

The Unbearable Lightness of Ignoring Axiomatic Principles – A Response to: “On Coping in a Non-Binary World: Rejoinder to Biedermann and Kotsoglou” (by Nicholas Scurich and Richard S. John, in: *Statistics and Public Policy*, 2024)

Alex Biedermann,^{1*} Kyriakos N. Kotsoglou²

¹*University of Lausanne, Faculty of Law, Criminal Justice and Public Administration
School of Criminal Justice, 1015 Lausanne-Dorigny (Switzerland)*

²*University of Northumbria, School of Law, Newcastle Upon Tyne, NE1 8ST (United Kingdom)*

February 14, 2024

Abstract

This paper supplements our *Commentary on “Three-Way ROCs for Forensic Decision Making” by Nicholas Scurich and Richard S. John (in: Statistics and Public Policy)* [6]. We reply to allegations made by Nicholas Scurich and Richard S. John in their rejoinder *On Coping in a Non-Binary World: Rejoinder to Biedermann and Kotsoglou* [19]. These allegations do not carry sufficient intellectual weight to warrant discussion in a published response. However, they are serious insofar as they misquote and misinterpret our *Commentary*, contain demonstrably false assertions, and thus amount to an attempt to mislead the scientific and legal communities. In this paper, we provide the context and background at the origin of this debate and examine each of Scurich and John’s allegations.

1 Background

In a paper published in *Forensic Science International: Synergy* [7], Dror and Scurich (DS) argued that the specific category of conclusions reached by experts, called “inconclusive”,¹ should be treated as an error. As a device to score “inconclusives” as errors, DS proposed a modified error rate study design in which “inconclusive” is not only used as a response category, but also as a ground truth state. In our paper *Forensic science and the principle of excluded middle: “Inconclusive” decisions and the structure of error rate studies* [5], we have rejected the account of DS [7] at all levels of analysis, conceptual, epistemological, and legal (procedural).

DS, as was their right, did not respond to our refutation of their treatment of “inconclusives”. Instead, Nicholas Scurich, writing with Richard S. John, continued to expound on the

*Corresponding author. Email: alex.biedermann@unil.ch

¹“Inconclusive” is a response category that some forensic experts use when they are unable to conclude whether or not a pair of items examined come from the same source, e.g., two fired bullets.

modified error rate study design of DS [7]. Scurich and John (SJ) moved their discussion to another journal, *Statistics and Public Policy*, which had previously been less involved in publishing on the subject in question. In their paper [18], SJ ignored virtually all of the fundamental criticisms that have been made of DS’s account [7] by us and others [e.g., 1]. In our letter to the editor of *Statistics and Public Policy* [6], we therefore highlighted and reiterated “our concern that the analyses presented by SJ [...] promote illogical and retrograde thinking in forensic science” [6, p. 1]. In particular, we pointed out that SJ [18] “use convoluted terminology, make assumptions that are known to be incorrect, and rely on a reporting format that is ubiquitous but that has long since been exposed as unscientific” [6, p. 1].

We can briefly summarise these points as follows. Regarding terminology, SJ [18] misuse the term (i) “match(ing)” (a descriptor for findings and observations) to denote propositions about the source of a pair of compared items (i.e., same or different source propositions), and (ii) “decision” as a descriptor for a forensic examiner’s conclusion. Regarding assumptions, SJ [18] treat “inconclusive” not only as a category of findings, but also as a ground truth state. Moreover, with regards to reporting formats, SJ [18] treat examiners’ opinions about ground truth, including source attribution determinations, as acceptable, for otherwise they would not commit to defending them as an object of study.

In their rejoinder entitled *On Coping in a Non-Binary World: Rejoinder to Biedermann and Kotsoglou* [19] (hereinafter “Rejoinder”), SJ address none of the points mentioned above. Instead, they resort to a series of complaints about our Commentary and our previous work. We examine and refute these complaints point by point in the next section.

