

Decision-theoretic and risk-based approaches to naked statistical evidence: some consequences and challenges

RAFAL URBANIAK[†]

University of Gdańsk, Jana Bazynskiego 4, 80-309, Gdańsk, Poland

ALICJA KOWALEWSKA

Gdańsk University of Technology, 80-233 Gdańsk, Poland

AND

PAVEL JANDA AND PATRYK DZIUROSZ-SERAFINOWICZ

University of Gdańsk, Jana Bazynskiego 4, 80-309, Gdańsk, Poland

[Received on 27 December 2019; accepted on 12 January 2018]

In the debate about the legal value of naked statistical evidence, Di Bello argues that (1) the likelihood ratio of such evidence is unknown, (2) the decision-theoretic considerations indicate that a conviction based on such evidence is unacceptable when expected utility maximization is combined with fairness constraints, and (3) the risk of mistaken conviction based on such evidence cannot be evaluated and is potentially too high. We argue that Di Bello's argument for (1) works in a rather narrow context, and that (1) is not exactly in line with the way expert witnesses are required to formulate their opinions. Consequently, Di Bello's argument for (2), which assumes (1), does not apply uniformly to all convictions based on naked evidence. Moreover, if Di Bello's analysis is correct, it applies also to eyewitness testimony, given empirical results about its quality, and so the distinctions drawn by DiBello cut across the distinction between naked statistical evidence and other types of evidence. Finally, if we weaken the rather strong requirement of precise measurability of the risk of mistaken conviction, to the availability of reasonable but imprecise and fallible estimates, many field and empirical studies show that often the risk of mistaken conviction based on naked statistical evidence can be estimated to a similar extent as the risk of mistaken conviction based on any other sort of evidence.

Keywords: naked statistical evidence; likelihood ratio; risk of mistaken conviction.

1. Introduction

In the development of a probabilistic approach to the necessary and sufficient conditions for binary rational beliefs, one traditional candidate, called the *Lockean thesis*, is that one should believe a given proposition just in case the probability of this proposition (given all available evidence) is

[†]Email: rfl.urbaniak@gmail.com

higher than a certain threshold value. But this, as it is well-known, leads to some problems, such as the ‘lottery paradox’ or the ‘preface paradox’.

An analogous view among some legal evidence scholars is the ‘threshold-based legal probabilism’ (TLP), according to which a court’s decision regarding a given fact is justified just in case the probability of this fact given the total evidence available is above a certain threshold.¹

Conceptual problems with TLP are to some extent analogous to epistemological paradoxes that result from attempts to reduce binary belief to degrees of belief. Quite a few difficulties for TLP have been put forward by Cohen (1977). One particular problem, which will be in focus of this article, is the notoriety of the task of indicating when statistical evidence is sufficient for conviction. It seems as tricky as explicating sufficient and necessary conditions for justified rational belief in probabilistic terms.

The challenge can be introduced by means of the following example. Consider the ‘prisoners problem’ (Nesson, 1979; Redmayne, 2008). There are 100 suspects, and we know that exactly 99 of them are responsible for an attack on a guard in the prison yard, and 1 person is innocent, but we cannot identify the innocent prisoner. Pick a random prisoner from the 100 suspects. Should we convict them? If we follow TLP in considering conviction as justified just in case the probability of guilt given the evidence is sufficiently high, purely statistical evidence available in this case should be sufficient for conviction.² Yet, even though the probability of guilt is high, the intuition that many share is that we should not convict without further evidence.³

The intuition that purely statistical evidence is somehow deficient is quite pervasive: judges are not usually willing to sentence based solely on evidence of this sort. This reluctance to find liability has been empirically investigated using various variants of the so-called ‘Blue Bus problem’ (Wells, 1992). Short stories, in which a dog was run over by a bus, and pieces of evidence of various types were available, were given to people and they were asked to decide whether the Blue Bus company was to be held liable. Although almost the same subjective probabilities were ascribed in various scenarios to the liability claim, the verdicts were significantly different. The subjects (including actual judges) who received purely statistical evidence based on the frequency of Blue Buses in the city, were much less likely to convict compared to those who were provided with testimonial evidence, despite declaring the same level of subjective belief in the company’s liability. This leads to questions: why is high probability sometimes not considered sufficient to form a decisive judgement? Is it justly so? If there is an additional requirement on evidence sufficient for conviction that purely statistical evidence fails to satisfy, what is it?

