Received: 8 April 2023 | Revised: 15 August 2023 | Accepted: 15 August 2023
DOI: 10.1111/1556-4029.15369

JOURNAL OF FORENSIC SCIENCES AAFS

COMMENTARY
General

The most *consistent* finding in forensic science is *inconsistency*

Itiel E. Dror PhD

Cognitive Consultants International (CCI-HQ), London, UK

Correspondence
Itiel E. Dror, Cognitive Consultants International (CCI-HQ), London, UK.
Email: itiel@cci-hq.com

Abstract

The most consistent finding in many forensic science domains is inconsistency (i.e., lack of reliability, reproducibility, repeatability, and replicability). The lack of consistency is a major problem, both from a scientific and a criminal justice point of view. Examining forensic conclusion data, from across many forensic domains, highlights the underlying cognitive issues and offers a better understanding of the issues and challenges. Such insights enable the development of ways to minimize these inconsistencies and move forward. The aim is to highlight the problem, so that it can be minimized and the reliability of forensic science evidence can be improved.

KEYWORDS
cognitive bias, conclusions, decision making, expertise, forensic conclusions, human factors, linear sequential unmasking, reliability, variability

Highlights

- Consistency in conclusions (between and within examiners) is critical for science and for justice.
- Across the majority of forensic domains, there is a lack of consistency in forensic decisions.
- Inconsistencies can emerge from differences in observations or/and differences in interpretations.
- Consistency can be enhanced by taking into account human factors involved in forensic decision making.
- Using procedures, such as linear sequential unmasking (LSU-E), can reduce inconsistencies.

1 | INTRODUCTION

What is the most consistent finding across many forensic domains? The most consistent finding is… inconsistency (i.e., lack of reliability, reproducibility, repeatability, and replicability). As I point out below, this is a major problem, both from a scientific and a criminal justice point of view. Then, I present and organize research data, so we better understand the problem. I conclude with ways to minimize these inconsistencies and move forward. In order to deal with this issue and improve the reliability of forensic science, we must first highlight and acknowledge the problem.

2 | THE PROBLEM

One of the most fundamental requirements of any scientific endeavor is consistency. Consistency—what is often termed in science as reliability (also sometimes referred to as variability)—is about getting the same results when examining the same evidence.

From a scientific point of view, it does not matter who or where an analysis is conducted, the same analysis must yield the same results. There can be no science if the same exact analysis gives different results. It is important to distinguish between reliability and validity: Validity refers to whether the results are *correct*; in contrast, reliability is concerned with whether results are consistent (regardless of whether they are correct). There can be no validity without reliability. In fact, one cannot even consider or examine validity if there is no reliability.

For example, one can ask whether a scale gives the correct weight (validity) only if the scale is reliable. That is, if I go on the scale, and it reliably (consistently) says I am 170 pounds, only then can one consider whether or not the scale is giving my correct weight (maybe it is my correct weight and the scale is giving a valid result, or I may actually be 180 pounds, in which case the result is not valid and the scale needs calibration or fixing). However, imagine I go on the scale, and it reads 170 pounds; I get off the scale and on again, but this time it reads 180 pounds; and then when I go off and on again, it now reads 190 pounds—the results are all over the place: there is no reliability (consistency) in results. In such a case, without reliability, one cannot even ask about validity (if the results are all over the place, inconsistent, then one cannot ask whether the scale is giving the correct result). Only when there is consistency, can one then ask whether the result is valid.

From a criminal justice point of view, it seems totally unacceptable that whether a person ends up being convicted, or not, may depend on the luck of who examined the forensic evidence. That is, without consistency different forensic experts reach different conclusions, then the results of the forensic examination (e.g., whether I am identified or the results are inconclusive) depends on which examiner analyses the evidence. If it is examiner X, then I am identified; if it is examiner Y, then I am not identified. Without consistency, the result of the forensic examination can depend on who does the analysis ("Eeny, Meeny, Miny, Moe, with which expert should we go?" [1])–that should not be the case.

Furthermore, of course, from a legal point of view, if another forensic examiner can legitimately have an alternative different conclusion about the same evidence, then 'reasonable doubt' has been demonstrated.

