methodological rigor).

164 *See* Nosek, Ebersole, DeHaven & Mellor, *supra* note 162, at 2604 ("The rules of statistical inference have no empathy for how hard it is to acquire the data.").

and that may be the most instructive in rethinking how the practice of litigation science should be reformed. They can be divided into three categories: planning, commitment, and presentation. I discuss each in turn.

1. *Planning*

Eliminating Analytical Flexibility from a research project begins at the planning stage. To start a research project in a manner that will generate results of adequate reliability, it is critical to (a) formulate an unambiguous hypothesis; (b) determine the scope of the project; and (c) design a detailed research protocol.

Formulating a Hypothesis. A major lesson offered by the Replication Crisis is the importance of separating prediction from postdiction.[165] When a researcher sets out on a research path without first formulating an unambiguous hypothesis (prediction) and instead constructs a hypothesis based on what she observes in the data (postdiction), she risks being captured by psychological phenomena such as hindsight bias or apophenia—the tendency to look for and find patterns in data where there are none.[166] HARKing (hypothesizing after results are known) is now generally recognized as a research practice that is inappropriate because it has a propensity to diminish the reliability of research results.[167] A result that was cherrypicked from a set of data generated with no particular hypothesis in mind—e.g., the green jelly bean result *supra* page 124–25—is less likely to be a "true" result, and more likely to be an irreproducible fluke than a result that generated with only a specific question in mind.[168]

---

165 *See id.* at 2600.

166 *See id.* (explaining that hindsight bias allows for an ex post facto construction of a narrative that imbues meaning to randomness); Munafò et al., *supra* note 94 (explaining that apophenia can lead to false conclusions).

167 *See, e.g.*, Allen & Mehler, *supra* note 137, at 3 (suggesting that restricting researchers' opportunities for continuous learning by digging into data without a formulated hypothesis "may be the price of unbiased science"); Gernsbacher, *supra* note 150 (characterizing HARKing as a "questionable research practice"); Munafò et al., *supra* note 94, at 2 (arguing that HARKing is a threat to reproducible science); Simons, *supra* note 151 (noting that *Perspectives on Psychological Science*s changed reporting practices in an effort to limit the practice).

168 *See* Shrout & Rodgers, *supra* note 150, at 491 (arguing that unexpected interactions are usually attributable to randomness and that even when support for the interaction can be found in literature after the fact, the researcher should remain skeptical when the result was not specifically anticipated).

Any researcher has the power to prevent HARKing: all the researcher needs to do is formulate and commit to[169] a hypothesis before starting a research project and then use any gathered data to test solely that hypothesis—and no additional hypotheses.[170]

Determining the Scope. The Replication Crisis revealed that many researchers traditionally have not planned their studies with enough care, tending to use sample sizes that are too small and therefore render a study "poorly powered,"[171] or even leaving the size of a study open.[172]

Selecting a sample size in a thoughtful manner should be a standard part of the planning of a research study.[173] Among proposed approaches are the adoption of minimum sample sizes[174] and setting sample sizes based on an a priori power analysis.[175] Not all researchers are convinced that these practices should be routine,[176] but there is broad support for the

---

169 A researcher can commit to a hypothesis through preregistration. *See infra* section III.B.2.

170 Testing only pre-formulated hypotheses requires clear differentiation between exploratory work (aimed at generating hypotheses) and confirmatory work (aimed at testing hypotheses). *See* Allen & Mehler, *supra* note 137, at 4; Gernsbacher, *supra* note 149, at 405.

171 The statistical concept of "power" refers to the likelihood that a study will detect an effect if an effect exists. The more powerful a study, the more likely it is to detect the existing effect and the less likely to conclude that there is no effect when in reality there is one. *See* Poldrack et al., *supra* note 96, at 116–17; s*ee also* Bakker, van Dijk & Wicherts, *supra* note 104, at 543 ("[T]he typical studies are insufficiently powerful . . . ."); Alexander Etz & Joachim Vandekerckhove, *A Bayesian Perspective on the Reproducibility Project: Psychology*, 11 PLoS ONE e0149794, at 10 (2016) (arguing that traditional sample sizes are too small); Open Sci. Collaboration, *supra* note 98, at 4 (asserting that low-powered research is "still the norm" as of 2013).

172 *See* Simmons, Nelson & Simonsohn, *supra* note 98, at 1361–62 (discussing the harm in ceasing data collection based on an interim analysis); *supra* subpart II.A & note 114.

173 *See* Bakker, van Dijk & Wicherts, *supra* note 104, at 552.

174 *See, e.g.*, *id.* (arguing that researchers should "end the pretense" that small sample sizes can ever be adequate); Poldrack et al., *supra* note 96, at 117 (arguing that inappropriately small sample sizes cannot be justified solely on the basis that lack of resources prevented the researchers from using larger samples); Simmons, Nelson & Simonsohn, *supra* note 98, at 1363 (recommending researchers either adhere to minimum sample sizes or justify why they are unable to do so).

175 *See* Bakker, van Dijk & Wicherts, *supra* note 104, at 552; Poldrack et al., *supra* note 96, at 117. A power analysis determines the minimum sample size needed to observe the desired minimum effect size with a desired minimum level of statistical significance. *See* Poldrack et al., *supra* note 96, at 117.

176 *See, e.g.*, Washburn et al., *supra* note 151, at 169–70 (finding that only forty percent of surveyed researchers believe that an a priori power analysis should be a required part of a standard research practice).

notion that sample sizes should be set before the start of the study, in a deliberate, purposeful way.[177]

Designing a Detailed Research Protocol. The flexibility researchers have to modify the way they gather or analyze data midstream can contribute greatly to the irreproducibility of research results.[178] The problem does not lie in the individual choices that researchers make, but rather in the existence of choice itself.[179] When this choice becomes invisible in the ultimate publication—i.e., when a researcher reports only those specific methodological steps that led directly to the reported, possibly cherrypicked results—it is not possible to tell whether the reported results are an artifact resulting from the presence of Analytical Flexibility or whether they were the result of a preplanned and faithfully executed research protocol (and therefore more likely to be a "true" result).[180] When Analytical Flexibility remains unreported, the statistical significance of reported results is likely to be overstated.[181]

It is impossible to provide an exhaustive list of types of Analytical Flexibility, because the ways in which a researcher can vary or improvise her approach are innumerable.[182] They

---

177 *See, e.g.*, Open Sci. Collaboration, *supra* note 98, at 6 (stating that *Psychological Science* requires authors to disclose how they selected their sample sizes and how large the samples were after exclusion of observations).

178 *See* Bakker, van Dijk & Wicherts, *supra* note 104 (demonstrating that commonly used "questionable research practices" can lead to false-positive findings up to 40 percent); Nelson, Simmons & Simonsohn, *supra* note 15, 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 and that, "[i]n truth, it is not that hard to get a study's false-positive rate to be very close to 100%").

179 *See supra* subpart II.A.

180 If the extent and nature of flexibility is known, it is often possible to adjust reported *p*-values and other measures of confidence to account for its presence. Typically, a reported *p*-value would increase significantly (i.e., the level of confidence it expresses would decrease) as a result of this adjustment. *See* Nosek, Ebersole, DeHaven & Mellor, *supra* note 162, at 2601 (correcting diagnosticity of *p*-values to account for the number of tests that was performed is possible, but rarely done).

181 *See* Simmons, Nelson & Simonsohn, *supra* note 98 (stating that the likelihood of finding a false-positive finding at the 5% level is necessarily greater than 5%).

182 *See, e.g.*, Simmons, Nelson & Simonsohn, *supra* note 98, at 1360 (noting that common "degrees of freedom" include choice of sample size and dependent variables and covariates, and selective reporting); Jelte M. Wicherts et al., *Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking*, 7 FRONTIERS PSYCHOL. 1, 3 (2016) (offering a checklist covering thirty-four types of Analytical Flexibility); *see also* Bakker, van Dijk & Wicherts, *supra* note 104, at 545 (stating that a majority of researchers admit to using "Questionable Research Practices" and many underestimate the consequences).

include decisions regarding data collection,[183] removal of outliers,[184] pre-analysis data cleanup,[185] and the analysis itself.[186]

Researchers may choose to explore different options for legitimate reasons, and digging into the data to explain unexpected results is not necessarily nefarious or an example of *p*-hacking.[187] However, by allowing flexibility in a research protocol, a researcher risks being "hijacked by the data" and obtaining results that other researchers would not be able to reproduce if they redrew and reanalyzed the sample.[188]

It is impossible to stamp out flexibility entirely without standing in the way of scientists' ability to do their job, but the more flexibility researchers can eliminate from research protocols, the more reliable the resulting outcomes are.[189] Designing a detailed research plan can prevent a researcher from falling prey to cognitive biases.[190] A plan should include a detailed description of how the researcher(s) will collect data, how they will clean data, what conditions they will apply, and which analyses they will run and in what manner.[191] The more a research protocol can be turned into a "script" setting out a predetermined set of steps, the narrower the opportunity for

---

183 When a study includes a pilot phase, a researcher can manipulate the end of the pilot phase and the start of the actual study. *See* Simmons, Nelson & Simonsohn, *supra* note 98, at 1361 (stating that approximately seventy percent of surveyed researchers reported having decided whether to continue data analysis based on an interim data analysis).

184 Outliers are data points that are out of line with the rest of the data. There are no set criteria for the removal of outliers and many reasonable options for doing so. *See id.* at 1360 (noting many options for the treatment of outliers are justifiable, but the availability of choice introduces problems).

185 *See generally* Open Sci. Collaboration, *supra* note 98, at 5 (noting that there are more than 200 ways to clean data).