The issue surfaces in various contexts. Consider a scenario in which a police officer knows that 99% of black residents in the neighbourhood have open arrest warrants.⁴ On a patrol he sees a black resident and believes that they have an open warrant. The belief seems to be highly probable, and a

¹ The assumption that the decision should be based on probabilities seems to be present in some legal decision rules. For instance, in civil cases in the Anglo-Saxon tradition, the case is said to be judged based on the preponderance of probabilities. It is somewhat less obvious whether the criminal standard of ‘proof beyond reasonable doubt’ involves merely probabilistic considerations, but the question of what else would it involve otherwise is notoriously difficult. So, often it is equated with the requirement that the probability of guilt given the evidence should be sufficiently high.

² If you think the threshold should be higher than 0.99, increase the numbers in the thought experiment.

³ One might think that naked statistical evidence is not really ‘evidence’, but rather part of background knowledge or reference-class fixing information. Indeed, we grant Di Bello’s assumption (which is quite common in the literature) that it is some sort of evidence, at least in a wide sense, and goes from there.

⁴ This is a modification of a case discussed in Basu (2018).

police officer acting on it might seem rational. However, at the same time, there seems to be something morally wrong in him doing so.

Many attempts to describe the missing component responsible for the differences in verdicts in spite of equal probabilities have been made without much agreement as to what the solution is.⁵ Most of them employed features of evidence that have not been explicated in probabilistic terms. For instance, evidence is claimed to be insufficient for conviction if it is not sensitive to the issue at hand: if it remained the same even if the accused was innocent (Enoch and Fisher, 2015). Or, to look at another approach, evidence is claimed to be insufficient for conviction if it does not normically support it: if—given the same evidence—less explanation would be needed if the accused was innocent than if they were guilty (Smith, 2017).

A Bayesian epistemologist, if she wants to remain one and agrees that at times strong, but purely statistical evidence indeed is insufficient for conviction, not only needs to indicate what makes evidence sufficient for conviction, but also to make sure that this explication is formulated in probabilistic terms. A recent promising attempt at achieving this has been developed by Di Bello (2019).

In Section 2, we present Di Bello's argument for the claim that the probative value of naked statistical evidence cannot be evaluated. In Section 3, we argue that the applicability of Di Bello's argument is limited. In Section 5, we describe Di Bello's decision-theoretic analysis, and criticize it in Section 6. In Section 7, we look at Di Bello's argument to the effect that his criticism does not apply to eyewitness testimony, and argue otherwise. Finally, in Section 8, we argue that if we weaken the rather strong requirement of precise measurability of the risk of mistaken conviction to the availability of reasonable but imprecise and fallible estimates, many field and empirical studies show that often the risk of mistaken conviction based on naked statistical evidence can be estimated to a similar extent as the risk of mistaken conviction based on any other sort of evidence.

2. The unknown likelihood ratio argument

Let us use the prisoners problem to discuss and evaluate the reason why Di Bello thinks that in that scenario we do not know the probative value of purely statistical evidence (and so a conviction is unjustified). Following Di Bello (2019, p. 5), we say that evidence E is probative of guilt over innocence provided the probability of E is higher given the hypothesis of guilt than given the hypothesis of innocence. Just because our evidence is probative, it does not mean we should convict. E can raise the probability of guilt only slightly, or E can be purely statistical evidence, which—if you share Di Bello's intuition—should not lead to conviction. So how do we measure the probative value of E in probabilistic terms?

One commonly proposed measure of probative value is the 'likelihood ratio' (see e.g. Finkelstein (2009) for a discussion): $P(E|G)/P(E|I)$. It is the ratio of the conditional probability of the evidence on the assumption of guilt to the conditional probability of the evidence on the assumption of innocence. It indicates which hypothesis supports the resulting evidence E better. If the ratio is greater than 1, then G makes the occurrence of E more likely than I : E is probative of guilt over innocence. If the ratio is lower than 1, I makes E more likely than G , and so E is probative of innocence over guilt. The likelihood ratio not only tells us whether the evidence is probative of guilt, its value also tells us how strongly probative it is. The higher the ratio (above 1), the stronger the support.

⁵ See Redmayne (2008) for a critical survey.

Now, to convict based on evidence, arguably, we should know what its probative value is. Intuitively, one needs to understand the evidence that one uses to convict, and if one does not know the likelihood ratio, one does not understand the extent to which the evidence contributes to the case, and so does not understand something important about the evidence.⁶ Di Bello argues that in the prisoners problem (and in cases with purely statistical evidence in general), it is impossible to evaluate the likelihood ratio. This epistemic deficiency of purely statistical evidence makes conviction unjustified. First, he expresses the likelihood ratio in terms of the ratios of prior and posterior probability distributions as follows (the rightmost identity is due to Bayes' Theorem):

$$LR(E|G) = \frac{P(E|G)}{P(E|I)} = \frac{P(G|E)}{P(I|E)} \times \frac{P(I)}{P(G)}. \quad (1)$$

Di Bello then says that in the prisoners problem, we can determine the posterior ratio $P(G|E)/P(I|E)$ easily; it happens to be 99 (Di Bello, 2019, p. 7). This, however, because we have not evaluated the probative value of the evidence so far, does not suffice for conviction.