Hence, both from a scientific point of view, as well as from a criminal justice point of view, inconsistency is a major problem. There are those who try to evade this serious concern by a misguided view that there can be two *different* valid conclusions about the *same* evidence, and hence that *different* conclusions can *all* be correct. For example, in many black box studies, when examiners reach different conclusions (e.g., an identification and an inconclusive), both conclusions are considered and calculated as correct [2].

Such a view undermines the very foundation of forensic (or any other) science, as it makes forensic science a matter of 'personal taste' (I can like sushi and you can like pasta; I can like my steak rare and you can like it well done…). That is okay when it comes to personal taste, but when it comes to science, if we have different conclusions about the same data, then one of us is right and one of us is wrong (actually, it may even be that both of us are wrong; what we know, as a matter of fact, is that at least one of us is wrong).

3 | THE DATA

What do the data say? Across many forensic domains, the data are clear: There is consistently a lack of consistency in forensic conclusions! The most consistent finding in forensic science is… inconsistency.

Research on fingerprinting (one of the most used forensic domains, that has been around for a very long time) has found that when the *same* examiner, examines the *same* pair of prints, 10% of the time they will reach different conclusions, that is, when they examine the same pair of prints on two separate occasions [3] (see also [4,5]).

There is lack of consistency not only in the conclusions (identification, exclusion, or inconclusive) but also about whether a fingermark is of sufficient quality for comparison [6]. Same story with DNA, where research shows "notable variation in whether labs would assess a given mixture as suitable or not" [7].

Further inconsistency is also apparent in what is observed: Different examiners analyzing the same fingermark see different minutia, and even the *same* examiner, examining the *same* fingermark, each time sees different minutia [8]. Findings in some studies (e.g., [3,7]) reveal inconsistencies that do not arise from bias due to task irrelevant contextual information. Thus, the aforementioned studies (and others) show that examiners' judgments were inconsistent even when the context in which those decisions were made was not changed (or contextual information was not given altogether).

DNA, often termed as the gold standard and praised for using statistics, also suffers from inconsistency. In 2011, the first publication revealing that different DNA analysts (from the same forensic laboratory) examining the same DNA mixture, reached different conclusions [9]. Since then, NIST published in 2018 the Mix05 and Mix13 studies [10], showing "variations observed among laboratory results". Further studies repeatedly show that forensic experts are inconsistent in DNA mixture conclusions, for example, "LR values obtained show a wide range of variation" [11].

If DNA and fingerprinting, two very established forensic domains, show that expert analysts lack consistency in their conclusions, experts in many other forensic domains surely suffer from the same problem. Indeed, lack of consistency is found across many other forensic domains: Digital forensic studies have shown, for example, "Results showed low reliability between DF examiners in observations, interpretations, and conclusions" [12]. Firearms studies have shown, for example, "that there are differences in examiner conclusions when examining the same evidence" [13]. Specifically, in firearms for example "when averaged over bullets and cartridge cases, the repeatability of comparison decisions (involving all levels of the AFTE Range) was 78.3% for known matches and 64.5% for known nonmatches. Similarly averaged reproducibility was 67.3% for known matches and 36.5% for known nonmatches" [14].

It is the same story across many other forensic domains: Bloodstain pattern analysis (BPA) studies have shown, for example, "The results show limited reproducibility of conclusions" [15]. It is also the same for footwear [16], forensic psychology [17], CSI [18], document examination [19], and the list goes on and on, covering many (not all, but many) forensic domains, showing lack of consistency. It is important to note that such inconsistencies are much more pronounced in the more difficult cases (e.g., near the threshold or close non-matches [20]) and that they vary across forensic domains (e.g., see misconception #3 in [21]).

4 | ORGANIZING AND UNDERSTANDING THE DATA

We need to make a number of distinctions so we can organize and understand the data about inconsistencies.

JOURNAL OF FORENSIC SCIENCES

First, it is important to distinguish the lack of consistency between and within examiners, that is, inter- and intra-consistency (some use the term "reproducibility" for between examiner consistency, and "repeatability" for within examiner consistency). When there is a lack of consistency between (inter) examiners, different examiners reach different conclusions about the same evidence (e.g., [9,12,16,19]).