186 Researchers can perform data analysis in a potentially infinite number of ways. *See, e.g.*, Wicherts et al., *supra* note 182, at 8 (explaining that even for a standard statistical analysis, commonly used statistics software offers multiple different options, but the choice of method is typically not described in articles).

187 For example, a researcher may explore a possible correlation between two variables or control for a variety of variables for legitimate reasons.

188 *See* Munafò et al., *supra* note 94, at 2 (observing that enthusiastic scientists motivated by discovery have a tendency to spot patterns in noise); Simmons, Nelson & Simonsohn, *supra* note 98 ("[I]t is common (and accepted practice) for researchers to explore various analytic alternatives . . . to then report only what 'worked.'").

189 *See* Simmons, Nelson & Simonsohn, *supra* note 98, at 1362–63; *see also* Allen & Mehler, *supra* note 137, at 3 (suggesting that a decrease in the efficiency of continuous learning "may be the price of unbiased science").

190 *E.g.*, Open Sci. Collaboration, *supra* note 98, at 6 ("By committing to a prespecified analysis plan, one can avoid common cognitive biases." (citation omitted)).

191 *See id.* at 11 (explaining that the plan should include all conditions, variables, covariates, as well as when data collection will stop and start).

intentional or inadvertent *p*-hacking toward a favorable result.[192]

2. *Commitment*

There is a simple solution to force researchers to commit to a predetermined hypothesis, scope, and research plan: preregistration.[193] Preregistration has been called "the most important outcome[ ] of the replication crisis"[194] and described as "the only way for authors to convincingly demonstrate that their key analyses were not *p*-hacked."[195] The use of preregistration has grown rapidly over the past few years, and some predict that it will become the norm in the next few years.[196] A sign of a cultural shift, some journals now make preregistration a precondition for publication.[197]

Preregistration forces researchers to think through possible contingencies and expected outcomes before starting a research project.[198] In doing so, it functions as a commitment device, reducing the risk of both "unconscious error"[199] and

---

192 *See* Marcus Munafò et al., *Scientific Rigor and the Art of Motorcycle Maintenance*, 32 NATURE BIOTECHNOLOGY 871, 872 (2014) (discussing guidelines in the form of checklists for preregistration of studies in medicine, animal studies, observational epidemiology, and other fields).

193 *See* Poldrack et al., *supra* note 96, at 120 (recommending a preregistration requirement to avoid HARKing in neuroimaging studies); Simons, *supra* note 151 (noting that *Perspectives on Psychological Science* has changed reporting requirements to avoid HARKing).

194 Shrout & Rodgers, *supra* note 150, at 493.

195 Nelson, Simmons & Simonsohn, *supra* note 15, at 519.

196 *See id.* at 520 (observing that preregistration is not yet the norm, but should become the norm in the next three to five years); Agnès Dechartres, Philippe Ravaud, Ignacio Atal, Carolina Riveros & Isabelle Boutron, *Association Between Trial Registration and Treatment Effect Estimates: A Meta-Epidemiological Study*, 14 BMC MED., July 2016, at 1, 1 (observing that preregistration has become the norm rather than the exception in medicine, and that the World Health Organization launched its own preregistration platform); Nosek, Ebersole, DeHaven & Mellor, *supra* note 162 (pointing at 8,000 preregistrations filed through the Open Science Framework as evidence of a cultural shift). *But see* Aba Szollosi et al., *Is Preregistration Worthwhile?*, 24 TRENDS COGNITIVE SCI. 94, 94–95 (2020) (arguing that while preregistration may increase transparency, it is unlikely to improve scientific reasoning and theory development).

197 *See, e.g.*, Christophe Bernard, Editorial, *Improving the Way Science Is Done, Evaluated, and Published*, 4 ENEURO e0373, at 2 (2017) (stating that *eNeuro* will soon implement preregistration); Leyser, Kingsley & Grange, *supra* note 128 ("[M]any psychology journals now recommend or require the preregistration of studies . . . ."); Brian A. Nosek & Daniël Lakens, Editorial, *Registered Reports: A Method to Increase the Credibility of Published Results*, 45 SOC. PSYCHOL. 137, 137 (2014) (stating that in 2014, *Social Psychology* published its first issue with only preregistered replication studies).

198 *See* Allen & Mehler, *supra* note 137, at 3.

199 *Id.* at 6.

cognitive bias,[200] and rescuing researchers from "being taken hostage by [their] own data."[201] It increases the credibility of research results.[202] The strongest form of preregistration includes a full specification of a researcher's hypothesis, study design, and analysis plan.[203] If the particulars of an analysis are dependent on the outcome of earlier steps, the plan can include a decision tree or list of operating procedures specifying which steps the researcher(s) will follow under each anticipated condition.[204] Preregistration can be done publicly[205] or confidentially;[206] its primary value lies not in the publication of the research protocol but in the binding of the researcher.[207] In addition to enforcing the use of predetermined research protocols, it offers an opportunity to separate scrutiny of research methodology from scrutiny of the results.[208]

Preregistration does not prevent authors from reporting unexpected findings.[209] Deviations from a preregistered research plan, necessitated by unexpected findings or unantici-

---

200 *See* Crüwell et al., *supra* note 144, at 17.

201 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/L69J-3PWZ].

202 *See* Nosek, Ebersole, DeHaven & Mellor, *supra* note 162.

203 *See id.* (suggesting that, ideally, a research plan includes all steps the researchers will take); Open Sci. Collaboration, *supra* note 98, at 7 (positing that preregistration should include full specification of study design and analysis plan).

204 *See, e.g.*, Munafò et al., *supra* note 192 (describing preregistration checklists used in medicine, animal studies, observational epidemiology, and other fields); Nosek, Ebersole, DeHaven & Mellor, *supra* note 162, at 2602 (discussing the preregistration of "decision trees"); Shrout & Rodgers, *supra* note 150, at 504 (lauding as "systematic and thoughtful" the Open Science Foundation's decision to allow researchers to file preregistration reports expressing uncertainty about what measure to use); Wicherts et al., *supra* note 182, at 1, 3 tbl.1 (suggesting a checklist covering thirty-four types of Analytical Flexibility).

205 *See, e.g.*, *What Is Preregistration?*, CTR. OPEN SCI., https://cos.io/prereg [https://perma.cc/ZR9V-FUG7] (describing preregistration through submission to a public registry).

206 *See* Crüwell et al., *supra* note 144, at 10 (describing private, self-archiving options); Nelson, Simmons & Simonsohn, *supra* note 15, at 519 (explaining that some preregistration platforms allow for researchers to lock preregistered studies from the public or share them anonymously). *But cf.* Washburn et al., *supra* note 151, at 169 (noting that some researchers avoid preregistration out of concern for being "scooped").

207 *See supra* subpart II.A.

208 *See* Bernard, *supra* note 197 (noting that *eNeuro* "nearly guarantees publication" of preregistered studies); Nosek & Lakens, *supra* note 197, at 138 (discussing the phenomenon of "CARKing": Critiquing After Results are Known); Wagenmakers & Dutilh, *supra* note 201 (stating that a journal can offer early in-principle acceptance of articles based on a preregistered research protocol).

209 *See* Gernsbacher, *supra* note 150, at 404.

pated violations of foundational assumptions, would not necessarily invalidate the study but would be expected to be reported clearly so that they can be taken into account when assessing the reliability of the study's findings.[210] Preregistration also does not take away researchers' "narrative license" in explaining and presenting the data.[211] It merely constrains their flexibility to adjust their hypothesis or methodology in the course of the research project.[212] Preregistration "does not eliminate the possibility of poor statistical practices, but it does make them detectable."[213]

3. *Presentation*

When research results are published with limited or selectively reported information about the process that produced them, readers cannot evaluate the extent to which their reliability has been impaired by Analytical Flexibility.[214] To allow readers to assess the presence of Analytical Flexibility, researchers should report all information necessary to allow others to replicate their study.[215] They should disclose all variables collected, as well as details about all analytical decisions, including descriptions of analyses that failed.[216] Some journals now require extensive disclosures as a precondition for publication, sometimes using checklists to ensure all necessary detail is reported.[217]

---

210 *See* Allen & Mehler, *supra* note 137, at 3–4 (arguing that researchers may make changes to a preregistered plan but should justify and discuss each change at publication); Nosek, Ebersole, DeHaven & Mellor, *supra* note 162, at 2602 (suggesting that for some types of study, preregistration could take place in stages, with each stage informing the steps to be registered, and then executed, in the next stage).

211 Nosek, Ebersole, DeHaven & Mellor, *supra* note 162, at 2604.

212 *See* Allen & Mehler, *supra* note 137, at 2–3. Preregistration also allows researchers to take credit for the quality of their predictions and can help build a researcher's reputation. *See* Wagenmakers & Dutilh, *supra* note 201.

213 Nosek, Ebersole, DeHaven & Mellor, *supra* note 162, at 2602.

214 *See supra* subpart II.A.

215 *See* Nelson, Simmons & Simonsohn, *supra* note 15, at 518–19 (noting that some peer reviewers now demand disclosure of all measures, conditions, exclusions, and other analytical decisions).

216 *See*, *e.g.*, Open Sci. Collaboration, *supra* note 98, at 11 (*Psychological Science* requires full disclosure of sample-size selection, exclusion of observations, and all experimental conditions, including failed manipulations); Simmons, *supra* note 157 (stating that *Behavioral Ecology* will only publish papers documenting all design decisions and analytical decisions); Wolf Vanpaemel, Maarten Vermorgen, Leen Deriemaecker & Gert Storms, *Are We Wasting a Good Crisis? The Availability of Psychological Research Data After the Storm*, 1 COLLABRA 1, 1 (2015) (noting that *Science* and *Nature* require openness of data).

