Beyond *CSI*: Calibrating public beliefs about the reliability of forensic science through openness
and transparency

Carlos Miguel Ibaviosa

School of Psychology, University of Adelaide

Jason M. Chin

School of Law, University of Sydney

*Please cite as: Carlos Miguel Ibaviosa & Jason M Chin “Beyond CSI: Calibrating public beliefs about the reliability of forensic science through openness and transparency” (accepted 2022) Science & Justice. <https://doi.org/10.31222/osf.io/tvcm6>.

**We thank the two anonymous reviewers for their insightful comments and feedback. We are also grateful to Kristy Martire for sharing some of the data visualized in Part II.

Abstract

What drives public beliefs about the credibility of a scientific field? This question is increasingly important, with recent discussion of a “reproducibility crisis” affecting many fields. Such discussions are vital in forensic science, a discipline that has experienced severe scrutiny from both the media and large oversight bodies. In this paper, we make three contributions to this discussion. First, we bring together and compare several studies in which laypeople were asked about the reliability of forensic science practices. This review suggests that forensic practices do not enjoy uniformly high reliability ratings from the public and these ratings are not calibrated with the scientific consensus. We then review three empirically-tested ways that other fields are dealing with their own crises, all centred around transparency and openness. Finally, we make recommendations for how forensic science can leverage transparency and openness to improve and maintain its long-term credibility. As part of these recommendations, we find that empirical research supports the Houston Forensic Science Center’s recent claims that it has improved its credibility through openness and transparency.

Part I. Introduction

If a critical mass of scientists become untrustworthy, a tipping point is possible in which the scientific enterprise itself becomes inherently corrupt and public trust is lost, risking a new dark age with devastating consequences to humanity. [1]

Courts and legislators often rely on openness and transparency to foster trust in their processes and decisions. For example, courtrooms in many jurisdictions are open to the public, and judges' reasons for decisions are regularly published. Forensic science is an important gear in the modern legal system, but its underlying research and practices remain largely opaque [2]. In this article, we will suggest that if forensic science wishes to retain trust and credibility in its work – especially in light of growing controversy and criticism (in forensic science and beyond, see the above epigraph) – it should seek to do so by embracing new norms of openness and transparency.

The stakes are high when it comes to the public's views on forensic science. For instance, courts have expressed concern that some jurors may view forensic science as more accurate than is warranted, producing outcomes that unfairly favour the party tendering the evidence, which in turn increases the risk of wrongful convictions [3,4]. Similarly, rejecting forensic scientific claims that have withstood substantial critical appraisal (e.g., DNA identification from a single source sample) is clearly inappropriate and could lead to wrongful acquittals [5]. Ultimately, the preferable outcome is calibration - where the public's views of forensic scientific research claims reflect the strength of the evidence behind them. In this article, we will rely on recent psychological and metascientific research to suggest that one path to this end – and one that has been sorely missing – is for forensic scientists to conduct their research and practice more openly and transparently. Indeed, as the National Academy of Sciences recently stated:

[...] research conducted openly and transparently leads to better science. Claims are more likely to be credible—or found wanting—when they can be reviewed, critiqued, extended, and reproduced by others. [6]

This article progresses in the following order. In Part II, we discuss earlier research on the CSI Effect, which generally theorizes that the media may have misled forensic drama viewers to believe an unrealistically rosy view of forensic science. We then synthesize more recent research on public perceptions of forensic science, which suggest that the actual state of affairs is more complex than CSI theories suggest, and that there is little evidence the public has a uniformly high view of forensic practices. Then, in Part III, we review three empirically-guided ways other scientific fields are attempting to restore and maintain their credibility. Part IV provides specific ways in which forensic science can implement the recommendations in Part III. Part V concludes.

Part II. The perceived credibility of forensic science

Forensic science evidence is at a crossroads. Over the last two decades, forensic science claims and methods have been subject to a growing chorus of academic and scientific criticism. [...] This attention and criticism have expanded public awareness and spurred nontrivial reforms and meaningful institutional and research engagements. [7]

We begin this part by providing the context: a well-publicized body of research that claimed that the general public may have labored under an unrealistically high view of forensic science because of how the field was portrayed in popular culture. We then turn to more recent studies indicating that the public's current view of forensic science is not uniformly positive and varies considerably by practice. As part of this section, we present data from several studies together for the first time, in some cases relying on raw data that has been shared with us. One

driver of perceptions of forensic science found in more recent studies, as Mnookin suggests in the above epigraph, may be widely publicized non-fictional critiques of the field. This shifting of the discussion around public views of forensic science from fictional to non-fictional influences is important because, as we see in the remainder of the article, it brings forensic science into a wider discussion about whether public trust in science is decreasing due to recent controversies, and how a field should react in the face of a widely-publicized crisis.

Forensic fictions

Some of the earliest research concerning the public's beliefs about forensic science responded to the rise of fictional television shows about forensic professionals (the "CSI Effect") [8,9]. The CSI Effect encompasses several (sometimes competing) predictions about the effect of shows like *CSI: Crime Scene Investigation* (we will refer to both this show and the genre it spawned as "*CSI*" for convenience) on the operation of the criminal legal system. All of these theories, however, flow from the governing premise that *CSI* depicts forensic science as unrealistically accurate and technologically sophisticated, and practiced by highly qualified individuals only motivated to get to the truth. This unrealistic depiction then supposedly impacts jurors and other members in the legal system [10].

Under one hypothesis, jurors' *CSI* viewership advantages defendants by creating an impossible-to-meet expectation of technologically sophisticated forensic science [10]. That is, if shows such as *CSI* lead jurors to believe that sophisticated forensic evidence will always be presented to them, then they may be inclined toward acquitting if no forensic evidence is present at trial. Another hypothesis suggests that *CSI* viewership influences juries in the opposite direction. This hypothesis indicates *CSI* leads juries to believe forensic science is more

sophisticated than it is in reality, which in turn influences them to overweight forensic evidence and convict defendants based on this evidence.

While early research surveying legal actors suggest that these *CSI* effects are widely believed to exist among the legal community and often influences how legal actors (particularly prosecutors) behave in trial [11], such concerns have not been supported by subsequent research [9]. Despite early empirical research finding associations between *CSI* viewership and juror decision making [12-14], studies in this area contain serious methodological limitations, such as small sample sizes and fundamental misunderstandings of statistics that call into question the validity of the results [15]. Such an effect may be hard to find because there are many factors that affect jurors' attitudes, such as other types of evidence presented or the science communication itself, with familiarity with a TV show being just one potential - and possibly small - contributor.

Other research has investigated the simpler question of whether *CSI* viewership affects the public's beliefs about forensic science credibility. In light of previous claims that various forensic science methods are infallible, it is reasonable to expect the general public to inflate estimates of forensic science credibility if exposed to mythic portrayals or false claims that have historically been associated with the field. However, studies attempting to measure this relationship have found inconsistent results, with some finding no relationship and others finding a weak one [14,16-19] (see footnotes for a brief summary).¹ This may be due to the fact that many people report that they do not think *CSI* is an accurate depiction of real life [20], as well as other factors, such as pre-existing beliefs about forensic science outside of *CSI* shows.

¹ [14] found that *CSI* viewership raises all evidentiary expectations. [16] and [17] found no effect. [18] found a correlation between *CSI* viewership and confidence placed on fingerprint evidence. [19] found that *CSI* viewership raises judgements of reliability of DNA evidence, though this is all they asked about.

Such results indicate that, despite the concerns of various legal actors, the public's views of forensic science are not strongly related to CSI shows and forensic myths. In fact, as we will discuss now, several studies have found that the public does not see forensic science practices as uniformly reliable.

Public beliefs about forensic science

Five studies over 11 years have measured the public's belief about the reliability of several forensic science practices using roughly the same methodology. Although these were not the only studies, as will be discussed, they did use comparable response scales and so will be the focus of our discussion (see Table 1 and Figure 1). We draw three main inferences from these studies. First, the public views many practices (even DNA in the most recent study) as quite fallible, which is inconsistent with a strong CSI Effect. Second, there is considerable miscalibration between public views of reliability and the scientific consensus. Third, there may be early indications that public perceptions of forensic science's reliability are falling. The data (some provided by the original authors) and code underlying our analysis and visualisations are publicly available (<https://osf.io/7k52r/>).

One of the earliest estimates of the public's beliefs about forensic science's reliability came from Lieberman and colleagues' 2008 study of students and jurors [16].² In the Lieberman study, the authors asked participants to rate the "accuracy" of several forensic methods on a scale from 0% to 100% accurate. Responses from students and jurors were consistent with each other

² While Lieberman et al. is not the first study to measure perceptions of forensic science, and several earlier studies have measured perceptions of forensic evidence in relation to juror decision making and evidence evaluation, e.g., [21-24], it appears to be the first to do so in a way that can be easily compared to several later studies because they all used very similar rating scales. For related studies, see [25-27] using categorical measurement, such as "Completely reliable" or "Somewhat reliable". These studies are difficult to place on the same scale as the ones we review in depth, but show a similar trend.

(see Table 1). They assigned high accuracy to DNA and fingerprint analysis (~90%) [16].³ Hair and fibre analysis was lower, in the high 80s, and drug and alcohol tests were rated around 80% accurate [16].

Similar surveys – albeit using slightly different wording and focused on different populations – have subsequently found what could cautiously be interpreted as a downward trend for some practices in a way that does not seem well-calibrated to the underlying research (see Figures 1 and 2). For instance, Martire and colleagues asked participants to rate the “reliability” of forensics methods from 0-10, collecting responses in 2011, 2012 and 2016 (the 2016 survey used a 0-100 scale that was later transformed to a 0-10 scale) [28]. While most of the forensic methods assessed showed no significant differences in perceived credibility over time, several findings are worth noting. Compared to the 2008 study, there may be a very slight downward trend with respect to DNA, which is one of the most well-tested forms of forensic evidence. The approximately 94% accuracy of DNA measured in the 2008 study dropped to approximately 88% in 2016 (see Figures 1 and 2). While this could be due to many factors, such as different wording across the Lieberman and Martire studies, an analysis of Martire’s data between 2011 (~93%) and 2016 (~88%) nevertheless reveals non-overlapping 95% confidence intervals when variables such as wording, participant demographic, lay sample, and measurement are held constant. Additionally, there were also statistically significant decreases with respect to dental/bitemark analysis and hair analysis (i.e., from ~80% to ~70% for each discipline) [28]. The authors note, however, that these decreases are very small.

