The Criminalist's Paradox as a Counterexample to the Principle of Total Evidence

Michał Sikorski and Alexanders Gebharter

Abstract

The principle of total evidence says that all relevant information should be considered when making an inference about a hypothesis. In this article, we argue that the criminalist's paradox from the literature on the methodology of forensic science constitutes a counterexample against the principle of total evidence. The paradox arises, for example, when a forensic scientist uses the results from other forensic procedures to inform their own analysis. In such cases, their results can become more reliable, but at the same time also dependent on those other forensic results and therefore less useful for decisions in court cases. Consequently, such influences are generally avoided in forensic science. We argue that structurally similar problems plague also other scientific disciplines, anticipate an objection to the counterexample, and propose two new versions of the principle of total evidence that do not fall prey to the criminalist's paradox.

1 Introduction

The principle of total evidence (PTE) is an important and popular principle in epistemology. In a nutshell, it says that one should always take all the available relevant evidence into account when making an inference about a hypothesis. One of the main driving factors behind PTE is the intuition that considering more relevant evidence should increase the reliability of one's inference. PTE is generally accepted by the majority of epistemologists. Nevertheless, we think that there is a wide range of cases to which it should not be applied. In this paper, we argue that the criminalist's paradox, a phenomenon described in the literature on the methodology of forensic science, constitutes a counterexample to PTE. We discuss a possible objection to this counterexample and argue that the range of problematic cases actually exceeds forensic science and that these cases can in principle be found in any scientific discipline whenever evidence is evaluated by some well-defined procedure. Finally, we propose two alternative versions of PTE that are also applicable to the problematic cases. As we will see, these alternative versions are not intended to generally replace PTE, but rather to supplement it. So the take home message is not that PTE should generally be replaced, but that it should be adopted only in certain cases while one should endorse the alternative versions we propose in other cases.

The paper is structured as follows: In section 2 we introduce PTE in more detail. In section 3 we then present the criminalist's paradox from the literature on the methodology of forensic science. In section 4 we construct a toy case from forensic science demonstrating that the criminalist's paradox constitutes a counterexample to PTE. In section 5 we discuss a possible objection. In section 6 we finally propose two alternative versions and argue that they should not replace PTE in general, but only in cases structurally similar to the criminalist's paradox. We conclude in section 7.

2 The principle of total evidence

In this section, we introduce PTE – which is also sometimes called the requirement of total evidence – in a bit more detail. The principle claims that one should take into account all available relevant information when making an inference about a hypothesis. More precisely, it can be stated as follows:

> "Suppose data d_1 are strictly logically stronger than data d_2, then an inference about hypothesis H should be based on d_1 if changing between d_1 and d_2 changes the evidential assessment." (Autzen 2016, p. 286)

The principle was first proposed by Bernoulli and is often attributed to Carnap (1947). While the principle is initially plausible, additional justification for it can also be found in the literature. Reichenbach (1949), for example, puts forward an epistemic argument for the principle while Carnap (1950) provides an example of a court case in order to justify PTE:[1]

> "If a judge in determining the probability of a defendant's guilt were to disregard some relevant facts brought to his knowledge; [...] then everybody would regard such a procedure as wrong." (Carnap 1950, pp. 211f)

Recently, two counterexamples to PTE have been presented by Epstein (2017). Draper (2020) and Barrett and Sober (2020) argued against Epstein's counterexamples and the principle remains popular and generally accepted in philosophy. Thus, we will not go into details about Epstein's counterexamples but will rather focus on a new one which we consider more promising: the criminalist's paradox.

3 The criminalist's paradox

The criminalist's paradox is a frequently debated topic in the literature on forensic science methodology. To the best of our knowledge, it has not yet been discussed within philosophy of science. To keep things simple, we start by presenting a concrete example of a specific forensic procedure: fingerprint analysis. We then present the criminalist's paradox and its connection to a generally accepted strategy to reduce a certain type of bias in forensic science.