Such between (inter) examiner inconsistencies are a problem, but even a greater problem is when there is a lack of consistency within (intra) examiner: when the same examiner reaches different conclusions about the same evidence (e.g., [3-5]).

It is important to note that when the same examiner reaches different conclusions about the same evidence, this cannot be easily explained by differences in laboratory policies, training, experience, and so forth. Such differences between forensic laboratories or between different forensic examiners ('Education & Training' and 'Personal Factors', see Levels 6 & 7 in [22]) cannot be factors in explaining inconsistencies that are found within (intra) examiner conclusions.

Second, it is also important to consider whether or not the inconsistencies arise from bias. Bias is a predictable and systematic variation (for a probabilistic formalization see [23]), whereas random variations can be regarded as 'noise' [24]. Inconsistencies due to bias (e.g., [4,5,9]), arise, for example, when examiners analyze the same evidence, but each time within different task-irrelevant contextual information (e.g., one examiner is told that the suspect confessed to the crime, whereas the other examiner is told that someone else confessed to the crime—see [22] for eight different sources of bias). Thus, distinguishing between inconsistencies that arise due to bias (e.g., [4,5,9]), versus inconsistencies that arise even without bias (e.g., [3,10,11]).

Inconsistencies that are not due to bias may arise from different analytical methods and laboratory policies, use of different statistical software, and a whole range of factors. However, differences in methods, software, SOPs, and so forth cannot explain result of studies which were conducted within a single forensic laboratory (e.g., in DNA [9]) and when inconsistencies (e.g., in fingerprinting [4,5]) were found within the same expert (for a review see [25]). It is important to note that regardless of the reasons for inconsistencies, they are unacceptable from a scientific and a criminal justice point of view (as outlined above).

Third, we need to distinguish between inconsistencies in conclusions and inconsistencies in observations. Consider two dentists, one says you need a root canal whereas the other says you are fine. The dentists—like forensic examiners—may "see" different things in the x-ray (observation) or they may see the same things in the x-ray but disagree over their meaning (conclusion). Thus, the two dentists may be inconsistent in interpreting the x-rays; however, their inconsistency in conclusions may actually arise from inconsistent in what they observe in the x-ray. In fingerprinting, for example, examiners can see the same friction ridge and minutia but nevertheless reach different conclusions. However, examiners may observe different minutia on the same fingermark, and even the *same* examiner may observe different minutia on the *same* fingermark [8].

Making these three distinctions, organizes the inconsistency data into eight categories [26]. Four of the eight relate to observations and four to conclusions, four relate to between examiners and four to within examiners, and four relate to inconsistencies without bias and four with bias (see Figure 1). The bottom of the hierarchy of inconsistencies has the most basic inconsistencies, within expert, no bias and in observation—thus inconsistencies that are not due to bias or to individual differences or even to the interpretation of the evidence; this level is purely observation inconsistencies. As we go up the hierarchy, the inconsistencies are more pronounced, as they include factors such as bias, individual differences and data interpretation.

5 | THE WAY FORWARD

Minimizing these inconsistencies is paramount, so as to strengthen forensic science as a science, as well as to establish fair justice that does not depend on luck of who will examine the evidence. To move forward, one has to first acknowledge the problem. Even if currently there is no way to move forward, one must at the very least be transparent about the problem [27].

The best way to increase consistencies is for forensic domains to establish standardized analytic procedures with objective and clear criteria (for conclusions as well as observations). "The more objective the approach, the more agreement there will be in the conclusions among examiners" [28] and "subjective methods require particularly