³ See also [26], who found that 64% of participants rated DNA evidence as ‘extremely reliable’.

Two recent studies reinforce the notion that forensic science does not carry the aura of infallibility that it may have once carried. The first of these, conducted by Ribeiro and colleagues in 2015 [17], studied Australians and asked them to rate the accuracy of several forensic methods in a manner very similar to the 2008 study (i.e., accuracy on a 0-100 scale, see Appendix A for a list of differences between studies). Ribeiro and colleagues expressly hypothesized that participants would hold a higher view of forensic science than their data ultimately supported:

We hypothesized [...] that the estimates of accuracy for each technique would be high (i.e., more than 90%). We found that participants' estimates of accuracy for each forensic technique ranged from 65.18% (document) up to 89.95% (DNA) [...] our predictions [...] were not supported as all forensic techniques were perceived to have an accuracy of less than 90%.

However, Ribeiro and colleagues did not have the benefit of the Martire study and may have formed lower hypotheses if they had.

The final study was performed in 2019 by a group of criminologists in the US. It used the exact same questions and wording as the Ribeiro study (see Appendix A). The responses, in many cases, were considerably lower than previous estimates. Compared to the 2008 study, which found perceived accuracy in the low 90% range for DNA and fingerprints, this study's participants reported 83% accuracy for DNA and 77% for fingerprints. Responses for the rest of the practices ranged from 76% to 55% (see Table 1) - all of which were rated significantly lower than the estimates provided in Ribeiro and colleagues' paper.

Given the variety of other differences between these studies that might also explain the pattern of results (see Appendix A), it is too early to say if the public is beginning to think less of some forensic practices. However, the results above should serve as a red flag for forensic scientists - particularly for practices such as DNA, fingerprints, hair and fibre analysis (see

Figures 1 and 2). Notably, DNA analysis, which is considered forensic science's "gold standard" for investigation and prosecution [5], did not fare well in recent studies.⁴ While DNA, along with fingerprints, is generally perceived as highly reliable in accordance with the current evidence base [5], it was not rated very differently to forensic dentistry/bitemark analysis, which may be cause for concern. While it is unclear whether participants in these studies distinguished between bitemark analysis used in suspect identification and dental examination used to identify dead bodies (the latter of which is considered much more reliable) [29], it would be concerning if participants' ratings reflected beliefs about bitemark analysis. Bitemark analysis has never been validated, likely has a high error rate, and has contributed to many wrongful convictions - and so a public perception that weighs DNA analysis similarly to bitemarks would be poorly calibrated to the evidence [5,20,30]. By that token, in the Ribeiro and colleagues study, forensic dentistry/bitemark evidence was rated quite similarly to fingerprint evidence (89.26% and 88.15% respectively) - despite fingerprint evidence having a much stronger research basis than bitemarks [17]. The studies also revealed other miscalibrations between respondent perceptions and what the evidence base currently suggests - such as lower than warranted views on document analysis, and possibly higher than warranted views on firearm/toolmark analysis [17,20,28].

Outside of these studies, similar concerns around miscalibrations can be found in research using different methodological approaches to measuring lay perceptions of forensic science. Lawson's 2014 study [31], for example, suggested that jurors found bitemark evidence as significantly more reliable than toolmark evidence, despite that proposition finding little support in extant research. Another study from Crozier and colleagues (discussed further in Part III)

⁴ It is possible respondents were thinking about DNA analysis from multiple-source samples, which can be less reliable. However, this is probably not common knowledge.

revealed that participants found fingerprint and bite mark evidence to be equally convincing and elicited similar conviction rates in a hypothetical scenario, echoing concerns from the Ribeiro et al study [32]. Similarly, Koehler's 2016 survey [33]⁵ suggested that bite mark and hair evidence were more persuasive to jurors than their evidence base warrants.⁶

Forensic non-fictions as potential influences for current attitudes towards forensic science

If the public's view of some forensic methods is not as positive as might be expected under a CSI Effect, then why might that be? One plausible reason is, as Mnookin suggests, that people are becoming sensitive to the reality of forensic science. Indeed, in the 1980s and 90s, DNA analysis fundamentally changed the legal system's view of wrongful convictions and forensic science. The technology, statistics, and methods underlying the comparison of two DNA profiles arose – unlike many other forensic practices – from a multidisciplinary academic scientific tradition [5]. The result is a method that, for the most part [5], has been rigorously tested and scrutinized in both courts and labs [5]. DNA analysis's robustness has, in turn, allowed for unprecedented error detection: post-conviction DNA testing has revealed over 350 wrongful convictions in the USA [36]. It has also found that DNA's less precise siblings were responsible for many of those miscarriages of justice [37]. As such, forensic science's role in producing miscarriages of justice attracted academic scrutiny and eventually critical reviews by institutional scientific bodies [5,38].

⁵ This study investigated how likely jurors believed that a false positive error would occur in various forensic science disciplines. The median estimates ranked DNA as the most reliable discipline (1 in 10,000,000), followed by fingerprints (1 in 5,500,000), then bite marks and microscopic hair evidence (both ranked as 1 in 1,000,000), then handwriting evidence (1 in 100,000).

⁶ See [27] for lawyer and juror perceptions of fingerprint evidence; see also [34] Murrie et al. (2019) for self-assessments from forensic analysts; and see [35] Garrett et al. (2021) for judges' perceptions of a range of forensic disciplines

These troubles in forensic science have attracted public awareness [7]. This coverage spans the various facets of the problem, from the failures to establish a research base to mistakes made in individual cases [39,40]. Coverage also takes different forms, from articles generated by newspapers of record, to evocative case studies in documentaries and in new media [39-43]. Moreover, media attention has focused both on the mistakes of individual forensic practitioners [44], and on the field as a whole [45].

We cannot say for sure that the results we observed above (i.e., lower than expected public views of forensic science in recent surveys, a lack of calibration with the scientific consensus) are related to well-publicized errors and shortcomings. Indeed, Martire and colleagues state that this hypothesis is constrained by the fact that not all controversial evidence types (i.e., confession evidence) experienced a decrease in public perception [28]. However, it seems plausible that wrongful convictions and negative media coverage would be at least somewhat influential in informing public beliefs about the field's reliability, given reports of changing perceptions among the legal community [46-48].

In any case, recent public controversies related to forensic science raise questions about how forensic scientific fields may seek to retain or regain their credibility, and how to communicate that credibility to lay or even skeptical audiences. Indeed, even as the field embarks on a path forward by conducting research to validate its methods, there is an appreciable risk that public perceptions will lag behind in those advances. To mitigate that risk, cross-disciplinary insights are valuable, especially those examining how laypeople react when hearing about uncertainties and errors in research, and how other scientific fields have sought to bolster their credibility. We now turn to such research.

Part III. The credibility of science

Outside of forensic science, the broader scientific enterprise is no stranger to public criticism and scandals. In many ways, such criticism has become more salient in recent years, with several fields experiencing what has been called a crisis of credibility following a litany of results that could not be replicated by independent labs [49-51]. This crisis reinvigorated theoretical and empirical work aimed at understanding why some research is credible and some is not, and how to communicate that uncertainty outside of scientific circles [52-54]. If forensic science wishes to improve its credibility as a field, it may wish to examine the reforms being developed in other fields, in which the response has centered around principles of humility, openness, and transparency.⁷

Specifically, this part reviews research concerning (1) the role of acknowledging mistakes and uncertainty in how scientists are perceived by the public, and in particular, the positive impacts of such disclosure on credibility. We then move on to (2) openness and transparency in research and practice and how these features promote public trust. The final section of this part discusses (3) the public's expectations that scientists will fully report their findings and share their data. This interdisciplinary perspective is useful in that it directly studies scientists' expectations about the effects of disclosing problems with their research, finding those expectations overstate the reality. In doing so, it bridges forensic science research with the larger research ecosystem and reinforces work within forensic science suggesting the field should be more willing to acknowledge error and uncertainty [32,56,57].⁸ It also introduces the construct of

⁷ These recommendations align well with longstanding recommendations for improving forensic science, see [55]. The new research we discuss, however, lends additional force to these suggestions.

⁸ "Based on our conversations with practitioners, they may fear that even a single error on a proficiency test might destroy an examiner's credibility in the eyes of the jury, or that information about such testing might induce general skepticism of their entire domain." [32]

“epistemic trust”, which goes beyond perceptions of a scientists’ expertise to lay people’s perceptions that scientists are working in the public’s interest.

Acknowledging uncertainty and mistakes, and the role of epistemic trust

Scientific progress often occurs through acknowledging previous mistakes and misapprehensions and correcting them [58]. In the field of forensic science, this might include previous claims of infallibility not holding up to modern scrutiny, or mistakes made by individual practitioners. When scientists are confronted with mistakes, prevailing scientific norms suggest that researchers should acknowledge such errors and update their conclusions in line with new evidence [59]. In reality, however, scientists may be reluctant to disclose their mistakes for fear of being perceived as a poor scientist [60]. This is a valid concern; however, as we will discuss, there may actually be reputational benefits to admitting personal error.

In one recent study, Fetterman and Sassenberg found that scientists overestimated the negative reputational consequences of their research failing to replicate [53]. Specifically, the authors asked a sample of scientists to imagine a set of scenarios, either concerning their finding that they were personally proud of or a famous finding from another scientist. In both conditions, they were asked to imagine that another group of researchers had attempted to replicate this work twice and had failed, with the failed replications being published and blogged about. Participants were also asked to imagine one of two responses: either a reply from the original author admitting that they were wrong or a non-admission dismissing the results of the non-replication. All participants were then asked to answer several questions about how the original author’s reputation would be affected by the non-replication. Fetterman and Sassenberg found that scientists perceived greater reputational damage to their own work as compared what they believed about others’ work. Or, in other words, that they overestimated the reputational

consequences of the replication failure. Moreover, the authors also found that admitting mistakes generally led to reputational benefits [53]. That is, participants who read about the admission rated the original finding as more likely to be an honest mistake and ascribed a higher reputation to the scientist who made the admission (as compared to scientists who refused to make such concessions).