2 Point-by-point examination of Scurich and John’s rejoinder [19]

1. The Rejoinder [19] begins with a quotation from an affidavit written by one of us (AB) and other co-authors, submitted in a real-world case. Introducing this quotation is misplaced for the following reasons:
 - Although the affidavit circulates in academic and legal circles, to the best of our knowledge, it is not publicly available anywhere on the internet (at the time of writing this response). It is, therefore, a poorly chosen source of information, not least because the reader has no way of realising that the quote that SJ present is taken out of context.
 - The quotation from the affidavit asserts that “inconclusives” can be statistically summarised and thereby serve as a source of information. It is a statement about facts in the world – as such, it is unsurprising and undisputed. We have never denied the reality that many, if not most, forensic pattern evidence examiners use “inconclusive” as a category of findings, and that this reality can be summarised using simple descriptive statistics. SJ [19] could have found ample statements of the same kind in our paper [5] (i.e., our critique of DS [7]). It was, therefore, unnecessary and pointless to refer to information given in the said affidavit.
 - If SJ introduced the said quotation from the affidavit in order to suggest a striking contradiction with the position defended in our Commentary [6], and sotto voce agreement with SJ, they fail for two reasons. *First*, as noted above, we have never denied the reality of “inconclusive” statements (in the sense of a conclusion reached by a forensic examiner). *Second*, the affidavit of interest clearly states that its authors take a balanced position, which means that they express both points of agreement *and* disagreement in the current debate about the forensic

field of feature comparison. It is not surprising, therefore, that the affidavit contains some statements that overlap *in part* with SJ’s position.

SJ’s opening paragraph thus represents an attempt of the most primitive kind to portray us as inconsistent and self-contradictory commentators.

2. In the third paragraph of the Rejoinder [19], SJ write: “In their letter to the editor, however, Biedermann and Kotsoglou (2023) advocate for an entirely different position regarding inconclusives” [19], and then go to quote from our Commentary [6] in which we call for the abandonment of “inconclusive” as a reporting format.

If SJ thereby suggest that we are contradicting ourselves, then they are confusing two ontologically and methodologically fundamental categories, the *Is* and the *Ought* [13]. As noted in point 1 above, we do not deny the *reality* that forensic examiners use the conclusion category “inconclusive”.² However, acknowledging the reality of a widely used current format does not in any way mean to suggest that this is how forensic examiners *should* report [2]. Indeed, it is a central point that we cannot derive an *Ought* from an *Is*. These are categorically different modes of thinking [11]. Arguing to the contrary would constitute a serious error of thinking known for centuries as the naturalistic fallacy. There is no contradiction between describing reality on the one hand and thinking about how *different* the world should or could be on the other hand.

3. In their fourth paragraph, SJ state that their “article provided a framework to evaluate how firearm examiners *do* make decisions rather than how Biedermann and Kotsoglou believe they *ought* to” [19, emphasis as in original]. SJ also note that “Biedermann and Kotsoglou’s opinion about how examiners ought to conduct their examinations does not address the substance of our article” [19].

Here SJ provide a telling example of descriptivism [2] and the ignorance that goes with it. In this context, it is worth remembering that Lindley succinctly put the problem of descriptivism as follows: “To spend too much time on description is unwise when a normative approach exists, for it is like asking people’s opinion of 2+2, obtaining an average of 4.31 and announcing this to be the sum. It would be better to teach them arithmetic” [15, p. ix].

We can extend this view to our context here as follows. Think, hypothetically, of traditional forensic feature comparison examiners (who report using the “identification – inconclusive – exclusion” framework) as arsonists who set fire to the courts. Then, SJ’s position amounts to descriptively summarising the arsonists’ work, rather than telling them they should not set fire to the system!

Put another way, SJ’s position amounts to saying something along the lines of ‘we just do some nice modelling of examiners’ responses; we don’t care about the rest’. This is vaguely reminiscent of the curve-fitting scientists described by Feyerabend : “They have a few results and they try to find the best curve between them. If you ask them if they think they are dealing with reality, some will say <<leave me alone, I’ve got nothing to do with that! I’ve got some nice numbers here and I am trying to connect them>>” [8, p. 57]. SJ might object that they *are* dealing with the reality of “inconclusive” statements, which is true, but misses the point entirely. By treating “inconclusive” as a ground truth state, SJ do *not* operate on the level of empirical reality. We will discuss this aspect further in point 4 below.

