

ARTICLE

HOW CAN A FORENSIC RESULT BE A “DECISION”?

A CRITICAL ANALYSIS OF ONGOING REFORMS OF FORENSIC REPORTING FORMATS FOR FEDERAL EXAMINERS

Simon A. Cole & Alex Biedermann***

ABSTRACT

The decade since the publication of the 2009 National Research Council report on forensic science has seen the increasing use of a new word to describe forensic results. What were once called “facts,” “determinations,” “conclusions,” or “opinions,” are increasingly described as “decisions.” Prior to 2009, however, the term “decision” was rarely used to describe forensic results. Lay audiences, such as lawyers, might be forgiven for perceiving this as a surprising turn. In its plain English meaning, a “decision” would seem to be a strange word choice to describe the

* Professor of Criminology, Law & Society, University of California, Irvine; Ph.D. (science & technology studies), Cornell University; A.B., Princeton University. This research was funded in part by the Swiss National Science Foundation through grant BSSG10_155809 and the Center for Statistical Applications in Forensic Evidence, sponsored by the National Institute of Standards and Technology. For comments on drafts of this Article, we are grateful to Timothy Lau, Julia Leighton, Amy Castillo, and the other participants in the University of Houston Law Center Symposium on the Future of Crime Labs and Forensic Science: Sandra Guerra Thompson, Jed Rakoff, Jessica Gabel Cino, Sarah Chu, Valena Beety, Bob Wicoff, Gerald Doyle, Nicole Cásarez, Edward Cheng, Brandon Garrett, and Nick Hughes. We are also grateful to the student editors of the *Houston Law Review* for their hard work on this Article.

** Associate Professor, University of Lausanne, Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, 1015 Lausanne-Dorigny (Switzerland); Ph.D., M.Sc. (forensic science), University of Lausanne. Visiting Researcher at Northwestern University Pritzker School of Law, Chicago, IL.

outcome of a scientific analysis, given its connotation of choice and preference. In this Article, we trace the recent history of the term “decision” in forensic analysis. We simply and clearly explain the scientific fields of “decision theory” and “decision analysis” and their application to forensic science. We then analyze the Department of Justice (DOJ) Uniform Language for Testimony and Reporting (ULTR) documents that use the term. We argue that these documents fail to articulate coherent frameworks for reporting forensic results. The Article identifies what we perceive to be some key stumbling blocks to developing such frameworks. These include a reluctance to observe decision theory principles, a reluctance to cohere with sound probabilistic principles, and a reluctance to conform to particular logical concepts associated with these theories, such as proper scoring rules. The Article elucidates each of these perceived stumbling blocks and proposes a way to move forward to more defensible reporting frameworks. Finally, we explain what the use of the term “decision” could accomplish for forensic science and what an appropriate deployment of the term would require.

TABLE OF CONTENTS

I. INTRODUCTION	553
II. COLLOQUIAL VERSUS FORMAL UNDERSTANDINGS OF THE NOTION OF DECISION	558
III. DECISION THEORY IN FORENSIC SCIENCE	563
A. <i>Decision Theory</i>	563
B. <i>Forensic Decisionalism</i>	566
C. <i>Decision Theoretic Comparison of Rival Decisions</i>	570
IV. CRITICAL ANALYSIS AND DISCUSSION OF ULTRs	574
A. <i>Preliminaries</i>	574
B. <i>Descriptive Analysis of ULTR Contents</i>	576
C. <i>ULTRs Read from a Decision-Analytic Viewpoint</i>	579
1. <i>Inverting Decision Theory</i>	580
2. <i>Deciding Probabilities?</i>	583
D. <i>A Coherent Way of Deciding Probabilities</i>	585

V. APPLYING DECISION THEORY	587
A. <i>Practitioner Uses of “Decision”</i>	590
B. <i>What Should Forensic Analysts Report?</i>	592

I. INTRODUCTION

What should forensic scientists call the results¹ of their analyses? A survey of forensic disciplines, providers, and practitioners would reveal a great variety of answers to this question. As an example, consider Table 1, a summary of the different terms suggested for reporting results for sixteen disciplines by one set of testimonial standards, the draft Uniform Language for Testimony and Reporting (ULTR), issued by the U.S. Department of Justice (DOJ) in 2016–2017. The draft ULTRs contain a wide variety of different words for results: conclusions, opinions, determinations, associations, findings, results, and classifications. What is the difference between these words?

Table 1. Proposed and Approved ULTRs Compared

Discipline	Draft ULTR (2016–2017)	Words used for output of analysis	Approved ULTR (2018–2019)	Basis for conclusion
Glass	X	Conclusion	X	Decision
Metallurgy	X	Determination Association Result	X	Decision
Geology	X	Conclusion	X	Decision
Anthropology	X	Determination	X	Decision
Hair	X	Consistency Classification	X	Decision
Latent Print	X	Determination	X	Decision
Fiber	X	Association	X	Decision
Firearms/ toolmark: pattern			X	Decision
Firearms/ toolmark: fracture			X	Decision
Serology	X	Result	X	Interpretation
MtDNA	X	Determination	X	Interpretation

1. In order to avoid playing favorites among competing terms, we use “results” as the most generic term in this Article, even though “results” counts among those competing terms.

ySTR			X	Interpretation
Autosomal DNA			X	Likelihood ratio
General Chemistry	X	Opinion Conclusion Result	X	Determination
Footwear/ Tire	X	Opinion		
Toxicology	X	Opinion Finding Result		
Fiber	X	Association		
Paint	X	Conclusion Association Determination Opinion Result		
Explosive Device	X	Determination Association		
Explosive Chemistry	X	Opinion Conclusion Determination Finding Result		
Handwriting	X	Opinion		
<p>— Highlighted rows indicate disciplines for which both a Proposed and an Approved ULTR exist.</p> <p>— As noted in the text, Approved ULTRs use the word decision, except the four ULTRs pertaining to biological evidence.</p> <p>—The third column indicates the terms used in the draft ULTR where the Approved ULTR uses the term “decision.”</p>				

Table 2. Summary of Table 1

Term	Number of mentions in draft or approved ULTRs
Decision	10
Determination	8
Opinion	6
Conclusion	5
Association	5
Result	5
Interpretation	3
Likelihood ratio	1
Finding	1
Consistency	1
Classification	1

Although it is possible to claim that all these words are essentially interchangeable, that seems far too simple. Each of these words conveys subtle differences in the epistemic strength that is attached to the scientific claim. A “determination,” for example, suggests that the result has been “determined” by the scientific evidence—that is, that no other interpretation of the evidence is reasonably conceivable. A “conclusion,” while not quite as strong, also conveys the notion that the result logically follows from the evidence. (“Conclusion” also has a weaker meaning, though, in the sense that it can merely refer descriptively to the final section of a scientific report.) “Opinion,” on the other hand, conveys that there are at least two possible interpretations of the evidence and that the expert is exercising some sort of judgment in advocating for one over the other(s). “Findings” and “results” are perhaps the most scientific-sounding terms and also seem the most neutral with regard to the above issues; these terms may be seen as descriptive of what has been observed during examination. They also differ from the other terms by *lacking* an implication of moving on to the step of drawing an inference from the evidence, much in the way that the “Results” section of a classically organized scientific paper merely reports the evidence, whereas the “Conclusion” section draws (sometimes speculative) inferences from that evidence.² “Association” and “classification” sound as if

2. See generally Frederic L. Holmes, *Argument and Narrative in Scientific Writing*, in *THE LITERARY STRUCTURE OF SCIENTIFIC ARGUMENT: HISTORICAL STUDIES* 164 (Peter Dear ed., 1991) (discussing the standard arrangement of scientific papers).

they are deliberately chosen for their weakness to temper the scientific claim with an appropriate dose of “epistemological humility.”³ If we were to order all these terms according to their perceived epistemic strength, we might have: determination, conclusion, opinion, classification, association, finding, and result.

As we pass the ten-year anniversary of the publication of the National Research Council (NRC) report, *Strengthening Forensic Science in the United States*,⁴ however, criminal lawyers may find themselves encountering a different term increasingly often. That term is “decision.” At first glance, to call the result of a forensic analysis a “decision,” rather than, say, a “conclusion,” seems decidedly strange. Historically, the results of forensic analyses have tended to be described as “opinions.”⁵ “Decision,” on the other hand, seems to imply a degree of choice and preference that would seem out of place in a scientific analysis. What could possibly be meant by calling the result of a forensic analysis a “decision”?

Another development over the decade since the publication of the NRC report (although it actually began earlier, around 2005) has been the publication of a body of scientific literature discussing the application of “decision theory,” or “decision analysis,” to forensic science problems at the reporting stage.⁶ The terms

3. See Jennifer L. Mnookin, *The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate*, 7 LAW PROBABILITY & RISK 127, 139 (2008).

4. NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD (2009), <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf> [<https://perma.cc/59PP-X864>].

5. See Paul L. Kirk, *The Ontogeny of Criminalistics*, 54 J. CRIM. L. CRIMINOLOGY & POLICE SCI. 235, 238 (1963).

6. E.g., FRANCO TARONI ET AL., DATA ANALYSIS IN FORENSIC SCIENCE: A BAYESIAN DECISION PERSPECTIVE (Stephen Senn & Vic Barnett eds., 2010); FRANCO TARONI ET AL., BAYESIAN NETWORKS FOR PROBABILISTIC INFERENCE AND DECISION ANALYSIS IN FORENSIC SCIENCE (2d ed. 2014); Alex Biedermann et al., *Analysing and Exemplifying Forensic Conclusion Criteria in Terms of Bayesian Decision Theory*, 58 SCI. & JUST. 159 (2018); Alex Biedermann et al., *Decision Theoretic Properties of Forensic Identification: Underlying Logic and Argumentative Implications*, 177 FORENSIC SCI. INT'L 120, 121–29 (2008) [hereinafter Biedermann et al., *Decision Theoretic Properties*]; Alex Biedermann et al., *The Decisionalization of Individualization*, 266 FORENSIC SCI. INT'L 29 (2016) [hereinafter Biedermann et al., *The Decisionalization*]; Alex Biedermann et al., *The Consequences of Understanding Expert Probability Reporting as a Decision*, 57 SCI. & JUST. 80, 83–84 (2017); Simone Gittelsohn et al., *Decision-Theoretic Reflections on Processing a Fingerprint*, 226 FORENSIC SCI. INT'L e42, e43–e44 (2013); Franco Taroni et al., *Decision Analysis in Forensic Science*, 50 J. FORENSIC SCI. 1 (2005) [hereinafter Taroni et al., *Decision*]; Franco Taroni et al., *Value of DNA Tests: A Decision Perspective*, 52 J. FORENSIC SCI. 31 (2007); Simone Gittelsohn, *Evolving from Inferences to Decisions in the Interpretation of Scientific Evidence* 1 (2013) (Ph.D. thesis, School of Criminal Justice, University of Lausanne), https://serval.unil.ch/resource/serval:BIB_620A73F01CCC.P001/REF.pdf [<https://perma.cc/AE94-4WQS>].