217 *See* Bernard, *supra* note 197, at 1 (stating that since 2017, *eNeuro* requires inclusion of computer code); Mercè Crosas et al., *Data Policies of Highly-Ranked*

Public preregistration can help create a culture that facilitates these disclosures. When all envisioned analytical steps have been preregistered, confirming the steps taken and reporting the outcome of *all* analyses—not just successful analyses—is a logical next step. At the publication stage, the researcher either confirms that she took the steps as envisioned or explains and justifies any deviations from the preregistered research plan.[218] In this context, incomplete or selective reporting of methodological steps taken would be against the author's interest; a reader could easily determine that the study was preregistered and the lack of a detailed description of methodology would raise suspicion.

III
LESSONS FOR LITIGATION SCIENCE

There is ample reason to believe that the problems that stood at the center of the Replication Crisis are at least as present in litigation science as it is in academic science. Subpart III.A explains why cultural and institutional norms in litigation science make it at least as likely as academic science to be vulnerable to the methodological flaws uncovered in the Crisis, and therefore at least as much in need of a critical look. Subpart III.B draws on lessons learned from the Replication Crisis to propose reforms to the way litigation science is created and assessed for admissibility.

A. Reliability Problems in Litigation Science

The defects of method uncovered in the course of the Replication Crisis and its aftermath are likely to be at least as pervasive in experiments conducted in connection with civil litigation as they have been in academic science. When expert witnesses are retained to conduct analyses or experiments, just like their colleagues in academic science, they face powerful incentives to produce a particular type of result: findings that can with-

---

Social Science Journals 8, SOCARXIV PAPERS (Mar. 30, 2018), https://osf.io/preprints/socarxiv/9h7ay [https://perma.cc/C3SB-A8EM] (noting that as of January 2018, 155 of 291 social sciences journals surveyed have a data-disclosure policy); Munafò, *supra* note 192 (noting that *Nature* has introduced reporting checklists).

218 *See* Allen & Mehler, *supra* note 137, at 3–4 (arguing that changes to preregistered plan may be made but should be justified and discussed); Open Sci. Collaboration, *supra* note 98, at 17 (When a study has been preregistered, "the 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.").

stand a *Daubert* (or equivalent) motion and be presented to a jury.[219] Research scientists aim for statistically significant results because it will help them get those results published;[220] expert witnesses aim for statistically significant results because they will enable them to pass the *Daubert* hurdle.[221] If the Replication Crisis has taught the world one thing, it is this: pressure to generate statistically significant findings is likely to provoke the type of behavior that leads to more reliable research results—HARKing, *p*-hacking, and other forms of exploitation of Analytical Flexibility.[222] There is no reason to believe that individuals performing research in support of litigation would be less influenced by this type of pressure than their colleagues on the academic side.[223] And Analytical Flexibility does not distinguish between academic science and litigation science; when present, it impairs the reliability of scientific findings regardless of the purpose for which they were generated.

Moreover, the undetectable nature of this impairment of reliability can be expected to be endemic in both modes of science. Just as in academic science, experimenters in litigation have traditionally been allowed to present research results selectively. As described in Part I, experts typically do not disclose their methodology until they submit a report in which they present the results of their completed study, typically toward the end of the expert-discovery period.[224] Until that time, they typically design and perform their experiments hidden from view, a cloak of litigant privileges protecting much of their work product from discovery and scrutiny, presenting selected outcomes only after experiments have been completed and re-

---

219 *See, e.g.*, Susan Haack, *Of Truth, in Science and in Law*, 73 BROOK. L. REV. 985, 988 (2008) ("[S]cientists complain about . . . the professional insult of being 'Dauberted' or 'dauberted out' . . . .").

220 *See supra* subpart II.A & note 100.

221 See George P. Lakoff, *A Cognitive Scientist Looks at* Daubert, 95 AM. J. PUB. HEALTH S114, S117 (2005) (An expert facing a *Daubert* challenge "has plenty to lose: his or her reputation for professional expertise . . . . When a scientist is 'Dauberted out' of a trial, the repercussions for the scientist are serious.").

222 *See supra* subpart II.A.

223 Indeed, many expert witnesses are academic scientists who occasionally branch out to litigation science.

224 *See supra* note 63 and accompanying text. Some jurisdictions do not require a report at any stage in the proceeding. *See, e.g.*, Beck v. Hirchag, No. G041955, 2011 Cal. App. Unpub. LEXIS 2649, at *15 (Cal. Ct. App. Apr. 11, 2011) (noting that expert reports are not required in California). In those jurisdictions, an expert's disclosures may be limited to information the expert (or retaining party) chooses to disclose, plus any disclosures in response to questions at a deposition or evidentiary hearing.

sults have been analyzed.[225] Just as in some of the academic sciences before the Replication Crisis, they are able to choose between fixing their experimental setup and analytical methodology up front and making it up as they go. No one watches over their shoulder to check the extent to which they exploit Analytical Flexibility to obtain favorable results. There is currently no mechanism to force an expert to preregister a study prior to conducting it. Just as in the academic sciences,[226] once an expert starts gathering and analyzing data, nothing stops him or her from modifying the protocol on the fly. Just as in the academic sciences before the Replication Crisis, once experiments have been completed, there is no pressure on the expert to disclose unfavorable results, and no one has access to the metaphorical file drawer containing undisclosed results.[227]

The corrosion in reliability that results from Analytical Flexibility is difficult to detect in civil actions that proceed according to the current standard procedure described in Part I: by the time the court considers the reliability of proffered expert testimony that is based on the expert's experimental results, it is impossible to determine whether a presented result was the result of unsavory forms of data manipulation or the result of reliable research methodology that was reliably applied.[228] The power of the adversarial process to bring to light flaws in proposed evidence is limited when an important cause of unreliability is undetectable.

Daubert motions (or equivalent motions under *Frye* or other standards), as currently implemented, have not proven to be adequate vehicles for screening testimony for this particular dimension of unreliability. This is not a flaw in the *Daubert* standard itself, but rather the result of an interplay between multiple factors: (1) courts and litigants lack awareness of the way Analytical Flexibility injects unreliability into an expert's process;[229] and (2) even if they were aware, by the time the

225 *See supra* Part I.

226 *See supra* subpart II.A.

227 *See, e.g.*, FED. R. CIV. P. 26(a)(2)(B)(i)–(ii) (providing that an expert report is required to include "all opinions the witness will express and the basis and reasons for them" and "the facts or data considered by the witness in forming them"); TEX. R. CIV. P. 194.2(f)(3) (providing that a party may request a disclosure of "the general substance of the expert's mental impressions and opinions and a brief summary of the basis for them").

228 *See supra* subpart II.A and note 180.

229 Major legal research databases Lexis and Westlaw contain only one court opinion in which the term "*p*-hacking" appears. *See In re* Roundup Prods. Liab. Litig., 390 F. Supp. 3d 1102, 1137 (N.D. Cal. 2018). The case poignantly illustrates the court's lack of awareness of problems relating to Analytical Flexibility.

expert presents her results—typically after she has completed her work and filed a report—there is generally no way to assess the role of Analytical Flexibility in the creation of the results (i.e., no way to assess whether Analytical Flexibility did in fact inject unreliability into the expert's process).[230] Opposing parties can question an expert about the expert's methodology and results, at a deposition or evidentiary hearing, but so long as crucial information about the research protocol and analytical steps the expert employed can be left out of an expert report, these adversaries will generally lack the information they need to demonstrate that the expert engaged in *p*-hacking or HARKing. In other words, Analytical Flexibility can severely undermine the reliability of scientific evidence, but is largely invisible under current gatekeeping proceedings.

Even if a court were able to detect or discover that Analytical Flexibility played a role in an expert's process, at the *Daubert* stage there is typically no way to predict—let alone quantify—the ways in which this flexibility has contaminated the resulting data.[231] There is also no way to account for the contamination in interpreting the data or to determine whether the data might be reliable *despite* the presence of Analytical Flexibility in the process of its creation. At the *Daubert* stage, the court's options are limited to admitting the proposed testimony—leaving the parties free to try to convince the finder of fact of the data's unreliability—or excluding it.

B. Proposals

Courts have an important role to play in addressing this problem. A rich scholarly literature examines the differences and similarities between "litigation science" and "academic science" and debates the extent to which litigation science ought to import the practices and standards that apply in academic science into its courtroom setting.[232] In adopting the *Daubert*

The defendant argued that a plaintiff's expert had engaged in *p*-hacking, performing tests using a number of different measures and reporting only those results that were favorable. The court rejected the argument on the ground that the defendant had not shown why the measure for which results were reported was an inappropriate choice. *Id.*

230 *See supra* subpart II.A.

231 *See id.*

232 *See generally* Leslie I. Boden & David Ozonoff, *Litigation-Generated Science: Why Should We Care?*, 116 ENVTL. HEALTH PERSP. 117 (2008) (arguing that "the problems with litigation-generated science . . . are very general and apply to much or most science"); Susan Haack, *A Match Made on Earth: Getting Real About Science and the Law*, 36 DALHOUSIE L.J. 39, 51 (2013) (describing "deep tensions between science and the culture of (at least U.S.) law"); Jasanoff, *supra* note 30, at

standard, the Supreme Court leaned on an assumption that litigation science should mimic academic science as much as possible.[233] In this view, correct scientific inquiry, is governed by the "scientific method," and any knowledge-generating endeavor should aim to follow the rules of this method.[234] Litigation science and academic science are simply two instantiations of a singular practice of science, with litigation science aiming to generate knowledge relating to a specific set of facts at issue in a case, and academic science aiming to generate knowledge for its own sake, typically not limited in application to any particular set of facts.[235]

The *Daubert* framework calls on judges to "think like scientists" and, in that role, to assess the validity of experimental and analytic methodology with reference to practices in established fields of science.[236] Only if the reference field has endowed the methodology with certain indicia of reliability—such as passing peer review or general acceptance in the field—will the methodology be considered reliable enough to be employed in connection with civil litigation.[237] While the *Daubert* standard leaves room for departures from laboratory standards as necessitated by litigation-specific constraints—time, cost, norms against the relitigation of disputes even after the emergence of new information—experts producing evidence for presentation in a civil proceeding are expected to hew as closely as possible to the practice of academic science.[238]

Sheila Jasanoff has argued that litigation science and academic science are too dissimilar to allow for a direct importation of scientific standards into a courtroom setting, and that any notion that such direct importation is possible rests on an

---

124–25; Sheila Jasanoff, *Law's Knowledge: Science for Justice in Legal Settings*, 95 AM. J. PUB. HEALTH S49, S56-S57 (2005) [hereinafter Jasanoff, *Law's Knowledge*] (listing "contextual factors" that "[may] lead to different evaluations of the same science in different contexts").