These findings about admissions are consistent with work on the factors that cause us to trust the knowledge provided by others, a construct known as “epistemic trust” [54,61]. Epistemic trust has three components: expertise, benevolence, and integrity. “Expertise” pertains to the researcher’s perceived competence (the focus of many mock juror studies in forensic science, as will be discussed), while “benevolence” refers to the researcher’s interest in improving society, and “integrity” describes whether the researcher follows prescriptive scientific norms and principles. Across two studies, Altenmuller and colleagues found that when scientists acknowledged prior errors in their research, they were perceived as being more benevolent and having greater integrity than when they denied making such errors (with either no change or a slight increase in perceived expertise) [54]. When compared to a control group where no comments were made at all, they found that admitting errors was no different on all facets of epistemic trust.

Similarly, another study by Hendriks and colleagues suggested that self-disclosing errors led to increased perceptions of benevolence and integrity compared to a no-comment control group, although there was a slight decrease in expertise (which, as Altenmuller and colleagues explain, may be due to methodological differences in framing the admission) [54,61]. The Hendriks et al. study also found that authors who self-disclosed were seen as more benevolent and had greater integrity than when external researchers criticized the work. Taken together,

these studies suggest that self-disclosure of errors generally does not harm scientists' reputations, contrary to widespread concerns, and may actually make scientists more trustworthy in important ways (i.e., benevolence and integrity) [54].

The value of acknowledging mistakes, and research on epistemic trust more generally, may be of great interest to forensic practitioners. Their role, in many cases, is to convey specialized knowledge about their field to laypeople. This process requires considerable epistemic trust because laypeople "usually have insufficient background knowledge to deal with scientific evidence directly, and assess, for example, its coherence, veracity, or adequacy" [61]. In light of recent controversies, forensic researchers and practitioners may therefore do themselves and their field well by acknowledging previous overstatements and mistakes as a way to advance the field's epistemic trustworthiness. A failure to do so may be perceived as denial of such errors, or as criticisms only coming from external actors - each of which seems to be perceived as less trustworthy than internal admissions.

Supporting this work on error admissions and epistemic trust in science generally, research in forensic science suggests practitioners may overestimate the negative effects of error and uncertainty disclosure. This research has occurred largely in the context of mock juror studies. These studies, focusing on feature comparison practices like firearms [62], bitemark [32,63], and fingerprint analysis [18,32,64,65], tend to find only a modest relationship between the stated limitations of the evidence and mock jurors' assessment of that evidence (and the associated expert).

Mock jurors will, for example, take into account a forensic examiner's proficiency when assessing their evidence and expertise [32,64]. In one set of studies, Crozier and colleagues informed some mock jurors that a firearms or fingerprint expert had received either perfect

results on blind proficiency tests or had made a substantial number of errors (the control group received no proficiency information) [32]. Compared to the control group, participants in the lower proficiency expert group did downgrade their ratings of experts' competency and skill, but still rated them highly on those metrics (roughly 6 on a 1-7 scale in which 7 indicated extremely competent and skilled, compared to roughly 6.5 for the control group). Indeed, the authors noted that their results suggested that concerns about reporting proficiency tests were "overblown" and that the effect of revealing proficiency tests was "objectively quite small" [32]. In a similar study [18], the examiner in the mock trial scenario acknowledged that recent studies found that errors occur in fingerprint examination. This reduced mock jurors' estimated probability that the defendant left the fingerprint, but they still judged it to be about 70% likely (when the expert evidence was the only evidence on this issue).⁹

Taken together with the broader research on epistemic trust, results from mock juror studies suggest that acknowledging error rates and proficiency information in the forensic sciences may both assist jurors in making more informed decisions about the quality of the evidence, while at the same time, not causing them to lose trust in the expert and their ability.

Promoting openness and transparency

Along with acknowledging mistakes and uncertainties, scientists are increasingly interested in improving the credibility of their work through transparency and openness [2,67]. For instance, many journal editors and funders have endorsed journal guidelines (the "Transparency and Openness Promotion" guidelines) that require or encourage authors to make

⁹ Converging with these findings, other work finds only a weak relationship between expressions of uncertainty in an expert's opinion and the weight mock jurors accord those opinions: [18,62,65]. On the importance of careful reporting generally, see [66].

their data and materials available for scrutiny and reuse [68]. In line with those guidelines, researchers across several social sciences are increasingly engaging in more transparent and open methods [68].

In her 2017 paper, Vazire suggested that adopting open and transparent scientific practices could prevent a ‘lemons market’ of scientific research [69, see also 70]. The idea of a lemons market was originally illustrated in the context of used cars [71]. In short, such a market can occur when consumers lack information about the quality of the product, and if that quality uncertainty is extreme enough, then buyers lose trust in all the products (even the high-quality ones) and the market collapses. In other words, if a consumer cannot tell the good from the bad, they will place a lower value on all such products. In this vein, the public may lose trust in science when there is a great deal of low-quality science that cannot be distinguished from high-quality work. Publishing raw data, analysis scripts, and methods along with studies, on the other hand, may alleviate negative market pressures by allowing the public (and other scientists) to examine the entire process.

There is emerging empirical data to suggest that embracing openness and transparency could improve credibility in this manner. First, consider the results of a large 2019 Pew survey in the United States. Troublingly, survey respondents generally did not think that scientists admit and take responsibility for their mistakes. Specifically, only 13% said medical researchers admitted and took responsibility for their mistakes all or most of the time [72]. That figure was 18% for nutrition researchers and 16% for environmental researchers. While the authors of this survey did not ask about forensic researchers, there is little reason to think they would have fared much better. There is hope however: when asked if they would be more apt to trust research when the data is openly available, 57% of respondents agreed with that statement.

These findings are in line with other recent research. When measuring perceptions of epistemic trust, Altenmuller and colleagues found that scientists who vocalised even minor intentions to commit to openness and transparency in their studies were perceived as more trustworthy on all aspects of epistemic trust compared to those who dismissed the need for such practices or were undecided about their implementation [54]. This was also the case for scientists who described how openness and transparency reforms might benefit the scientific enterprise [54]. Similarly, Song and colleagues performed a recent study in which they first provided lay participants with a primer on several open and transparent research practices (e.g., registering a study's methods and hypotheses on a public database prior to conducting the study, see Part IV), and then randomly assigned them to read and assess abstracts that, in some conditions, contained such practices [73]. Participants reading the abstracts containing open and transparent practice rated the associated scientist as more credible and trustworthy.

These results provide early evidence that being open and transparent in forensic science research could improve credibility in a similar manner. However, it is important to note that these studies have somewhat contradictory results to other studies, which found smaller or null effects [see 74,75]. Such research focused on more expert audiences and required more technical knowledge. The studies that we have mentioned, however, pertain to lay people's reactions to public reporting of science (e.g., the science blog studies, the Pew survey), and would be more applicable to lay people interpreting forensic science claims (in court or otherwise). Nevertheless, we encourage further studies to focus on forensic science in order to fully investigate the effects of openness and transparency within this field.

Aligning with public expectations

Finally, forensic science may enhance its public standing by aligning its practices with the public's expectations about how research ought to be conducted and applied. Consider, for instance, recent work by Pickett and Roche [76]. In this study, participants were asked what they thought about fraudulent practices (i.e., data falsification/fabrication) that are widely condemned within science, compared to 'questionable' practices (i.e., selective reporting) that are widely used but can demonstrably bias the scientific literature [77,78]. The examples of the questionable practices the researchers provided were:

- (1) A scientist runs several experiments, but only reports those experiments that show what he/she wants to see
- (2) A scientist analyzes several outcomes (e.g., lung cancer, liver cancer, throat cancer) in a single study, but only reports the outcome that shows the results that he/she wants to see
- (3) A scientist tries several different methods for analyzing the data, but only reports the results for the method that show what he/she wants to see [76]

Not surprisingly, Pickett and Roche found that 96% of participants reported that research fraud was unacceptable and 93% said that researchers who commit fraud should lose their funding. However, they also found that 71% said that the widely used *questionable* practices reporting were morally unacceptable and 73% said that researchers who use such practices should lose their funding. In both cases, participants agreed that such practices should be met with sanctions - with the majority of participants stating that authors who commit such practices should be fined, and some participants suggesting incarceration.

While we found no similar research that focused on public expectations towards forensic science practices, there is little reason to believe that forensic scientists would be perceived any differently. There are clear analogues to such questionable practices among forensic scientists - for example, conducting multiple tests but only reporting those that confirm the prosecution's theory. Additionally, the punitive attitudes recorded in Pickett and Roche's study also seem to be linked to several widely held beliefs: that science can be incredibly influential to society and people's wellbeing, that such practices are a waste of taxpayer money, and that they undermine public trust [79,80]. Each of these beliefs also apply to forensic sciences.

These findings about what scientific behaviour the public finds condemnable complements surveys about what the public *wishes* scientists would do when handling data. For example, one recent study conducted by Bottesini and Vazire asked over 1,200 people about open data practices [81]. The researchers asked if de-identified data used in one publication should 'be shared with others who wish to verify the validity of the scientific findings' and 'be shared with others who wish to use the data to test different hypotheses'. They found that 93% thought data should be shared for verification and 90% thought it should be shared for reuse. In another study, when debriefing actual participants about how they wished their data would be used, 54.4% of participants ranked publishing on an open database as their first choice, 32.5% ranked sharing upon request as their first choice, and 13.2% ranked no sharing as their first choice.

Taken together with the previous studies, these results suggest that the public expects scientists to be honest and open with their data, while condemning fraudulent or misleading claims. When considered among the other benefits of transparency and epistemic trust, it seems clear that the more reliably scientific values are upheld, the more credible and trustworthy the

field becomes - to both the public and other scientists. However, given individual practitioners' reluctance to admit mistakes, it seems that if forensic science is to live up to these values, we must rely less on the virtue of individuals, and more on implementing proper reforms and incentives towards these norms on a systems level. What such reforms may look like in practice will now be discussed.