Di Bello further argues that the lack of background information does not allow us to determine the prior probabilities of guilt and innocence and so the prior ratio remains unknown (Di Bello, 2019, p. 7). Therefore, by (1), we do not know the value of the likelihood ratio, and consequently, cannot justifiably convict.

3. The argument for unknown likelihood ratio fails

This, however, seems hasty. Is it really the case that always just because we cannot find the prior odds, we also cannot find the likelihood ratio? In general, if you know that $a = b \times c$ and you know b , but not c , it does not follow that you do not know a . It would follow, if your only way of calculating a was to multiply b by c . But in the present case, it is at least not obvious that this assumption holds.

For instance, 'Law Commission's 2011' report, *Expert Evidence in Criminal Proceedings in England and Wales* explicitly recommends that an expert witness must report a reasoned opinion on the likelihood of the item of evidence under one or more alternative propositions. At the same time, the expert is expected to withhold their opinion (if any) about the priors. In general, while the assessment of likelihood ratio in principle lies within the competence of an appropriate expert witness, their testimony should neither be based on nor reveal any opinion about the priors. Here is a representative selection of recommendations from Aitken et al. (2010) on this topic:

- For reasons that will become more apparent as we proceed, it is often illuminating and sometimes essential to express the extent to which evidence supports a particular proposition relative to another proposition in terms of the ratio of two likelihoods: (1) the likelihood of the evidence if one proposition is true; and (2) the likelihood of the evidence if the other proposition is true. (p. 35)
- This is the type of question [about the priors] which fact-finders, rather than expert witnesses, should be left to resolve in contested criminal trials. (p. 38)
- Expert witnesses must not trespass on the province of the jury by commenting directly on the accused's guilt or innocence, and should generally confine their testimony to presenting the likelihood of their evidence under competing propositions. (p. 42)

⁶ Di Bello later on connects the knowledge of likelihood ratio with understanding the risks of mistaken decisions. We will get to this argument later on.

- ... evidence evaluation is always a fundamentally comparative enterprise. At all levels of proposition, the scientist needs to consider the likelihood ratio for the evidence. (p. 59)
- Ideally, expert witnesses should testify to the likelihood of the evidence under two competing propositions (or assumptions), the prosecution's proposition and the competing proposition advanced by the defence (which may simply be the negation of the prosecution's proposition in the absence of fuller pre-trial defence disclosure). (p. 83)

If, as Di Bello's argument assumes, the only way to form an opinion about likelihood ratio was to rely on priors, this would mean expert witnesses are systematically asked to achieve the impossible.⁷

So how do expert witnesses come to estimate likelihood ratios if not in the way Di Bello imagines? For instance, [Sottas et al. \(2008\)](#) discuss the interpretation of blood doping markers, and clearly make the distinction that 'the prior on the blood doping variable is obtained from population statistics, whereas a likelihood ratio is derived from individual test results of the marker' (p. 198). Another example is the interpretation of shoe marks from crime scenes, discussed by [Lucy \(2013\)](#). What is the probability that the shoe marks would fit the suspect's if he left the trace? Very close to 1. What's the probability that they would fit if someone else left the shoe mark? Well, here we need to know more about the distribution and the frequency of shoes of the relevant type in a given region. While the latter estimate is tricky, none of the estimations require the expert to know the prior probability of the suspect's leaving the trace.

Another example, also discussed by Lucy, is that of blood type evidence in New Zealand. Say the blood sample is of type 0. Since the frequency of this type of blood in New Zealand is 68.3%, the likelihood ratio is $1/0.683 \approx 1.464$ (1 in the numerator comes from the probability of blood type match if the suspect was the source). On the other hand, if the blood type is B, the likelihood ratio is $1/0.063 \approx 15.87$, since the frequency of this type is 6.3%. Again, the estimation does not require the knowledge of the prior probability of the suspect being the source of the sample.

Notice also that Di Bello's resolution, if successful, would work only against statistical evidence in cases in which no reasonable estimate of the prior probability of guilt can be made. This, perhaps, is true in a made-up thought experiment, but it is far from obvious that the same holds for real-life cases (for a recent proposal, see [Fenton et al. \(2019\)](#)).