3.1 An exemplary procedure: fingerprint analysis

Fingerprint analysis is a forensic discipline during which experts identify the source of two studied fingerprints. Fingerprints are impressions of the skin of the palms of the hands and fingers and the soles of the feet and toes on touched surfaces. Typically one of the analyzed fingerprints is secured at the crime scene and the second one is collected from the person of interest (e.g., the suspect). The approach most commonly used in fingerprint analysis is the so-called ACE-V approach. It consists of four stages corresponding to the letters in the acronym: analysis, comparison, evaluation, and validation. During the first step, an expert analyzes the features of both fingerprints. Those features are typically divided into three levels of detail. This distinction was introduced in (Ashbaugh 1999). The first level is the overall pattern of the ridges. There are three basic patterns: loops, whirls, and arches. The second level features

[1] For a discussion of those and other arguments for PTE see, for example, (Roush 2020).

include the configurations of individual ridges, their length, position, and configuration. The third level consists of the more fine-grained details of individual ridges such as edges, textures, and pore positions. During the comparison stage, the expert compares the features. In the third step, she formulates her decision on the basis of the results of the previous stages.

After these three steps, the experts must come to a qualitative verdict. The three possible conclusions are (cf. OSAC 2017): agreement (there is a sufficient agreement to determine that the source of both fingerprints is the same), disagreement (there is a sufficient level of difference to determine that the sources of the two fingerprints are different), and inconclusive (there were not enough clear details to determine the source). During the validation stage, another expert repeats the whole procedure to validate the result. Since the way in which each of the stages should be conducted is not specified, fingerprint analysis allows for a significant degree of flexibility. Until recently, it was believed to be highly reliable. This confidence was to some degree shaken by the results suggesting that fingerprint analysis and other forensic disciplines are susceptible to the biasing influence of contextual information.

3.2 The criminalist's paradox and bias in forensic science

In response to worries raised in an influential report presented by the National Research Council (NRC 2009), a number of studies were conducted to explore the extent to which external factors (e.g., information concerning the subject's admission of guilt or results of other forensic procedures) can bias forensic analysis and their results. The results of the majority of those experiments show that external factors have the potential to change the results of forensic analyses (see Cooper and Meterko 2019 for systematic review). Consequently, a number of authors and organizations proposed to enforce strict rules concerning which information should be used in forensic practice (e.g., NRC 2009, Thompson 2010, Quigley-Mcbride et al. 2022). The best-developed version of such a rule is the following relevance principle:

> RP "Forensic scientists should rely solely on task-relevant information when performing forensic analyses." (NCFS 2016, p. 1)

Task-relevant information is information that changes the probability of collected evidence (e.g., the fingerprint collected at the crime scene) given the hypothesis (e.g., the person of interest is the source of the fingerprint collected at the crime scene) is true or false. In other words, task-relevant information is any information that would still have a probabilistic impact on the evidence even if one would learn that the hypothesis were true or false. An example of task-relevant information is information concerning some of the conditions at the crime scene, for example, the surface from which the fingerprint was collected. A typical example of information that is not relevant and should not influence the analysis is information concerning the results of other forensic procedures.

Unsurprisingly, some more traditionally-minded forensic experts opposed the principle (see, e.g., Butt 2013, Elaad 2013, Curley et al. 2019, Curley, Munro, and Lages 2020). According to them, an experienced forensic expert can use task-irrelevant information to increase the reliability of her analysis. For example, Curley, Munro, and Lages (2020) cite results of Stevenage and Bennett (2017) which suggest that in some cases knowledge of the results of the corresponding DNA analysis makes fingerprint experts more reliable. Consequently, they claim that task-irrelevant information can and even should be used in forensic examinations. Such an increase in reliability is both plausible and unsurprising. The positive results of the DNA analysis indicate the presence of the DNA of the person of interest at the crime scene. The presence at the crime scene can result in the deposition of her fingerprints. Therefore, the positive results of DNA analysis are typically a good predictor of positive results of fingerprint analysis. Additionally, DNA analysis is more reliable than fingerprint analysis. In light of that it is expected that knowledge of the results of the DNA analysis will increase the reliability of the fingerprint analysis. The least controversial example of such positive influence is a case in which the same trace is analyzed by both methods. Due to the ever-increasing sensitivity of DNA analysis, DNA from the particles of sweat and skin contained in a fingerprint can now be tested. Consequently, both DNA and fingerprint analysis can be used to analyze the same fingerprint. In the vast majority of cases, they will deliver consistent results. Once again, given the superior reliability of DNA analysis and the expected consistency of results, the reliability of fingerprint analysis would benefit from access to previous results of DNA analysis.