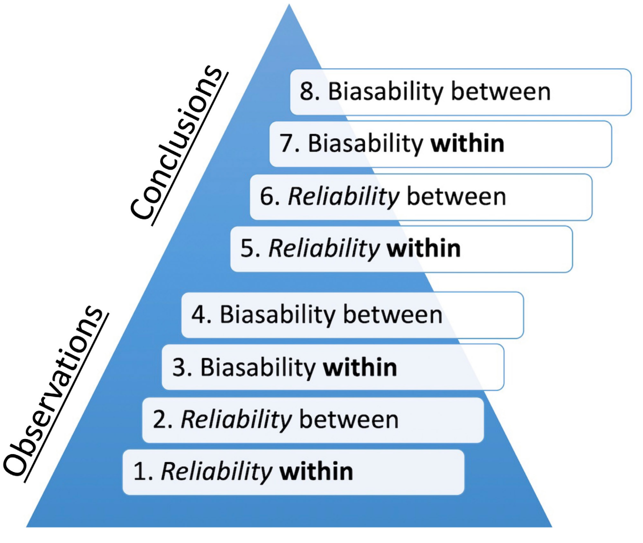

FIGURE 1 The hierarchy of expert performance (HEP), organizing inconsistency into eight categories, depending whether the inconsistency is in observations or in conclusions, is within or between examiners, and with or without bias [17,26]. At the very bottom is the most pure and basic inconsistency: even without bias, within the same examiner, and in the observations. As we go up the hierarchy, more factors contribute to the inconsistency: individual differences between examiners (e.g., training and experience), biases, and interpretations as per the meaning of the observations.

careful scrutiny because their heavy reliance on human judgment means they are especially vulnerable to human error, *inconsistency across examiners*, and cognitive bias" (emphasis added) [29]. See PCAST recommendation #2 [29], as well as [30,31].

Conversely, a word of caution about the danger in the ambition to try and be objective in forensic science [32]. It is also important to note that statistics, per se, does not give consistency, as evident by various DNA studies (e.g., [10,11]).

Until such a time that objective criteria are developed (if ever—there may be a question whether all forensic science domains can have such objective criteria—a topic beyond the scope of this paper), the question is what can be done to minimize such inconsistencies when there is no objective criteria.

Research in cognitive and decision sciences has repeatedly shown that the order in which information is examined has far reaching implications. For example, the first information is not only remembered well, but it also generates ideas and hypotheses that impact how subsequent information is examined. It may impact selective cognitive attention (what to pay attention to, what to ignore); it can form an initial impression, which can lead to confirmation bias and escalation of commitment; and so forth. Thus, the same set of information, presented in different sequences, gives rise to different conclusions (e.g., [33-37]), including in forensic science [38]. By standardizing the order of examining information in forensic analysis, for example, by using linear sequential unmasking (LSU-E), results will be more consistent [39] and make forensic science more reliable.

Such standardized order should be determined by the objectivity, relevance and biasability of the information—starting off with the more objective, more relevant, and less biasing information. Determining the order can be done by the case manager, supervisor, or the examiners themselves (for practical details, see [39-42]).

However, since many cases have typical sets of information that often repeat themselves, deciding the order does not need to be done again and again for each case, the laboratory's SOP can specify the best order for typical sets of information. And, of course, forensic examiners should not be exposed at all to information that is totally task irrelevant [22,39,40].

LSU-E is not only about minimizing bias but also about reducing noise and increasing consistencies in conclusions so as to enhance the reliability of forensic science evidence. This and other such procedures will minimize the problem highlighted in this paper: Inconsistencies are the most consistent finding in many forensic domains.

CONFLICT OF INTEREST STATEMENT

The author has no conflicts of interest to declare.