4. A remark on background information

Di Bello argues that the probative value of the 99 : 1 statistics, as measured by the likelihood ratio $P(\text{stat}[99 : 1] | G \wedge B) / P(\text{stat}[99 : 1] | I \wedge B)$, varies depending on background information B , and since in this case we do not know which B holds, we cannot estimate the likelihood ratio as well. For instance, if we knew that the prisoners, whenever a prison guard is killed, invariably commit a collective killing with a participation ratio of 99 : 1, the likelihood ratio would be 1 and the statistical evidence would not be probative.

However, one can reasonably argue that some pieces of background information should *not* be taken into account when assessing the likelihood that some hypothesis confers on some evidence. This in turn indicates that including background information in the estimation of likelihood in the prisoners scenario might have been too hasty.

⁷ See the treatment of likelihood in [Aitken et al. \(2010\)](#) or [Finkelstein \(2009\)](#).

Why should we think that the likelihood ratio for any pair of hypotheses depends on background information? Consider the following example due to Eddington (1939). Suppose you use a net to fish in a lake and observe that all the fish in the net are large (E). At first blush, this evidence seems to favour the hypothesis that the lake consists of mostly large fish (L) over the hypothesis that it consists of mostly small fish (S). So it seems that we have that $P(E|L) > P(E|S)$ and the likelihood ratio, $P(E|L)/P(E|S)$, is greater than 1. But then suppose that you look down at your net and learn that the holes in your net are too large to catch small fish (B). So you realize that you were bound to obtain the evidence E , regardless of whether L or S is true. Hence, to ignore the background information B would seem to make your assessment that $P(E|L) > P(E|S)$ epistemically defective. For if we take B into account, then it should be that $P(E|L \wedge B) = P(E|S \wedge B)$. After all, you are just as likely to catch a large fish in your net regardless of whether L or S is true.

It seems that the above supplies a compelling reason for why a hypothesis's likelihood should depend on background information. But one might go further and think that cases like the Eddington's fish show that, when assessing the likelihood of a given hypothesis, we should take into account 'any' background information that is available to us. But, as it has been argued by Weisberg (2005) and Kotzen (2012), this way of thinking can lead to some unwelcome consequences. To see their point, let us get back to Di Bello's analysis of the prisoners scenario. Di Bello (2019, p. 7) illustrates the variability of the likelihood ratio by considering various possible information that might be included in B . In particular, he considers the possibility that B says that the prisoners, whenever a prison guard is killed, invariably commit a collective killing with a participation ratio of 99 : 1. But if we include this information in the assessment of G 's and I 's likelihood, then $\text{stat}[99 : 1]$ is no evidence for G over I , for $P(\text{stat}[99 : 1]|G \wedge B) = P(\text{stat}[99 : 1]|I \wedge B) = 1$. After all, we are certain that ninety-nine of one hundred prisoners would kill collectively the guard, regardless of whether G or I is true. But this point generalizes: if we include any information in B which entails $\text{stat}[99 : 1]$, it follows that $\text{stat}[99 : 1]$ would never be any evidence for 'any' hypothesis (and so not only G) over 'any other' hypothesis (and so not only I). However, estimating the likelihood of a piece of evidence by considering a potential background assumption which we do not really have a reason to believe and which entails the hypothesis under consideration seems rather hasty. In general, the question that pieces of background knowledge should be considered in the process of likelihood estimation is non-trivial (non-existent background knowledge entailing the evidence being one extreme example of the phenomenon). So the likelihood in the prisoners scenario might not be as sensitive to background information as Di Bello would like it to be and an independent argument for the admissibility of the pieces of background knowledge that he considers for the task at hand is needed. Let us now turn to Di Bello's another argument.

5. The decision-theoretic argument

Di Bello's decision-theoretic argument for the insufficiency of naked statistical evidence for conviction relies on a well-known decision-theoretic perspective: one should choose an action that minimizes the expected cost or disutility. On this approach, we should prefer conviction to acquittal if the expected cost of convicting an innocent defendant is strictly lower than the expected cost of acquitting a guilty defendant.

Let us think about the prisoners problem as a decision situation in more detail. We seem to have two possible states of the world. The defendant is either guilty or innocent. We also have two possible actions. We can either acquit, or we can convict the defendant. We can then write the payoff

matrix in terms of the costs as follows; let $cost(AG)$ be the cost of acquitting a guilty defendant, $cost(CI)$ the cost of convicting an innocent defendant, and so on.