“decision” and “decision-making” (by experts) have also been studied from a forensic psychological perspective.⁷

What, if anything, do these converging historical phenomena involving the same word have to do with one another? It is possible, of course, that these are two unrelated phenomena which happen to employ the same word. This seems unlikely. Although the ULTRs do not explicitly refer to any of the decision theory literature, it would be strange indeed if these parallel developments were a mere coincidence—that is, to suppose that the appearance of the term “decision” in forensic reporting standards has nothing to do with the scholarly literature that discusses that same term.

What explains the proliferation of this term, and what does it mean for forensic science and for law? In Part II, we introduce and analyze the increasing use of the term “decision” in contemporary forensic reporting standards, especially the recently issued ULTRs regarding forensic reporting formats for federal examiners. In Part III, we provide a brief introduction to decision theory and its application to legal problems and, in particular, those involving forensic evidence. In Part IV, we discuss, and critically analyze in detail, the use of the term “decision” in the Approved ULTRs. Although the ULTRs remain ambiguous as to whether the term is intended to invoke formal decision theory or not, we argue that the ULTRs misuse the notion of decision—whether meant formally or colloquially—in important ways. In Part V, we discuss the way in which decision theory, properly applied, can be useful in framing legal problems involving forensic evidence. Ultimately, we argue that while decision theory may be useful to forensic practitioners, primarily in educating them about what *not* to do, it should be of great benefit to legal practitioners in understanding the requirements of making legal decisions based on forensic evidence and in better understanding where scientific analysis should end and legal analysis should begin.

7. E.g., Itiel E. Dror et al., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 FORENSIC SCI. INT'L 74, 75–77 (2006); William C. Thompson, *Determining the Proper Evidentiary Basis for an Expert Opinion: What Do Experts Need to Know and When Do They Know Too Much?*, in BLINDING AS A SOLUTION TO BIAS: STRENGTHENING BIOMEDICAL SCIENCE, FORENSIC SCIENCE, AND LAW 133, 147–48 (2016).

II. COLLOQUIAL VERSUS FORMAL UNDERSTANDINGS OF THE NOTION OF DECISION

The proliferation of forensic experts characterizing the results of their analyses as “decisions” appears to be a relatively recent phenomenon. We have been unable to find any forensic standards document that uses the term “decision” prior to 2011.⁸ And yet, the use of the term “decision” appears to be on the rise. In 2011 the U.S. standard-setting body for friction ridge (“latent print” or “fingerprint”) analysis changed the word “conclusion,” which it had used in its 2009 standard, to “decision” in describing the results of analyses.⁹ Observe below the change in the description of the “evaluation” step of friction ridge analysis¹⁰ between the 2009 NRC report (that is the subject of this Symposium) and the 2012 National Institute of Standard and Technology/National Institute of Justice (NIST/NIJ) report on human factors in friction ridge analysis only three years later:

NRC (2009)

Source **determination** is made when the examiner **concludes**, based on his or her experience, that sufficient quantity and quality of friction ridge detail is in agreement between the latent print and the known print.¹¹

NIST/NIJ (2012)

In the Evaluation phase, the examiner makes the ultimate **decision** regarding source attribution.¹²

8. See SCI. WORKING GRP. ON FRICTION RIDGE ANALYSIS, STUDY & TECH., STANDARDS FOR EXAMINING FRICTION RIDGE IMPRESSIONS AND RESULTING CONCLUSIONS (2011), http://clpex.com/swgfast/documents/examinations-conclusions/111026_Examinations-Conclusions_1.0.pdf [<https://perma.cc/Y278-GJGC>] (using the term “decision” to refer to examiners’ conclusions about fingerprints).

9. *Id.*; SCI. WORKING GRP. ON FRICTION RIDGE ANALYSIS, STUDY & TECH., FRICTION RIDGE EXAMINATION METHODOLOGY FOR LATENT PRINT EXAMINERS (2009), <http://clpex.com/swgfast/documents/methodology/100506-Methodology-Reformatted-1.01-Archived.pdf> [<https://perma.cc/X4SS-KE9M>].

10. This is the “E” in the notorious “ACE-V methodology.” Sandy L. Zabell, *Fingerprint Evidence*, 13 J.L. & POLY 143, 154, 178 (2005) (characterizing ACE-V as “an acronym, not a methodology”).

11. NAT’L RESEARCH COUNCIL, *supra* note 4, at 138 (emphasis added).

12. EXPERT WORKING GRP. ON HUMAN FACTORS IN LATENT PRINT ANALYSIS, LATENT PRINT EXAMINATION AND HUMAN FACTORS: IMPROVING THE PRACTICE THROUGH A SYSTEMS APPROACH 7 (2012) (emphasis added), https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=910745 [<https://perma.cc/GLS9-HPTE>].

The change in other disciplines has been slower, but draft standards documents used the term “decision” for drug analysis in 2014¹³ and glass analysis in 2017.¹⁴

The explosion of the use of “decision” across the forensic disciplines, however—and the primary subject of this Article—dates to the publication of the ULTR documents by the DOJ in 2018 and 2019, almost exactly a decade after the publication of the NRC report.

In February 2016, the Deputy Attorney General of the United States announced that the DOJ would develop what would later become the ULTRs, but were then called “Approved Scientific Standards for Testimony and Reports” (ASSTRs), in many forensic disciplines.¹⁵ Describing the ASSTRs, the Deputy Attorney General said, “We hope this effort will serve as a model for demonstrating our commitment to strengthening forensic science, now and in the future.”¹⁶ The first draft documents, by then renamed ULTRs, were published for public comment in June 2016.¹⁷

In April 2017, the U.S. Attorney General sunsetted the National Commission on Forensic Science (NCFS).¹⁸ This was an important development, because the NCFS had been created in 2013 “to provide recommendations and advice to the Department of Justice (DOJ) concerning national methods and strategies for: strengthening the validity and reliability of the forensic sciences.”¹⁹ The Attorney General replaced the NCFS with a

13. SCI. WORKING GRP. FOR THE ANALYSIS OF SEIZED DRUGS, RECOMMENDATIONS 9, 28, 53 (2014), http://www.swgdrug.org/Documents/SWGDRUG%20Recommendations%20Version%207-0_Archived.pdf [<https://perma.cc/LQ82-6LHQ>].

14. Overseas Sec. Advisory Council, Standard Practice for Interpretation and Report Writing in Forensic Comparisons of Trace Materials 16 (May 4, 2017) (unpublished draft) (on file with the first Author).

15. Sally Q. Yates, Deputy Attorney Gen., Remarks During the 68th Annual Science Meeting Hosted by the American Academy of Forensic Science (Feb. 24, 2016), <https://www.justice.gov/opa/speech/deputy-attorney-general-sally-q-yates-delivers-remarks-during-68th-annual-scientific> [<https://perma.cc/DC58-CRBX>] (“To address this problem [of testimonial overstatement revealed by the microscopic hair comparison review], the FBI is close to finalizing new internal standards for testimony and reporting—which they’re calling ‘Approved Scientific Standards for Testimony and Reports,’ or ASSTR.”).

16. *Id.*

17. *Forensic Science*, U.S. DEP’T JUST. ARCHIVES, <https://www.justice.gov/archives/dag/forensic-science> [<https://perma.cc/N4V2-J32J>] (last visited Jan. 31, 2020).

18. Spencer S. Hsu, *Sessions Orders Justice Dept. to End Forensic Science Commission, Suspend Review Policy*, WASH. POST (Apr. 10, 2017), https://www.washingtonpost.com/local/public-safety/sessions-orders-justice-dept-to-end-forensic-science-commission-suspend-review-policy/2017/04/10/2dada0ca-1c96-11e7-9887-1a5314b56a08_story.html [<https://perma.cc/4286-HNRG>].

19. NAT’L COMM’N ON FORENSIC SCI., U.S. DEP’T OF JUSTICE, CHARTER 1 (2015), <http://www.justice.gov/ncfs>.

Forensic Science Working Group.²⁰ In the announcement of this new organization, the ULTRs were among only two specific projects mentioned that the DOJ would pursue “aimed at ensuring that the testimony of the Justice Department’s forensic examiners is consistent with sound scientific principles and just outcomes.”²¹

ULTRs are part of the DOJ’s “quality assurance measures to help ensure that the results of forensic analyses are properly qualified and appropriately communicated in both reports and testimony.”²² In essence, an ULTR purportedly “reflects the range of appropriate conclusions that Department examiners may provide in reports and testimony. It also sets forth important scientific limitations on those conclusions and other testimonial assertions.”²³

In 2018, at the annual meeting of the American Academy of Forensic Science, Deputy Attorney General Rod Rosenstein announced the DOJ’s “[p]lans to [a]dvance [f]orensic [s]cience.”²⁴ In fleshing out these “[p]lans,” the DOJ’s press release listed four specific actions.²⁵ First among these was the publication of the ULTR for the latent print discipline.²⁶ As Rosenstein noted, the latent print ULTR “is the first approved Uniform Language document.”²⁷ Consistent with the increasing use of the term “decision” in that discipline, the latent print ULTR used the term “decision” to characterize the result of friction ridge analyses.²⁸ In

ps://www.justice.gov/archives/ncfs/file/624216/download [https://perma.cc/U2H6-KEWQ].

20. Beth Reinhard, *Sessions Scuttles Forensics Team*, WALL ST. J., Aug. 8, 2017, at A5; Press Release, U.S. Dep’t of Justice, Justice Dep’t Announces Plans to Advance Forensic Sci. (Aug. 7, 2017), <https://www.justice.gov/opa/pr/justice-department-announces-plans-advance-forensic-science> [https://perma.cc/2W5B-6Z8K].

21. Press Release, U.S. Dep’t of Justice, *supra* note 20.

22. Memorandum from the Deputy Attorney Gen. to Heads of Dep’t Components 2 (Jan. 18, 2018), <https://www.justice.gov/file/1036796/download> [https://perma.cc/MD32-2V8F].

23. *Id.* at 1.

24. Press Release, U.S. Dep’t of Justice, *supra* note 20.

25. *Id.*

26. *See id.* The other three were: (1) a testimony monitoring framework; (2) plans (as yet unfulfilled) to publish documents such as quality management documents and internal validation studies (and, presumably, standard operating procedures); and (3) “the rechartering of the Council of Federal Forensic Laboratory Directors.” *Id.* For more on standard operation procedures on the FBI’s Latent Print Unit, see Simon A. Cole, *Implementing Counter-Measures Against Confirmation Bias in Forensic Science*, 2 J. APPLIED RES. MEMORY & COGNITION 61 (2013).

27. Rod J. Rosenstein, Deputy Attorney Gen., U.S. Dep’t of Justice, Remarks at the American Academy of Forensic Science (Feb. 21, 2018), <https://www.justice.gov/opa/speech/deputy-attorney-general-rostein-delivers-remarks-american-academy-forensic-sciences> [https://perma.cc/R8HF-62C2].