233 *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 592–93 (1993); Jasanoff, *supra* note 30, at 123–24 ("Judges, in this view, should serve as inscription . . . devices, automatically writing scientists' standards of reliability and validity into their assessments of the evidence . . . .").

234 *See* Jasanoff, *supra* note 30, at 123–24.

235 *See* Jasanoff, *Law's Knowledge*, *supra* note 232, at S51 ("[T]he law needs facts as necessary adjuncts to doing justice; science seeks facts more as an end in itself.").

236 Jasanoff, *supra* note 30, at 123–24 ("In *Daubert*'s epistemological framework, scientists establish the criteria for what counts as science, and judges are charged with importing these into admissibility decisions . . . .").

237 *See Daubert*, 509 U.S. at 593–94 (holding that in making a reliability determination, a court may consider factors such as peer review, publication, and general acceptance).

238 *See* Jasanoff, *supra* note 30, at 124.

overly idealized view of the scientific method.[239] Jasanoff argues that when judges purport to apply scientific norms in their evaluation of scientific evidence, they necessarily apply *their own idea* of what those scientific norms are, rather than scientific norms that would actually apply in the relevant field of academic science.[240] Litigation science and academic science operate within bounds and contexts that are different enough from one another that interpretations of *Daubert* that ask judges to "think like scientists" have led to the creation of a "'junk science' of how science works."[241] While Jasanoff appears to accept the *Daubert* court's notion that the ideal practice space for litigation science should "imitate the ideal scientific laboratory," she has argued persuasively that *Daubert*, at least as currently applied, does not allow courts to approach that ideal.[242]

As described in subpart II.A, there are research practices that render research results fundamentally unreliable, in ways that are undetectable to anyone examining the reported findings, and that will remain undetected unless we rethink the procedures by which scientific evidence is created and evaluated, and the role that courts play in those procedures. Jasanoff's argument that in evaluating scientific data, courts should focus on the process of their creation as an indicator of their reliability[243] is a step in the right direction, but a focus on process is not enough. To make visible the "invisible contamination" caused by Analytical Flexibility in litigation science, the creation process should be structured in new ways, and up-

---

239 *See id.* (arguing that courts' application of the *Daubert* framework "rest[s] on idealized, misleading, or misinformed assumptions about the scientific method"); Jasanoff, *Law's Knowledge*, *supra* note 232, at S53 (arguing that the *Daubert* majority "assumed that there is a well-articulated model of 'good science' whose standards can be objectively applied to offers of scientific evidence (the myth of 'scientific method')").

240 *See* Jasanoff, *supra* note 30 ("The mistake, in law[,] . . . is to believe that there are pregiven, determinate standards, extrinsic to what scientists themselves would invoke when confronted by specific, competing claims.").

241 *Id.* at 123–24; *see also* Jasanoff, *supra* note 232, *Law's Knowledge*, at S50 ("[I]n an ironic turn, the 'science' that the Court officially embraced remained profoundly a creation of the law's own biases, needs, and misconceptions concerning scientific inquiry . . . .").

242 Jasanoff, *supra* note 30; *see also* James R. Dillon, *Expertise on Trial*, 19 COLUM. SCI. & TECH. L. REV. 247, 250–51 (2018) (arguing that judges cannot possess substantive expertise across all possible fields of scientific evidence and therefore cannot be expected to interpret scientific expert testimony effectively).

243 Jasanoff, *supra* note 30, at 129 ("As referees of science-in-the-making, judges would focus on the process through which litigation science is generated rather than on its validity or invalidity.").

dated norms should guide the type of information that is gathered and disclosed with respect to the creation process.

It is now understood that research results can be relied upon only if their creation is governed by a well-defined set of procedures and guided by a predetermined hypothesis, if their reporting is accompanied by certain types of information that allow a reader to assess their reliability, and if their evaluation takes into account indicia of Analytical Flexibility.[244] This improved understanding of what is needed to allow a reviewer to assess a study's validity should inform what information should be available to courts evaluating litigation science for admission into evidence. Making this information available to courts and opposing parties, in turn, requires the generation and preservation of the required information earlier in the research process. Even under the traditional view, in which a judge's role is to "think like [a] scientist[ ]" and adjudicate the reliability of scientific evidence through the application of norms and criteria from the corresponding field of academic science,[245] the lessons from the Replication Crisis should be instructive. When the reference field of academic science experiences a knowledge crisis and undergoes a large-scale cultural shift, participants to the process of litigation science ought to take note.

The social academic sciences are reforming the way in which science is being conducted by redesigning the chain of creation starting from the back end.[246] Just as academic sciences are redesigning the research process with publication in mind, working back from publication to creation, so too should litigation science processes be rethought to ensure that the process generates not only the desired research results, but also the information necessary to assess their reliability. In light of the invisibility of Analytical Flexibility and the importance of screening scientific evidence for reliability, there should be a reevaluation of the way in which research is conducted in connection with litigation.

The institutional framework in which litigation science is created is in some ways uniquely suited to bring about this cultural shift—in some respects more suited than academic science. While academic scientists remain largely free to conduct their work any way they want and journals and organizations can at most *encourage* adherence to updated norms and

244 *See supra* subpart II.A.
245 Jasanoff, *supra* note 30, at 123.
246 *See supra* subpart II.B.

standards, when it comes to litigation science, courts can affirmatively *require* certain conduct.[247] Litigation science also holds a greater potential for peer enforcement: while in academic science some enforcement of norms is allocated to peer reviewers—typically unpaid, anonymous, and therefore likely varying in their motivation to perform scrupulous review—in litigation science, some of the enforcement of norms is allocated to opposing litigants—a group uniquely motivated to find flaws in their adversaries' experts' work and bring them to light.[248]

The consequences of contaminated methodology forming the basis of expert testimony can be serious. If the contamination in a presented methodology cannot be detected, a court may determine that the methodology meets the jurisdiction's reliability standard, and that testimony based on the methodology—testimony that may present fundamentally unreliable experimental results as if they were reliable—is therefore admissible.

This subpart begins a rethinking of the process by which scientific evidence is created, presented, and evaluated for admissibility. It describes a number of measures that could be taken within the existing (*Daubert*, *Frye*, and other) frameworks for the evaluation of the reliability of proposed expert evidence: (1) early proceedings aimed at ensuring the rigorous planning of proposed litigation science; (2) updated and more detailed expert disclosure requirements; and (3) consideration of Analytical Flexibility as part of a court's evaluation of the reliability of proposed expert evidence.

1. *Early Proceedings on Proposed Litigation Science*

In current litigation practice, courts typically assess the admissibility of expert evidence shortly before or during trial, through a review of the expert's fully formed opinions and the expert's stated bases for those opinions.[249] Conducting some or all of the reliability assessment much earlier in the process would be the most impactful change that courts could make in recognition of the problems uncovered in the Replication Crisis. By adopting an early and/or staged process for admissibil-

---

247 *See infra* subpart III.B.

248 *See, e.g.*, Boden & Ozonoff, *supra* note 232, at 120 ("A competent attorney, aided by competent experts, should be in a better position to expose the flaws in [an opposing side's expert's] research than is the peer reviewer, who often takes less time than the expert in a legal case and has more limited resources to probe than does the cross-examining attorney." (citation omitted)).

249 *See supra* Part I and note 54.

ity determinations, courts could contain the currently largely invisible dimension of Analytical Flexibility. Requiring an expert to commit to (a) an unambiguous hypothesis; (b) a scope for the project; and (c) a detailed research protocol would radically limit the expert's Analytical Flexibility, prevent HARKing, and significantly limit the opportunities for *p*-hacking.[250] To the extent that any aspects of the protocol were left open for decision later,[251] this remaining Analytical Flexibility would become visible.

In substance, early admissibility proceedings could mirror the preregistration practices that are being adopted in many fields of academic science.[252] By requiring an expert to describe her intended study in detail, including every variable she intends to collect, every envisioned analytical step, and the criteria she is going to use in excluding data and in interpreting the results, the court gains visibility that enables it to determine whether results are statistical flukes or "real" results.[253]

Early proceedings on the reliability of proposed expert methodology could take a variety of forms. They could be conducted by the court or happen entirely between the parties without the court's involvement. They could be voluntary or mandatory. They could involve an adjudicative element or not, with a presumption of admissibility or not. The main purpose of early proceedings is as a commitment device: just as in academic science preregistration forces a researcher to commit to a predetermined hypothesis and research plan, so would early proceedings relating to expert evidence force an expert witness to make the same commitment.

