Letter to the Editor: Commentary on "Strategic choice in linear sequential unmasking, Roger Koppl, *Science & Justice*, https://doi.org/10.1016/j.scijus.2018.10.010"

Alex Biedermann[a,*], Simone Gittelson[b]

[a]*University of Lausanne, School of Criminal Justice, 1015 Lausanne-Dorigny, Switzerland*
[b]*University of Technology Sydney, Centre for Forensic Science, Broadway NSW 2007, Australia*

Abstract

This letter to the Editor comments on the paper 'Strategic choice in linear sequential unmasking' by Roger Koppl (*Science & Justice*, https://doi.org/10.1016/j.scijus.2018.10.010).

Dear Editor,

In his paper 'Strategic choice in linear sequential unmasking', Koppl [10] proposes a way to describe how analysts formulate conclusions about peaks in electropherograms (EPGs), a task more generally known as allele designation. The author uses a hypothetical example of a peak labeled 12 at locus D3S1358, and an examiner's conclusion that this peak is an artifact. In addition, the author supposes that examiners will be required to provide an expression of the extent to which they are convinced that their determination is accurate. He assumes this is to be done by means of a color label, either green (for certainty), yellow (meaning low certainty) or red (meaning questionable). More specifically, Koppl inquires about how to assign such labels in a way that is optimal with respect to some criterion that is at the heart of his discussion. Specifically, he considers a quantification of the desirability of the various consequences of the color label determinations, using a numerical score, weighted by probabilities of (future) unmasking information to impact – in one way or another – the determination made. Koppl asserts that his approach "(...) is descriptive and not normative. It does not attempt to represent the problems forensic scientists should solve or how they should solve them" [10, at p. 7]. Notwithstanding, he is drawing conclusions about so-called (laboratory) 'error tolerance regimes'.

We greatly welcome Koppl's view that conclusions of forensic practitioners are actually decisions, a view that we have supported previously [e.g., 3, 9, 14]. However, we find that Koppl's decision model leads to concerns. Further, we disagree with his definition of error and find that his divide between description and normativism is vague. In this comment we will argue that genuine normative analysis also has powerful descriptive character regarding the decision problem – with the additional advantage of addressing real problems that matter to forensic scientists. Below, we address four points in turn: the object of study, modeling allele designation decisions, the notion of error, and the descriptive capacity of normative analyses.

1. *The object of study.* – Koppl states that his analysis does not attempt to represent problems forensic scientists should solve. Instead, he asserts, his interest focuses on the following question: "Will LSU [linear sequential unmasking] confidence attributions *in fact* improve forensic science relative to sequential unmasking (SU)?" [10, at p. 7]. From his analyses he concludes "(...) that a positive answer is contingent on the overarching organization of the crime laboratory in general and its error tolerance regime in particular" [10, at p. 7, emphasis as in original]. The way

---

*Corresponding author
Email address: alex.biedermann@unil.ch (Alex Biedermann)

in which this answer is worked out for the initial question raises two issues. First, if the question is whether an improvement 'will in fact' be obtained, then this is, in essence, an empirical question for which actual experiments under controlled conditions are most suitable to provide an answer.[1] Second, if – nevertheless – the question of interest is to be answered with a conceptual analysis, then the suitability of the conclusion is dependent on how well the analysis captures the object of study. But what exactly is the object of study? Is it allele designation by practising forensic scientists, the performance (or added value) of procedural prescriptions (such as LSU and SU), or both? Whichever is the case, the fact is that Koppl chooses a particular way of looking at how forensic practitioners presumably formulate conclusions (i.e., make decisions) about allele designations. We will comment on this choice in the point 2 below.