28. Simon A. Cole, *A Discouraging Omen: A Critical Evaluation of the Approved Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline*, 34

2019, twelve more Approved ULTRs were published. As noted *supra*, nearly all of the Approved ULTRs use the term “decision,” replacing words like “determination,” “conclusion,” or “result.”²⁹

How does “decision” differ from these discarded terms? “Decision” can be distinguished from these terms by a certain sense of choice or free will on the part of the expert. Formal decision theory uses the term *preference*, that is the decision-maker’s expression of preferences among decision options (possible courses of action), and that term works quite well for the colloquial usage of the word “decision” as well. To be sure, the word “opinion” seems to contain the notion of choice as well. But “opinion” suggests that the expert is choosing the most plausible of the possible interpretations. A scientist’s “opinion” (about a scientific matter) must necessarily—according to common understandings—be that which she believes most likely to be true. But a *decision* does not necessarily seem to require that. And, indeed, as we shall see below, decision theory can offer examples of cases in which one can rationally make a “decision” that contradicts one’s “opinion.”³⁰

To illustrate the differences among these words, we attach each of them to a scientific result in Table 3.

GA. ST. U. L. REV. 1103, 1119–20 (2018).

29. See *supra* Table 1.

30. See *infra* Part III.

Table 3. *Illustration of the Differences Between the Terms “Determination,” “Conclusion,” “Opinion” and “Decision”*

Suppose a physician says . . .	Possible interpretation of the physician’s statement
. . . she determined that the patient has a medical condition.	This would seem to convey that the physician has gotten a positive result from some sort of test that strongly indicates the target condition, though it conveys nothing about what, if anything, will be communicated to the patient.
. . . she concluded that the patient has a medical condition.	This sounds less like the physician is relying on a single definitive test, but rather is relying on a heterogeneous assemblage of evidence, but nonetheless the physician seems to be claiming that the diagnosis (probably) follows from that evidence.
. . . it is her opinion that the patient has a medical condition.	Now the physician sounds a bit less certain of the diagnosis. She seems to be acknowledging the possibility of alternative interpretations of the evidence but still believes that the diagnosis is the most plausible interpretation.
. . . she has decided that the patient has a medical condition.	This would sound strange to most people’s ears. Why is the physician using the term “decide” instead of one of the above terms? It sounds like some deliberate element, i.e. <i>the physician’s</i> preference, has been incorporated into the analysis, and this does not seem appropriate. ³¹

31. Notice that we are talking only about “deciding” on the presence of the medical condition. For a physician to decide whether *to tell* the patient they have a medical condition or how *to treat* the patient for the condition do not sound nearly as strange, are different matters entirely, and, in fact, as we shall see *infra*, are good illustrations of the proper application of decision theory.

Consider likewise the following examples:

- A theoretical physicist says she has “determined” that the Higgs boson exists.
- A theoretical physicist says she has “concluded” that the Higgs boson exists.
- A theoretical physicist says it is her “opinion” that the Higgs boson exists.
- A theoretical physicist says she has “decided” that the Higgs boson exists.

- A climate scientist has “determined” that anthropogenic climate change is occurring.
- A climate scientist has “concluded” that anthropogenic climate change is occurring.
- A climate scientist says it is her “opinion” that anthropogenic climate change is occurring.
- A climate scientist has “decided” that anthropogenic climate change is occurring.

In both of these illustrations, the “decision” example stands out as something that sounds unsuitable for an expert to be saying about a scientific matter. Again, it sounds like it contains some element that should not be there. That element, we argue, is *preference*. Making that very notion of preference transparent and logically coherent is one of the aims of decision theory as we shall explain in the next Part.

III. DECISION THEORY IN FORENSIC SCIENCE

A. *Decision Theory*

The 1940s, and the decades that followed, marked a time characterized by an increasing interest in systematic methods of problem analysis, problem solving, and decision-making.³² At the same time, the development of the computer favored fields of study with a common interest in decision-making and decision analysis, in particular artificial intelligence.³³ The notion of decision also

32. Ralph F. Miles, Jr., *The Emergence of Decision Analysis*, in *ADVANCES IN DECISION ANALYSIS: FROM FOUNDATIONS TO APPLICATIONS* 13, 22–25 (Ward Edwards et al. eds., 2007).

33. Eric J. Horvitz et al., *Decision Theory in Expert Systems and Artificial Intelligence*, 2 *INT’L J. APPROXIMATE REASONING* 247, 248 (1988).

attracted researchers in disciplines such as mathematics, statistics, philosophy of science, and psychology, among others, sometimes referred to as *decision sciences*.³⁴ Economists, for example, were interested in applications where decisions had monetary consequences, and the aim was to make decisions that reflected coherent behavior with regard to the decision-maker's preferences among decision consequences and judgments about uncertain events upon which decision consequences depend.³⁵ Not surprisingly, such considerations also found appeal among legal scholars.³⁶

Generally speaking, decision theory combines probability theory with utility theory. It is a mathematical theory for analyzing decision problems under uncertainty, providing criteria for the comparison of rival decisions. The probability component of decision theory should be well-known to at least some legal readers: probability provides a way to coherently assign beliefs about *propositions* (i.e., assertions about the real-world)³⁷ when knowledge and information are incomplete. In turn, utility theory provides a framework for appraising the relative desirability of the various possible decision consequences.³⁸ Here, a decision consequence (also called outcome) is understood as the result of choosing a particular action when a particular state of the world applies.

Decision theory as considered here focuses on an individual's point of view; it is an individualistic theory in that it supposes a single decision-maker or a group of persons who act in common (i.e., express a common opinion).³⁹ For our present purposes, we

34. See Howard Raiffa, *Decision Analysis: A Personal Account of How It Got Started and Evolved*, 50 OPERATIONS RES. 179, 184 (2002).

35. E.g., John W. Pratt et al., *The Foundations of Decision Under Uncertainty: An Elementary Exposition*, 59 J. AM. STAT. ASS'N 353, 356 (1964).

36. E.g., Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1030–31 (1977) (applying decision theory in analyzing courts' discretion to exclude relevant evidence).

37. In the context here, a proposition is a "statement that is true or false, that can be affirmed or denied." TERENCE ANDERSON ET AL., ANALYSIS OF EVIDENCE 385 (Cambridge Univ. Press 2d ed. 2005) (1991).

38. The notion of preference among decision consequences is well-known in legal scholarship and often associated with, for example, Blackstone's preference for "missed convictions" (i.e., wrongly freeing defendants) over wrongful convictions. It should be noted, however, that Blackstone's preference statement seems to relate to error *ratios* across multiples cases, rather than the expression of relative losses for adverse outcomes in a given single case. E.g., D.H. Kaye, *Clarifying the Burden of Persuasion: What Bayesian Decision Rules Do and Do Not Do*, 3 INT'L J. EVIDENCE & PROOF 1, 4–5 (1999); Larry Laudan & Harry D. Saunders, *Re-Thinking the Criminal Standard of Proof: Seeking Consensus About the Utilities of Trial Outcomes*, 7 INT'L COMMENT. ON EVIDENCE, no. 2, 2009, at 1, 12–13.

39. See Dennis V. Lindley, *The Philosophy of Statistics*, 49 STATISTICIAN 293, 313

leave aside additional complications such as situations of conflict and competition. As we will see below, decision theory in the classic individualistic sense is powerful enough to substantially clarify and sharpen our understanding of the subject of this Article. Methodologically, it is a prerequisite for any future, more advanced levels of analysis.

For illustration, consider a simplified application of the elements of decision theory. Its reasoning also complies with common sense⁴⁰ and could readily be applied to the colloquial meaning of the word “decision.” To avoid idiosyncrasies associated with forensic science examples, consider the following simple two-decision/two-state-of-nature problem.⁴¹ Imagine the President of a large university. A tropical depression is bearing down on the city in which the university is located, along with the possibility of flooding on the following day. The President must *decide* whether: (1) to close the university; or (2) NOT close the university.⁴² The President reasons as follows:

1. I **believe** that it is more probable that it will NOT flood than that it will flood. I believe the probability that it will flood is 1/3 and the probability that it will NOT flood is 2/3.

2. If I am correct either way, I will incur no additional penalty attributable to my decision (though, of course, there may be monetary losses resulting from physical damages to campus facilities). If I close the university and it floods, I will be lauded for my prudence. If I keep the university open and it does not flood, I will be commended for my mettle.

3. However, if I am wrong either way, I will suffer a consequence that, for me, represents a **loss**. If I keep the university open and it floods, the **consequence** is that people may be placed in dangerous situations, and my acting may critically be exposed as reckless, possibly damaging my reputation.

(2000).

40. *E.g.*, Alex Biedermann et al., *Normative Decision Analysis in Forensic Science, ARTIFICIAL INTELLIGENCE & L.* (forthcoming 2020), <https://link.springer.com/article/10.1007%2Fs10506-018-9232-2> [<https://perma.cc/7MMZ-JBDH>].

41. A similar example is described in Taroni et al., *Decision*, *supra* note 6, at 5–9.

42. Participants in, and prospective attendees of, the University of Houston Law Center Symposium on the Future of Crime Labs and Forensic Science may find this illustration eerily familiar. For an example regarding the decision to order an evacuation in the context of earthquakes, see Stephen S. Hall, *At Fault?*, 477 NATURE 264, 265, 269 (2011), for a report on the devastating earthquake in the area around the Italian city of L'Aquila, causing over 300 fatalities and leading to subsequent trials for manslaughter of individuals involved in assessing whether an earthquake was imminent.

4. On the other hand, if I close the university and it does not flood, the **consequence** is that people will lose out on a variety of educational activities for no reason. As just one of many examples, academic symposia scheduled for that day, for which scholars may have traveled great distances and endured extensive delays, may be cancelled.

5. However, the **losses** associated with (3) are much *greater* than the **losses** associated with (4) because they involve physical danger, rather than mere inconvenience. Stated otherwise, I greatly prefer the consequences associated with (4) to those associated with (3).⁴³

6. Hence, I **decide** to close the university, even though my **opinion** is that it is more probable that it will NOT flood, than that it will flood (see point (1)).

In summary, the example illustrates a situation in which each of the two decisions can lead to a specific adverse outcome, or to a nonadverse outcome. If one way of deciding can lead to a much more adverse consequence than the alternative way of deciding, then we should weigh the stakes involved against our beliefs of how the world will turn out.⁴⁴ For the example considered here, even though there is a preponderant probability that it will *not* flood, the more severe potential consequences of flooding deter the President from keeping the university open. Instead, one ought to opt for the decision to close the university. Decision theory allows the President to think rationally about the decision problem she faces, prior to acting. It allows her to take into consideration not only the probabilities of the two possible outcomes but also her preferences regarding the consequences of those outcomes (i.e., the losses they represent for her).

B. Forensic Decisionalism

The applicability of decision theory to legal problems involving forensic science should be apparent. A fact-finder may be faced with the question of whether Mr. A is the source of a fingerprint⁴⁵ or whether Mr. A signed a document. Such questions may be conceptualized as decisions.

43. It is possible, of course, to relax this assumption and consider more general examples.

44. See Richard D. Friedman, *The Persistence of the Probabilistic Perspective*, 48 SETON HALL L. REV. 1589, 1590 (2018); *infra* Section III.C.

45. We will use the term “fingerprint” in this Article to denote a trace (possibly

In principle, such questions are not very different from the issues encountered at other decision points in the legal process, including advanced stages of legal proceedings concerned with the question of what verdict to render (i.e., conviction or acquittal). In legal scholarship, it has long been recognized that such questions of decision, in particular their logical underpinnings, can be critically analyzed and discussed using formal methods of analysis based on, for example, decision theory.⁴⁶ In this Section, we will briefly and informally outline the tenets of decision theory and elements of decision analysis as applied to forensic science.