Kukucka (2020) responded to these reactionist voices by claiming that because of the role forensic analysis plays in the legal context, the independence of forensic evidence is more important than its reliability:[2]

> [...] "any effect of task-irrelevant information on an examiner's judgment – whether "positive or negative" – is problematic. For instance, a person with zero training or experience in latent fingerprint identification could likely achieve a respectable accuracy rate if they always knew of DNA results before comparing two fingerprints and simply followed whichever conclusion the DNA suggested – but it would be absurd to call this person a fingerprint expert or to ascribe any independent value to their opinion." (Kukucka 2020, p. 112)

Forensic results are useful during a trial only to the degree to which they are independent[3] of other forensic results and other evidence considered by the jury. Consequently, any degree of dependence compromises the value of forensic evidence. This constitutes the criminalist's paradox:

> **CP** Some of the information typically available to a forensic scientist can (in some cases) improve the reliability of forensic results, but also compromise their independence, making them less useful for court cases.

From this forensic scientists typically draw the conclusion that such information should be avoided:

> "The conclusion that less information is better may be difficult for some forensic scientists to accept, however, due to what I will call 'the criminalist's paradox'. By considering contextual information, analysts may well become more likely to interpret their evidence correctly – that is, to reach conclusions that correspond to what actually happened. Yet by doing so, they also (paradoxically) undermine the ability of the trier-of-fact to determine the truth, and thereby reduce the likelihood the legal system will reach a just outcome. This is the paradox: by helping themselves be 'right' such analysts make it more likely that the justice system will go wrong. By trying to give the 'right' answer, they prevent themselves from providing the best evidence." (Thompson 2010, p. 130)

[2] See also the response by Thompson 2020 to Curley, Munro, and Lages 2020.

[3] Note that there are different notions of dependence. In one sense, two forensic results can be dependent if they aim at establishing the same hypothesis which, if true, would generate the outcome of both forensic procedures. In such a case, one positive result would increase the likelihood of the other one being positive too. But this is not the dependence in which we are interested in this paper. When saying that two forensic results are independent, we mean that the expert carrying out one of the forensic procedures did not take the outcome of the other forensic procedure into account in her own inquiry.

The postulated negative effects of task-irrelevant information used during forensic procedures on criminal verdicts have not been empirically tested. Given the challenges involved in assessing the correctness of verdicts, controlling which information was used by forensic experts and the many bureaucratic obstacles, it may be impossible to conduct such a study. At the same time, there is significant evidence showing that such influences compromise the reliability of forensic results which have been shown to be important factors contributing to convictions (see e.g., Nir and Griffiths 2018, Peterson et al. 2013, Ling, Kaplan, and Berryessa 2021, or Smit, Morgan, and Lagnado 2018) and, more crucially, wrongful convictions (see Morgan 2023, Bonventre 2020, Garrett and Neufeld 2009, and Sikorski 2022 for discussion). Finally, the potential negative effect of task-irrelevant information being used in forensic examinations on verdicts was validated in a recent simulation study presented in (Cuéllar, Mauro, and Luby 2022). All this evidence contributes to the consensus in forensic science methodology that forensic experts should – in line with RP – rely solely on task-relevant factors to ensure that their results are independent of other forensic results (see e.g., Bird, Jones, and Ballantyne 2024, Kunkler and Roy 2023, Koehler, Mnookin, and Saks 2023, or PCAST 2016).

RP became popular in forensic science and inspired important methodological reforms, for example, sequential unmasking (Krane et al. 2008, Robertson and Kesselheim 2016). Sequential unmasking consists of hiding information that is inessential at a given stage of the analysis from the expert. If an expert does not know, for example, what the results of other forensic procedures were, this knowledge cannot bias her own investigation. As we will see in section 4, the criminalist's paradox and the intuitive plausibility and wide acceptance of RP constitute a strong case against the universality of PTE.