Part IV. How to improve the openness and transparency of forensic science

Unbiased and impartial science is facilitated through transparent relationships. Within laboratory operations, quality assurance protocols and accreditations are forms of ensuring and documenting/presenting organizational transparency publicly. The importance of transparency within forensic science cannot be discounted as it ensures trust in the organization-stakeholder (i.e., consumer) relationship via disclosure, clarity, and accuracy. A lack of transparency and clear working relationships can lead to conflicts of interest, which are the greatest threats to the maintenance of scientific integrity.¹⁰ [82]

As we have discussed, scientists have several reforms available to them to successfully earn and communicate trust and credibility of their discipline. Many of these reforms are also applicable to forensic practitioners and researchers. In this section, we will examine how such reforms have been enacted in the context of a revolutionary forensic crime lab in response to prior scandals, and their role in improving public trust and perceived credibility (as noted in the above epigraph). We then make specific recommendations for forensic researchers to consider that have yet to gain momentum.

In line with the themes discussed in Part III, it is important to note that our recommendations will primarily focus on reforms concerning transparency and error admission.

¹⁰ References from original excerpt omitted.

This should not distract from other related reforms concerning improved quality control and validation research that are already underway in the forensic sciences [83]. As Heavey and colleagues recently noted, these reforms are related in that transparency promotes the identification of risks and errors and thus is a key component of quality control systems in laboratories [84]. Moreover, beyond its value in promoting credibility, sharing data and materials – and generally conducting research more transparently – accelerates foundational forensic science research and improves its reliability [2].

Forensic science practice

Although forensic science laboratories have historically struggled with scientific principles of openness and transparency, one promising case study involves the Houston Forensic Science Center (HFSC), a lab that has been more aggressive and inventive than most in opening itself up to scrutiny [84-86]. In response to previous controversies associated with their lab, the HFSC reports that greater transparency through admitting mistakes, opening up their systems, and publishing error rates has only improved their credibility. As we will discuss, and building upon the findings in Part III, their experience is consistent with systematic research on what the public expects of science.¹¹

Following several controversies, including forensic scientific testing errors contributing to wrongful convictions, the Houston Police Department's crime lab became the first U.S. lab to detach itself from law enforcement and establish an independent crime lab. Led by scientists at the Houston Forensic Science Center, it took several steps to advance its transparency and

¹¹ Again, note that we do not aim to provide a thorough accounting of quality control and transparency's role in it (but see a recent discussion by Heavey and colleagues [84] for a more detailed investigation), but rather a brief case study of the HFSC's experience as it relates to improving forensic science's credibility through openness and transparency.

openness, such as making its standard operating procedures, incident reports, and corrective action reports available on its website [85]. It also introduced blind proficiency testing into its examiners' casework, with the results of those tests recently being published. Moreover, the HFSC commenced recording and uploading its meetings, which were opened up to the public and criminal justice stakeholders to allow all parties to voice their goals, interests and concerns transparently [85,86].

Since these reforms were implemented, the HFSC has reported several benefits related to its credibility [86]. By detaching from law enforcement and embracing new scientifically-driven norms of openness and transparency, the lab's directors report that their reputation has transformed from one of subservience to law enforcement to one of trust and independence, with a more equal and legitimate standing with police and prosecutors [87]. The HFSC directors suggest that evidence for their improved credibility can be found in changes in their media coverage. During the initial scandal, where the crime lab was rife with mismanagement, incompetence, and fraud (see [88,89]), the media and public responded with years of harsh criticism. For instance, one *New York Times* article called the lab "the worst crime lab in the country" [90]. However, when contrasted with a more recent incident, where a 2017 audit revealed numerous evidence errors across sixty-five cases, the HFSC's commitment to transparency in reporting the results straight to the media helped retain their credibility: '... [the] publicity associated with the incident began and ended in one news cycle; without a cover-up, the laboratory could hardly be faulted for taking steps to improve its operations' [86]. Indeed, such changes even prompted an oversight body to propose working with them as a model of accountability - a vastly different response to the initial scandal [91].

In addition to these direct improvements to the HFSC's public image, other credibility benefits flow indirectly from improvements to the reliability of its work. For example, with the implementation of published blind proficiency testing that can be publicly scrutinized, the HFSC is able to provide estimates of its examiners' proficiency that are better calibrated to their underlying ability and performance [92]. This knowledge, valuable in its own right, seems to have benefited the lab's general credibility. For instance, examiners at the HFSC report increased confidence on the witness stand flowing from the blind testing data providing an evidence base for their opinions [93]. Additionally, in the long run, blind testing provides clarity on areas for improvement and allows for more targeted training in response [85].¹² Such measures allow for continuous improvement for practitioner performance, heading off future mistakes and improving its legitimacy and reliability in the eyes of the public (see [94,95]).¹³

Other research summarized in Part III suggests that the HFSC's impression of its improving credibility is accurate. For instance, its reforms seem to align with each criterion of building long-term epistemic trust. Blind proficiency testing to provide an accurate measure of performance bolsters the criterion of "expertise", and the various ways in which the HFSC has implemented transparency in its work aligns with the criterion of 'integrity'. In terms of 'benevolence', it seems that the public disclosure of mistakes and opening HFSC meetings to the public and criminal justice stakeholders helps satisfy this aspect of epistemic trust as well.

Several other components of the HFSC's reform efforts exemplify the theory and research on credibility. For instance, these reforms have the additional benefit of helping to avoid

¹² [94] noted that "[i]n most forensic science disciplines, there simply are no data pertaining to the rates at which forensic science procedures and forensic scientists err"; [95] noted that blind testing "will provide valuable information on the reliability and accuracy of laboratory methods"

¹³ See [85] for practitioner performance data.

a lemons market. By engaging the public and showing its work, the HFSC is demonstrating that it can be easily distinguished from other labs that have been widely criticized and less transparent. The HFSC's move towards operational independence is also likely to improve its credibility. This is because, in the aforementioned 2019 Pew survey, 52% said they would trust science more if findings were "reviewed by an independent committee" and 58% would trust science less if they know it was "funded by industry groups" [72]. Importantly, the Pew survey also investigated how these attitudes display in minority groups, with Black and Hispanic adults being more likely than white adults to be critical of both practitioners and researchers - presumably because they have historically not benefited equally from scientific advances (see [72,96]) This suggests that transparency and independence in forensic science may be especially important for such groups, as they have also been disproportionately impacted by failures of forensic science, wrongful convictions and incarceration (see [97]). If this is true, then it seems the HFSC is heading in the right direction for re-establishing trust in crime lab operations.

Forensic science research

Complementing the above practices modeled by the HFSC, enhanced openness and transparency can also improve the forensic science research base. We will now review three steps that researchers and journal editors can take: preregistration, registered reports, and open data, analysis scripts, and materials [98]. As with our discussion of forensic science practice, the reforms discussed below promote epistemic trust and align with public expectations (see Part III). And, while this is not our focus, they can also assist in improving the field's long term reliability by producing research claims that are more likely to withstand scrutiny and facilitating

error detection where necessary.¹⁴ We will focus on practices that have received empirical support and that have been recent targets of reform efforts within forensic science.

First, forensic science researchers should consider preregistering their research on one of many public registries [99,100]. Preregistration involves placing a timestamped version of a research protocol online (it can be embargoed until the study is complete) prior to the data being collected. Unlike traditional methods of conducting research, this combats the file drawer problem (i.e., important findings not being published) because even if the study is never formally published, other researchers can find it and include it in meta-analyses and systematic reviews (among other reviews). Preregistration also improves transparency in that others can compare the final published protocol to what was initially planned to see what changes were made after the data was collected (e.g., exclusion of outliers) - and to see if any deviations from the initial preregistration were sufficiently justifiable.

As we have discussed, several fields (e.g., psychology) have seen their credibility challenged for practices like excluding outliers to make their results more speciously convincing. Although the results have been mixed [101], at least one study has found that preregistrations in an abstract improve the perceived integrity of the authors [73]. Moreover, the general move in science towards preregistration [102,103] suggests that if forensic science does not keep up, it may be seen as less credible than other fields.¹⁵

While preregistration has started to become the norm in the social sciences in recent years [103,105], it has been required by law in clinical medical research for decades [106,107]. The reason legislatures mandated preregistration for medical research is the importance of this field's

¹⁴ For a fuller analysis of how openness and transparency can improve the reliability of forensic science, see [2,116].

¹⁵ But see [104] for an opposing perspective.

claims being open to scrutiny and ultimately calibrated to the strength of the evidence. Such changes were also spurred by legal controversies surrounding undisclosed null results [108]. When considering forensic science research, there are clear analogues to this. Given forensic science's ability to influence criminal trials, and controversies surrounding wrongful convictions, forensic science is no less important; and therefore, forensic science researchers should be encouraged, if not required, to preregister their studies as well. Accordingly, preregistration of forensic science research was recently recommended by OSAC [109].

Second, forensic science journals should consider adopting the registered report journal format [110]. Registered reports are a journal article format in which it is the research plan (i.e., hypotheses, methodology, materials) that is peer-reviewed, prior to data collection, rather than the final article. This allows peer-reviewers to help improve the plan before the study has actually commenced. Once this plan is accepted, publication is guaranteed as long as it is followed, regardless of whether the results are “desired” or not.

The registered report format has several benefits, each of which is as applicable to forensic science research as it is in other fields. One benefit of registered reports is that they encourage increased transparency when reporting a study's findings. Indeed, because publication is not contingent on a positive result, researchers have no reason to selectively report findings. Furthermore, the registered report workflow itself, whereby the peer-reviewed research plan is published alongside the final manuscript containing the results, deters selective reporting by making any data-contingent changes to the initial plan more detectable. As noted in Part III, selective reporting is a practice that the public equates with research misconduct. Registered reports may therefore prove valuable in preventing such behaviour and bringing the field more in line with public expectations.

Another benefit is that the registered report format seems to increase the quality of the paper generally. In a recent study, blind coders read and rated both standard studies and those published under the registered report framework [111]. Not only was the methodological quality of registered reports perceived as higher, but there was also better perceived alignment between the question and methodology, as well as the findings and interpretation. This suggests that this format is effective at increasing the quality of the most important aspects of the paper for peer review [111]. These benefits also remained without undermining other aspects of research - namely novelty, creativity and innovation - which some scientists worried registered reports might reduce [112,113]. Additionally, as with many of the other practices we have discussed, the improved quality of registered reports will promote the field's long term credibility, with its claims being better calibrated to observable outcomes.