For example, when a fact-finder *decides* that the questioned handwriting is that of Mr. A,⁴⁷ but in reality, the handwriting is from an unknown person (i.e., the proposition that Mr. A is the writer⁴⁸ is false), then the consequence of the fact-finder's decision will be a false association of Mr. A with the questioned handwriting (i.e., an erroneous outcome). Tables 4(i) and 4(ii) further illustrate these notions using two examples: an analogy between a hypothetical medical diagnosis problem and the "problem" of inferring the source of a questioned fingermark (i.e., forensic individualization).⁴⁹ Note that the logic of Table 4 readily generalizes to any forensic inference of source problem, for example in the controversial tool-/impression- and bite-mark examination disciplines, but also in other areas, such as digital evidence.⁵⁰ Classification tasks, too, can be seen as an instance of application of Table 4. An example for classification, i.e.,

incomplete and of limited quality) left under unknown conditions, as compared to a "fingerprint" taken from a known person under controlled (laboratory) conditions. *E.g.*, CHRISTOPHE CHAMPOD ET AL., FINGERPRINTS AND OTHER RIDGE SKIN IMPRESSIONS 317 (2d ed. 2016).

46. The first detailed decision theoretic account for legal applications is widely attributed to John Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065 (1968).

47. We do not intend to suggest that forensic examiners should express themselves in this way. We merely use this example because forensic document examiners commonly do express direct opinions about particular propositions regarding the source of handwritten items, though as we will point out in later parts of this Article, such statements require assessments and assumptions that go beyond the scope of the forensic examiners' area of competence. Judges and jurors with knowledge of the entire case file are more suitably positioned to give such statements.

48. In later parts of this Article, the proposition that the handwriting is from Mr. A will be denoted H_p for short, and the proposition that an unknown person is the writer, H_u .

49. Note that for simplicity, only two decisions are considered here, "identifying" Mr. A as the source of the fingermark and "not identifying" Mr. A. For a more general decision theoretic development, allowing for more than two decisions, see Biedermann et al., *Decision Theoretic Properties*, *supra* note 6.

50. For example, propositions of interest (i.e., states of nature) in digital evidence could be "this digital video was recorded using Mr. A's mobile phone" versus "this digital video was recorded with an unknown digital device."

“identifying” an object as belonging to a particular class,⁵¹ is the determination of the nature of examined material (e.g., a scientist may “identify” tiny transparent fragments as glass, or a yellowish-white powder as cocaine, etc.).

Table 4. *Illustration of the Notions of Decision, States of Nature, and Decision Consequences (Which May Be Correct or Erroneous)*

<i>Decisions</i>	<i>States of Nature</i>		<i>Decisions</i>	<i>States of Nature</i>	
	Patient is infected	Patient is not infected		Fingermark comes from Mr. A	Fingermark comes from unknown person
Diagnose infection	correct	error	Conclude that fingermark comes from Mr. A	correct	error
Do not diagnose infection	error	correct	Do not conclude that fingermark comes from Mr. A	error	correct

(i) (ii)

Table 4 may be familiar to some readers as a “confusion matrix” used in the approach to identification known as “signal detection” analysis.⁵² Numerous commentators, most recently the PCAST⁵³ report, have noted the importance of data regarding the

51. Kirk, *supra* note 5, at 236.

52. See generally Victoria L. Phillips et al., *The Application of Signal Detection Theory to Decision-Making in Forensic Science*, 46 J. FORENSIC SCI. 294 (2001); EXPERT WORKING GRP. ON HUMAN FACTORS IN LATENT PRINT ANALYSIS, *supra* note 12, at 26–27 (explaining signal detection theory).

53. The President’s Council of Advisors on Science and Technology (PCAST) is “an advisory group of the Nation’s leading scientists and engineers” who directly advise the President of the United States and the Executive Office of the President. PCAST makes policy recommendations in the many areas where understanding of science, technology, and innovation is key to strengthening our economy and forming policy that works for the American people. The PCAST report was published in September 2016 at the request of President Obama. The report reviewed several fields of forensic science for the purpose of strengthening the various fields and clarifying the requirements for what it called “foundational validity” and “validity as applied.” PRESIDENT’S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS, at x–xi (2016) [hereinafter PCAST REPORT], https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf [https://perma.cc/YCY4-3T7Z].

frequency of the four outcomes depicted in Table 4 in controlled studies in which ground truth is known.⁵⁴ They have also noted the general absence of such data for many forensic disciplines. This criticism has become increasingly well-known and understood in legal and forensic circles. What is less well-known and understood—and glossed over by the PCAST report—however, is that data about accuracy is necessary but still not sufficient. More is needed to enable sound decision-making.

At this juncture, the reader might wonder why it is not sufficient, as suggested by PCAST and Professor Kafadar, for example, to obtain data on false positives and negatives, obtained from multiple training cases (with known ground truth), in order to derive general performance measures such as sensitivity and specificity. The point is that while such measures may be informative about a particular method or technique *in general* (e.g., in discourses about admissibility), they do not directly address what decision to make in a given case at hand, or whether a particular decision made in a given case is suitable.⁵⁵ Although it is true that there are standard probabilistic procedures available for using probabilities for false positive and false negative results to arrive at inductive probabilistic conclusions about propositions of interest,⁵⁶ this solves only half of the decision problem. The reason for this is that we do not only decide based solely on what we believe to be true (i.e., the extent to which we think that a particular proposition is true), but also based on our preferences among the possible decision consequences!

Decision theory explicitly acknowledges these preferences by attaching utilities (or losses) to decision consequences. These utilities (losses) express the relative (un-) desirability of the various decision consequences. For example, in the case of a fingerprint examination as depicted in Table 4(ii), a decision theoretic analysis requires one to express a ranking among the different decision outcomes. Following common understanding in

54. *Id.* at 5–6; Gary Edmond et al., *A Guide to Interpreting Forensic Testimony: Scientific Approaches to Fingerprint Evidence*, 13 LAW PROBABILITY & RISK 1, 9–11 (2014); Karen Kafadar, *Statistical Issues in Assessing Forensic Evidence*, 83 INT'L STAT. REV. 111, 113–16 (2015); NAT'L RESEARCH COUNCIL, *supra* note 4, at 118. As part of its call for the scrutiny of foundational validity, PCAST encourages the establishment of more studies “in which many examiners render *decisions* about many independent tests (typically, involving ‘questioned’ samples and one or more ‘known’ samples) and the error rates are determined.” PCAST REPORT, *supra* note 53, at 5–6 (emphasis added).

55. Alex Biedermann et al., *A Formal Approach to Qualifying and Quantifying the ‘Goodness’ of Forensic Identification Decisions*, 17 LAW PROBABILITY & RISK 295, 305 (2018).

56. To pass from a probability (Pr) of a finding (*E*) given a proposition (*H*), $\Pr(E|H)$, to a probability for the proposition *given* the evidence, $\Pr(H|E)$, Bayes’ theorem needs to be invoked.

forensic science, this would mean preferring accurate outcomes over erroneous outcomes. That much is obvious. More subtly, the decision-maker may also have preferences even between different erroneous outcomes. For example, it is common in forensic science to claim that a false exclusion should be regarded as less adverse than a false identification. For example, one possible ranking of the relative desirability of outcomes for fingerprint analysis might be:

1. Correct decision same source (tie)
Correct decision not to identify (tie)
2. Incorrect decision not to identify
3. Incorrect decision same source

The situation in (2) is often considered less adverse than (3) because it would tend to lead to the exculpation of a guilty person. The situation in (3) is often considered worse because it would tend to incriminate an innocent person.⁵⁷ Legal philosophy generally views the latter as worse than the former, invoking arguments such as “Blackstone’s ratio,” even though it is often overlooked that the latter refers to ratio across multiple cases, rather than a relative assessment within a given case.⁵⁸ In other applications, e.g., security screening, the preferences between (2) and (3) are commonly inverted.⁵⁹

C. *Decision Theoretic Comparison of Rival Decisions*

The above features of decision theory are mainly descriptive. They lay out in a clear and transparent way the inevitable and fundamental ingredients of any problem of decision under uncertainty. But there is more to decision theory. Decision theory also provides guidance on how to coherently combine probabilities for each state of nature with utilities (or losses) for possible decision consequences in order to provide a criterion for comparing the relative merit of the rival decisions.

57. Biedermann et al., *The Decisionalization*, *supra* note 6, at 32.

58. See also *supra* note 38 (discussing Blackstone’s preference for “missed convictions”).

59. Because it is considered better, for example, to incorrectly suspect someone of having a prohibited weapon and subject them to further screening, than to miss a prohibited weapon.

To illustrate this, consider again Table 4(ii) for a classic problem of forensic inference of source. The two available courses of action (i.e., decisions) can be denoted as follows:

d_1 : “report that the fingerprint comes from Mr. A”

d_2 : “do not report that the fingerprint comes from Mr. A”

Each of these decisions can lead to exactly one desirable and one undesirable outcome. That is:

- when deciding d_1 , the conclusion that the fingerprint comes from Mr. A is correct if the fingerprint truly comes from Mr. A (i.e., proposition H_p is true); it is erroneous when in fact an unknown person is the source of the fingerprint (i.e., proposition H_d is true).⁶⁰
- when deciding d_2 , not concluding that the fingerprint comes from Mr. A is correct if the fingerprint truly comes from an unknown person (i.e., proposition H_d is true); it is erroneous when in fact Mr. A is the source of the fingerprint (i.e., proposition H_p is true).

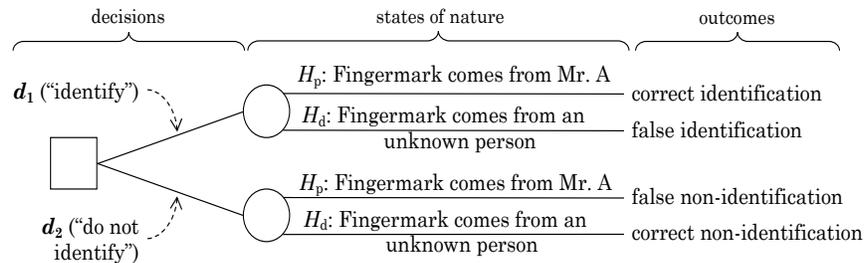


Figure 1. Graphical Summary⁶¹ of the Notions of Decisions, States of Nature, and Decision Consequences (Outcomes) for a Two-Decision/Two-State-of-Nature Problem, Such As Forensic Identification, Sketched in Table 4(ii).

Because all evidence is inherently probabilistic, the evidence (in the case of the fingerprint images themselves) alone does not give us enough information to decide to behave as if that Mr. A

60. For an explanation of the nomenclature using H_p and H_d , see *supra* note 48.

61. This kind of representation is also known as decision tree. HOWARD RAIFFA, DECISION ANALYSIS: INTRODUCTORY LECTURES ON CHOICES UNDER UNCERTAINTY 10 (Frederick Mosteller ed., 1968). The particular example shown here in Figure 1 is adapted with slight modification from Alex Biedermann & Joëlle Vuille, *Understanding the Logic of Forensic Identification Decisions (Without Numbers)*, 83 SUI GENERIS 397, 402 (2018).