4 The criminalist's paradox as a counterexample to the principle of total evidence

When all the elements are set up it is easy to see how the situation described in the criminalist's paradox constitutes a counterexample to the PTE. As we have seen, there is consensus in forensic science that task-irrelevant information should not be used by forensic experts. At the same time, it is clear that some of the task-irrelevant information constitutes evidence in the sense of PTE. Thus, in accordance with RP and contrary to PTE, there is evidence that should not be considered by forensic scientists. Here is a fully-fledged counterexample:

C A forensic expert conducts a fingerprint analysis. The tested hypothesis is that the examined fingerprints were deposited by the same person. The expert works in a conservative institution and therefore sequential masking was not implemented. Consequently, she has access to the results of DNA analysis performed on the fingerprint and other task-irrelevant information. She can base her decision concerning the source of examined fingerprints on (among others) two sets of evidence. Firstly, she can use set d_3 containing only task-relevant evidence. Secondly, she can use set d_4 that contains task-relevant evidence and the result of the DNA analysis. d_4 is logically stronger than d_3 and, in this case, using d_4 instead of d_3 would change the evidential assessment. The task-relevant evidence does not support a positive identification as the fingerprint collected from the crime scene is not of good quality. On the other hand, when the expert learns that the DNA result implicates the person of interest it will inform her analysis and lead her to a consistent, positive

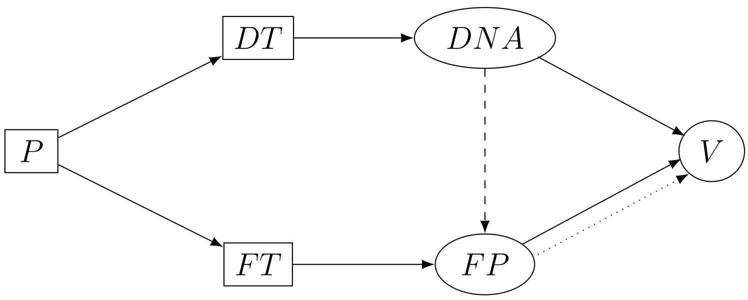

Figure 1: The relation between results of fingerprint and DNA analyses in C. P is the presence of the suspect at the crime scene. DT and FT are DNA material and fingerprints deposited at the crime scene. DNA and FP are the results of DNA and the fingerprint analysis and V is the verdict in the court case. The lines indicate causal relations, the dashed lines indicate a problematic influence of the results of DNA analysis on the results of fingerprint analysis, and the dotted line is an indirect but equally problematic influence of DNA results on the verdict through the results of fingerprint analysis.

> conclusion. Consequently, PTE requires her to use d_4. On the other hand and in line with **PR**, she should not use it because it will make the conclusion dependent on the result of the DNA analysis and therefore less useful for the court than it would be otherwise.

C clearly constitutes a counterexample against PTE. Further illustration is provided in Figure 1. At this point, is is important to note that the hypothesis mentioned in PTE can in principle express any descriptive statement and that what should be considered as hypothesis and what as evidence can change when changing one's perspective. In C, the hypothesis is the fingerprint expert's potential result that the fingerprints at the crime scene match the suspect's fingerprints (FP in the figure) and possible pieces of evidence are the fingerprints left at the crime scene (FT) and the DNA analysis' result (DNA). Taking also the DNA analysis' results into account, the fingerprint expert's result would indeed become more reliable and there is nothing wrong with that when looking at the fingerprint expert's work in isolation. But this is not the end of the story. If we now take the judge's perspective, we see that she has another inference target. The hypothesis she needs to evaluate is whether the suspect is guilty or not (V) and the evidence available to her is the fingerprint expert's result (FP) and the DNA expert's result (DNA). To arrive at an unbiased decision, she needs both results FP and DNA to be independent. Otherwise, she would fall prey to double-counting since she would count the part of the fingerprint expert's reported result due to the influence of the DNA expert's result twice in her evaluation of whether the suspect is guilty.

It is important to also highlight that C is not driven by the defectiveness of the evidence in question (the result of DNA analysis). Plausibly there is plenty of defective evidence we would prefer not to include in our reasoning. Such evidence may be epistemically defective (e.g., results of unreliable studies) or morally defective (e.g., scientific results reached in unethical experiments). Use of such defective evidence is typically avoided and therefore they may be considered to constitute a counter-example to PTE. On the other hand, given that this defectiveness makes those potential counterexamples unconvincing, as PTE seems to presuppose that we do not know if a given piece of evidence is defective or not. In the case described in C the evidence in question (the result of DNA analysis) is reliable and not problematic in any other way. This is realistic since the results of DNA analysis are typically very reliable and more reliable than the results of fingerprint analysis. The counterexample is driven by the role of forensic analysis in a criminal trial. It shows that no matter how reliable the given evidence is, there may be contexts in which it should not be used. In the next section, we discuss a possible objection to our counterexample.