Of course, SJ are free to confine themselves to pure descriptivism. However, asserting that we refrain from engaging with the substance of their work is incorrect. As noted

²See also point 4 below for further explanation of this aspect.

in Section 1, we have formulated specific criticism of SJ’s work, both terminological (i.e., about the terms “match(ing)” and “decision (making)”) and conceptual (i.e., SJ’s modified error rate study design in which “inconclusive” is treated as a ground truth state). Our Commentary [6] did not (re-)explain these critiques in full detail, as this would have been redundant; all the points have been made repeatedly and at length in the existing literature to which our Commentary refers. Again, SJ may avoid the effort to track down these first-hand sources, but it is inconsistent to then claim that we do not engage with their work. The fact remains that SJ [19] do not respond to the fundamental problems we raised in [6] and earlier in [5].

4. In paragraph five of the Rejoinder [19], SJ write: “Most of the letter to the editor is devoted to repeating (“reiterating”) Biedermann and Kotsoglou’s (2021) contention that there is no category of inconclusive evidence” [19]. **This is, we cannot stress enough, a false statement.** We have never denied the existence of “inconclusive” as a *reporting category* (i.e., a statement issued by a forensic examiner), neither in [5] nor any of our previous publications on the decision-theoretic treatment of forensic examiners’ conclusions [e.g., 3, 4, 20]. This would be equivalent to a street preacher denying the reality of knife crime. Here is an excerpt from our paper [5] that unambiguously ascertains, contrary to SJ’s complaint, the reality of “inconclusive” as a reporting category, to be distinguished from binary ground truth states:

“In a forensic identification setting, there are (...) two mutually exclusive and exhaustive propositions: either the examined trace (e.g., a fingerprint) comes from the person of interest, or the trace comes from an unknown person. These two possibilities are commonly referred to as same- and different source propositions (or, specific versus unknown source propositions). Thus far, the problem is in line with the medical test example (...). The analogy also holds when it comes to the result (test outcome), i.e. examiners’ conclusions (i.e., responses): forensic scientists do not necessarily state an identification or an exclusion. *Many forensic scientists operate on a tripartite framework including the response “inconclusive” (...).*” [5, p. 2, emphasis added]

Thus, either SJ still have not grasped the difference between categories of *expert utterances* and *ground truth states*, they refer to our previous work in ignorance of its contents, or they make a deliberately false allegation.

5. In paragraph six, SJ write: “Rather than engage our article and explain why these examples are misguided or could be addressed by different statistical techniques, Biedermann and Kotsoglou (2023) merely state:

In forensic science comparison contexts, there are simply no ground truth states other than ‘same source’ and ‘different sources’. To claim or suggest otherwise would violate the principle of the excluded middle (citing themselves).” [19]

This is a further false and misleading statement made by SJ because in their quotation from our Commentary [6], they omit the part where we directly address their flawed analogy to medical applications. For the sake of completeness, we reproduce here our point in full length, highlighting in *italics* the part omitted by SJ:

“*There is a further instance where SJ (2023) carry a descriptor for findings over to a ground truth state, but in a way that raises fundamental conceptual*

issues that undermine the rationality of the analyses proposed by the authors. In fact, one of the key assumptions made by SJ (2023) is that the category of results termed “inconclusive” could serve as a ground truth state, thus enabling a three-way ROC. As an analogy, SJ (2023) mention medical applications, such as when “deciding whether a film displays malignant lesions, benign lesions, or no lesions.” Clearly, the latter categories of medical conditions denote real(istic) ground truth states, but this cannot serve as an analogy in the way suggested by SJ (2023). Why not? In forensic science comparison contexts, there are simply no ground truth states other than “same source” and “different sources.” To claim or suggest otherwise would violate the principle of the excluded middle (Biedermann and Kotsoglou 2021).” [6, p. 1]

As noted above, our Commentary [6] did not reiterate our arguments from [5] in full detail, because this would go beyond the scope of a concise letter to the editor. The above explanation should be self-explanatory for readers even mildly aware of the principles of (medical) diagnostic testing and forensic inference of source problems (i.e., identification / individualization). Even if that should not be the case, the quote offers relevant guidance to further literature, where we make our point in full detail.