2. *Modeling allele designation decisions.* – As a preliminary, let us note that there are different types of artifacts one can observe on an EPG. We differentiate between the stutter artifacts and the other, non-stutter artifacts. The decision of calling a peak allelic or a stutter artifact is outdated today because it has been replaced with probabilistic methods taking the form of continuous DNA interpretation models [e.g., 15]. These probabilistic methods take into account the uncertainty related to the nature of the peak (allelic, stutter, or allelic and stutter) by quantifying the probability of the observed peak heights based on a model for expected allele and stutter peak heights [6]. They do not require any categorical decision on whether a peak is allelic or stutter. Clearly, the author is perfectly aware of these developments and his example was chosen only for its clarity, specificity, and simplicity. Also, there is some benefit in Koppl's view if we assume – as we will do in the rest of this letter – that "artifact" refers to the other artifacts such as dye blobs, spikes, pull-ups, and off ladder (OL) alleles, for which DNA analysts do need to make decisions, regardless of what DNA interpretation model they subsequently apply for the evaluation of their observations.

Koppl's analysis supposes that when scientists attribute a color label to a given statement (e.g., 'the peak marked 12 is an artifact'), they do so by thinking about whether at the end of the analytical process their conclusion will be supported or reversed by information revealed through the unmasking process.[2] Decision theory is a liberal concept in the sense that it does not dictate what exactly we should consider as conditioning states of nature. However, the states of nature should provide a reasonable coverage of the various possibilities (i.e., events of the present, past or future). So, strictly speaking, Koppl's analysis (i.e., Koppl's Table 1) should also allow for the possibility that the unmasked information has *no* effect, neither supporting nor undermining the original decision. This will affect the computation of the expected score (or, in Koppl's terminology, 'payoff').

More fundamentally, Koppl's definition of the states of nature as the occurrence or non-occurrence of a situation of 'reversal' raises intricate concerns, because it conditions the scientist's decision on particular factors that are external to the observations on the questioned EPG. In Koppl's example of stating that the peak labeled '12' at locus D3S1358 is an artifact, the scientist using Koppl's states of nature must make the decision by taking into account the probability distribution over the possible traits of the information that will subsequently be revealed in the unmasking process. This information is, for example, the genotype(s) of the reference EPG(s) of one or several potential, contested donor(s). In other words, the 'reversal risk', as Koppl calls it, depends on the probability that the genotype of (one of) the reference EPG(s) of a potential, contested donor has an allele 12 at locus D3S1358, since such an observation might lead to a reversal according to the LSU procedure [e.g., 8]. We find that a decision model forcing the scientist to make decisions about the nature of the peaks in the questioned EPG based on his or her expectations of the features of the additional, potentially biasing, masked information would encourage questionable practice, and raise issues such as confirmatory bias [11, 16]. On the contrary, current best practice actually dictates that such decisions should be made without any knowledge (or presumptions thereof), of the reference EPG(s) of the contested donors [12, guidelines 3.1.1 and 3.1.2.1].

In our opinion, the scientist should be concerned about the accuracy of his or her statement, rather than about whether this statement will be reversed at the end of the process. By 'accuracy', we mean how closely the statement reflects the true state of nature, which in this case is whether the peak of interest is genuine or an artifact. In other words, we believe that the probability to take into consideration when making this decision should be the probability that this peak is allelic, rather than the risk that this decision will be reversed. Consider, by analogy, the decision of

---

[1] See also [5] on the difference between decision empiricism and decision logic.

[2] It is not clear to what kind of unmasking information Koppl exactly refers to, i.e. whether it is information that reveals the true state of nature with certainty (e.g., in case of a proficiency test), or whether it is merely imperfect evidence on whether or not a given peak is genuine or an artifact, respectively.

designating the DNA donor's genotype for a low-template DNA profile [9], or a fingerprint examiner's conclusion as to whether a given finger of interest is or is not identified (or individualized) as the source of a questioned mark [e.g., 2]. The scientist draws these conclusions in the light of the uncertainty about the true donor's genotype for the DNA case, and in light of the uncertainty about whether or not the finger of interest truly is the source of the crime mark in the fingerprint case.