The number of journals that accept registered reports is growing with over 300 journals now offering a registered report track for submissions. Part of the reason for this is that the barriers to entry are decreasing. For example, many journals accept registered reports through the 'Peer Community in' initiative whereby journals do not directly review registered reports, but rather rely on the expertise of editors with more experience with the format [114]. This may be especially attractive to forensic science journals with high workloads that wish to rely on established guidelines and editors with more familiarity with the format. In any case, we are only aware of one forensic science journal that has signalled adoption of the format [115]. This was the result of a grassroots campaign and open letter to forensic journals signed by many researchers in the field [116]. You are welcome to join as a signatory.¹⁶

¹⁶ <https://forms.gle/DN7qmQgCv1KHr4326>.

Third, we recommend that - subject to privacy and ethical constraints - forensic science researchers make their data, analytic code, and digital materials available. Several free open access repositories, such as the Open Science Framework [117], are available to host these outputs. Journals wishing to advance these research practices can adopt the Transparency and Openness Promotion guidelines, which are modular standards (applying, e.g., to data, code, preregistration, etc) that can be adopted at three levels of stringency (from encouraging to requiring) [118].¹⁷ Level 2 (requiring the open practice subject to the authors demonstrating a legal or ethical reason it cannot be fulfilled) is likely appropriate for forensic science journals [119].¹⁸

As previously mentioned, a Pew survey found that respondents were substantially more likely to report trust in science when data are publicly accessible and independently reviewed (an increase in public perception of 57 and 52 percent respectively) [72]. Beyond bolstering public perception, however, sharing data, code, and other materials openly allows other researchers to more easily verify existing findings [120]. In taking such steps to increase transparency, researchers wishing to criticize, investigate, or build upon previous findings, can more readily attempt to reproduce another paper's claims and assess its methodology.

Open science 'badges' represent an additional mechanism to encourage open and transparent research practices that forensic science journals may wish to implement. Open science badges are a relatively new intervention where appropriate papers are published with up to three badges: earned for preregistration, open data, or open materials. These badges allow

¹⁷ In forensic science (see NAS Report, p. 15 [6]), as in other fields, funders play an important role in setting the research agenda. Accordingly, funders, such as the National Institute of Standards and Technology, may wish to require the research it funds meets Transparency and Openness Promotion guidelines: [118].

¹⁸ Sample level two language can be found at the journal Cortex: [119].

other scientists - and even those without specialized scientific knowledge - to quickly understand when and how authors are being transparent with their data and materials, allowing greater scrutiny for papers that have adopted such practices [121]. Early evidence suggests that papers that earned badges for data sharing contained data and materials that were more ‘available, correct, usable, and complete’, compared with papers that had merely reported openness without earning badges [122]. Encouragingly, since one journal adopted open science badges, trends for reporting openly available data seemed to be on a sharp rise, with an approximately ten-fold increase in such reporting within the span of two years, and an approximately three-fold increase in materials sharing [122]. As these practices become normalised and refined, such initiatives offer accessible ways for the public and other scientists to quickly assess the contents of previously more opaque research papers and distinguish between transparent and less transparent research. This allows for any benefits for transparency to quickly be gleaned [122]. By signalling transparency in this way and proactively providing skeptics with ammunition to critique one’s work, open science badges are a highly visible mechanism by which forensic science researchers can demonstrate their commitment to self-correction - one of the core mechanisms by which a scientific field earns its credibility among the public [123].

Accordingly, our final recommendation is that forensic science researchers continue to develop a culture of critical appraisal - that is, actively scrutinizing others’ work and criticizing where necessary. While this is already one aspect of the scientific process, traditional methods of error detection in science, such as peer review, seem to be less effective than they could be (see [124]). Several studies have demonstrated that peer reviewers can miss the majority of errors and discrepancies presented in a paper [125,126], and meta-analyses have revealed that agreement between peer reviewers is barely higher than chance level [127]. Peer reviewer decisions can

also be biased by factors that are irrelevant to the actual quality of the paper [128]. Such findings suggest that new mechanisms are required to facilitate error detection among papers.

Vazire and Holcombe have suggested several ways to increase critical appraisal [123]. These include new tools for assessing the quality, reliability and reproducibility of a paper's statistics, increased attempts to replicate previous studies, more accessible avenues for post-publication peer review, and other recommendations. One promising mechanism for critical appraisal is a journal article format known as a 'verification report', focused on computationally reproducing the results of prior studies [129]. Still, critical appraisal appears to be lagging behind transparency reforms across the scientific ecosystem [123]. While this is understandable because critical appraisal is effortful and requires rethinking historic practices, an ethic of self-scrutiny has an important role to play in a field's credibility, especially ones facing the degree of public scrutiny and accountability that forensic science faces.

Part V. Conclusion

While some forensic practitioners have historically claimed infallibility in their methods, emerging research not only demonstrates that this is not the case, but also that the public does not uniformly believe this to be the case (if it ever did). If forensic science is to maintain its reputation or even improve it, we recommend that forensic scientists critically examine their research and practices and consider aligning them with the move towards openness and transparency occurring elsewhere within science. In this path, credibility is earned because the research and work underlying forensic scientific claims are transparently reported such that they can be tested by other scientists. Such claims will also be more credible in the long run because they will more likely be consistent with observable reality, thus reducing forensic science's

contributions to wrongful convictions. Moreover, recent empirical research suggests that taking these steps aligns with what the public expects of science. Forensic science practitioners wishing to take up these principles of openness and transparency have a model to follow in the HFSC. Additionally, in the wake of the credibility revolution, researchers similarly have many well-established tools available to them to make their research easier to access and scrutinize.

While we have focused on forensic science retaining (or recovering, depending on your view of recent studies) its credibility, there are other benefits to openness and transparency in forensic science [2]. We will end by discussing two others: that transparency and openness can lead to more efficient practice and research, and that they ensure forensic science evidence comports with the rules and principles of evidence law.

Beginning with efficiency, while it is true that some of what we have discussed will require some initial costs, such costs are not insurmountable, and will likely be outweighed by benefits in the long run. For instance, the HFSC noted that critics were concerned their blind testing program would be ‘burdensome and expensive’ [95]. They found, however, that the upfront costs to that and other aspects of their new systems were minimal and are already producing benefits that outweigh the costs [85,86]. Further efficiencies are possible. The HFSC notes, for instance, that sharing of samples used for testing between labs (i.e., ‘open materials’) would further reduce costs and they are actively seeking such collaborations [85]. The same is true in the context of research [2,130]. Sharing data and materials (when appropriate) can accelerate discovery by allowing researchers to build off of each other’s work to test new questions and weed out spurious claims [130].

Openness and transparency also help ensure that forensic scientific evidence comports with the rules of evidence and the principles behind them [131,132]. These rules seek to regulate

evidence such that factfinders (judges or jurors) can evaluate the evidence for themselves and need not rely on the say-so of the expert witness (i.e., *ipse dixit*) or extrinsic factors, such as the eminence of the witness. Evidence rules also exist to ensure that the parties are able to fully examine and test the evidence tendered against them. Transparency and openness advance these ideals in many ways, providing crucial information on the validity of research behind the scientific practice and the proficiency of the examiner.

References

- [1] M.A. Edwards, S. Roy, Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition, *Environ. Eng. Sci.* 34 (1) (2017) 51-61. <https://doi.org/10.1089/ees.2016.0223>.
- [2] J.M., Chin, G. Ribeiro, A. Rairden, Open forensic science, *J. Law Biosci.* (2019) 255-288. <https://doi.org/10.1093/jlb/lbz009>.
- [3] *Director of Public Prosecutions (DPP) v Wise (a pseudonym)* [2016] VSCA 173.
- [4] *Director of Public Prosecutions (DPP) v Paulino* [2017] VSCA 38.
- [5] President's Council of Advisors on Science and Technology, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, Executive Office of the President, Washington, D.C., 2016.
- [6] National Academies of Sciences, Engineering, and Medicine, Open Science by Design: Realizing a Vision for 21st Century Research, The National Academies Press, Washington, D.C., 2018.
- [7] J.L., Mnookin, The uncertain future of forensic science, *Daedalus* 147 (4) (2018) 99-118. https://doi.org/10.1162/daed_a_00523.
- [8] K. Podlas, The "CSI Effect", *Criminol. Crim. Justice* (2017) 1-20. <https://doi.org/10.1093/acrefore/9780190264079.013.40>.
- [9] J.M. Chin, L. Workewych, The CSI Effect, in: *Oxford Handbooks Online*. Oxford University Press, 2016, pp. 1–28.
- [10] S.A. Cole, R. Dioso-Villa, CSI and its effects: media, juries, and the burden of proof, *N. Engl. Law Rev.* 41 (2007) 435-469.

- [11] S.A. Cole, R. Dioso-Villa, Investigating the ‘CSI Effect’ effect: Media and litigation crisis in criminal law, *Stanford Law Rev.* 61 (6) (2009) 1335-1374.
- [12] K. Podlas, “The CSI Effect”: Exposing the media myth, *Fordham Intelect. Prop. Media Entertain. Law J.* 16 (2) (2005) 429-465.
- [13] N.J. Schweitzer, M.J. Saks, The CSI Effect: Popular fiction about forensic science affects the public expectations about real forensic science, *Jurimetr. J. Law Sci. Technol.* 47 (3) (2007) 357-364.
- [14] D.E. Shelton, Y.S. Kim, G. Barak, A study of juror expectations and demands concerning scientific evidence: Does the ‘CSI Effect’ exist? *Vanderbilt J. Entertain. Technol. Law* 9 (2) (2007) 331-368.
- [15] Statistical Modeling, Causal Inference, and Social Science, The purported CSI effect and the retroactive precision fallacy.
<https://statmodeling.stat.columbia.edu/2018/11/05/purported-csi-effect-retroactive-precision-fallacy/> (accessed 18 August 2021).
- [16] J.D. Lieberman, C.A. Carrell, T.D. Miethe, D.A. Krauss, Gold versus platinum: Do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence? *Psychol. Public Policy Law* 14 (1) (2008) 27-62.
<https://doi.org/10.1037/1076-8971.14.1.27>.
- [17] G. Ribeiro, J.M. Tangen, B.M. McKimmie, Beliefs about error rates and human judgment in forensic science, *Forensic Sci. Int.* 297 (2019) (2019) 138-147.
<https://doi.org/10.1016/j.forsciint.2019.01.034>.