Sheila Jasanoff has argued for a process whereby the parties engage in discussions or negotiations about protocols to be followed, with the judge acting as a "referee[ ]" not of the validity or correctness of the resulting agreed-upon methodology but of the process the parties employed to arrive at it.[254] The interventions suggested here follow that approach, but go further. The court's role during expert discovery, in this model, is not limited to that of a referee adjudicating discovery disputes between litigants, but lies primarily in ensuring that experts disclose and commit to a hypothesis and research plan up front. By routinely facilitating early proceedings relating to litigation

250 *See supra* notes 195–204 and sources cited therein.
251 *See supra* subpart II.B and note 204.
252 *See supra* subpart II.B.
253 *See* Nosek, Ebersole, DeHaven & Mellor, *supra* note 162.
254 *See* Jasanoff, *supra* note 30, at 129.

science, courts can bring about a cultural shift: an *expectation* that experts make the necessary commitments and disclosures.[255]

In a process directed by the court, a party engaging an expert witness would submit a proposed plan to the court at an early stage in discovery, before the expert has carried out the work. Similar to a preregistration report, the plan would include a detailed description of the expert's hypothesis and every step the expert plans to take in the course of data collection, cleanup, and analysis.[256] In practice, this plan would likely resemble the description of the methodology in the expert report that the expert would ordinarily be submitting later in the process.[257] Indeed, the report could take the form of a typical expert report with the "results" and "conclusions" sections omitted.

Submitting a report stripped of a results and conclusions section early in the process (before experiments have commenced) rather than later (after experiments and analysis have been completed) changes more than just the timing. In submitting a methodological description up front, the expert commits to the described methodology and binds her own hands against the ability to improvise or experiment with different methodological steps. From the court's perspective, a presubmitted plan ensures that the expert followed a predetermined set of steps to test a preselected hypothesis, whereas a postsubmitted report leaves the court guessing at how much of the expert's methodology and theory was determined up front and how much was decided along the way.

Adjudication is an optional element of early proceedings relating to litigation science. Following submission of the expert's proposed plan, the court could hold a *Daubert*-type hearing to evaluate the plan. In federal courts and the state courts applying the *Daubert* standard, these hearings would take the same shape as ordinary hearings on the admissibility of scientific testimony. Just as in a regular *Daubert* hearing, the court would evaluate the expert's (here: proposed) methodology for

---

255 *Cf.* notes 196–197 and sources cited therein (indicating cultural shift toward preregistration in social sciences).

256 *See* Nosek, Ebersole, DeHaven & Mellor, *supra* note 162, at 2602 (an ideal research plan should include all steps to be taken); Open Sci. Collaboration, *supra* note 98, at 11 (preregistration should include full specification of study design and analysis plan); *see also* Chin, Growns & Mellor, *supra* note 25, at 390 (discussing voluntary preregistration by expert witnesses).

257 *See, e.g.*, FED. R. CIV. P. 26(a)(2)(B)(i)–(ii); N.Y. C.P.L.R. 3101(4)(d)(1) (McKinney 2014); TEX. R. CIV. P. 194.2(f)(3).

indicators of reliability, such as the methodology's known or potential error rate, and whether the methodology can be tested, has been tested, has been subjected to peer review, has been published, and/or is "generally accepted" in the relevant scientific community.[258]

By conducting gatekeeping proceedings early in the process, a court can evaluate the methodology separately from the results and before the process of executing the methodology has had a chance to *alter* the methodology and inject unreliability into the process.[259] This opportunity typically will have disappeared by the time an ordinary admissibility hearing takes place: after data have been gathered and fully analyzed.[260] A second, generally much more limited, hearing would take place at the end of expert discovery, after all data have been gathered and analyzed.[261] At this stage, the only questions before the court would be whether the expert applied the pre-approved methodology exactly as described, and whether any departures from the pre-approved methodology reduced the reliability of the methodology to a sufficiently severe degree as to render the expert's conclusions inadmissible under the *Daubert* standard.[262]

A court issuing an early ruling on a proposed methodology with a well-defined set of steps and goals would do well also to offer a precise description of the types of arguments still available at a later stage to parties opposing an expert's proposed plan. For example, the court might rule that any argument that the party could have made at the early hearing is waived, but that any argument that was not made and could not have been made—because it relies on information that came to light later—will remain available.[263] A ruling that incorporates this

---

258 *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593–94 (1993).

259 *See supra* subpart II.A.

260 *See id.*

261 The later-stage hearing could be termed a *Joiner* hearing, after the second case in the *Daubert* trilogy. General Elec. Co. v. Joiner, 522 U.S. 136 (1997). In *Joiner*, the Court walked back a full separation between methodology and conclusions that had been read into *Daubert*, holding that part of a trial court's gatekeeping role is to be on guard for conclusions that have too tenuous a connection to the methodology, where there is "too great an analytical gap between the data and the opinion offered." *Id.* at 146; *see also* Haack, *supra* note 232, at 43 (arguing that *Joiner* "quietly jettisoned *Daubert*'s key distinction between the methodology an expert uses and the conclusion he reaches").

262 A party opposing the admission of the expert's opinions would, of course, still be free to make objections on the basis of relevance under the jurisdiction's applicable rules of evidence.

263 *See, e.g.*, Fed. Hous. Fin. Agency v. JPMorgan Chase & Co., 11 Civ. 6189 (DLC), 2012 U.S. Dist. LEXIS 173768, at *37–38 (S.D.N.Y. Dec. 3, 2012) (granting

type of specification would not visit injustice upon litigants on either side. While it forces the opposing party to be diligent in defending its interest at an early stage of expert discovery, rather than keep its powder dry and preserve some of its arguments for a subsequent round of briefing, it stops clear of fully "blocking the inquiry" into the expert's methodology and its application at a later stage in light of subsequent developments.[264] In addition, requiring the opposing party to make any argument available to it at this early stage (or waive it forever) balances full disclosure of methodology with full disclosure of arguments in opposition. It increases the transparency and efficiency of the process and the utility of the early evidentiary battle. A court choosing this option should describe its decision and the arguments foreclosed by it very precisely, to prevent the ruling from being rendered meaningless by a skilled litigant's attempt to reargue its position by dressing an old argument in new facts later.

Early hearings relating to the admissibility of expert opinions are not a novel idea, though thus far they have been requested—typically by plaintiffs—on different grounds. In a handful of cases arising out of the 2008 Financial Crisis, plaintiffs facing an expensive expert-discovery process sought advance approval from the court of their expert's proposed methodology, arguing that it would be unreasonable and unduly burdensome for these plaintiffs to incur the expense of an unusually extensive and time-consuming expert analysis without some assurance that the resulting findings would end up being admissible.[265] In a small number of other financial-crisis

---

preapproval to the plaintiff's expert's proposed methodology, but reserving defendants' rights to file later challenges that had not yet been available to it at the preapproval stage).

264 Susan Haack, *Do Not Block the Way of Inquiry*, 50 TRANSACTIONS CHARLES S. PEIRCE SOC'Y 319, 322 (2014) ("[B]locking the way of inquiry is much worse than mere inefficiency . . . to set up a philosophy which barricades the road of further advance toward the truth is the one unpardonable offense in reasoning . . . .").

265 *See, e.g.*, Nat'l Credit Union Admin. Bd. v. RBS Sec., No. 11-2340-JWL, 2014 U.S. Dist. LEXIS 59893, at *7 (D. Kan. Apr. 30, 2014) (plaintiff submitted early in limine motion on reliability of expert testimony based on proposed sampling scheme); MBIA Ins. Corp. v. Countrywide Home Loans, Inc., No. 602825/08, 2010 N.Y. Misc. LEXIS 6182, at *9 (N.Y. Sup. Ct. Dec. 22, 2010) (same); The People's Notice of Motion in Limine Re Sampling Methodology, People v. Morgan Stanley & Co., No. CGC-16-551238 (Cal. Super. Ct. Feb. 15, 2018) (same); *see also* MASTR Adjustable Rate Mortgs. Tr. 2006-OA2 v. UBS Real Estate Sec., Inc., No. 12-CV-7322 at *1 (S.D.N.Y. Mar. 11, 2015) (plaintiffs submitted a letter application for early proceedings regarding their expert's proposed sampling study); Fed. Home Loan Bank of Bos. v. Ally Fin., Inc., No. 11-10952-GAO, at *3 (D. Mass. May. 1, 2014) (plaintiff filed a "Motion for Approval of Statistical Sampling Methodology"); Mass. Mut. Life Ins. v. Residential Funding Co., 989 F. Supp. 2d 165,

cases, similar early proceedings took place at the initiative of the court.[266] Most courts faced with early evidentiary motions of this ilk[267] have been willing to engage in early proceedings on the admissibility of proposed expert methodology.[268]

As an alternative to court-run proceedings, a mechanism for early exchanges of information could be orchestrated by the parties. Parties could meet and confer without the court's involvement, possibly at the court's encouragement, to discuss proposed work to be performed by any experts they plan to retain. As discussed in more detail below, both proponents and opponents of an expert's testimony tend to gain from early formal or informal proceedings relating proposed expert work. The opponent benefits from an opportunity to pin down a proponent's expert's methodology at an early stage, blocking the

---

172 (D. Mass. 2013) (magistrate judge recommending that the court set a briefing schedule for early-determination proceedings, following a petition by the plaintiff).

266 *See, e.g.*, *Fed. Hous. Fin. Agency*, 2012 U.S. Dist. LEXIS 173768, at *33 (ordering the parties to submit early expert-study proposals after defendants brought a *Daubert* challenge to plaintiff's expert's proposed sampling methodology); *In re* Countrywide Fin. Corp. Mortg.-Backed Sec. Litig., 984 F. Supp. 2d 1021, 1025 (C.D. Cal. 2013) (same).