More generally, especially given [5], we reject SJ’s complaint that we do not explain “why [their] examples are misguided or could be addressed by different statistical techniques” [19], based on the following two grounds:

- The central problem of DS’s [7] modified error rate study design, adopted by SJ [18], is to treat the possible outcome “inconclusive” as a ground truth state. To reiterate the point: we are *not* denying, contrary to SJ’s allegations, that forensic examiners use *non-binary reporting categories*; what we *are* contesting, instead, is that *ground truth states* are non-binary (in the context of the application discussed here). In [5], we have explained in simple terms why SJ’s carrying over of conclusion categories to ground truth states is misguided:

“ «inconclusiveness» (...) does not represent a ground truth state, in the same way that being unsure about how to get to London does not mean that London’s existence is uncertain.” [5, p. 6]

Put differently, a state of human ignorance (i.e., uncertainty due to incomplete knowledge) does not and cannot change anything about the state of the real world. This should suffice to explain why DS’s [7] assumptions, reused by SJ [18], are misconceived.

In any case, the fact that we have to argue with trained quantitative psychologists (SJ) with, for one, a training in statistics, that ground truth in forensic inference of source problems is binary, is, by itself, a major concern. Normative claims presuppose a firm grasp of reality. Empirical claims, at the same time, cannot serve as their own justification.

- Regarding statistical techniques, our paper [5] acknowledges the use of summary statistics for examiners’ responses in the various conclusion categories, including “inconclusive”. Our paper [5] also refers to similar applications in the medical statistics literature. It is, therefore, plainly incorrect to assert that we do not acknowledge (other) statistical means of dealing with data regarding examiners’ conclusions.

6. In the seventh paragraph, SJ lament that our “letter to the editor does not define «the principle of excluded middle»” [19] and that “the unelaborated citation to Aristotle is merely a rhetorical device with no substantive contribution” [19].

These statements are deeply regrettable and symptomatic of a widespread, lackadaisical indeed anti-intellectual and counter-productive tendency and attitude observable in modern higher education. SJ’s statement implies that some academics can no longer be assumed to be willing or be able to familiarise themselves with, let alone appreciate, foundational pillars of scientific knowledge. It is in a way true that we did not define the principle of excluded middle. Doing that would be reminiscent of a scientist who amidst an experiment, starts questioning the existence of his or her apparatus or defining the meaning of numbers. This would engineer a change in the direction of inquiry.

Notwithstanding, we did say (SJ conveniently left that part out) that Aristotle’s principle has served for the past two and half millennia as the “metaphysical substrate” for empirical research. How could we, or anyone, define whether a statement is true or not? Should we then articulate and make explicit all methodological and logical presuppositions necessary for meaningful and (concise) research in the field of evidence theory? Furthermore, what form would such a definition take?

We mention that because it becomes apparent that SJ fail to understand that we cannot argue for logical principles without changing the context, i.e., without start doing philosophy. Neither we nor SJ are philosophers. The difference is that we, as far as we can tell, do not make philosophical claims masqueraded as (wobbly) forensic science. For obvious reasons, we cannot climb the steep learning curve on behalf of SJ. We dismiss, therefore, SJ’s call for explaining first principles and cordially invite them to immerse themselves in foundational topics.

7. Despite our comments in the previous point, we respond to the following, more specific, contention of SJ:

“Biedermann and Kotsoglou do not explain how Aristotle’s metaphysical concept applies to the instant matter, nor do they address the possibility that objective facts may include not only the *source* but also the *condition* of the items being compared” [19, emphasis as in original].

Here, SJ bring up a new conjecture: “the possibility that objective facts may include not only the *source* but also the *condition* of the items being compared” [19, emphasis as in original]. Hence, it is neither surprising nor a failure that we have not addressed this point in our Commentary [6]. However, we can readily deal with it here.