As yet another example, consider a weather forecaster who announces rain tomorrow, with some probability. The 'goodness' of this prevision is scored against whether or not it will rain tomorrow. It will not be scored against, using Koppl's words, the event called 'subsequently unmasked information tends to support (undermine) the statement'. Table 1 illustrates the weather forecaster's decision model (columns 1 through 3), and aligns it with the analogous example of our allele designation decision model with the conditioning on the true state of nature (columns 4 through 6). More specifically, we list the weather forecaster's possible probability assignments for the statement that it will rain tomorrow (column 1), followed by the scores obtained when there will in fact be rain (column 2) and no rain (column 3), respectively. By analogy, column 4 contains the possible probability assignments for the statement that the peak labelled 12 is an artifact. Note that here the assignment of a probability for such a statement is just a numerical equivalent to Koppl's qualitative color label. Columns 5 and 6 give the scores incurred when the peak is indeed an artifact or genuine, respectively. There is substantial literature on how to specify the scores for the various probability assignments [e.g., 13]. Among the well known concepts are so-called proper scoring rules, especially the quadratic scoring rule (also known as the Brier score). It consists of assigning the square of the difference between the assigned probability and the actual truth value of the conditioning state of nature, where 1 is the truth value if the state is true, and 0 is the truth value when the state is false. For example, if the weather forecaster assigns the value 0.8 for the proposition 'Rain tomorrow', then he will incur the score $(1 - 0.8)^2 = 0.04$ if it will rain tomorrow, and the score $(0.8)^2 = 0.64$ if it will not rain tomorrow. Scoring rules, in particular proper scoring rules, have strong and deep conceptual properties that are beyond the scope of what we can cover here. It suffices to notice that proper scoring rules[3] imply a minimisation of the *expected penalty* when scientists state their actual probabilities. That is, the scoring rule encourages the elicitation of sincere probability assertions. We argue that this is the incentive that shall drive the scientist's assessment. We do *not* want scientists to state a probability (or, using Koppl's qualitative color label, a color) *different* from that they actually think is suitable (or, scientifically defensible). Let us restate this result: when we call experts we want them to give their best assessment regarding the nature of an observable peak in an EPG. And, we want them to give these assessments in an honest and transparent way, by which we mean that the statement must not be skewed, distorted or exaggerated (e.g., by rounding a small probability down to zero). The 'goodness' of such asserted probabilities is most naturally scored against the true state of nature. At the center of this analysis is the actual nature of a phenomenological observation – the EPG peak. The center of attention is not the scientists concern about their statements being discredited or not.

Table 1: Illustration of the quadratic scoring rule for (i) the case of a weather forecaster providing a probability for the event 'Rain tomorrow' (columns 1 to 3), and (ii) the case of a forensic expert providing a probability for the statement 'The peak labelled 12 is an artifact' (columns 4 to 6). The scores in columns 2 and 3 (5 and 6) are given by the squared difference between the value in column 1 (4) and the truth values in columns 2 and 3 (5 and 6), respectively.

Asserted probability for 'Rain tomorrow'	Rain tomorrow		Asserted probability for 'Peak is an artifact'	Peak is an artifact	
	True	False		True	False
0	1	0	0	1	0
0.01	0.9801	0.0001	0.01	0.9801	0.0001
...	...	...	...	...	...
0.99	0.0001	0.9801	0.99	0.0001	0.9801
1	0	1	1	0	1

3. *The notion of error.* – Koppl frames the notion of error as follows: "Information revealed subsequently may induce the examiner to change their determinations. Such a change could be construed as showing that the earlier

---

[3]Note that there exist proper scoring rules other than the quadratic scoring rule, though the latter is probable among the simplest ones.

determination was an error" [10, at p. 17]. This account of the notion of error raises two discussion points.

First, unless the unmasking information reveals the true state of nature (which may be the case in an experiment under controlled conditions), the suitability of considering a notion of error under states of nature defined as 'obtaining confirming/undermining information' hinges upon the capacity of such information to point to the true state of nature. Specifically, the true state of nature may – after all – be *different* from what the confirming/undermining information suggests. What is more, it is not sufficient to consider only the inferential direction in which unmasking information points, it is equally relevant to account for any information regarding the true state of nature *prior* to considering the unmasking information. Indeed, prior information may be so strong that even though unmasking information points in a new direction, this does not mean that one should entirely overturn or give up on a current direction of reasoning.[4] Both unmasking and other (i.e., prior) information logically shapes one's belief about the true state of nature, suggesting thus again that the relevant consideration is a conditioning on the true state of nature, as sketched in Table 1 and point 2 above. According to this view, the proper use of the notion of inaccurate (or, erroneous) outcomes requires configurations where an expert asserts 'A' ('not-A'), for exemple, when *in reality* not-A (A) is true.