267 These cases all involved disputes about investments in securities backed by large pools of mortgage loans and experts retained by the plaintiff(s) to examine a representative sample of loans from these pools—a process that was commonly understood to be expensive and time-consuming even when applied only to a limited sample. Although I am not aware of early-approval requests made in the context of different types of disputes, the potential utility of an early admission determination is not unique to mortgage-backed-securities cases. Even aside from reliability concerns, any case involving costly and/or time-consuming expert analysis could be a venue for a similar request. Similarly, no procedural mechanism bars *defendants* from making similar motions, but I am not aware of any case in which defendants have taken this step. *See also infra* note 270.

268 *See, e.g.*, *MBIA Ins. Corp.*, 2010 N.Y. Misc. LEXIS 6182, at *15 (granting plaintiff's early motion in limine, noting that a court has the ability to "regulate the conduct of the trial" to promote an expedient resolution of the case before it and that deciding the motion before it was "within [that] inherent power"); *see also Nat'l Credit Union Admin. Bd.*, 2014 U.S. Dist. LEXIS 59893 at *1 (granting preapproval to the plaintiff's expert's proposed methodology, but reserving defendants' rights to file later challenges); *Fed. Hous. Fin. Agency*, 2012 U.S. Dist. LEXIS 173768, at *36–37, 61–62 (pre-approving plaintiff's proposed sampling plan before it had been executed); *Mass. Mut. Life Ins.*, 989 F. Supp. 2d at 165 (denying, at the proposal stage, motion to exclude expert testimony, but without prejudice); *In re Countrywide Fin. Corp. Mortg.-Backed Sec. Litig.*, 984 F. Supp. 2d at 1021 (ordering parties to submit proposed sampling plans). *But cf. MASTR Adjustable Rate Mortgs. Tr. 2006-OA2*, at *1–2 (declining to grant early approval because it "would either be a non-binding, advisory opinion . . . or it would be a binding ruling made on an inadequate record at the wrong juncture in the case, a shot from the judge's hip"); *Fed. Home Loan Bank of Bos.* at *2 (acknowledging that the court had the power to grant "*Daubert* pre-clearance" as "primarily [a matter] of case management," but declining to do so in this instance, because the court "likely . . . would still need to revisit the methodology, and its application" after the expert had issued a final report).

expert's ability to exploit Analytical Flexibility to his or her benefit later. The proponent may be amenable to hammering out an agreed-upon research protocol in exchange for the opponent's commitment not to oppose the expert on methodological grounds later. The parties could memorialize their agreement in a stipulation, optionally to be filed with the court.

Participation in court-run proceedings or party-orchestrated exchanges could be voluntary or mandated by the court. Courts have broad case-management powers that allow them to mandate an early process regarding expert methodology.[269] But as discussed below, parties may well be willing to submit to such a process voluntarily, especially when the process includes an adjudicative step.

There are powerful structural reasons for a party who has retained an expert to seek an early (partial) ruling on the admissibility of future expert testimony that relies on extensive or expensive analysis, in particular for plaintiffs.[270] *First*, a favorable decision on the motion reduces the risk that the expert performs expensive, time-consuming analytical work, only to be barred from testifying about it at the end of the proceeding. For a plaintiff, a favorable ruling immediately reduces the risk it faces of having to abandon its suit close to trial (or continuing its suit in substantially weakened form), as a consequence of a crucial expert being barred from testifying.[271] An unfavorable early ruling, on the other hand, might inform a plaintiff that its planned strategy for litigating its case is fatally flawed, enabling it to make informed decisions about its contin-

---

269 *See, e.g.*, FED. R. CIV. P. 16(b)(3)(B) (providing that a scheduling order may modify the timing and nature of disclosures and the scope of discovery); *Fed. Home Loan Bank of Bos.*, at *2 (stating that the court has the power to grant "*Daubert* pre-clearance" as "primarily [a matter] of case management").

270 In general, challenges of proposed expert testimony are brought much more frequently by defendants than by plaintiffs. *See* Jasanoff, *Law's Knowledge*, *supra* note 232, at S50 ("*Daubert* has been invoked most often to exclude plaintiffs' testimony . . . ."); D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALB. L. REV. 99, 108 (2000) (observing that almost ninety percent of expert challenges in civil cases are brought by defendants); *see also* Berger, *supra* note 56, at S64 ("[T]here is little point in plaintiffs going to the expense of *Daubert* motions to exclude defendant's experts until they know if their case will proceed.").

271 *See* Weisgram v. Marley Co., 528 U.S. 440, 455–56 (2000) (stating that after expert has been excluded shortly before trial, plaintiffs cannot salvage their case by substituting a new expert). While under *Weisgram v. Marley* late-stage exclusion of an essential expert is typically fatal to the plaintiff's ability to prevail on a claim that relies on the expert's testimony, substitution of a new expert (or methodology) following exclusion at an *early* stage is not precluded. *Id.*

ued litigation of the action.[272] *Second*, a favorable ruling on an early expert motion immediately strengthens a party's position in settlement negotiations. The settlement value of a case correlates with the likelihood that the plaintiff will prevail at trial, which in turn depends on the admissibility of critical expert evidence. *Third*, a plaintiff with a pre-approved expert-discovery plan may have easier access to litigation finance options. Litigation financiers may be more willing to take on the litigation risk when one significant component of that risk—a critically important expert being barred from testifying at trial—has already been eliminated.[273] *Fourth*, early proceedings relating to an expert's proposed methodology may afford a party a low-risk opportunity to gain approval of a methodology that is on the less rigorous end of the spectrum (and therefore perhaps less expensive and/or time-consuming), and opt for a more rigorous methodology only if the less rigorous option ends up being rejected first. This incremental approach toward the selection of a methodology is not available when admissibility decisions are made at the end of expert discovery, shortly before trial.[274]

Opposing parties stand to gain from early exchanges relating to expert discovery as well. *First*, an early exchange might help the party refine its strategy. It may also lead to productive negotiations between the parties about the way discovery will be structured. *Second*, by opposing an expert's proposed methodology at an early stage, the party has an opportunity to influence the work to be performed by the expert. In particular, it has an opportunity to limit the expert's ability to deploy Analytical Flexibility to his or her advantage. The more a party can limit this ability, the less likely the expert is to obtain results that can pass through the *Daubert* gate, but that in fact cannot

---

272 An illustrative example is *Homeward Residential Inc. v. Sand Canyon Corp.*, No. 12 Civ. 5067 (JFK), 2017 WL 5256760 (S.D.N.Y. Nov. 13, 2017). The plaintiff petitioned the court for permission to prove its contract claims by analyzing a sample of loans from a large pool of loans. *Id.* at *10. The court reviewed the contractual provisions at issue and concluded that a finding of liability would require a breach determination for each individual loan at issue. *Id.* at *7. Because the proposed sampling analysis would yield only aggregate information, the court rejected the sampling proposal. *Id.* at *7, *10. The plaintiff now could either proceed with an analysis of *all* loans at issue in the case or abandon its claims. *Id.* at *10.

273 On the other hand, it is possible that, on the whole, litigation funders will be willing to fund a smaller number of cases, choosing to wait and see which cases survive the discovery-plan hurdle.

274 *See Weisgram*, 528 U.S. at 455–56 (holding that plaintiffs cannot salvage their case by submitting an updated expert report or substituting a new expert).

be relied upon.[275] *Third*, as with the party proffering the expert, the party's position in settlement negotiations would improve considerably if it obtained a favorable ruling.

A court could impose conditions on a party seeking an early admissibility ruling: it could require a commitment by the party to confine its proposed expert evidence to the methodology and analysis described: that is, (1) not to expand the scope of the expert's proposed testimony, and (2) not to retain a different expert to testify on the same or similar topics. Such a restriction on the party's ability to surprise the other side with unexpected expert evidence at a later stage of the litigation may make the tradeoff appealing to the other side as well, and the parties might be able to reach mutual agreement to submit the expert's proposal to the court on conditions along these lines. In the case of a voluntary process, a party that prefers not to bind itself to such a compromise, preferring instead to retain its ability to expand the scope of its expert testimony in the future, is free to forego an early motion and take its chances at an admissibility hearing at the end of discovery, per the more traditional procedural sequence. This party would risk being faced with a *Daubert* (or equivalent) motion later, on grounds that the results its expert obtained were contaminated by Analytical Flexibility and therefore unreliable. A party that is cost-conscious, on the other hand, may well be willing to take this bargain, even if it results in limits on how it can prosecute its case.

To protect an expert's ability to plan multiple experiments but choose to present to the finder of fact only some results but not others, courts could limit an opposing party's ability to reference precleared experiments dropped from the expert's presentation. Such a limitation would have to be carefully considered, however, in light of the room it provides for file-drawer problems.[276]

---

275 *See supra* subpart II.A.

276 A limitation along these lines could be imposed in conjunction with the expert's early disclosure of a research protocol, with metes and bounds carefully delineated based on the expert's specific plans. An expert's selective reporting could be relatively harmless in a scenario where the expert performed three different experiments, each with a different goal and different experimental approach, and chose to present just one result at trial. In this scenario, barring the opposing party from inquiring about the two unreported experiments does not necessarily raise publication bias or *p*-hacking concerns; the three experiments are likely sufficiently independent that selective reporting would not falsely inflate the confidence level for the single reported study. In contrast, a scenario whereby an expert tried 100 different analyses, all aimed at proving the same hypothesis, and

An early ruling by the court could be binding or nonbinding and could result in a presumption of future admissibility or not. Just as some journals will presumptively publish a preregistered study,[277] a court could consider expert testimony based on a presubmitted protocol to be presumptively admissible, provided the expert had carried out the plan as described and intervening developments had not rendered the study irrelevant. An early ruling does not reduce the parties' freedom and narrative license to present their expert's opinions to the finder of fact in a manner of their choosing.[278]

A more limited intervention would have a party "preregister" its expert's plans with the court before execution, under seal and without disclosure to opposing parties (or even necessarily the court), and without any adjudication or other feedback process.[279] While this option would deny the parties the attendant benefits of early exchanges of information, it would serve the most important function of early proceedings: to be a commitment device. By the time the expert presented their completed work for admissibility, the party could rely on the sealed filing to demonstrate that the expert's study tested a predetermined hypothesis and had been conducted according to a predetermined protocol.