5 A possible objection

PTE remains very popular in epistemology and is rarely contested. In light of that, some push-back against the counterexample presented in section 4 can be expected. In this section, we anticipate the most obvious objection.

One might worry that forensic analysis is a very specific case that is not representative of scientific practice and that, thus, the counterexample is not a noteworthy threat for PTE in the majority of scientific disciplines. We do not think that this is the case. The general structure of the situation described in C is not unique to forensic science. Consider, for example, the case of a randomized controlled trial (RCT). Let us say that a scientist is performing an RCT to test the toxicity of a certain substance. She is aware of the results of earlier studies testing the same hypothesis. Should the scientist use her knowledge of these results of previous experiments during the design and conduction of the new experiment? Firstly, it seems that those results cannot be used directly as data points in an RCT as it is not designed to accommodate them. But perhaps the scientist in question can use her knowledge of the previous experiment in an indirect way. For example, it can inform her methodological decisions such as the choice of the population from which participants are to be recruited or the choice of the primary outcome. Such influence may also be problematic, the methodological inspiration may result in reproducing problematic features present in the original study. This may result in confirmation bias, a situation in which knowledge of previous results makes the scientist more likely to find support for those results (e.g., Fries and Krishnan 2004 or Montaner, O'Shaughnessy, and Schechter 2001).

In light of that, it seems that in the case of RCTs some of the evidence relevant for the tested hypothesis should not be taken under consideration. This example is less obvious and the conclusion we drew from it is less entrenched in scientific methodology than RP is in the methodology of forensic science. Nonetheless, it seems that it suggests that some of the available evidence should be excluded for reasons analogous to those raised in the discussion of the criminalist's paradox. It seems that some scientific procedures are designed with a particular kind of data as input and there seems to be no reason or even a way to incorporate other evidence into them. Moreover, the inclusion of additional information may even be harmful.

Glymour realizes this dependence of what should be taken as total evidence on the procedure utilized. He uses the example of causal search algorithms to demonstrate that different epistemic procedures take different inputs:

> "For example, the PC algorithm (Spirtes et al., 2001) uses only the covariance matrix of the data and the sample size to derive conditional independence estimates and outputs an equivalence class of causal Directed Acyclic Graphs (DAGs); however, methods based on Linear, Non-Gaussian, Acyclic Models (LinGAMs) (Shimizu et al., 2006) further exploit certain non-Gaussianity aspects of the data distribution as part of the total evidence, leading to the ability to estimate a unique causal DAG." (Glymour 2017, p. 112)

Those different algorithms differ in accuracy and efficiency. In light of that Glymour claims that what constitutes total evidence depends on our epistemic goal and the methodology we use to achieve it. He proposes to include the used methodology as a conditional variable in the definition of total evidence. Though in Glymour's particular case, taking additional information into account will not threaten the reliability of the inference, we believe that relativizing what should be taken as total evidence to the specific procedure used can help to overcome our problem as well. We thus follow his recommendation when presenting two new versions of PTE in the next section.

6 Two alternative versions of the principle of total evidence

One of the most important roles PTE plays in epistemology is to increase the reliability of the inference establishing or disconfirming a hypothesis. The intuitive idea is that the inference should become the more reliable the more relevant evidence one takes into consideration. As we saw in section 4 and section 5, this intuition seems misplaced in a wide range of scientific contexts. In particular, these cases are those in which one's research outcome (e.g., the fingerprint analysis) is used to confirm or disconfirm another hypothesis (e.g., whether the suspect is guilty). In such cases information that would make one's results more reliable might, at the same time, compromise the inference to the other hypothesis. Nevertheless, the basic epistemic goal to increase reliability is intuitively plausible and undoubtedly worth preserving. Thus, it will also be the main driving factor underlying our alternative versions of the principle of total evidence. The main challenge is to formulate these principles in such a way that they take all relevant information except the problematic one into account. In the light of the discussion of the counterexamples in section 4 and section 5, it seems that when a well-defined method is used, only the evidence that can be taken as input by that procedure should be incorporated. This will translate to the following first version of the principle:

PTE' If sets d_1 and d_2 contain only data and evidence which can be incorporated by procedure p (conducted to test hypothesis H) and data d_1 are strictly logically stronger than data d_2, then procedure p should be based on d_1 if changing between d_1 and d_2 changes the evidential assessment.