- [18] B.L. Garrett, G. Mitchell, How jurors evaluate fingerprint evidence: The relative importance of match language, method information, and error acknowledgement, *J. Empir. Leg. Stud.* 10 (3) (2013) 484-511.
- [19] P.R. Brewer, B.L. Ley, Media use and public perceptions of DNA evidence, *Sci. Commun.* 32 (1) (2010) 93-117. <https://doi.org/10.1177/1075547009340343>.
- [20] J. Kaplan, S. Ling, M. Cuellar, Public beliefs about the accuracy and importance of forensic evidence in the United States, *Sci. Justice* 60 (2020) (2020) 263-272.
<https://doi.org/10.1016/j.scijus.2020.01.001>.
- [21] D.A. Nance, S.B. Morris, Jury understanding of DNA evidence: An empirical assessment of presentation formats for trace evidence with a relatively small random match probability, *SSRN* (2003) 1-36. <http://dx.doi.org/10.2139/ssrn.462880>.
- [22] J. Schklar, S.S. Diamond, Juror reactions to DNA evidence: Errors and expectancies, *Law Hum. Behav.* 23 (2) (1999) 59-184. <https://doi.org/10.1023/A:1022368801333>.
- [23] G. Jenkins, R.A. Schuller, The impact of negative forensic evidence on mock jurors' perceptions of a trial of drug-facilitated sexual assault, *Law Hum. Behav.* 31 (4) (2007) 369-380. <https://doi.org/10.1007/s10979-006-9068-2>.
- [24] S. Dartnall, J. Goodman-Delahunty, Enhancing juror understanding of probabilistic DNA evidence, *Aust. J. Forensic. Sci.* 38 (2) (2006) 85-96.
<https://doi.org/10.1080/00450610609410635>.
- [25] D.K. Carlson, Americans conclusive about DNA evidence: Fewer confident in reliability of fingerprints, *Gallup* (2005) <https://news.gallup.com/poll/19915/americans-conclusive-about-dna-evidence.aspx>.

- [26] V.P. Hans, D.H. Kaye, B.M. Dann, E.J. Farley, S. Albertson, Science in the jury box: Jurors' comprehension of mitochondrial DNA evidence, *Law Hum. Behav.* 35 (1) (2011) 60-71. <https://doi.org/10.1007/s10979-010-9222-8>.
- [27] B.L. Garrett, G. Mitchell, Forensics and fallibility: Comparing the views of lawyers and jurors, *W. Va. Law Rev.* 119 (2016) 621-637.
- [28] K.A. Martire, K.N. Ballantyne, A. Bali, G. Edmond, R.I. Kemp, B. Found, Forensic science evidence: Naive estimates of false positive error rates and reliability, *Forensic Sci. Int.* 302 (2019) (2019) 1-9. <https://doi.org/10.1016/j.forsciint.2019.109877>.
- [29] C.M. Bowers, Identification From Bitemarks, in: D.L. Faigman, D.K. Kaye, M.J. Saks, J. Sanders (Eds.), *Modern Scientific Evidence: The Law and Science of Expert Testimony*, Thomson-West St. Paul, MN, 2010, pp. 1–96.
- [30] J.M. Chin, D. White, Forensic bitemark identification evidence in Canada, *U.B.C. Law Rev.* 52 (1) (2019) 1-50.
- [31] V.Z. Lawson, The influence of naive and media-informed beliefs on juror evaluations of forensic science evidence, *CUNY Academic Works* (2014) 1-180.
- [32] W.E. Crozier, J. Kukucka, B.L. Garrett, Juror appraisals of forensic evidence: Effects of blind proficiency and cross-examination, *Forensic Sci. Int.* 315 (2020) 1-11. <https://doi.org/10.1016/j.forsciint.2020.110433>.
- [33] J.J. Koehler, Intuitive error rate estimates for the forensic sciences, *Jurimetrics* 57 (2016) 153-168. <http://dx.doi.org/10.2139/ssrn.2817443>
- [34] D.C. Murrie, B.O. Gardner, S. Kelly, I.E. Dror, Perceptions and estimates of error rates in forensic science: A survey of forensic analysts, *Forensic Sci. Int.* 302 (2019) 1-9. <https://doi.org/10.1016/j.forsciint.2019.109887>.

- [35] B.L. Garrett, B.O. Gardner, E. Murphy, P. Grimes, Judges and forensic science education: A national survey, *Forensic Sci. Int.* 321 (2021) 1-7.
<https://doi.org/10.1016/j.forsciint.2021.110714>.
- [36] Innocence Project, DNA Exonerations in the United States.
<https://innocenceproject.org/dna-exonerations-in-the-united-states/> (accessed 20 July 2021).
- [37] B.L. Garrett, P.J. Neufeld, Invalid forensic science testimony and wrongful convictions, *Va. Law Rev.* 95 (1) (2009) 1-97.
- [38] Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, *Strengthening Forensic Science In the United States: A Path Forward*, The National Academies Press, Washington, D.C., 2009.
- [39] L. Ricciardi, M. Demos, *Making a Murderer*, Synthesis Films (2015).
- [40] J.-X. de Lestrade, *The Staircase*, Maha Productions (2018).
- [41] T. Closson, They Spent 24 Years Behind Bars. Then the Case Fell Apart, *New York Times* (2021) <https://www.nytimes.com/2021/03/05/nyregion/queens-wrongful-convictions.html>.
- [42] J. Flom, *Wrongful Conviction with Jason Flom*, Lava for Good Podcasts (2019)
<https://www.wrongfulconvictionpodcast.com/with-jason-flom>.
- [43] *Last Week Tonight with John Oliver*, Forensic Science, HBO (2017).
- [44] C. Prosser, Push for royal commission following High Court finding on pathologist's flawed testimony, *ABC News* (2017).

- [45] E. Humes, Bad forensic science is putting innocent people in prison, Los Angel. Times (2019). <https://www.latimes.com/opinion/op-ed/la-oe-humes-forensic-evidence-20190113-story.html>.
- [46] L. Mannix, Top judge worried forensic evidence putting innocent people behind bars, The Age (2019). <https://www.theage.com.au/national/top-judge-worried-forensic-evidence-putting-innocent-people-behind-bars-20190823-p52k3l.html>.
- [47] D. Cornwall, Courts' use of forensic evidence called into question, ABC News (2012). <https://www.abc.net.au/news/2012-05-03/courts-use-of-forensic-evidence-called-into-question/3989860>.
- [48] J. MacDonald, How scientific is forensic science? JSTOR Daily (2019). <https://daily.jstor.org/how-scientific-is-forensic-science/>.
- [49] J.D. Angrist, J.-S. Pischke, The credibility revolution in empirical economics: How better research design is taking the con out of econometrics, J. Econ. Perspect. 24 (2) (2010) 3-30. <https://doi.org/10.1257/jep.24.2.3>.
- [50] S. Vazire, Implications of the credibility revolution for productivity, creativity, and progress, Perspect. Psychol. Sci. 13 (4) (2018) 411-417. <https://doi.org/10.1177/1745691617751884>.
- [51] M. Baker, 1,500 scientists lift the lid on reproducibility, Nature 533 (2016) 452-454. <https://doi.org/10.1038/533452a>.
- [52] B.T. Rutjens, S.J. Heine, R.M. Sutton, F. van Harreveld, Attitudes towards science. Adv. Exp. Soc. Psychol. 57 (2018) 125-165. <https://doi.org/10.1016/BS.AESP.2017.08.001>

- [53] A.K. Fetterman, K. Sassenberg, The reputational consequences of failed replications and wrongness admission among scientists, *PLoS ONE* 10 (12) (2015) 1-13.
<https://doi.org/10.1371/journal.pone.0143723>.
- [54] M.S. Altenmuller, S. Nuding, M. Gollwitzer, No harm in being self-corrective: Self-criticism and reform intentions increase researchers' epistemic trustworthiness and credibility in the eyes of the public, *Public Underst. Sci.* (2021) 1-15.
<https://doi.org/10.1177/09636625211022181>.
- [55] J.L. Mnookin, S.A. Cole, I.E. Dror, B.A.J. Fisher, M.M. Houck, K. Inman, D.H. Kaye, J.J. Koehler, G. Langenburg, D.M. Risinger, N. Rudin, J. Siegel, D.A. Stoney, The need for a research culture in the forensic sciences, *UCLA Law Rev.* (2011) 1-60.
- [56] H. Swofford, S. Cole, Mt. Everest—we are going to lose many: A survey of fingerprint examiners' attitudes towards probabilistic reporting, *Law Probab. Risk* 19 (3-4) (2020) 255-291. <https://doi.org/10.1093/lpr/mgab003>.
- [57] A.S. Bali, G. Edmond, K.N. Ballantyne, R.I. Kemp, K.A. Martire, Communicating forensic science opinion: An examination of expert reporting practices, *Sci. Justice* 60 (3) (2020) 216-224. <https://doi.org/10.1016/j.scijus.2019.12.005>.
- [58] M.R. Munafo, B.A. Nosek, D.V.M. Bishop, K.S. Button, C.D. Chambers, N.P. du Sert, U. Simonsohn, E.-J. Wagenmakers, J.J. Ware, J.P.A. Ioannidis, A manifesto for reproducible science, *Nat. Hum. Behav.* 1 (2017) 1-9. <https://doi.org/10.1038/s41562-016-0021>.
- [59] K. Popper, *The Logic of Scientific Discovery*, Routledge Classics, New York, 1992.
- [60] J. Bohannon, Replication effort provokes praise—and 'bullying' charges, *Science* 334 (2014) 788-789. <https://doi.org/10.1126/science.344.6186.788>

- [61] F. Hendriks, D. Kienhues, R. Bromme, Disclose your flaws! Admission positively affects the perceived trustworthiness of an expert science blogger, *Stud. Commun. Sci.* 16 (2) (2016) 124-131. <https://doi.org/10.1016/j.scoms.2016.10.003>.
- [62] B.L. Garrett, N. Scurich, W.E. Crozier, Mock jurors' evaluation of firearm examiner testimony, *Law Hum. Behav.* 44 (5) (2020) 412-423.
<https://psycnet.apa.org/doi/10.1037/lhb0000423>.
- [63] W.C. Thompson, N. Scurich, How cross-examination on subjectivity and bias affects jurors' evaluations of forensic science evidence, *J. Forensic Sci.* 64 (5) (2019) 1379-1388.
<https://doi.org/10.1111/1556-4029.14031>.
- [64] G. Mitchell, B.L. Garrett, The impact of proficiency testing information and error aversions on the weight given to fingerprint evidence, *Behav. Sci. Law* 37 (2) (2019) 195-210.
<https://doi.org/10.1002/bsl.2402>.
- [65] B. Garrett, G. Mitchell, N. Scurich, Comparing categorical and probabilistic fingerprint evidence, *J. Forensic Sci.* 63 (6) (2018) 1712-1717. <https://doi.org/10.1111/1556-4029.13797>.
- [66] R. Marquis, A. Biedermann, L. Cadola, C. Champod, L. Gueissaz, G. Massonnet, W.D. Mazzella, F. Taroni, T. Hicks, Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings, *Sci. Justice* 56 (5) (2016) 364-370. <https://doi.org/10.1016/j.scijus.2016.05.009>.
- [67] C. Allen, D.M.A. Mehler, Open science challenges, benefits and tips in early career and beyond, *PLOS Biol.* 17 (12) (2019). <https://doi.org/10.1371/journal.pbio.3000587>.
- [68] Center for Open Science, TOP Guidelines. <https://www.cos.io/initiatives/top-guidelines> (accessed 20 July 2021).