(H_p) rather than an unknown person (H_d) is the source of the fingerprint. Virtually all real-world evidence is inherently incomplete, hence incapable to support one hypothesis over the other with total certainty.

However, as we discussed *supra*, there is an additional consideration that is often overlooked. The potentially adverse consequence associated with d_1 , an incorrect decision of same source, is considered by many worse than the potentially adverse consequence associated with d_2 , an incorrect decision not to identify Mr. A. Because the decision-maker does not know the ground truth, the decision-maker does not know which of the two possible consequences, depicted in Figure 1, will follow each of the two possible decisions. So, what can decision-makers do when they cannot be sure about the actual consequence incurred following decisions d_1 and d_2 ?

The first point to note is that not knowing which consequence will be incurred means that one also does not know—at the time of making the decision—the *actual reward* (or loss) incurred. What decision-makers *can* instead consider—and decision theory formalizes this explicitly—is the *expected* utility (loss). The expected utility EU (loss, EL) is obtained by weighing the utility U (loss, L) of each decision consequence by the probability Pr of its occurrence and summing these probability-weighted utilities (losses). Slightly more formally, we can write this as follows:

$$EU(d_1) = U(\text{correct identification}) \times Pr(\text{Fingerprint comes from Mr. A}) + U(\text{false identification}) \times Pr(\text{Fingerprint comes from an unknown person}), \text{ and}$$

$$EU(d_2) = U(\text{missed identification}) \times Pr(\text{Fingerprint comes from Mr. A}) + U(\text{correct nonidentification}) \times Pr(\text{Fingerprint comes from an unknown person}).$$

The expected utility (loss) thus characterizes the rival decisions and provides a criterion for their comparison. Further, the so-called maximum (minimum) expected utility (loss) principle says that the optimal decision is the one which has the maximum (minimum) expected utility (loss).⁶²

The full numerical development of the above notions is not necessary for the general arguments we seek to advance and goes beyond the scope of this Article.⁶³ The essential points of the

62. Despite the use of the same words, this should not be confused with the well-known minimax criterion. See *infra* note 65.

63. As emphasized by Howard, “The overall aim of decision analysis is insight, not numbers.” Ronald A. Howard, *An Assessment of Decision Analysis*, 28 OPERATIONS RES. 4,

decision theoretic analysis of the “problem” of identification are the following:⁶⁴

- Decision-makers are only in control of the decisions to be made (i.e., the branches on the left-hand side in Figure 1); they are *not* in control over decision outcomes (i.e., the end-points on the right-hand side in Figure 1).
- A decision is made not only on the basis of what one thinks is (most) probably true, but also based on one’s preferences among decision outcomes.
- There is a difference between a good outcome and a good decision: since one cannot decide such that the outcome will be optimal (i.e., the best outcome for a given decision) for sure, one can at best select the decision that offers the best *prospect* with regards to the relative desirability of the various decision outcomes (i.e., the decision with the maximum/minimum expected utility/loss).⁶⁵

The operational precept that derives from the above principles can, without going into the details of the full numerical development, be summarized as follows:

“Suppose that Option One has far worse consequences if wrong than does Option Two. Then a sensible decision-maker will choose Option One rather than Option Two only if she has a high degree of confidence that Option One rather than Option Two is correct, or, put another way, only if she thinks Option One is far more probable than Option Two.”⁶⁶

As an illustration, consider again the case of a fingerprint found to show similarities and differences with respect to a fingerprint of Mr. A. Even though (i) one may have strong evidence in support of the proposition that Mr. A is the source of the fingerprint (H_p), rather than an unknown person (H_d); and (ii) after consideration of all the evidence, the decision-maker’s probability for H_p is much higher than that for the proposition H_d , this does *not* imply or suggest that the optimal decision is to identify Mr. A as the source of the fingerprint (i.e., decide d_1). Indeed, if the loss associated with an erroneous identification of

9 (1980).

64. *E.g.*, Biedermann et al., *supra* note 40.

65. Note that other, nonprobabilistic decision criteria exist, such as the minimax decision rule, though they may render forensic identification unworkable. Biedermann et al., *supra* note 55, at 300–03.

66. Friedman, *supra* note 44, at 1590.

Mr. A as the source of the fingerprint is considerably greater than the loss associated with a missed identification (i.e., deciding d_2 when in fact H_p is true), identifying Mr. A as the source of the fingerprint (decision d_1) may not be the optimal decision.⁶⁷

Put another way, using the terminology discussed in Part I, the decision-maker's *opinion* may be that Mr. A is the source of the fingerprint. But even a decision-maker holding such an *opinion* might *decide* to behave as if Mr. A is not the source of the mark, solely because the potential losses for an incorrect same-source decision are so great.

These are examples of the possibility we mentioned *supra*,⁶⁸ in which it may be rational for a decision-maker to decide to behave as if a proposition is true even when the decision-maker considers the proposition less probable than its alternative.

These qualitative statements of decision theoretic advice for deciding between competing courses of action should sound uncontroversial to scientists and legal specialists. Indeed, even without referring explicitly to decision theory, it is commonly upheld as a precept that relative losses associated with adverse decision consequences should be “weighted” against one's strengths of beliefs about what the actual state of the world is (or will turn out to be).⁶⁹ Decision theory provides a formal justification for this intuition.

IV. CRITICAL ANALYSIS AND DISCUSSION OF ULTRS

A. Preliminaries

While the draft ULTRs were quite heterogeneous, the Approved ULTRs are remarkably similar to one another. It is reasonable to suppose that the first published ULTR, for latent prints, provided the template from which the others were adapted.⁷⁰ Below is a typical passage from an ULTR with the relationship between the words “decision” and “conclusion”

67. For numerical examples, see, for example, Biedermann et al., *The Decisionalization*, *supra* note 6, at 35–36.

68. See *supra* note 30 and accompanying text.

69. See STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 481 (Marcia J. Horton et al. eds., 3d ed. 2010) (“*The right thing to do—the rational decision—therefore depends on both the relative importance of various goals and the likelihood that, and degree to which, they will be achieved.*”).

70. See Simon A. Cole, *Forensics, Justice, and the Case for Science-Based Decision Making*, UNION CONCERNED SCIENTISTS (Nov. 14, 2018, 10:01 AM), <https://blog.ucsusa.org/science-blogger/forensics-justice-and-the-case-for-science-based-decision-making> [<https://perma.cc/YCY4-3T7Z>].

highlighted. For illustrative purposes, we have chosen one of the two ULTRs for the forensic firearm and toolmark discipline.

‘Source identification’ is an examiner’s **conclusion** that two toolmarks originated from the same source. This **conclusion is** an examiner’s **decision** that all observed class characteristics are in agreement and the quality and quantity of corresponding individual characteristics is such that the examiner would not expect to find that same combination of individual characteristics repeated in another source and has found insufficient disagreement of individual characteristics to **conclude** they originated from different sources.

The **basis for** a ‘source identification’ **conclusion** is an examiner’s **decision** that the observed class characteristics and corresponding individual characteristics provide extremely strong support for the proposition that the two toolmarks came from the same source and extremely weak support for the proposition that the two toolmarks came from different sources.

A ‘source identification’ is the statement of an examiner’s **opinion** (an inductive **inference**) that the probability that the two toolmarks were made by different sources is so small that it is negligible. A ‘source identification’ is not based upon a statistically-derived or verified measurement or an actual comparison to all firearms or toolmarks in the world.⁷¹

Nine of the ten ULTRs that use the term “decision” posit the same structural relationship between the “decision” and the “conclusion” using almost the exact same words.⁷² This common structure consists of three paragraphs. It is not clear to us whether the paragraphs are supposed to represent a progression of arguments or to offer three different ways of saying the same thing, but we tend to think the latter is more plausible.

Elsewhere, one Author has already criticized this language as it appeared in the latent print ULTR.⁷³ The primary criticism was that the ULTR remained a “categorical” (or nonprobabilistic) statement: “two toolmarks originated from the same source.”⁷⁴ However, this nonprobabilistic statement was followed by a

71. U.S. DEP’T OF JUSTICE, UNIFORM LANGUAGE FOR TESTIMONY AND REPORTS FOR THE FORENSIC FIREARMS/TOOLMARKS DISCIPLINE – PATTERN MATCH EXAMINATION 2 (2018) (emphasis added).

72. See *supra* Tables 1 and 2. The exception is chemistry. See U.S. DEP’T OF JUSTICE, UNIFORM LANGUAGE FOR TESTIMONY AND REPORTS FOR GENERAL FORENSIC CHEMISTRY AND SEIZED DRUG EXAMINATIONS (2019).

73. Cole, *supra* note 28, at 1113–19.

74. U.S. DEP’T OF JUSTICE, *supra* note 71, at 2.

number of probabilistic statements that purported to support the nonprobabilistic statement.⁷⁵ That doesn't make sense.⁷⁶

Although broad criticisms of the latent print ULTR have already been made, in this Article we go more deeply into the issues with the use of the word "decision" in that ULTR and the eight additional ULTRs that closely follow its wording. We cannot, however, analyze and discuss the notion of decision in a deliberate way without a (scientific) reference point. With respect to the notion of decision, the reference framework on which we shall rely is given by the decision theoretic notions introduced in Sections III.A–III.C.

B. Descriptive Analysis of ULTR Contents

We start by unraveling the contents of the ULTRs regarding source identification. In essence, the first paragraph makes two assertions, referred to as decisions. First, it is said that the conclusion "is" a decision that the observed features are "in agreement."⁷⁷ The second assertion is more intricate and convoluted: it is the decision that the "examiner would not expect to find that same combination of individual characteristics repeated in another source."⁷⁸ What is not clear with this second assertion is whether it means to say:

75. *Id.*

76. Cole, *supra* note 28, at 1113–19. We suspect that the authors of the ULTRs are "probabilistically aware," making an oblique reference to the notion of criminals who are "forensically aware." BOB WOFFINDEN, *THE NICHOLAS CASES: CASUALTIES OF JUSTICE* 179, 181, 185–89 (2016). They are aware of statisticians' criticisms of categorical statements. Therefore, they include some statements alluding to probabilistic notions, but still are not willing to delete the categorical statements. As we shall see, the author(s) of the ULTRs also appear to be "decisionally aware."

77. U.S. DEP'T OF JUSTICE, *supra* note 71, at 2.

78. *Id.* As an aside, the expression "individual characteristics" is problematic for two reasons: first, one may wonder how one can know that characteristics are individual (unless one treats this assertion as an assumption), and second, if characteristics are indeed individual, then the question is why there is discussion about duplication in another source.

- i.* that the probability⁷⁹ of seeing the features on the questioned item given that they have been left by another source is (very) low (i.e., paraphrasing the expression “would not expect to find”), or
- ii.* that, given the observed features on the questioned item, the probability that an unknown source would be found to leave features of the observed kind is (very) low.