Different procedures may be able to incorporate different data. For example, fingerprint analysis can accommodate features of compared fingerprints such as size or shape and pieces of information concerning conditions in which the prints were deposited. In the case of RCTs, it will be occurrences of tested properties in the control and therapy group, and in the case of meta-analysis, they are results of relevant studies and their statistical properties. Another version of the principle can be reformulated in terms of task-relevancy:

PTE" If sets d_1 and d_2 contain only data and evidence that are task-relevant for procedure p (conducted to test hypothesis H) and data d_1 are strictly logically stronger than data d_2, then procedure p should be based on d_1 if changing between d_1 and d_2 changes the evidential assessment.[4]

Both PTE' and PTE" make the evidence recommended for inclusion dependent on the methodology used. Consequently, they will not recommend using the result of DNA analysis while conducting fingerprint analysis in C, which is a welcome result.

[4]In case of scientific procedures, task-relevant information is information that changes the probability of the collected data (e.g., occurrences of tested properties in the control and therapy group) given the hypothesis (e.g., the tested substance is toxic) is true or false.

We would like to emphasize that the notion of a scientific procedure used in PTE' and PTE" is kept general and informal on purpose in order to give scientists the needed flexibility to cover all types of procedures across the sciences – depending on the specific scientific discipline and task at hand, procedures can vary drastically. What evidence is intended as "input" and, consequently, what information is task-relevant seems to be intuitively clear for many procedures (e.g., a meta-analyses of several RCTs). On the other hand, other methodological details are often not predetermined. For example, the methodology of an RCT is compatible with both Bayesian and classical statistics (see e.g., Brocklehurst and Hoare 2017 or Charlesworth and Carlisle 2022) and, thus, RCTs do not come with a built-in criterion for evidential assessment. Finally, the intended input may be underdefined, and consequently, it is not clear which information will be task-relevant for other scientific procedures (e.g., interviews or case studies). In such cases of underdetermination, PTE' and PTE" may be hard to apply and may produce different results. However, both versions PTE' and PTE" come with restrictions that clearly avoid information giving rise to problematic cases discussed in this paper such as C. A more thorough investigation of when better to apply which of the two principles must await future research.

Before concluding, let us take a step back and have a brief view at the larger picture. Do our arguments show that PTE should generally be replaced by the two alternatives PTE' or PTE"? We believe that the answer to this questions has to be negative. First of all, since PTE' and PTE" are both formulated relative to a procedure p, they can only be applied to cases where a well-defined procedure p is involved in establishing an inference about a hypothesis. Paradigmatic cases are the ones we discussed: fingerprint analysis, DNA analysis, and RCTs. The first point to note here is that this characteristic mark can be used to generalize the domain of problematic cases: For any case where a specific procedure p should be used to make an inference about a hypothesis on the basis of well-defined type of evidence, a hypothetical counterexample like C can be constructed. Thus, in all of these cases one needs to endorse PTE' or PTE" instead of the original PTE. The second point is that it is not obvious that there is a unique and well-defined procedure for every problem. Take the judge's hypothesis of interest, the suspect's guilt, as an example. There seems to be no unique and clearly defined procedure p she is supposed to use and consequently, in line with Carnap's quote from section 2, she should follow PTE and consider all the relevant evidence. The upshot is that when there is no generally accepted and well-defined procedure for a given task, it may make sense to rely on the original PTE. The judge, for example, should take all the results provided to her by forensic reports into account as well as the existence of an alibi, the presence of a clear motive, witness reports, the consistency of the suspect's story, etc.