As understanding of Analytical Flexibility expands, courts could require proponents of scientific evidence to submit to early reliability proceedings as a prerequisite for having the expert's opinion admissible into evidence later. Participation in early proceedings could boost an expert's credibility and might protect the expert's opinions from being attacked on grounds of *p*-hacking and HARKing later.

It is important to recognize that early proceedings are not a panacea. Some research projects are simply too simple to be amenable to this process. In cases where the proposed study is neither time consuming nor expensive, an expert could make an end-run around the preregistration process by performing the study as many times as needed to figure out the parameters that yield favorable results, then submit a plan consisting

---

chose to present only the most favorable analysis, raises serious reliability concerns.

277 *See supra* note 208 and sources cited therein.

278 *Cf.* Nosek, Ebersole, DeHaven & Mellor, *supra* note 162, at 2604 (explaining that preregistration does not deprive researchers of their "narrative license" in explaining and presenting their data).

279 *Cf. supra* note 206 and sources cited therein (describing private, locked, and "self-archiving" options for preregistration).

of those parameters to the court.[280] Yet the fact that early proceedings are not useful in all cases should not prevent them from becoming a feature in many cases.[281]

2. *Updated Disclosure Requirements*

A second intervention that would improve the court's ability to accurately determine the reliability of litigation science could be implemented at the disclosure stage: just as journals do in academic science, the court could require more detailed disclosures in expert reports. Just as in academic science, determining whether a study has been contaminated by Analytical Flexibility requires detailed knowledge of the process by which the results were produced.[282] To determine whether an expert's opinions are reliable enough to be presented to the finder of fact, the court needs considerably more detail than is typically disclosed in an expert report.[283] Courts should demand detailed disclosures of all analytical steps performed by the expert and all deviations from a predetermined research plan.[284] Courts could use checklists to guide these disclosures.[285] If increased disclosure requirements are combined with early proceedings in which the expert commits to a hypothesis and research plan, the disclosures would be fairly straightforward: the expert would either confirm that he performed the plan exactly as described at the outset or provide detail and justification for any departures from the plan.[286]

---

280 When a study is truly inexpensive and quick to perform, a party might even ask multiple experts to perform an analysis and retain the expert who performed the most favorable analysis. *See, e.g.*, Jonah B. Gelbach, *Expert Mining and Required Disclosure*, 81 U. CHI. L. REV. 131, 131 (2014) ("[A]ttorneys can do data mining's dirty work by hiring multiple experts, asking each to provide an expert report on the same issue, and then put on the stand only the one who provides the most favorable report.").

281 Courts could require experts and retaining parties to sign a statement disclosing what expert work had been performed prior to the filing of the expert's research plan. This, too, would present opportunities for evasion, however, for example through the use of consulting experts, whose identity, retention, and work product are typically not discoverable.

282 *See supra* section II.B.2.

283 *See, e.g.*, FED. R. CIV. P. 26(a)(2)(B)(i)–(ii); N.Y. C.P.L.R. 1301(4)(d)(1) (McKinney 2014); TEX. R. CIV. P. 194.2(f)(3).

284 *Cf. supra* notes 216–217 and sources cited therein (arguing that researchers should disclose all data collected, all analytical steps made, and all departures from their preregistered plan).

285 *Cf. supra* note 217 and sources cited therein (describing use of checklists by scientific journals).

286 *Cf.* Allen & Mehler, *supra* note 137, at 3–4 (noting that a researcher can make changes to a preregistered plan, but should justify them and discuss them in the resulting publication); Open Sci. Collaboration, *supra* note 98, at 17 ("[D]etail what was changed and why.").

Courts could even choose to go so far as to require the reporting of null results—results of tests conducted that did not yield statistically significant results or that yielded results that were unfavorable to the retaining party. While this would result in a more transparent, more reliable process, it is hard to imagine that courts and litigants would welcome this significant disruption of the adversarial system.

3. *Consideration of Analytical Flexibility*

Finally, in making reliability determinations, courts need to heed the lessons of the Replication Crisis and take into account the role Analytical Flexibility is playing in research results. At the gatekeeping stage, relying on descriptions of methodology and on reported *p*-values or other indicators of confidence is insufficient to assess the reliability of research results.[287] Courts evaluating scientific evidence for admissibility should focus not just on the methodological choices, but also on the timing and manner by which those choices were made, and in particular the extent to which the process that produced the evidence left room for Analytical Flexibility. Parties have a role to play as well by bringing contamination by Analytical Flexibility to the court's attention.

The prevailing legal frameworks used to assess the reliability of scientific evidence can easily accommodate this dimension of inquiry. The *Daubert* standard instructs the court to assess whether the methodology that produced scientific evidence is "scientifically valid" by looking at a "flexible" set of indicia of reliability.[288] The *Daubert* court offered a number of factors a court might consider but stressed that they were not intended as a "definitive checklist or test." Analytical Flexibility is a consideration that could readily be added to the set.[289] The *Frye* standard that is still applied in the courts of a small number of states focuses on a narrower inquiry: whether the methods used by the expert have gained "general acceptance in the particular field to which [they] belong[ ]."[290] As measures to contain Analytical Flexibility are becoming increasingly common in academic science, an inquiry regarding their presence or absence should organically become a part of a *Frye* inquiry.

---

287 *See supra* subpart II.A.

288 *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593–94 (1993); *see also supra* Part I.

289 Jason Chin has argued that Analytical Flexibility issues can be assessed as part of *Daubert*'s "generally accepted" prong. *See* Chin, *Psychological Science's Replicability Crisis*, *supra* note 25.

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

If a court invites parties to engage in voluntary early proceedings regarding expert work and the party proffering an expert declines to participate, the court could draw an inference that the results may have been tainted by *p*-hacking or HARKing, and may therefore not be sufficiently reliable to be admissible. Without precommitment to a hypothesis and research plan, at the gatekeeping stage the party would have a hard time rebutting that inference.[291] A party that did participate in early proceedings but then had its expert present additional data that had not been precommitted should raise the court's suspicion in equal measure on the basis that there may be additional work that the expert performed but did not choose to present.[292]

IV
IMPLICATIONS

The lessons from the Replication Crisis, if heeded by the litigation community in the way suggested above, have far-reaching implications for the role of a judge, the balance between the parties, and the standards by which we judge admissibility. I discuss each in turn.

A. The Role of the Judge

Ensuring the reliability of expert evidence requires an updated conceptualization of a judge's role in the expert discovery process. By becoming involved in the process of creation of litigation science at an early stage, the judge's role evolves from that of a gatekeeper to that of a facilitator. Unlike a gatekeeper who waits until evidence is presented in its completed form and adjudicates its admissibility,[293] and unlike a referee who monitors the process of its creation,[294] a facilitator creates conditions for its generation. When a court mandates or invites filing of an expert's plan before execution—whether publicly or sealed, and whether combined with adjudication or not—the most important role it plays is as a facilitator of a commitment device. Whether the court issues a ruling on the expert's proposed plan is of secondary importance—of interest from a viewpoint of efficiency and case management, but not from the

291 *See supra* subpart II.A.
292 *Id.*
293 *See Daubert*, 509 U.S. at 579–80.
294 *See* Jasanoff, *supra* note 30, at 129.

perspective of ensuring reliability.[295] By (1) creating a commitment device and (2) demanding more extensive disclosures, the court facilitates a process that allows the type of information to be generated that is necessary to make a reliable assessment of reliability later.[296]

Judges have the power to create a cultural shift toward the use of more reliable methodology. They can encourage early information exchanges and broader disclosures of information. By acknowledging that current practice does not yield the information they need, and by facilitating procedures that allow for the creation and exchange of that information at appropriate junctures in a litigation, judges can bring into being a process that will generate the information they need to fulfill the gatekeeping function they will ultimately be serving. And by doing so, they can ensure that scientific evidence that reaches the finder of fact meets required standards of reliability.[297]

If early proceedings include an adjudicative component, the eventual gatekeeping moment becomes a less momentous occasion. When the reliability of the methodology has been affirmed at an early stage, pretrial admissibility proceedings will be limited to confirming that the expert carried out the plan as intended or to examining any deviations from the preapproved plan.[298] When the early exchange of expert plans is structured as a voluntary process between the parties without the court's involvement, then gatekeeping proceedings will still have some bite. Even in that scenario, however, the arguments available to a party opposing the testimony may be significantly curtailed: disclosure of an expert's plan to an opposing party at an early stage, followed by execution of that exact plan, should take any arguments sounding in *p*-hacking and HARKing off the table.

B. The Balance Between Plaintiffs and Defendants

If early proceedings include an adjudicative component, this might affect the power balance between plaintiffs and defendants. Much has been written about the impact the *Daubert*

---

295 To create reliable, reproducible scientific knowledge, it is essential to constrain Analytical Flexibility to the extent possible. *See* Munafò et al., *supra* note 94, at 2; Simmons, Nelson & Simonsohn, *supra* note 98, at 1362–63.

296 *See supra* subpart II.B.

297 *E.g.*, *Daubert*, 509 U.S. at 579–80; Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

298 *See supra* subpart III.B.

decision has had on parties' use of expert evidence.[299] The decision is often said to have shifted some litigation risk toward plaintiffs: since plaintiffs tend to be more reliant on expert evidence to establish their claims than defendants are in defending against them, having to meet higher standards of reliability of expert testimony tends to affect plaintiffs more than defendants.[300] Some commentators have argued that the pendulum has swung too far: that *Daubert* has made it prohibitively difficult to pursue certain types of cases, including environmental tort cases that cannot be brought successfully without time-consuming and expensive expert studies.[301] Furthermore, *Daubert* was decided at the early dawn of large-scale electronic discovery, and technological developments that followed the decision have sharpened its bite. Early adjudication proceedings in more civil actions would represent a swing back of the pendulum.