The reader might be tempted to think that interpretations *i* and *ii* are virtually the same and that we are splitting hairs here. However, note that strictly speaking, they express different aspects. Assertion *i* is focusing on the probability \Pr of the observations E given that an unknown source left the mark (proposition H_d): i.e., $\Pr(E|H_d)$. Assertion *ii* refers to the probability of an unknown source being able to leave marks of the observed kind. But note that this is not a classic alternative source proposition of the kind “the mark comes from an unknown source.” It is a proposition that incorporates observations made (i.e., a proposition of the kind “an unknown source presents the features of interest”): this is an amalgam of a source level proposition and observations made.⁸⁰ So, while interpretation *i* focuses on the findings given the proposition, interpretation *ii* pertains to a proposition colored by knowledge of the findings (i.e., the focus is on another source with said features). The “would not expect” expression thus is, at least, unclear. In the worst case, it is confusing because the probability of an unknown source having said features is prone to be misinterpreted as the probability of an unknown source having left the observed features on the questioned item.

As an aside, the expression “not expect to find” itself raises a host of questions, in particular how scientists come to think that they do “not expect to find,” an expression that calls for empirical and quantifiable grounds.⁸¹

In the second paragraph it is said that a decision that the likelihood ratio⁸² (LR) is some very high number (i.e., that it is far

79. Using phraseology centered on the term “expectation,” the first ULTR paragraph eloquently avoids the term probability. Here we make the assumption that expectation, in essence, invokes the notion of probability for we do not see what else it could invoke.

80. For a discussion of the problematic nature of such propositions, see Tacha Hicks et al., *The Importance of Distinguishing Information from Evidence/Observations When Formulating Propositions*, 55 *SCI. & JUST.* 520 (2015).

81. See, for example, concerns expressed in the letter from Rush D. Holt, Chief Exec. Officer, Am. Ass’n for the Advancement of Sci., to Rod Rosenstein, Deputy Attorney Gen., U.S. Dep’t of Justice (Mar. 26, 2018).

82. Here we make the assumption that the notion of “[strong] support” refers to a

more probable to see the observed evidence if the marks come from the same source than if they come from different sources) forms the “basis” for the conclusion.⁸³

In the third paragraph, it is said that a source identification “is” an opinion (which is an inference) that the probability of the defense hypothesis is very low. More formally, this is an assertion of the kind $\Pr(H_d | \bullet)$, that is the probability \Pr of the proposition H_d , stating that “the two toolmarks were made by different sources.”⁸⁴ This looks incomplete, however, because one does not usually hold a probability (or state of mind, belief, etc.) in isolation. Instead, a probabilistic statement is conditioned upon information, knowledge, and evidence available at the time a person issues a probabilistic statement. Presumably, thus, in the case here, the probability is conditioned on the observations made (E , short for evidence in our notation) and task-relevant conditioning information I . But, given the unclear meaning of the ULTR on this aspect, we use the generic notation “ \bullet .” The formal notation helps us clarify that paragraph three, with its focus on $\Pr(H_d | \bullet)$, is fundamentally different from the focus of paragraphs one (with its tentative interpretation as $\Pr(E | H_d)$) and two (LR). A further aspect of concern with paragraph three, besides its focus on a proposition H rather than on the finding E , is that it makes an unsubstantiated assessment and deliberate choice: that is, it asserts not only that the probability $\Pr(H_d | \bullet)$ is “small”—an assertion openly devoid of a statistical basis and not “verified” by any other means—but also small enough to be considered “negligible.”⁸⁵ The latter assertion opens a host of interrogations, such as how an examiner may come to the conclusion that a probability is sufficiently small to be considered negligible and justify such a conclusion. Vagueness on this point also raises doubts as to the possibility of ensuring that such assessments will be made consistently across different examiners. Paradoxically, this tends to compromise the ULTR as a whole: its intention to ensure the eponymous “uniformity” in reporting language does not have any counterpart in assuring consistency in the process of figuring out *when exactly* a given uniform reporting expression is

likelihood ratio, which is a reference measure for strength of evidential support. *E.g.*, COLIN G.G. AITKEN & FRANCO TARONI, STATISTICS AND THE EVALUATION OF EVIDENCE FOR FORENSIC SCIENTISTS 7 (2d ed. 2004).

83. U.S. DEPT OF JUSTICE, *supra* note 71, at 2.

84. *Id.*

85. *Id.*

to be used in any given case at hand.⁸⁶ We will elaborate further on this aspect in the next Section.

Lastly, paragraph three does not use the term “decision.” We will not discuss it further at this point, but we will, however, note that it follows the same form as the other two paragraphs: it claims that a nonprobabilistic statement “is” a statement that a probability⁸⁷ is very low.⁸⁸ So, it follows the same pattern of advocating that experts round purportedly low probabilities down to zero probabilities for fact-finder consumption.⁸⁹

The ULTRs, then, display a lack of clarity that makes them difficult to interpret from a statistical point of view. This might not be viewed as a problem if comprehensibility to that particular discipline is not perceived as a problem. But the ULTRs appear to invoke statistical concepts and use statistical terms. If so, statisticians should be able to easily understand at least what the ULTRs are *claiming*. But our argument in the next Section is that no matter how one interprets these semantic ambiguities, the ULTRs suffer from fundamental flaws in reasoning.

C. *ULTRs Read from a Decision-Analytic Viewpoint*

We will argue that the ULTRs’ use of the term “decision” is problematic for two primary reasons.

First, they invert decision theory by treating decisions as inputs into the reasoning process, rather than outputs. The outputs of their proposed reasoning processes are phrased as “conclusions,” rather than “decisions.” This falsely suggests that examiners can know the true state of nature—whether an object is, or is not, the source of a trace.

Second, they have examiners “deciding” probabilities, rather than “assigning” them. Though there is, in theory, a decision theoretic sense in which a probability can be understood as a decision,⁹⁰ it inevitably invokes the notion of preference (i.e., a score) and ULTR provides no guidance as to whether, and if so, how, examiners shall cope with this notion.

86. Defenders of ULTR may argue that consistency in proceedings across examiners is not a stated aim of ULTR.

87. We use the expression “a probability” here because, following our comments above, the ULTR refers to different types of probability (e.g., probability of the evidence as compared to probability of a proposition).

88. U.S. DEP’T OF JUSTICE, *supra* note 71, at 2.

89. Cole, *supra* note 28, at 1116–18.

90. Alex Biedermann & Joëlle Vuille, *The Decisional Nature of Probability and Plausibility Assessments in Juridical Evidence and Proof*, 16 INT’L COMMENT. ON EVIDENCE 1, 25–31 (2018).

1. *Inverting Decision Theory.* Table 5 graphically compares the structural position of “decision” in the ULTRs’ template format and posited reasoning process with classical decision theory. Stages in roman type are assigned to the expert and stages in *italic* type are assigned to the fact-finder. Squared brackets are used to designate stages of the reasoning process that we consider to be stipulated by ULTR, though not actually formally stated.⁹¹

Table 5. *Comparison of the Posited ULTR Reasoning Process (Columns 1 and 2) and Inference of Source in the Classic Decision Theoretic Sense (Column 3).*

ULTR ¶1	ULTR ¶2	Decision Theory
[Express expectation (i.e., assign probability)]	[Assign probabilities	Assign probabilities to findings
→ decide expectation (probability) + decide agreement	→ obtain likelihood ratio]	→ obtain likelihood ratio
→ conclude source	→ decide likelihood ratio	→ <i>consider probability of competing propositions of interest (given all the evidence and information) and possible decision consequences</i>
	→ conclude source	→ <i>consider preferences among decision consequences</i>
		→ decide source [<i>i.e., same vs. different source</i>]

— Steps in squared brackets are inferred from but not explicitly mentioned in ULTRs.

— Roman type steps designate tasks commonly considered to be in the realm of the experts’ competence.

— Steps in italics are tasks that we consider to be in the fact-finders’ area of competence. The term “decide” is bolded whenever used.

Table 5 shows that in the ULTRs, the role of decisions in the reasoning process is completely inverted from their role in decision theory. For decision theory, a decision is the final step—the

91. Note that ULTRs are limited to descriptions of (reporting) conclusions. ULTRs provide no directions as to the (reasoning) processes that lead to particular conclusions.

output—of the decision-analytic process:⁹² one cannot come to an appraisal of the decision options, and choose among them, before thinking through the probabilities and assessing the relative (un)desirability of the various possible decision outcomes.⁹³ The ULTRs, however, assert that a decision forms the “basis” for the conclusion. The output of the ULTRs are conclusions, and decisions are inputs. Contrast this with the view of decision theory where the output is decisions. The whole purpose of decision theory is to elaborate procedures that specify how to arrive at decisions. The crucial point of the decision theoretic view is that one does not “conclude”—if by “conclude,” one means “know”—that something is the source of a given toolmark: one can only “decide” to *behave or proceed as if* something is the source of the toolmark. The ULTRs ignore this essential insight of decision theory and instead claim not only that conclusions about source based solely on the evidence are possible, but that they can be “based on” of all things: decisions.

If we refer back to the simple flooding illustration we outlined above,⁹⁴ the ULTRs would seem to invert the decision-making process as follows:

1. I **decide** that it is twice as likely that it will NOT flood as that it will flood.
2. I **conclude** that it will NOT flood.

Notice that here, in point (2), the ULTR has claimed to arrive at a state of complete knowledge about the ground-truth proposition: it will not flood. Consider also the strangeness of thinking that one can *decide* whether it will flood, as opposed to either (i) deciding whether to *behave as if* it will flood; or (ii) assigning a probability to the event of flooding! In a decision theoretic perspective, one will show far more humility. The decision-maker will acknowledge that she cannot know for certain whether it will flood or not, but, nonetheless, the decision-maker can rationally decide whether to close the university even under

92. See *supra* Part III. Of course, a decision regarding source is rarely the end of the matter. There are a host of further (even concurrent) decision points following up. For example, at the stage of the verdict, a decision regarding the guilt or innocence of the defendant needs to be made, followed by a decision regarding the type of sentence (fine, etc.).

93. See *supra* Part III. It is possible, though, to analyze a decision ex-post and make statements about the decision ingredients (i.e., probabilities and utilities/losses) that a coherent decision-maker ought to have (had), but this is of little practical interest. All practical decision problems are ex-ante decision problems.

94. See *supra* Section III.A.

this condition of uncertainty. She will do so by giving due consideration to her strength of belief in the event of flooding and to the relative desirability of the decision consequences.

Notice also that in the ULTRs' process, the action—opening or closing the university—seems to have vanished. This illustrates an important aspect of decision theory: it is driven by the need for action (keep the university open or close it) rather than the quest for truth itself. The relative probability of the propositions is of interest in decision theory only in the service of choosing an action. Decision theory begins with the behavioral problem—what is the action I shall take? It then tries to gather information about the truth of propositions in order to solve that behavioral problem of how to act in the light of inevitable uncertainty.

The ULTRs, on the other hand, display the traditional mindset that the goal of analysis is to determine which proposition is true. If one truly has determined the truth, then one does not need decision theory: the behavioral choice is obvious.⁹⁵ But the problem is that in real-world circumstances regarding legal disputes one can rarely determine the truth of contested events (i.e., competing propositions). The above flooding example makes this particularly clear. We all understand intuitively that we cannot “decide” that it will not flood and expect nature always to comply. We even know that we cannot “conclude” that it will not flood—even based on rich scientific information—and expect nature always to comply, hence the familiar phenomenon of closing universities on days on which it never does flood. The ULTRs have lost sight of the behavioral problem that stimulated our interest in truth in the first place. The decision we have to make is not whether it will flood—we understand intuitively that is out of our control. The decision we have to make is whether we should close the university. To make this decision rationally, we need to make our best assessment regarding the truth of the proposition of interest (whether or not it will flood). In decision theory, asserting or pinning down the probability of flooding is not the end but rather a means toward making a rational decision.