7 Conclusion

In this paper, we argued that the criminalist's paradox constitutes a counterexample to PTE. We argued that the counterexample is not driven by the defectiveness of the evidence in question. To the best of our knowledge, the criminalist's paradox was not yet discussed in the philosophy of science literature and it may be an interesting case study for other contexts as well, as the problematic cases are clearly not limited to forensic science. Finally, we anticipated a possible objection and proposed two alternative versions of the principle of total evidence relativized to a procedure p. These alternative versions are able to accommodate the counterexample and are consistent with the recommendation from Glymour (2017) that whenever we are dealing with cases involving a well-defined procedure p, what counts as total evidence needs to be restricted to p. They should, however, not replace PTE in general, but only in cases where a unique and well-defined procedure p is involved. In this sense, the original PTE and our two alternative versions can be understood as supplementing each other as principles for scientific inference. Though both can clearly handle the problematic cases, they are not logically equivalent. The possibly relevant differences and implications of these principles as well as the question of when to prefer which one need to be further investigated in future research.

Funding: Funded by the European Union - Next Generation EU, Mission 4, Component 1, CUP I53D23006890001.

Acknowledgements: We would like to thank, among others, Rafał Urbniak, Pedro Naranjo, Antonio Vassallo, Hans Joachim Greif, Frank Zenker, Marcin Koszowy, Jonah N. Schupbach, Jan Sprenger, Matteo Plebani, T.Y. Branch, and anonymous reviewers for their useful comments.

Michał Sikorski
https://orcid.org/0000-0002-6251-0183
Center for Philosophy, Science, and Policy (CPSP)
Marche Polytechnic University
60126 Ancona AN, Italy
michalpsikorski@gmail.com

Alexander Gebharter
https://orcid.org/0000-0003-1708-2949
Center for Philosophy, Science, and Policy (CPSP)
Marche Polytechnic University
60126 Ancona AN, Italy
alexander.gebharter@gmail.com

References

Ashbaugh, D.R. 1999. *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. Practical Aspects of Criminal and Forensic Investigations. Taylor & Francis.

Autzen, Bengt. 2016. “Significance Testing, P-Values and the Principle of Total Evidence.” *European Journal for Philosophy of Science* 6 (2): 281–295.

Barrett, Martin, and Elliott Sober. 2020. “The Requirement of Total Evidence: A Reply to Epstein’s Critique.” *Philosophy of Science* 87 (1): 191–203.

Bird, Carolyne, Kylie Jones, and Kaye N. Ballantyne. 2024. “Cognitive bias and contextual information management: Considerations for forensic handwriting examinations.” *WIREs Forensic Science*.

Bonventre, Catherine L. 2020. “Wrongful convictions and forensic science.” *WIREs Forensic Science*.

Brocklehurst, Paul, and Zoe Hoare. 2017. “How to design a randomised controlled trial.” *BDJ* 222:721–726.

Butt, Leonard. 2013. "The forensic confirmation bias: Problems, perspectives, and proposed solutions—Commentary by a forensic examiner." *Journal of applied research in memory and cognition* 2:59–60.

Carnap, Rudolf. 1947. "On the Application of Inductive Logic." *Philosophy and Phenomenological Research* 8 (1): 133–148.

— . 1950. "Logical Foundations of Probability." *Mind* 62 (245): 86–99.

Charlesworth, Michael, and J B Carlisle. 2022. "How to design and interpret a randomised controlled trial using Bayesian statistics." *Anaesthesia* 78.

Cooper, G., and Vanessa Meterko. 2019. "Cognitive bias research in forensic science: A systematic review." *Forensic science international* 297:35–46.

Cuéllar, Mariano-Florentino, Jacqueline A. Mauro, and Amanda Luby. 2022. "A probabilistic formalisation of contextual bias: From forensic analysis to systemic bias in the criminal justice system." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 185:S620–S643.

Curley, Lee John, James Munro, and Martin Lages. 2020. "An inconvenient truth: More rigorous and ecologically valid research is needed to properly understand cognitive bias in forensic decisions." *Forensic Science International: Synergy* 2:107–109.

Curley, Lee John, et al. 2019. "Assessing Cognitive Bias in Forensic Decisions: A Review and Outlook." *Journal of Forensic Sciences* 65.

Draper, Paul. 2020. "In Defense of the Requirement of Total Evidence." *Philosophy of Science* 87 (1): 179–190.

Elaad, Eitan. 2013. "Psychological contamination in forensic decisions." *Journal of applied research in memory and cognition* 2:76–77.