Second, the *ex post* qualification of *ex ante* assessments as errors is delicate, raising the fundamental question of whether it is sensible to consider that probability (or, in Koppl's case, colour label) assignments could be erroneous, in the light of new information. De Finetti's writings can help us reflect on this question:

> "If, on the basis of observations (...) one formulates new and different previsions (...) for events whose outcome is unknown, it is *not* a question of a *correction*. It is simply a question of a new evaluation, *cohering with the previous one*, and making use – by means of Bayes's theorem – of the new results which enrich one's state of information, drawing out of this the evaluations *corresponding to this new state of information*. For the person making them (You, me, some other individual), these evaluations are as correct *now*, as were, and are, the preceding ones, though of *then*. There is no contradiction in saying that my watch is correct because it now says 10.05 p.m., and that it was also correct four hours ago, although it then said 6.05 p.m." [7, at p. 176]

According to this view, a given probability is a function of the information available at the time when the probability is asserted. Although there may, ex post, be a mismatch between the true state of nature and the proposition initially supported, this would be a case of misleading evidence. But there is no point in blaming a probability assessment based on information that was unavailable at the time when the original assessment was made. Instead, the important point is to recognize that one's assessment needs to be reviewed. This perspective goes some way along the idea of encouraging what Koppl's refers to as "error tolerance regimes" [10, at p. 2] to adopt more constructive ways of dealing with new information.

4. *Descriptive capacity of normative analyses*. – Koppl distinguishes between normative and descriptive analyses, and insists on focusing on the latter. We find this vague because Koppl's account, too, contains prescriptive elements. An example is the following assertion: "The examiner will choose the confidence attribution that maximizes their expected payoff" [10, at p. 14]. As we understand it, the suggested procedure aims to show scientists which action is optimal *for them* with respect to a specified criterion. Implicit, then, is the stipulation that scientists *should* select the action that the procedure has singled out as optimal. This goes some way along classic normative decision-theoretic analysis that focuses on minimizing (maximizing) expected loss (utility).

There is, of course, a difference between descriptive and normative claims, but this does not mean that normative decision theory could not have descriptive potential, too. Normative decision theory may be descriptively poor with respect to observable behavior, that is the empirical observation that practical decision makers, unaided by formal approaches to reasoning, do not behave such as to minimize (maximize) expected loss (utility) [e.g., 1]. Notwithstanding, it cannot be denied that normative decision theory is descriptive with respect to the *problem* faced by scientists in the first place, that is the decision to be made under uncertainty – and its fundamental ingredients. Those ingredients are actions, states of nature (with associated probabilities) and possible outcomes (with associated utilities/losses).

What is more, even on the level of normative decision criteria we can, under certain circumstances, observe descriptive capacity. As an example, consider a reduced version of the example sketched in the right-hand side of Table

---

[4]The foundations of these arguments rely on Bayesian inference. See also the discussion in the paragraph below with the quote from De Finetti [7].

1. For the sake of simplicity, consider only two decisions, the statements 'The peak labeled 12 is an artifact' (action 1) and 'The peak labeled 12 is genuine' (action 2), and leave the definition of the states of nature as indicated on top of columns 5 and 6 (calling them propositions 1 and 2 for brevity). Suppose further that we use a $(0, k)$ loss function which assigns a 0 loss to accurate outcomes (e.g., concluding 'The peak is an artifact' when in fact the peak is an artifact), and a loss equal to k to erroneous outcomes (e.g., wrongly concluding that a peak is an artifact). Then it can be shown that the decision-theoretic criterion amounts to a comparison between the odds in favor of the first against the second proposition, and the relative losses associated with the two possible ways of deciding erroneously (for details see, e.g., [5]). Without going into numerical details, this result says that the more adverse a wrong outcome from action 1 is, compared to a wrong outcome resulting from action 2, the more we shall be sure that proposition 1 rather than proposition 2 is true, in order for action 1 to be warranted. This normative statement is descriptive in the sense that it reflects how we commonly think about decision making under uncertainty: indeed, we often say 'The more is at stake, the more you shall be sure before you decide' [4].