Carrying this difference of approach over to forensic identification, we might observe the following: the goal of the ULTRs is to determine whether the defendant is the source of the trace. The goal of using decision theory is to decide whether or not to behave as if the defendant is the source of the trace, followed by

95. That is, if we *knew* there would be flooding, we would close the university; if we knew there would not be flooding, we would leave the university open. Stated otherwise, if we *knew* which state of nature would come about, we would know which decision to make in order to obtain the best consequence possible under the respective state of nature.

further action of legal nature. The ULTRs describe a method in which the forensic scientist would spuriously convince herself that a state of certainty has been achieved about the source of the trace. Decision theory, in contrast, seeks to develop a way to act “under uncertainty,” as decision theorists say—that is, despite one’s awareness that one is uncertain—and will always be uncertain—about the true source of the trace. Thus, decision theory would approach the problem as follows: “I must decide whether to consider the defendant being the source of the fingerprint, rather than an unknown person. In order to make this decision, it is useful to try to assess, as best I can, the relative probabilities with which each of the two possible decisions may lead to undesirable outcomes.”

Decision theory describes a process of using *information* to make a *decision* under *uncertainty*. The ULTRs, in contrast, describe a process of making *decisions* to (purportedly) achieve a state of *certainty*, which makes no sense.

2. Deciding Probabilities? In addition, the ULTRs twice suggest that examiners “decide” probabilities. This is peculiar wording. A suitable term that statisticians may use here is “assign,” although other terms may also be used. Examples depend heavily on the context of applications, but may include “assess,” “define,” “ascribe,” “estimate,” or “compute.” However, “decide” fits uneasily with probability. To “assign” a probability is a statement about one’s belief about the truth of a proposition. To decide is to make a choice. It makes sense for me to decide whether to eat a burger made of meat or a vegetable burger. That decision does not really require any cognition about the truth of any proposition. It is merely choice. It also makes sense for me, when presented with a burger to assign a probability to the proposition that the burger contains meat. But it makes little linguistic sense for me to say that I am going to “decide” the truth of the proposition that a burger contains meat merely by looking at it.

In paragraph 1, the ULTR has the examiner deciding whether the class features are in agreement. This raises the question of why the examiner should “decide” this instead of, say, “observing,” “judging” or, if she has the tools to do so, “determining” it. Though it is not impossible, in principle, to consider the assessment of “agreement” as a decision, doing so formally would require—following the theory exposed in Section III.C—value judgments for the various consequences of deciding agreement (i.e., accurate and inaccurate determinations) and, again, probabilities. These

requirements offer as much room for debate as their counterparts do in the context of considering source (attribution) as a decision.⁹⁶

Then the ULTR has the examiner deciding that the probability of the defense hypothesis is low. At first glance, it sounds improper to “decide” a probability. A probability is what it is. Generally, it is not subject to preference, and therefore, it is not “decided.” Nonetheless, there is a formal decision theoretic approach to understanding probability,⁹⁷ though there is no indication in the ULTR that presupposes this formal approach.

In paragraph 2, the ULTR has the examiner deciding that the likelihood ratio⁹⁸ is enormously high or, in the words of ULTR, that the findings “provide extremely strong support.”⁹⁹ Again, the word “decision” looks unsuitable in this context, for at least two reasons.

First, because a likelihood ratio is not to be decided about. A likelihood ratio is obtained as a direct result of two conditional probabilities that have been assigned individually. Once these two component probabilities have been set, the likelihood ratio (i.e., its order of magnitude) follows by definition. There is nothing else to be decided. What is more, the term “deciding” in this context suggests that the value of evidence (i.e., likelihood ratio) assessments are one-off conclusions; they are not. It would be a misconception to think that an examiner could come up, out of the void, with a statement of the kind “I decide that my likelihood ratio is so and so” (or “I decide that the findings provide such and such strength of support”). Instead, the proper application of a logical approach to evaluative reporting requires the scientist to think about the observations given each of the competing propositions in turn. It is only afterwards that, by combining the component assessments, the likelihood ratio (or strength of the evidence) is found. But this latter stage follows merely by necessity (i.e., as the ratio of the two component assessments); it requires no further intervention or a decision of any kind.

Second, a likelihood ratio (or, more generally, an expression of strength of support) *by itself* cannot logically warrant a source identification conclusion. It is possible, however, to consider whether the likelihood ratio at hand exceeds the minimum value necessary to warrant identification decisions *given* particular assumptions about other factors, such as probabilities for

96. See *supra* Section III.B.

97. See *supra* note 90; *infra* Section IV.D.

98. Recall that we interpret here the ULTR expression “provid[ing] . . . support” as referring to the likelihood ratio. See *supra* note 82.

99. U.S. DEP’T OF JUSTICE, *supra* note 71, at 2.

competing propositions based on considerations other than the scientist's evidence and loss ratios for adverse consequences of identification decisions.¹⁰⁰ But, again, the rationale for these considerations derives from the full decision theoretic approach and there are no indications in ULTR that “deciding” the likelihood ratio is understood in this way.

Of course, decision theory does not possess a monopoly over the word “decision.” It is, of course, possible—and following our analysis the most likely explanation—that the ULTRs are using the term “decision” in its colloquial sense. Given the extensive forensic literature on the use of the decision theory, we believe the ULTRs have a responsibility to at least clarify whether they intend to invoke that literature or not. But even using “decision” in its colloquial sense, it is improper for the expert to be “deciding” probabilities without a clear view of what exactly this means, and what “deciding probabilities” means in other, more formal theoretical frameworks (i.e., decision theory).

D. A Coherent Way of Deciding Probabilities

We have argued in the preceding Section that it is peculiar to think of deciding probabilities. Strictly speaking, it is possible to decide probabilities in a coherent way. However, the theory of considering probability as a decision is rather specialized and advanced, which makes it doubtful that ULTRs intend to embrace it.¹⁰¹ Specifically, if the theory of understanding of probability as a decision were indeed endorsed, then examiners should *not* be allowed to make assertions that imply (or suggest) that a small probability can be rounded off to zero, as stipulated for example by ULTR ¶1.¹⁰² Simply put, examiners could not defensibly claim that it is suitable to intentionally *report* a smaller probability than the one they actually have in mind. The reason for this stems from the concept of proper scoring rules, that is a type of score¹⁰³ function that measures the “goodness” of probability assertions with the particular property that it is optimal for a probability assessor to

100. See, e.g., Biedermann et al., *The Decisionalization*, *supra* note 6, at 36 tbl. 4.

101. See, e.g., Biedermann & Vuille, *supra* note 90 *passim* (discussing the relevance of this theory for applications in forensic science and the law).

102. For a full development of this argument, see, for example, Alex Biedermann et al., *The Subjectivist Interpretation of Probability and the Problem of Individualisation in Forensic Science*, 53 SCI. & JUST. 192 (2013).

103. A score can be thought of as a penalty that is smaller (greater) the closer (farther away) an asserted probability lies from the actual truth-value of a proposition. For example, the closer an asserted probability is to one (zero) for a proposition that is actually true (false), the smaller (greater) the score (penalty).

report probabilities that correspond to what the person actually believes. Stated otherwise, under a proper scoring rule the question of which probability to report is treated as a decision, and the theory shows that the optimal¹⁰⁴ decision (i.e., the probability to be reported) is none other than the one that corresponds to one's actual probabilistic belief. Since the middle of the last century, this feature has been largely explored in fields involving expert assessments made under uncertainty, using probabilities.¹⁰⁵ It has also been the object of intense study by statisticians.¹⁰⁶ The scoring rule scheme is of interest from an operational point of view because it encourages experts to state their actual beliefs as probabilities, i.e., encourages honest probability reporting and discourages the reporting of distorted probabilities. In this sense, the theory that views probability assertion as a decision looks relevant to the ULTRs' suggestion that examiners "decide probabilities," though there are hurdles from an applied perspective, because scoring rules are subtle and intricate.¹⁰⁷ The challenging nature of the scoring rule is even admitted by founding writers on the topic, such as Savage: "[T]he subject must understand the scoring rule. . . . [M]ost ordinary people will not understand it at all; and even those with mathematical training may not be nearly apt enough at calculation to use the rule effectively."¹⁰⁸ For these reasons, it appears nonfeasible to require examiners to strictly adhere to understanding probability assignment as a decision. For practical purposes, the idea that a probability relies, inherently, on a personalized assessment—given the best knowledge, information and evidence available at the time an assessment needs to be made—can suitably be expressed by the notion of "probability assignment" (i.e., "assigning" a probability). At the same time, the important point to keep in mind is that a probability is supposed to reflect the assigner's true beliefs about the truth or falsity of the event.

104. In the context here, a decision is optimal if it has the minimum expected score.

105. *E.g.*, Glenn W. Brier, *Verification of Forecasts Expressed in Terms of Probability*, 78 MONTHLY WEATHER REV. 1 (1950).

106. *E.g.*, Bruno de Finetti, *The Proper Approach to Probability*, in EXCHANGEABILITY IN PROBABILITY AND STATISTICS 1, 1–3 (G. Koch & K. Spizzichino eds., 1982); Leonard J. Savage, *Elicitation of Personal Probabilities and Expectations*, 66 J. AM. STAT. ASS'N 783 (1971).

107. Further, the understanding of probability as a decision may be considered debatable in this context of application because it requires value judgments for decision consequences. Here, the consequence of a "decided" probability is its distance to the truth-value of the uncertain proposition, and the value judgment is operationalized through the score. *See supra* text accompanying note 103.

108. Savage, *supra* note 106, at 799.

In short, we argue that the ULTRs misuse the term “decision,” whether it is intended in the technical or the colloquial sense. But that does not mean that the term cannot be useful in thinking about forensic problems. Indeed, it can be. In the next Part, we explain how.

V. APPLYING DECISION THEORY

We argued above that decision theory can be useful in conceptualizing key features of legal problems involving forensic evidence. In this Part, we expand on that argument.

As we argued in Part III, decision theory offers a way of coherently thinking about making a decision when presented with competing propositions, such as: (i) Mr. A is the source of this fingermark; and (ii) an unknown person is the source of this fingermark. Decision theory is typically applied in situations in which the available evidence is insufficient to know for certain which of these propositions is true. But decision theory offers a way to logically decide to behave as if one of them is true. In other words, decision theory offers a way to decide which hypothesis (or proposition) to *endorse*.

Decision theory demonstrates what would be required to make such a rational decision. Not surprisingly, some kind of analysis of the evidence is required. In addition, as statisticians have long pointed out, prior odds are required too.¹⁰⁹ The primary contribution of decision theory then is to articulate that some sort of statement of preferences between different consequences, leading to expected utilities or losses (allowing the decision-maker to qualify and compare rival decisions), is also required.¹¹⁰

For legal actors, decision theory can thus be useful in laying out the logical requirements necessary to *behave as if* a statement such as “Mr. A is the source of this fingermark” is true. The decision to behave as if such a statement is true is of obvious pertinence in legal proceedings. Decision theory can, in theory, articulate the steps toward getting there, though this way is paved with a host of questions that we briefly address below.