Epstein, Peter Fisher. 2017. "The Fine-Tuning Argument and the Requirement of Total Evidence." *Philosophy of Science* 84 (4): 639–658.

Fries, James F., and Eswar Krishnan. 2004. "Equipoise, design bias, and randomized controlled trials: the elusive ethics of new drug development." *Arthritis Research & Therapy* 6:R250–R255.

Garrett, Brandon, and Peter Neufeld. 2009. "Invalid Forensic Science Testimony and Wrongful Convictions." *Virginia Law Rev.* 95 ().

Glymour, Clark. 2017. "What Evidence is Total?"

Koehler, Jonathan J, Jennifer L. Mnookin, and Michael J Saks. 2023. "The scientific reinvention of forensic science." *Proceedings of the National Academy of Sciences of the United States of America* 120.

Krane, Dan E., et al. 2008. "Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation." *Journal of Forensic Sciences* 53.

Kukucka, Jeff. 2020. "People who live in ivory towers shouldn't throw stones: A refutation of Curley et al." *Forensic Science International: Synergy* 2 ().

Kunkler, Kimberly S., and Tiffany Roy. 2023. "Reducing the impact of cognitive bias in decision making: Practical actions for forensic science practitioners." *Forensic Science International: Synergy* 7.

Ling, Shichun, Jacob Kaplan, and Colleen M. Berryessa. 2021. "The importance of forensic evidence for decisions on criminal guilt." *Science & Justice* 61 (2): 142–149.

Montaner, Julio S. G., Michael V. O'Shaughnessy, and Martin T Schechter. 2001. "Industry-sponsored clinical research: a double-edged sword." *The Lancet* 358:1893–1895.

Morgan, John. 2023. "Wrongful convictions and claims of false or misleading forensic evidence." *Journal of Forensic Sciences* 68:908–961.

National Commission on Forensic Science. 2016. *Ensuring That Forensic Analysis is Based Upon Task-Relevant Information.*

National Research Council. 2009. *Strengthening forensic science in the United States: A path forward.* 1–328.

Nir, Esther, and Elizabeth Griffiths. 2018. "Sentencing on the Evidence." *Criminal Justice Policy Review* 29 (4): 365–390.

Organization of Scientific Area Committees for Forensic Science. 2017. *Guideline for the Articulation of the Decision-Making Process Leading to an Expert Opinion of Source Identification in Friction Ridge Examinations.*

Peterson, Joseph L., et al. 2013. "Effect of Forensic Evidence on Criminal Justice Case Processing." *Journal of Forensic Sciences* 58.

President's Council of Advisors on Science and Technology. 2016. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods.*

Quigley-Mcbride, Adele, et al. 2022. "A practical tool for information management in forensic decisions: Using Linear Sequential Unmasking-Expanded (LSU-E) in casework." *Forensic Science International: Synergy* 4.

Reichenbach, Hans. 1949. *The Theory of Probability: An Inquiry Into the Logical and Mathematical Foundations of the Calculus of Probability.* Berkeley: University of California Press.

Robertson, C. T., and A. Kesselheim. 2016. "Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law."

Roush, Sherrilyn. 2020. "Epistemic Justice and the Principle of Total Evidence."

Sikorski, Michał. 2022. "Is Forensic Science in Crisis?" *Synthese* 200 (3): 1–34.

Smit, N., R. Morgan, and D. Lagnado. 2018. "A systematic analysis of misleading evidence in unsafe rulings in England and Wales." *Science & justice : journal of the Forensic Science Society* 58 2:128–137.

Stevenage, Sarah V., and Alice Bennett. 2017. "A biased opinion: Demonstration of cognitive bias on a fingerprint matching task through knowledge of DNA test results." *Forensic science international* 276:93–106.

Thompson, William. 2010. "What Role Should Investigative Facts Play in the Evaluation of Scientific Evidence." *Australian Journal of Forensic Sciences* 43 ().

Thompson, William C. 2020. "Commentary on: Curley LJ, Munro J, Lages M, MacLean R, Murray J. Assessing cognitive bias in forensic decisions: a review and outlook. J Forensic Sci doi: 10.1111/1556-4029.14220. Epub 2019 Nov 6." *Journal of Forensic Sciences* 65.