- [69] S. Vazire, Quality uncertainty erodes trust in science, *Collabra Psychol.* 3 (2017) 1-5.
<https://doi.org/10.1525/collabra.74>.
- [70] R.C. Cottrell, Scientific integrity and the market for lemons, *Res. Ethics* 10 (2013) 17-28.
<https://doi.org/10.1177/1747016113494651>.
- [71] G.A. Akerlof, The market for “lemons”: Quality uncertainty and the market mechanism, *Q. J. Econ.* 84 (3) (1970) 488-500. <https://doi.org/10.2307/1879431>.
- [72] C. Funk, M. Hefferon, B. Kennedy, C. Johnson, Trust and mistrust in american’s views of scientific experts. *Pew Res. Cent.* (2019).
<https://www.pewresearch.org/science/2019/08/02/trust-and-mistrust-in-americans-views-of-scientific-experts/>.
- [73] H. Song, D.M. Markowitz, S.H. Taylor, Trusting on the shoulders of open giants? Open science increases trust in science for the public and academics, *OSF Preprints* (2021) 1-46. <https://doi.org/10.31219/osf.io/g328c>.
- [74] F. Anvari, D. Lakens, The replicability crisis and public trust in psychological science, *Compr. Results Soc. Psychol.* 3 (3) (2018) 266-286.
<https://doi.org/10.1080/23743603.2019.1684822>.
- [75] T. Wingen, J.B. Berkessel, B. English, No Replication, No trust? How low replicability influences trust in psychology, *Soc. Psychol. Personal. Sci.* 11 (4) (2020) 454-463.
<https://doi.org/10.1177%2F1948550619877412>.
- [76] J.T. Pickett, S.P. Roche, Questionable, objectionable or criminal? Public opinion on data fraud and selective reporting in science, *Sci. Eng. Ethics* 24 (1) (2018) 151-171.
<https://doi.org/10.1007/s11948-017-9886-2>.

- [77] L.K. John, G. Loewenstein, D. Prelec, Measuring the prevalence of questionable research practices with incentives for truth telling, *Psychol. Sci.* 23 (5) (2012) 524-532.
<https://doi.org/10.1177/0956797611430953>.
- [78] H. Fraser, T. Parker, S. Nakagawa, A. Barnett, F. Fidler, Questionable research practices in ecology and evolution, *PLoS ONE* 13 (7) (2018) 1-16.
<https://doi.org/10.1371/journal.pone.0200303>.
- [79] Pew Research Center, Public Esteem for Military Still High, *Pew Res. Cent.* (2013).
<http://www.pewforum.org/2013/07/11/public-esteem-for-military-still-high/>.
- [80] Pew Research Center, Public and Scientists' Views on Science and Society, *Pew Res. Cent.* (2015). <https://www.pewresearch.org/science/2015/01/29/public-and-scientists-views-on-science-and-society/>.
- [81] J.G. Bottesini, S. Vazire, Do participants want their data to be shared? Preregistered study, available at Open Science Framework (2019). <https://osf.io/tcqyd/>.
- [82] N. Passalacqua, M.A. Pilloud, W.R. Belcher, Scientific integrity in the forensic sciences: Consumerism, conflicts of interest, and transparency, *Sci. Justice* 59 (5) (2019) 573-579.
<https://doi.org/10.1016/j.scijus.2019.06.010>.
- [83] C. Roux, W. Céline, From research integrity to research relevance to advance forensic science, *Forensic Sci. Res.* (2021) 1-3. <https://doi.org/10.1080/20961790.2021.1977480>.
- [84] A.L. Heavey, G.R. Turbett, M.M. Houck, S.W. Lewis, Toward a common language for quality issues in forensic science, *WIREs Forensic Sci.* (2021) 1-9.
<https://doi.org/10.1002/wfs2.1452>
- [85] S.G. Thompson, N.B. Casarez, Solving Daubert's dilemma for the forensic sciences through blind testing, *Houst. Law Rev.* 57 (3) (2020) 617-670.

- [86] N.B. Casarez, S.G. Thompson, Three transformative ideals to build a better crime lab, Ga. State Univ. Law Rev. 34 (4) (2018) 1007-1072.
- [87] P. Stout, The Recovery of Houston, Forensic Science Error Management Symposium (2017). <https://www.nist.gov/news-events/events/2017/07/2017-international-forensic-science-error-management-symposium>.
- [88] S.G. Thompson, Cops in Lab Coats: Curbing Wrongful Convictions Through Independent Forensic Laboratories, Carolina Academic Press, North Carolina, 2015.
- [89] S.G. Thompson, N.B. Casarez, Building the infrastructure for “justice through science”: The Texas model, W. Va. Law Rev. 119 (2) (2016) 711-748.
- [90] A. Liptak, Worst Crime Lab in the Country—Or is Houston Typical? New York Times (2003).
- [91] R. Plushnick-Masti, HFSC to Work With TFSC to Create Best Practices for Crime Scene Investigations, Houston Forensic Science Center (2017).
<https://houstonforensicscience.org/news/592efaaeqTZKQ%20Panel.pdf>.
- [92] D.J. Boone, H.J. Hansen, T.L. Hearn, D.S. Lewis, D. Dudley, Laboratory evaluation and assistance efforts: Mailed, on-site and blind proficiency testing surveys conducted by the Centers for Disease Control, Am. J. Public Health 72 (12) (1982) 1364-1368.
<https://doi.org/10.2105/ajph.72.12.1364>.
- [93] Forensic Technology Center of Excellence, Just Blind Proficiency Testing, Forensic Advancement (2019). <https://forensiccoe.org/js7-e10/>.
- [94] J.J. Koehler, Forensics or fauxrensic? Ascertaining accuracy in the forensic sciences, Ariz. State Law J. 49 (2018) 1369-1416. <http://dx.doi.org/10.2139/ssrn.2773255>.

- [95] National Commission on Forensic Science, 2016. Views of the Commission Facilitating Research on Laboratory Performance. National Institute of Standards and Technology.
- [96] S. Toy, 45 years ago, the nation learned about the Tuskegee Syphilis study. Its repercussions are still felt today, USA Today, (2017).
<https://www.usatoday.com/story/news/2017/07/25/tuskegee-syphilis-study-its-repercussions-still-felt-today/506507001/>.
- [97] T. Ward, The less equal we become, the less we trust science, and that's a problem, The Conversation (2020) <https://theconversation.com/the-less-equal-we-become-the-less-we-trust-science-and-thats-a-problem-151691>.
- [98] V. Denault, P. Plusquellec, L.M. Jupe, M. St-Yves, N.E. Dunbar, M. Hartwig, S.L. Sporer, J. Rioux-Turcotte, J. Jarry, D. Walsh, H. Otgaar, A. Viziteu, V. Talwar, D.A. Keatley, I. Blandon-Gitlin, C. Townson, N. Deslauriers-Varin, S.O. Lilienfeld, M.L. Patterson, I. Areh, A. Allan, H.E. Cameron, R. Boivin, L. ten Brinke, J. Masip, R. Bull, M. Cyr, L. Hope, L.A. Stromwall, S.J. Bennett, F.A. Menaiya, R.A. Leo, A. Vredeveltdt, M. Laforest, C.R. Honts, A.L. Manzanero, S. Mann, P.-A. Granhag, K. Ask, F. Gabbert, J.-P. Guay, A. Coutant, J. Hancock, V. Manusov, J.K. Burgoon, S.M. Kleinman, G. Wright, S. Landstrom, L. Freckelton, Z. Vernham, P.J. van Koppen, The analysis of nonverbal communication: The dangers of pseudoscience in security and justice contexts, *Anu. Psicol. Juridica* 30 (2020) 1–12. <https://doi.org/10.5093/apj2019a9>.
- [99] Open Science Framework, OSF Registries. <https://osf.io/registries> (accessed 19 August 2021).
- [100] As Predicted, As Predicted Home. <https://aspredicted.org/> (accessed 19 August 2021).

- [101] S.M. Field, E.-J. Wagenmakers, H.A.L. Kiers, R. Hoekstra, A.F. Ernst, D. van Ravenzwaaij, The effect of preregistration on trust in empirical research findings: results of a registered report, *R. Soc. Open Sci.* 7 (2020) 1-16.
<http://dx.doi.org/10.1098/rsos.181351>.
- [102] B.A. Nosek, C.R. Ebersole, A.C. DeHaven, D.T. Mellor, The preregistration revolution, *Proc. Natl. Acad. Sci. U. S. A.* 115 (11) (2018) 2600.
<https://doi.org/10.1073/pnas.1708274114>.
- [103] B.A. Nosek, D.S. Lindsay, Preregistration becoming the norm in psychological science, *Assoc. Psychol. Sci.* (2018)
<<https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science>>.
- [104] M. Rubin, Does preregistration improve the credibility of research findings? *Quant. Methods Psychol.* 16 (4) (2020) 376-390. <https://doi.org/10.20982/tqmp.16.4.p376>.
- [105] B.A. Nosek, E.D. Beck, L. Campbell, J.K. Flake, T.E. Hardwicke, D.T. Mellor, A.E. van't Veer, S. Vazire, Preregistration is hard, and worthwhile, *Trends Cogn. Sci.* 23 (10) (2019) 815-818. <https://doi.org/10.1016/j.tics.2019.07.009>.
- [106] K. Dickersin, I. Chalmers, Recognizing, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the WHO, *J. R. Soc. Med.* 104 (12) (2011) 532-538. <https://doi.org/10.1258/jrsm.2011.11k042>.
- [107] B. Goldacre, N.J. DeVito, C. Heneghan, F. Irving, S. Bacon, J. Fleminger, H. Curtis, Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource, *Br. Med. J.* (2018) 1-10. <http://dx.doi.org/10.1136/bmj.k3218>.