In articulating the reasoning process behind such decision statements, decision theory exposes some serious concerns with common legal practice in handling forensic evidence in the United States and abroad. The principal such concern is that decision statements such as “Mr. A is the source of this fingermark” are

109. *E.g.*, G. Parmigiani, *Decision Theory: Bayesian*, in INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL & BEHAVIORAL SCIENCES 3327, 3330–31 (Neil J. Smelser & Paul B. Baltes eds., 2001).

110. Biedermann et al., *supra* note 55, at 302–04.

often—in some cases always—made by experts. This practice has been criticized based on a number of different arguments. Some legal actors may have a vague sense that it is wrong but have difficulty articulating what is wrong about it. We feel that decision theory has the potential to help more clearly articulate what is wrong with this practice.

As explained above, a decision is based on preferences among the possible decision outcomes (consequences). In forensic problems, these consequences tend to lie in the realm of adducing evidence against an innocent person or failing to adduce evidence against a liable person. A moment's reflection should quickly lead to the realization that the relevant preferences ought not to be *the expert's* preferences. The question of how much we prefer failing to adduce evidence against a guilty person versus adducing evidence against an innocent person (or vice versa) is clearly better suited to a legal actor, such as the fact-finder, than to the forensic expert. This is because a forensic scientist does not have expertise in the moral considerations that would underlie these preferences. Nor does the forensic scientist have the moral authority to set these preferences. One can imagine these preferences being made society-wide, and one can imagine them being made in individual cases by juries.¹¹¹ But it is clear that forensic scientists possess no special expertise or moral authority that justifies using their preferences rather than the fact-finder's. As Dr. Stoney has noted, conventional practice in fingerprint analysis has included “assuming priors and including decision-making preferences. This created an overwhelming and unrealistic burden, asking fingerprint examiners, in the name of science, for something that science cannot provide. As a necessary consequence, fingerprint examiners became unscientific.”¹¹²

The logical consequence of this is that experts *ought not make decisions*. And, the logical consequences of *that*, by definition, is that they ought not make categorical statements like “Mr. A. is the source of this fingerprint.”

What, then, are experts supposed to do? They are, first of all, supposed to focus on the immediate results of their practical work, which are—strictly speaking—observations. They are not supposed to focus on statements about propositions regarding the source of marks and traces. They are supposed to evaluate their observations, if introduced as evidence, using their particular

111. But even on these general levels, it is far from clear how those preferences ought to be framed, let alone whether—for practical purposes—they actually can be framed.

112. David A. Stoney, Discussion, *Quantifying the Weight of Evidence from a Forensic Fingerprint Comparison: A New Paradigm*, 175 J. ROYAL STAT. SOC'Y 371, 400 (2012).

expertise, and inform the decision-maker about the probity of that evidence with respect to the different competing propositions. To be clear, the emphasis here is on the value of the findings with respect to propositions, not the reverse. It is the decision-maker who then needs to take that evidence further to render a decision to behave as if a proposition is true, based on *their* appreciation of what is at stake in the case at hand (i.e., their preference structure).¹¹³

We are aware that our argument may sound disempowering to forensic experts. There is ample evidence that forensic scientists believe that failing to report what forensic statisticians call “posterior probabilities” (e.g., Mr. A is the source of the trace) renders them less useful to the justice system. As one example, consider this remark by a forensic hair analyst reported in the root cause analysis review of overstatement of the value of microscopic hair comparison analysis evidence at the FBI: “[y]ou cannot leave the jury hanging or you sound like a meteorologist—it could rain tomorrow.”¹¹⁴ The hair analyst seems to perceive the meteorologist as useless to the decision-maker unless she reports posterior probabilities. But, as we showed in our illustration above (Section III.A), we do recommend that the role of the forensic scientist is more like that of the meteorologist than that of the university President who must decide what action to take.¹¹⁵

We do not think it is necessary for forensic scientists to perceive decision theory as disempowering. On the contrary, decision theory helps articulate reasons for experts to remain in their area of expertise and not be lured outside it by the pressures of the adversarial system. That is, in essence, a call to cut off the

113. Some readers may recognize this argument as similar to the arguments advanced by forensic statisticians based on Bayes’ Theorem as to why forensic experts ought not report about posterior probabilities. *E.g.*, AITKEN & TARONI, *supra* note 82. The two arguments are indeed similar—and consistent. Decision theory simply offers yet another reason why statements like, “Mr. A is the source of this fingerprint,” ought not be made by experts.

114. ABS GRP., ROOT AND CULTURAL CAUSE ANALYSIS OF REPORT AND TESTIMONY ERRORS BY FBI MHCA EXAMINERS 138 (2018), <https://vault.fbi.gov/root-cause-analysis-of-microscopic-hair-comparison-analysis/root-cause-analysis-of-microscopic-hair-comparison-analysis-part-01-of-01/view> [<https://perma.cc/546Z-NDFV>].

115. *Supra* Section III.A. The analogy is not entirely apt. Meteorologists do give probabilities for possible states of nature when they report, for example, “there is a 70% probability of rain tomorrow.” To be entirely in line with the context of reporting in forensic science, the meteorologist would have to give probabilities for the meteorological data conditioned on, for example, “rain” or “not rain” (though these states of nature may be further refined)—thus, something along the lines of “I consider it ten times more probable to observe the meteorological data if it will rain than if it will not rain” (admittedly not a statement well suited to mass media communications).

rotten branches of a tree to help it concentrate efforts to the growing of the healthy parts. All that we are arguing is that forensic scientists should not offer statements that logically necessitate conceptual assessments that they cannot defensibly make—or, worse, they may not even be aware that they are making.¹¹⁶ In combination, one could just as easily perceive these reflections as empowering.

We are also aware that our argument might make it hard for forensic scientists to perceive decision theory as “useful” to them. Indeed, it is true that decision theory is not a statistical tool that will directly help forensic scientists communicate the results of their analyses. Decision theory is “useful” to forensic scientists primarily in educating them about what *not* to do. Nonetheless, for forensic scientists puzzled or frustrated by arguments that they should not make categorical identification statements, we hope that decision theory can offer another way of understanding the reasoning behind such arguments.

For legal actors, however, we think the usefulness of decision theory is clearer. It should allow judges and attorneys to more clearly see the appropriate roles for experts and fact-finders and how they can work together to produce a rational decision.

A. *Practitioner Uses of “Decision”*

In some forensic circles, it has become fashionable to use the term “decision” to describe the output of their analyses. For example, an increasing number of friction ridge examiners are now saying, “I made an identification decision.” Likewise, some researchers refer to forensic “decisions.”¹¹⁷ For these researchers, “decision” does not invoke decision theory; “decisions” are simply “responses” whose accuracy can be scored. This trend emerged independent of the ULTRs, although there is undoubtedly some influence in both directions.

We have mixed feelings about this trend. On the one hand, the replacement of terms like “determination” or “conclusion” with “decision” can reasonably be interpreted as an effort toward

116. The underlying idea here is what in other contexts de Finetti has expressed with the sentence: “Nothing is lost what was a mere illusion.” Bruno de Finetti, *Bayesianism: Its Unifying Role for Both the Foundations and Applications of Statistics*, 42 INT’L STAT. REV. 117, 121 (1974).

117. See, e.g., Amanda Luby, *Decision-Making in Forensic Identification Tasks*, in OPEN FORENSIC SCIENCE IN R, ch. 8 (Sam Tyner ed., 2019) (ebook); Bradford T. Ulery et al., *Accuracy and Reliability of Forensic Latent Fingerprint Decisions*, 108 PROC. NAT’L ACAD. SCI. 7733, 7733 (2011).

greater epistemological humility.¹¹⁸ “Identification decision” might convey to the fact-finder that the expert does not *know* the truth with certainty. The word “decision” seems to connote that some manner of uncertainty is involved. Thus, the use of the term “decision” may be seen as one of many arguments that forensic statisticians use to try to persuade forensic scientists not to report posterior probabilities. Likewise, when used by researchers, the use of the term “decision” may indicate an insistence on the uncertainty of forensic results.

However, the word “decision” does not connote uncertainty so much as choice. Introducing the notion of choice into the reports of forensic experts ought to unsettle legal actors. Why, after all, is the expert “deciding” the meaning of the evidence, rather than simply reporting “the findings”? As discussed above, the word “decision” sounds strange in this context. For all the reasons discussed, it seems inappropriate to include the expert’s choices and/or preferences in a report about the evidence. Likewise, researchers’ use of the term “decision” could be misinterpreted as conceding that it is appropriate for forensic scientists to make “decisions”—which, as we have discussed, would entail making assumptions about preferences.

For other forensic scientists, the use of the term “decision” might be an effort to signal awareness of decision theory and some of the arguments outlined in this Article. For an expert witness in such a posture, the statement “I rendered an identification decision” would implicitly contain the following qualifications:

- “I am fully aware that in making ‘an identification decision,’ I am making an assumption about the prior odds that the defendant is the source of the mark.”
- “I am fully aware that in making ‘an identification decision,’ I am either applying my own preferences about the consequences of the various possible decisions or I am making an assumption about what someone else’s preferences would be.”

If assumed to contain these qualifications, the statement—“I rendered an identification decision”—might, in principle, be considered logically acceptable. However, it might also be considered lacking in transparency: it would be better to make the qualifications explicit, rather than assume that legal audiences understand these very subtle qualifications to be implied merely

118. See generally Mnookin, *supra* note 3, at 139 (discussing epistemological humility and the use of more modest claims).

by use of the word “decision,” let alone agree that these qualifications are being made by the scientist.

B. What Should Forensic Analysts Report?

While we have been critical of many uses of the term “decision,” as a general matter we recognize that the growing use of the term may well mark an important step in the evolution of forensic reports from statements of certainty to statements of uncertainty, from “determination” to something more defensible. But what is that more defensible “something”? Is “decision” enough?

Based on this analysis, and especially on Table 1, we tend to think that “findings” is the most appropriate of all the reporting terms floating around. “Findings” does the best job of conveying—to the expert and customer alike—that the report concerns the evidence alone. Not the evidence combined with other evidence. And, not the evidence combined with preferences. “Findings” helps more clearly distinguish between the analysis of the evidence and the inference to be drawn from that analysis. And, “findings” is commonly used in other fields of science to describe the analysis of (empirical) evidence.

In the long run, we hope to see the term “decision” have a long life in the law where it can be used to properly reconceptualize decision statements as decisions to behave as if a proposition is true, rather than claims that a proposition actually *is* true. Leaving identification *decision* authority to scientists would mean to let them continue to impose, implicitly, their unsubstantiated value assessments on a legal system that operates in deferential mode. By taking ownership of the term “decision,” the law could seize the opportunity to take control over a process that by its nature, i.e., a decision, lies in its area of competence. Genuinely understanding forensic identification as a decision, and controlling it, would also offer the opportunity to all parties to whom the evidence is of concern—most importantly defendants, who are the primary impacted subjects of identification conclusions—to have their say.

And, finally, we hope for the term “decision” to have a useful, but brief life in forensic science, a stepping stone on the journey rather than the end of the journey, as forensic analysts increasingly turn their attention to where their expertise lies—in the analysis of evidence.