In conclusion, while we agree with Koppl that allele designation is a problem of decision making under uncertainty, our main difference in view regards the definition of the states of nature. As mentioned, Koppl focuses on the event of future information tending to undermine or confirm the examiner's determination. In turn, we argue in favor of conditioning on the true state of nature, that is the reality underlying the phenomenological observation made (i.e., peak in an EPG). This way of modeling the decision problem makes it clear that new information pointing in a different direction does not necessarily reveal an error. We also emphasize that assigning a probability cannot be erroneous in the first place because probabilities are a function of the available information at the time probabilities are asserted – though probabilities may, of course, require *revision* as knowledge and information evolve and accumulate.

Acknowledgments

References

[1] J Baron. *Thinking and Deciding*. Cambridge University Press, New York, 4th edition, 2008.

[2] A Biedermann, S Bozza, and F Taroni. Decision theoretic properties of forensic identification: Underlying logic and argumentative implications. *Forensic Science International*, 177:120–132, 2008.

[3] A Biedermann, S Bozza, and F Taroni. The decisionalization of individualization. *Forensic Science International*, 266:29–38, 2016.

[4] A Biedermann, S Bozza, and F Taroni. Normative decision analysis in forensic science. *Artificial Intelligence and Law*, in press, 2018.

[5] A Biedermann, S Bozza, F Taroni, and P Garbolino. A formal approach to qualifying and quantifying the 'goodness' of forensic identification decisions. *Law, Probability and Risk*, 17:295–310, 2018.

[6] J-A Bright, D Taylor, JM Curran, and JS Buckleton. Developing allelic and stutter peak height models for a continuous method of DNA interpretation. *Forensic Science International: Genetics*, 7:296–304, 2013.

[7] B de Finetti. *Theory of Probability, A Critical Introductory Treatment*. John Wiley & Sons, Chichester, reprint edition, 2017.

[8] IE Dror, WC Thompson, CA Meissner, IK Kornfield, D Krane, M Saks, and M Risinger. Letter to the editor–context management toolbox: A linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Sciences*, 60(4): 1111–1112, 2015.

[9] S Gittelson, A Biedermann, S Bozza, and F Taroni. Decision analysis for the genotype designation in low-template-DNA profiles. *Forensic Science International: Genetics*, 9:118–133, 2014.

[10] R Koppl. Strategic choice in linear sequential unmasking. *Science & Justice*, in press, 2018.

[11] DE Krane, S Ford, J Gilder, K Inman, A Jamieson, R Koppl, I Kornfield, DM Risinger, N Rudin, MS Taylor, and WC Thompson. Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. *Journal of Forensic Sciences*, 53(4):1006–1007, 2008.

[12] Scientific Working Group on DNA Analysis Methods (SWGDAM). Interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. https://www.swgdam.org/publications, 2017. [last accessed on the 26th of November 2018].

[13] G Parmigiani and L Inoue. *Decision Theory: Principles and Approaches*. John Wiley & Sons, Chichester, 2009.

[14] F Taroni, S Bozza, and C G G Aitken. Decision analysis in forensic science. *Journal of Forensic Sciences*, 50:894–905, 2005.

[15] D Taylor, J-A Bright, and Buckleton JS. The interpretation of single and mixed DNA profiles. *Forensic Science International: Genetics*, 7: 516–528, 2013.

[16] WC Thompson. Painting the target around the matching profile: the texas sharpshooter fallacy in forensic DNA interpretation. *Law, Probability and Risk*, 8:257–276, 2009.