SJ’s suggestion that the issues of *source* and *condition* could be aggregated, and thereby seemingly justify the modeling assumption of tripartite ground truth states, is unfounded because this would amount to mixing up findings with propositions. This is a common error in evidence evaluation (see, e.g. [10] for a detailed discussion): findings, observations, and/or experts’ conclusions, i.e., *evidence* (or: *observables*), must be distinguished from the propositions of interest, i.e., the *probandum* or the *probanda* (or: *unobservables*).

In making the above suggestion, SJ may have been misled by their rather careless use of terminology. As we have mentioned in Section 1, our Commentary [6] pointed out that SJ used the term “match(ing)”, a descriptor for findings and observations, to denote propositions about the source of a pair of compared items (i.e., same or different source propositions). In [6], we also emphasised that “[t]his is not just semantic pedantry” [p.

1]. Instead, it is about properly distinguishing fundamental underlying concepts. Thus, by proposing the aggregation of observations and ground truth conditions, SJ merely demonstrate that they have, once again, either ignored or misunderstood our insistence on semantic and conceptual distinctions.

8. In their last paragraph, SJ write:

“Our paper described one approach that can be used to evaluate those decisions as part of overall examiner performance, and we demonstrated why other treatments of inconclusives lead to implausible results. Biedermann and Kotsoglou’s letter fails to engage our analysis or provide any practical solution to the current situation.” [19]

We have already responded in point five above to the allegation that we did not engage in SJ’s analysis and have shown why this allegation is unfounded. Our previous paper [5] refutes DS’s [7] account, adopted by SJ [18], in detail. If SJ find our basic arguments³ in [5] for *binary* truth states in forensic reporting incomprehensible, then we cannot help further.

SJ’s last complaint that we do not provide “any practical solution to the current situation” [19], is, ultimately, self-defeating. If SJ’s salient suggestion here is that *they* provide one such solution, and we think they make this claim (SJ write: “Our paper described one approach that can be used to evaluate (...)” [19]), they err fundamentally. They err because nothing is solved by applying an illogical approach (DS’s and SJ’s modified error rate framework) to something unscientific in the first place (i.e., the default forensic examiners’ reporting categories for source attribution). Their ‘cure’ is merely another manifestation of the disease plaguing forensic examiners.

3 Conclusions

To conclude, let us restate our position. The mainstream forensic reporting scheme, whereby forensic examiners express direct opinions about source level propositions, is unscientific and obsolete [e.g., 12]. Once we realise this, the obvious consequence to be adopted – *as scientists* – should be to stop using those unscientific practices [14]. This is what we have called for in [6] and elsewhere before. This call is neither new nor original [e.g., 17]. The field of forensic genetics has mostly understood this position ever since.⁴ Other fields, including academics and practitioners, should follow suit. There is no excuse for complacency here. SJ may not like this stance, because abandoning the traditional reporting scheme, as we propose, would mean removing their cherished object of study altogether, which can explain their vehement attachment to their theory, even at the price of irrationality.

Unfortunately, SJ chose sides with the prevalent tendency among forensic scientists. By presenting a paper that takes the traditional reporting format of “identification – inconclusive – exclusion” as the object of study, SJ contribute to perpetuating, if not purporting to legitimise, this unscientific reporting format [2]. We do wonder what is worse: forensic examiners who report in terms “identification – inconclusive – exclusion”, or illogical remedies as suggested by SJ, despite overwhelming argument to the contrary. But together they represent the worst of all possible worlds, the one in which SJ operate. The “paradigm

³See, for example, the “London example” reproduced here in point 5.

⁴There are sporadic exceptions in the form of proposals that seek to resurrect source attribution conclusions in forensic genetics. See, e.g. [9] for a recent example.

shift” postulated by Saks and Koehler [17] or the “path forward” outlined in NRC’s 2009 report [16] is undoubtedly not akin to SJ’s apology for unscientific forensic reporting.