ORCID

Itiel E. Dror https://orcid.org/0000-0003-4866-209X

REFERENCES

1. Lidén M, Dror IE. Expert reliability in legal proceedings: "Eeny, Meeny, Miny, Moe, with which expert should we go?". Sci Justice. 2021;61:37–46. https://doi.org/10.1016/j.scijus.2020.09.006
2. Dror IE, Scurich N. (Mis)use of scientific measurements in forensic science. Forensic Sci Int Synergy. 2020;2:333–8. https://doi.org/10.1016/j.fsisyn.2020.08.006
3. Ulery BT, Hicklin RA, Buscaglia J, Roberts MA. Repeatability and reproducibility of decisions by latent fingerprint examiners. PLoS One. 2012;7(3):e32800. https://doi.org/10.1371/journal.pone.0032800
4. Dror IE, Charlton D. Why experts make errors. J Forensic Identif. 2006;56(4):600–16.
5. Dror IE, Charlton D, Peron AE. Contextual information renders experts vulnerable to make erroneous identifications. Forensic Sci Int. 2006;156(1):74–8. https://doi.org/10.1016/j.forsciint.2005.10.017
6. Earwaker H, Morgan RM, Harris AJ, Hall LA. Fingermark submission decision-making within a UK fingerprint laboratory: do experts get the marks that they need? Sci Justice. 2015;55:239–47. https://doi.org/10.1016/j.scijus.2015.01.007
7. Hicklin A, Richetelli N, Emerick B, Vever R, Davoren J. Variation in assessments of suitability and number of contributors for DNA mixtures. Forensic Sci Int Genet. 2023;65:102892. https://doi.org/10.1016/j.fsigen.2023.102892
8. Dror IE, Champod C, Langenburg G, Charlton D, Hunt H, Rosenthal R. Cognitive issues in fingerprint analysis: inter- and intra-expert consistency and the effect of a 'target' comparison. Forensic Sci Int. 2011;208(1–3):10–7. https://doi.org/10.1016/j.forsciint.2010.10.013
9. Dror IE, Hampikian G. Subjectivity and bias in forensic DNA mixture interpretation. Sci Justice. 2011;51(4):204–8. https://doi.org/10.1016/j.scijus.2011.08.004
10. Butler JM, Kline MC, Coble MD. NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. Forensic Sci Int Genet. 2018;37:81–94. https://doi.org/10.1016/j.fsigen.2018.07.024
11. Barrio PA, Crespillo M, Luque JA, Aler M, Baeza-Richer C, Baldassarri CL, et al. GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: results and evaluation. Forensic Sci Int Genet. 2018;35:156–63. https://doi.org/10.1016/j.fsigen.2018.05.005
12. Sunde N, Dror IE. A hierarchy of expert performance (HEP) applied to digital forensics: reliability and biasability in digital forensics decision making. Forensic Sci Int Digit Investig. 2021;37:301175. https://doi.org/10.1016/j.fsidi.2021.301175
13. Law EF, Morris KB. Evaluating firearm examiner conclusion variability using cartridge case reproduction. J Forensic Sci. 2021;66(5):1704–20. https://doi.org/10.1111/1556-4029.14758
14. Monson K, Smith E, Peters E. Repeatability and reproducibility of comparison decisions by firearms examiners. J Forensic Sci. https://doi.org/10.1111/1556-4029.15318 Epub 2023 July 02.
15. Hicklin A, Winer K, Kish P, Parks C, Chapman W, Dunagan K, et al. Accuracy and reproducibility of conclusions by forensic bloodstain pattern analysts. Forensic Sci Int. 2021;325:110856. https://doi.org/10.1016/j.forsciint.2021.110856
16. Majamaa H, Ytti A. A survey of the conclusions drawn of similar footwear cases in various crime laboratories. Forensic Sci Int. 1996;82:109–20. https://doi.org/10.1016/0379-0738(96)01972-X
17. Dror IE, Murrie D. A hierarchy of expert performance (HEP) applied to forensic psychological assessments. Psychol Public Policy Law. 2018;24(1):11–23. https://doi.org/10.1037/law0000140
18. Liden M, Almazrouei M. "Blood, bucks and bias": reliability and biasability of crime scene investigators' selection and prioritization of blood traces. Sci Justice. 2023;63:276–93. https://doi.org/10.1016/j.scijus.2023.01.005
19. Dror IE, Scherr S, Mohammed L, MacLean C, Cunningham L. Biasability and reliability of expert forensic document examiners. Forensic Sci Int. 2021;318:110610. https://doi.org/10.1016/j.forsciint.2020.110610
20. Koehler J, Liu S. Fingerprint error rate on close non-matches. J Forensic Sci. 2021;66(1):129–34. https://doi.org/10.1111/1556-4029.14580