- [108] M. Wadman, Spitzer sues drug giant for deceiving doctors, *Nature* 429 (2004) 589.
<https://www.nature.com/articles/429589a>.
- [109] The Organization of Scientific Area Committees (OSAC) for Forensic Science, Human Factors in Validation and Performance Testing of Forensic Science, OSAC Human Factors Committee (2020) 1-35. <https://doi.org/10.29325/OSAC.TC.0004>.
- [110] C. Chambers, L. Tzavella, The past, present and future of Registered Reports, *Nat. Hum. Behav.* (2021) 1-14. <https://doi.org/10.1038/s41562-021-01193-7>.
- [111] C.K. Soderberg, T.M. Errington, S.R. Schiavone, J. Bottesini, F.S. Thorn, S. Vazire, K.M. Esterling, B.A. Nosek, Initial evidence of research quality of registered reports compared to the standard publishing model, Unpublished, Archived at MetaArXiv Prepr. (2021) 1--65. <<https://osf.io/preprints/metaarxiv/7x9vy/>>.
- [112] A. Cropley, Research as artisanship versus research as generation of novelty: The march to nowhere, *Creat. Res. J.* 30 (4) (2018) 323-328.
- [113] R.F. Baumeister, Charting the future of social psychology on stormy seas: Winners, losers, and recommendations, *J. Exp. Soc. Psychol.* 66 (2016) 153-158.
<https://doi.org/10.1016/J.JESP.2016.02.003>.
- [114] Peer Community In, About PCI Registered Reports.
<https://rr.peercommunityin.org/about/about>, 2022 (accessed 11 Jan 2022).
- [115] Elsevier, Coming soon to FSI Synergy: Registered Reports.
<https://www.journals.elsevier.com/forensic-science-international-synergy/news/coming-soon-to-fsi-synergy-registered-reports> 2021 (accessed 19 August 2021).
- [116] J.M. Chin, R. McFadden, G. Edmond, Forensic science needs registered reports, *Forensic Sci. Int.* 2 (2020) (2020) 1-41. <https://doi.org/10.1016/j.fsisyn.2019.10.005>.

- [117] O. Klein, T.E. Hardwicke, F. Aust, J. Breuer, H. Danielsson, A.H. Mohr, H. IJzerman, G. Nilsson, W. Vanpaemel, M.C. Frank, A practical guide for transparency in psychological science, *Collabra Psychol.* 4 (1) (2018) 1-15. : <https://doi.org/10.1525/collabra.158>.
- [118] R. Masden, Funders must mandate and reward open research records, *Nature* (2020). <https://www.nature.com/articles/d41586-020-02395-1>.
- [119] Transparency and Openness Promotion (TOP) Guidelines. https://www.elsevier.com/__data/promis_misc/Cortex-TOP-author-guidelines.pdf (accessed 11 Jan 2022).
- [120] M. S. Anderson, B.C. Martinson, R. DeVries, Normative dissonance in science: results from a national survey of U.S. scientists, *J. Empir. Res. Hum. Res. Ethics.* 2 (4) (2007) 3-14. <https://doi.org/10.1525/jer.2007.2.4.3>.
- [121] Center for Open Science, Open Science Badges enhance openness, a core value of scientific practice. <https://www.cos.io/initiatives/badges> (accessed 20 July 2021).
- [122] M.C. Kidwell, L.B. Lazarevic, E. Baranski, T.E. Hardwicke, S. Piechowski, L.-S. Falkenberg, C. Kennett, A. Slowik, C. Sonnleitner, C. Hess-Holden, T.M. Errington, S. Fiedler, B.A. Nosek, Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency, *PLOS Biol.* 14 (5) (2016) 1-15. <https://doi.org/10.1371/journal.pbio.1002456>.
- [123] S. Vazire, A.O. Holcombe, Where are the self-correcting mechanisms in science? Unpublished, Archived at PsyArXiv Prepr. (2021) 1–23. <https://psyarxiv.com/kgqzt/>.
- [124] R. Smith, Classical peer review: an empty gun, *Breast Cancer Res.* 12 (2010) 1-4. <https://doi.org/10.1186/bcr2742>. for a review.

- [125] S. Schroter, N. Black, S. Evans, F. Godlee, L. Osorio, R. Smith, What errors do peer reviewers detect, and does training improve their ability to detect them? *J. R. Soc. Med.* 101 (10) (2008) 507-514. <https://doi.org/10.1258/jrsm.2008.080062>.
- [126] G.D. Cole, M.J. Shun-Shin, A.N. Nowbar, K.G. Buell, F. Al-Mayahi, D. Zargar, S. Mahmood, B. Singh, M. Mielewczik, D.P. Francis, Difficulty in detecting discrepancies in a clinical trial report: 260-reader evaluation, *Int. J. Epidemiol.* 44 (3) (2015) 862-869. <https://doi.org/10.1093/ije/dyv114>.
- [127] R.L. Kravitz, P. Franks, M.D. Feldman, M. Gerrity, C. Byrne, W.M. Tierney, Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? *PLoS ONE* 5 (4) (2010) 1-5. <https://doi.org/10.1371/journal.pone.0010072>.
- [128] D.P. Peters, S.J. Ceci, Peer-review practices of psychological journals: The fate of published articles, submitted again, *Behav. Brain Sci.* 5 (1982) 187-195. <https://doi.org/10.1017/S0140525X00011183>.
- [129] C.D. Chambers, Verification Reports: A new article type at Cortex, *Cortex* 129 (2020) A1-A3. <https://doi.org/10.1016/j.cortex.2020.04.020>.
- [130] M.R. Munafo, B.A. Nosek, D.V.M. Bishop, K.S. Button, C.D. Chambers, N.P. du Sert, U. Simonsohn, E.-J. Wagenmakers, J.J. Ware, J.P.A. Ioannidis, A manifesto for reproducible science, *Nat. Hum. Behav.* 1 (1) (2017) 1-9. <https://doi.org/10.1038/s41562-016-0021>.
- [131] J.M. Chin, D.T. Mellor, B. Grown, Improving Expert Evidence: The role of open science and transparency, *Ottawa L. Rev.* 50 (2) (2019) 1-52.
- [132] E.K. Cheng, G.A. Nunn, Beyond the Witness: Bringing a process perspective to modern evidence law, *Tex. Law Rev.* 97 (6) (2019) 1077-1124.

Figure 1

Figure 1. Dispersed data visualisation representing recorded trends in perceived accuracy and reliability across forensic evidence disciplines over time. All measurements have been converted to a 100-point scale. Error bars display 95% CIs.

Figure 2

Figure 2. Compressed data visualisation representing recorded trends in perceived accuracy and reliability across biometric forensic sciences over time. All measurements have been converted to a 100-point scale. Error bars display 95% CIs.

Table 1

Table 1. See Appendix A for methodological differences.

	Lieberman (~2008) student	Lieberman (~2008) juror	Martire (2011)	Martire (2012)	Ribeiro (2015)	Martire (2016)	Kaplan (2019)
DNA	93.7	94.9	9.292	8.952	89.95	8.769	83.09
Fingerprint	90.1	91.4	8.652	8.227	88.15	8.862	78.62
Hair*	88.1	89.2	8.004	7.226		7.016	
Fibre*	88.1	89.2		6.734			
Alcohol and drug	79.2	78.4			86.66	6.764	76.12
Dental				8.198	89.26	7.101	75.88
Blood pattern				7.329	78.53	7.155	64.28
Firearm*				6.71	79.63	7.323	68.15
Toolmark*					79.63	5.71	68.15
Explosive substance*				6.604	74.56		
Fire investigation*					74.56	6.524	
Document examination					65.18	6.414	
Voice identification			6.669	6.106	71.47	6.287	55.30
Trace						6.237	
Image comparison				6.411	78.21	5.825	
Tire impression						5.861	
Handwriting			6.138	5.464		5.78	
Shoeprint comparison				5.522			56.98
Damage examination						5.496	
Face					74.03		
Materials					79.37		
Gunshot residue					78.87		67.98
Geological materials					77.04		
Wildlife					47.77		
Anthropological					81.33		

Appendix A

Appendix A. Methodological differences in the studies reviewed in Part II.

Study	Characteristics	Geography	Size	Year	Question wording
Lieberman et al (1)	Student	USA	383	~2008	'Rate the following types of evidence for accuracy' from 0% ('not accurate at all') to 100% ('completely accurate').
Lieberman et al (2)	Recent jurors	USA	233	~2008	'Rate the following types of evidence for accuracy' from 0% ('not accurate at all') to 100% ('completely accurate').
Martire et al (1)	General public (MTurk) and student	Aus/USA	276	2011	'How reliable do you think [practice] is?' Rated from 0 ('not at all reliable') to 10 ('extremely reliable').
Martire et al (2)	General public (MTurk) and student	Aus/USA	207	2012	'How reliable do you think [practice] is?' Rated from 0 ('not at all reliable') to 10 ('extremely reliable').
Martire et al (3)	General public (MTurk)	USA	191	2016	'How reliable do you think [practice] is?' Rated from 0 ('not at all reliable') to 100 ('extremely reliable'). *The authors transformed this to a 0-10 scale to match their prior studies
Ribeiro et al	Student	Aus	101	2015	'What are your overall impressions about the accuracy of the different types of forensic evidence listed below?' Participants were provided a slider to move from 0-100%.
Kaplan et al	General public (MTurk)	USA	155	2019	'What are your overall impressions about the accuracy of the different types of forensic evidence listed below?' Participants were provided a slider to move from 0-100%. *Replicating Ribeiro et al