Conflict of Interest Statement

The authors declare that this response was written in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- [1] Arkes, H. R. and J. J. Koehler (2022). Inconclusives and error rates in forensic science: a signal detection theory approach. *Law, Probability and Risk* 20, 153–168.
- [2] Biedermann, A. (2022). The strange persistence of (source) “identification” claims in forensic literature through descriptivism, diagnosticism and machinism. *Forensic Science International: Synergy* 4, 100222.
- [3] Biedermann, A., S. Bozza, and F. Taroni (2008). Decision theoretic properties of forensic identification: underlying logic and argumentative implications. *Forensic Science International* 177, 120–132.
- [4] Biedermann, A., S. Bozza, F. Taroni, and J. Vuille (2019). Are inconclusive decisions in forensic science as deficient as they are said to be? *Frontiers in Psychology (Cognition)* 10, 1–9.
- [5] Biedermann, A. and K. N. Kotsoglou (2021). Forensic science and the principle of excluded middle: “inconclusive” decisions and the structure of error rate studies. *Forensic Science International: Synergy* 3, 100147.
- [6] Biedermann, A. and K. N. Kotsoglou (2024). Commentary on “Three-Way ROCs for Forensic Decision Making” by Nicholas Scurich and Richard S. John (in: *Statistics and Public Policy*). *Statistics and Public Policy* 11, 1–2.
- [7] Dror, I. E. and N. Scurich (2020). (Mis)use of scientific measurements in forensic science. *Forensic Science International: Synergy* 2, 333–338.
- [8] Feyerabend, P. (2011). *The Tyranny of Science*. Cambridge: Polity Press.
- [9] Hahn, M., K. Anslinger, M. Eckert, R. Fimmers, S. Grethe, C. Hohoff, S. Kranz, C. Leuker, C. Oppelt, S. Razbin, T. Rothämel, H. Schneider, M. Templin, M. Vennemann, A. Wächter, V. Weirich, P. Zimmermann, and P. M. Scheider (2023). Joint recommendations of the project group “Biostatistical DNA Calculations” and the Stain Commission on the Biostatistical Evaluation of Forensic DNA Analytical Findings with Fully Continuous Models (FCM). *Rechtsmedizin* 33, 3–12.
- [10] Hicks, T., A. Biedermann, J. A. de Koeijer, F. Taroni, C. Champod, and I. W. Evett (2015). The importance of distinguishing information from evidence/observations when formulating propositions. *Science & Justice* 55, 520–525.

-
- [11] Kelsen, H. (1992). *Introduction to the Problems of Legal Theory. A Translation of the First Edition of the Reine Rechtslehre or Pure Theory of Law, translated by B. Litschewski Paulson and S. L. Paulson.* Oxford: Clarendon Press.
- [12] Koehler, J. J., J. L. Mnookin, and M. J. Saks (2023). The scientific reinvention of forensic science. *Proceedings of the National Academy of Sciences* 120, 1–24.
- [13] Kotsoglou, K. N. (2014). Der normative Zugang zur Welt. Zur Theorie rechtsdogmatischer Strukturen. In J. C. Schuhr (Ed.), *Rechtssicherheit durch Rechtswissenschaft*, pp. 73–122. Tübingen: Mohr Siebeck.
- [14] Kotsoglou, K. N. and A. Biedermann (2022). Inroads into the ultimate issue rule? structural elements of communication between experts and fact finders. *The Journal of Criminal Law* 86, 223–240.
- [15] Lindley, D. V. (2017). Foreword. In B. de Finetti, *Theory of Probability, A Critical Introductory Treatment* (Reprint ed.), pp. ix–xi. Chichester: John Wiley & Sons.
- [16] National Research Council (2009). *Strengthening Forensic Science in the United States: A Path Forward.* Washington, D.C.: The National Academies Press.
- [17] Saks, M. J. and J. J. Koehler (2005). The coming paradigm shift in forensic identification science. *Science* 309, 892–895.
- [18] Scurich, N. and R. S. John (2023). Three-way ROCs for forensic decision making. *Statistics and Public Policy* 10, 1–10.
- [19] Scurich, N. and R. S. John (2024). On coping in a non-binary world: Rejoinder to Biedermann and Kotsoglou. *Statistics and Public Policy* 11, 1–2.
- [20] Taroni, F., S. Bozza, and A. Biedermann (2021). Decision theory. In D. L. Banks, K. Kafadar, D. H. Kaye, and M. Tackett (Eds.), *Handbook of Forensic Statistics*, pp. 103–130. Boca Raton: CRC Press.