21. Spellman BA, Quigley-McBride A. Reasoning about forensic science evidence. In: Brigaglia M, Roversi C, editors. Legal reasoning and cognitive science: Topics and perspectives. Diritto & Questioni Pubbliche; In press.
22. Dror IE. Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. Anal Chem. 2020;92(12):7998–8004. https://doi.org/10.1021/acs.analchem.0c00704
23. Cuellar M, Mauro J, Luby A. A probabilistic formalisation of contextual bias: from forensic analysis to systemic bias in the criminal justice system. J R Stat Soc Ser A Stat Soc. 2022;185:S620–43. https://doi.org/10.1111/rssa.12962
24. Kahneman D, Sibony O, Sunstein CR. Noise: a flaw in human judgment. London: William Collins; 2021. https://doi.org/10.1111/jols.12345
25. Kukucka J, Dror IE. Human factors in forensic science: psychological causes of bias and error. In: DeMatteo D, Scherr K, editors. The Oxford handbook of psychology and law. Oxford: Oxford University Press; 2023. p. 621–42. https://doi.org/10.1093/oxfordhb/9780197649138.013.36
26. Dror IE. A hierarchy of expert performance (HEP). J Appl Res Mem Cogn. 2016;5(2):121–7. https://doi.org/10.1016/j.jarmac.2016.03.001
27. Almazrouei MA, Dror IE, Morgan R. The forensic disclosure model: what should be disclosed to, and by, forensic experts? Int J Law Crime Justice. 2019;59:100330. https://doi.org/10.1016/j.ijlcj.2019.05.003
28. Kafadar K. The need for objective measures in forensic evidence. Significance. 2019;16:16–20. https://doi.org/10.1111/j.1740-9713.2019.01249.x
29. President's Council of Advisors on Science and Technology (PCAST). Report to the President—Forensic science in criminal courts: ensuring scientific validity of feature-comparison methods. 2016. Available from: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf. Accessed 15 Aug 2023.
30. Albright TD. How to make better forensic decisions. PNAS. 2022;119:e2206567119. https://doi.org/10.1073/pnas.2206567119
31. Dorfman AH, Valliant R. Inconclusives, errors, and error rates in forensic firearms analysis: three statistical perspectives. Forensic Sci Int Synergy. 2022;5:100273. https://doi.org/10.1016/j.fsisyn.2022.100273
32. Dror IE. The ambition to be scientific: human expert performance and objectivity. Sci Justice. 2013;53(2):81–2. https://doi.org/10.1016/j.scijus.2013.03.002
33. Dean ML. Presentation order effects in product taste tests. J Psychol. 1980;105:107–10. https://doi.org/10.1080/00223980.1980.9915138
34. Asche SE. Forming impressions of personality. J Abnorm Soc Psychol. 1946;41:258–90. https://doi.org/10.1037/h0055756
35. Hovland CI, editor. The order of presentation in persuasion. London: Yale University Press; 1957.
36. Darley JM, Gross PH. A hypothesis-confirming bias in labeling effects. J Pers Soc Psychol. 1983;44:20–33. https://doi.org/10.1037/0022-3514.44.1.20
37. Barry C, Halfmann K. The effect of mindset on decision-making. J Integr Soc Sci. 2016;6:49–74.
38. Davidson M, Nakhaeizadeh S, Rando C. Cognitive bias and the order of examination in forensic anthropological non-metric methods: a pilot study. Aust J Forensic Sci. 2023;5:255–71. https://doi.org/10.1080/00450618.2021.1998625
39. Dror IE, Kukucka J. Linear Sequential Unmasking–*Expanded* (LSU-E): a general approach for improving decision making as well as minimizing noise and bias. Forensic Sci Int Synergy. 2021;3:100161. https://doi.org/10.1016/j.fsisyn.2021.100161
40. Kunkler KS, Roy T. Reducing the impact of cognitive bias in decision making: practical actions for forensic science practitioners. Forensic Sci Int Synergy. 2023;7:100341. https://doi.org/10.1016/j.fsisyn.2023.100341
41. Quigley-McBride A, Dror IE, Roy T, Garrett BL, Kukucka J. A practical tool for information management in forensic decisions: using Linear Sequential Unmasking-Expanded (LSU-E) in casework. Forensic Sci Int Synergy. 2022;4:100216. https://doi.org/10.1016/j.fsisyn.2022.100216
42. Whitehead FA, Mr W, Sigman ME. Decision theory and linear sequential unmasking in forensic fire debris analysis: a proposed workflow. Forensic Chem. 2022;29:100426. https://doi.org/10.1016/j.forc.2022.100426

How to cite this article: Dror IE. The most *consistent* finding in forensic science is *inconsistency*. J Forensic Sci. 2023;00:1–5. https://doi.org/10.1111/1556-4029.15369