Early information exchanges would offer an opportunity for cooperative, cost-effective approaches to expert discovery, as well as early opportunities for negotiations on the scope of discovery or even settlement.[302] As discussed in subpart III.B,

---

299 *See, e.g.*, Berger, *supra* note 56, at S61–S63 (surveying the impact of *Daubert* on toxic-tort litigation); David L. Faigman, *Admissibility Regimes: The "Opinion Rule" and Other Oddities and Exceptions to Scientific Evidence, the Scientific Revolution, and Common Sense*, 36 SW. U. L. REV. 699, 700 (2008) (contrasting *Daubert* and *Frye* admissibility regimes); Haack, *supra* note 219, at 988–92 (describing the historical development of reliability standards).

300 *See generally* Berger, *supra* note 56, at S64 ("*Daubert* has undoubtedly shifted the balance between plaintiffs and defendants and made it more difficult for plaintiffs to litigate successfully."); Haack, *supra* note 219, at 990 ("[O]n the whole, at least in civil cases, *Daubert* has made it harder . . . to get scientific testimony admitted."); Risinger, *supra* note 270 (almost ninety percent of expert challenges in civil cases are brought by defendants).

301 *See* Berger & Twerski, *supra* note 36 (arguing that as standards of admissibility have become more exacting, it has become "nigh impossible" for certain plaintiffs to maintain a civil action); *see also* Carol Krafka, Meghan A. Dunn, Molly Treadway Johnson, Joe S. Cecil & Dean Miletich, *Judge and Attorney Experiences, Practices, and Concerns Regarding Expert Testimony in Federal Civil Trials*, 8 PSYCHOL. PUB. POL'Y & L. 309, 322 (2002) (noting that judges reported having excluded or limited the testimony of 25% of challenged experts pre-*Daubert*, but 41% post-*Daubert*).

302 These exchanges may be envisioned by the 2015 amendments to Federal Rules of Civil Procedure 1 and 16. *See* John Roberts, 2015 Year-End Report on the Federal Judiciary 5–7 (2015), https://www.supremecourt.gov/publicinfo/year-end/2015year-endreport.pdf [https://perma.cc/Y7LG-C4PJ] (Amended Rule 1 "highlights the point that lawyers . . . have an affirmative duty to work together, and with the court, to achieve prompt and efficient resolutions of disputes."); *see also* The Sedona Conference, *Cooperation Proclamation*, 10 SEDONA CONF. J. 331, 331–32 (2009) (urging a "paradigm shift for the discovery process" toward a culture that promotes "open and forthright information sharing" and "the development of practical tools to facilitate cooperative, collaborative, trans-

when there is an adjudicative component to early proceedings regarding the admissibility of expert testimony, plaintiffs may be willing to provide more transparency in exchange for the prospect of an early determination on admissibility. A favorable early ruling would immediately improve their litigation position, and an unfavorable early ruling would help them decide their litigation strategy—including whether to consider abandoning the case. Defendants stand to gain from early proceedings as well, by receiving information about an opposing expert at an early stage and the potential early exclusion of proposed expert methodology.

Up-front clarity about the admissibility of expensive or time-consuming methodologies may lower the barrier to litigation for a putative plaintiff who is considering whether to bring a lawsuit. A plaintiff might hesitate to bring suit when the expert evidence required to bring a claim is very costly to acquire and the risk of the evidence being excluded is significant.[303] An early determination on the admissibility of an expert's proposed methodology lowers that risk and may unlock access to justice to a larger group of aspiring plaintiffs with potentially meritorious claims who might not otherwise be able to bring suit: if the court determines that the proposed methodology cannot support admissible evidence, it does so before the plaintiff has spent considerable resources on procuring the evidence.[304] Additionally, as discussed above, a plaintiff seeking to engage in a costly litigation may have better access to litigation finance if a major source of litigation risk has been removed at an early stage in the proceeding.[305] When the risk is reduced that they will be unable to present expert testimony at trial or use it at the summary-judgment stage, plaintiffs may be able to obtain litigation finance more easily and at more favorable terms.[306]

Separating assessments of methodology from assessments of application may change parties' approach to objections to expert evidence: when results of a study are not yet known, the methodology that will produce them may be more capable of being judged on the merits.

---

parent discovery" and arguing that "[c]ooperation does not conflict with the advancement of [lawyers'] clients' interests [but rather] enhances it").

303 Berger & Twerski, *supra* note 36.

304 *See supra* text accompanying notes 270–272.

305 *See supra* subpart III.B.

306 *See id.*

If early proceedings relating to expert evidence with an adjudicatory component were to become more prevalent, over time courts would create a valuable library of precedents that currently does not exist. In light of the "Vanishing Trial" and the reality that the vast majority of civil litigations settle before they reach trial or even the summary-judgment stage, the current custom of adjudicating evidentiary issues toward the very end of proceedings limits the development of a body of precedent—the vast majority of cases are terminated before there ever is a ruling on the admissibility of any of the available evidence.[307] If courts instead routinely ruled on the admissibility of expert evidence at the outset of discovery, the body of relevant case law would grow at a more rapid pace.[308] A more robust record of evidentiary decisions is beneficial to all players in the field: plaintiffs and defendants both benefit from increased predictability of the outcome of evidentiary motions. Parties to disputes that still linger in the pre-litigation stage would also benefit from access to a broader pool of decisions. Having a more developed record of the types of evidence that courts have been willing to admit in support of the contemplated claims would allow parties to predict with more accuracy their ability to produce evidence that will meet the court's requirements, and better assess their chances in litigation.

CONCLUSION

The frameworks that apply to the admissibility of scientific evidence in civil litigation rely on an implicit assumption: that the court assessing the reliability of the evidence will have the information needed to make that assessment at the time it is called upon to do so.[309] The knowledge crisis that roiled the scientific community has cast serious doubt on the validity of this assumption. It has thrown the spotlight on a dimension of reliability that is typically overlooked in admissibility proceed-

---

307 *See* John H. Langbein, *The Disappearance of Civil Trial in the United States*, 122 YALE L.J. 522, 522 (2012) ("Since the 1930s, the proportion of civil cases concluded at trial has declined from about 20% to below 2% in the federal courts and below 1% in state courts."); Nat'l Ctr. for State Courts, *supra* note 58 ("[T]he failure to 'survive' the *Daubert* challenge is contributing to the Vanishing Trial . . . ."); Frederick Schauer, *On the Supposed Jury-Dependence of Evidence Law*, 155 U. PENN. L. REV. 165, 172 (2006) ("[T]he number of criminal and civil jury trials has declined substantially in recent years.").

308 Similar arguments might well be made with respect to other types of evidence. Indeed, perhaps early proceedings on the admissibility of *any* type of evidence would be a welcome development. But such arguments are beyond the scope of this paper.

309 *See supra* Part I.

ings and in fact cannot be evaluated in these proceedings as they are currently conducted.[310]

The Analytical Flexibility that stands at the root of the problem cannot be eradicated entirely, but as the recent systematic methodological inquiry in the social sciences has proven, a legal community that is motivated to curb its impact has a wealth of tools at its disposal.[311]

This Article aims to start a conversation about the tools that might work in a civil-litigation context, suggesting a reframing of the courts' role in the creation of scientific evidence and a series of procedural and methodological reforms for the treatment of science in the courts.[312] If courts are to transition to the role of a facilitator, in addition to their role as gatekeeper or referee, further work is necessary to explore optimal ways of making this transition—ways that promote both transparency and efficiency, while preserving the parties' rights to fundamental fairness.

A broader conceptual discussion may be called for as well. Almost three decades have passed since the *Daubert* decision, and during that time, the architecture of scientific practice has changed dramatically. In addition to the developments at the center of this Article, novel mechanisms for the sharing and reviewing of ideas and data, such as open-data platforms[313] and post-publication or "continuous" peer review,[314] have changed the nature of the interactions between researchers, the science they produce, and those who consume it. This Article focused on procedural and cultural changes that are possible within the dominant legal frameworks for the creation and evaluation of scientific evidence in the civil context. It is worth exploring whether even greater progress could be made if

---

310 *See supra* subparts II.A & III.A.

311 *See supra* subpart II.B.

312 *See supra* subpart III.B.

313 *See, e.g.*, Crüwell et al., *supra* note 144, at 11–13 (describing open data practices aimed at improving reliability and reproducibility); Rafael C. Jiménez et al., *Four Simple Recommendations to Encourage Best Practices in Research Software*, F1000RESEARCH (June 13, 2017), https://f1000research.com/articles/6-876 [https://perma.cc/Q54G-THLJ] (recommending practices "designed around Open Source values" for the development of software used in research); Washburn et al., *supra* note 151, at 169 (reporting researchers' views on open-data practices); Vanpaemel, Vermorgen, Deriemaecker & Storms, *supra* note 216 (assessing empirically whether researchers are willing to provide their research data).

314 *See* Nosek, Spies & Motyl, *supra* note 98, at 623; Jonathan P. Tennant et al., *A Multi-Disciplinary Perspective on Emergent and Future Innovations in Peer Review*, F1000RESEARCH (Nov. 29, 2017), https://f1000research.com/articles/6-1151 [https://perma.cc/ST4H-QHWZ].

those frameworks were to be replaced by a new structure—one that reflects the open and iterative nature of scientific practice today. The just outcome of litigations that depend on the reliability of scientific evidence demands a continued engagement with these questions.