	Innocent	Guilty
acquit	$cost(AI)$	$cost(AG)$
convict	$cost(CI)$	$cost(CG)$

Di Bello works with the assumption that the costs of acquitting an innocent defendant and the cost of convicting a guilty defendant are 0. We will work with this assumption. This will allow us, in the expectation formula, to consider only $cost(AG)$ and $cost(CI)$.

Di Bello uses standard expectation methods, so we also need probabilities to find the expected cost of each action. The probabilities that he uses are $P(G|E)$ and $P(I|E)$, which are the probability of guilt and innocence given the evidence E , respectively. We should then prefer conviction to acquittal if the expected cost of convicting an innocent defendant is strictly lower than the expected cost of acquitting a guilty defendant, that is, when:

$$P(I|E) \times cost(CI) + P(G|E) \times 0 < P(G|E) \times cost(AG) + P(I|E) \times 0. \quad (2)$$

The inequality (2) can be easily rewritten to the following form; we suppose that costs are positive numbers,⁸ so the inequality sign does not change:

$$\begin{aligned} P(I|E) \times cost(CI) &< P(G|E) \times cost(AG) \\ \frac{cost(CI)}{cost(AG)} &< \frac{P(G|E)}{P(I|E)}. \end{aligned} \quad (3)$$

We reached (3), which is the formula with which Di Bello starts his decision-theoretic argument.

Once we fix the cost ratio, we have a threshold for conviction and to decide whether to convict, we now need to determine the ratio of posterior probabilities. This task, according to Di Bello, proves to be very difficult, because we usually cannot directly assess those—the prisoners scenario is an unusual case in which the posterior probability is simply given, but in more realistic scenarios, we need to evaluate the evidence before we estimate the posteriors. Di Bello uses Bayes' Theorem to rewrite the posterior probabilities in terms of likelihoods and prior distributions:

$$P(I|E) = \frac{P(E|I)P(I)}{P(E)} \text{ and } P(G|E) = \frac{P(E|G)P(G)}{P(E)} \quad (4)$$

Using (4), we can rewrite (3) to avoid the need to directly determine the posterior probabilities:

$$\frac{cost(CI)}{cost(AG)} < \frac{P(E|G)}{P(E|I)} \times \frac{P(G)}{P(I)}. \quad (5)$$

The question now is how to find, or at least estimate, the likelihood ratio and the prior ratio. Let us start with the prior ratio. After going through a couple of suggestions for the prior ratio and reasons why they are unsatisfactory for the task, Di Bello decides to make the following step:

⁸ This is just a slightly unusual way of speaking, on which the greater the positive cost, the greater the negative impact of the action.

... my suggestion is that we move away from the epistemic task of finding out the accurate value of the prior ratio—a task that, we have seen, is riddled with difficulties—and turn instead to the legal task of respecting the normative constraints that make certain prior ratios acceptable and others unacceptable. (Di Bello, 2019, p. 14)

The way Di Bello proposes to accomplish this switch from a purely epistemic task to a legal one is by changing the decision rule (5). Instead of the epistemically motivated prior ratio, he proposes to use a non-specific low value k , so that the decision rule now looks as follows:

$$\text{convict iff } \frac{\text{cost}(CI)}{\text{cost}(AG)} < \frac{P(E|G)}{P(E|I)} \times k. \quad (6)$$

The normative reason for choosing a low k is to secure the fair application of the presumption of innocence and provide strong protection to defendants against false conviction. In other words, the smaller k is, the more difficult it is for the right side of (6) to be greater than the left side, and by the same token, to convict the defendant.

The last element of the inequality that needs to be determined is the likelihood ratio $P(E|G)/P(E|I)$. Di Bello admits that, in many cases, the likelihood ratio can be estimated, for example, by cross-examining an eyewitness or with the use of statistical data accompanying DNA-based evidence. According to Di Bello, however, this is not the case for purely statistical evidence. He believes that the value of the ratio for purely statistical evidence remains unknown. Consequently, we do not know whether the right side of the inequality is greater than the left side (especially, given that we assume k to be low). Therefore, using the decision rule (6), we have no grounds to convict the defendant with purely statistical evidence, for example, in the prisoners problem.

Working with the claim that the likelihood ratio for purely statistical evidence is unknown, Di Bello relies on the earlier argument, which we have already discussed. At this stage, let us just note that the decision-theoretic argument is parasitic on the likelihood ratio argument: if the latter falls, so does the former. Having said this, let us look at further details.

6. Decision-theoretic argument and unknown likelihood

For the sake of brevity and clarity, we will concentrate on the application of Di Bello's approach to the prisoners problem. Let us start by re-writing (6) for this particular scenario:

$$\text{convict iff } \frac{\text{cost}(CI)}{\text{cost}(AG)} < \frac{P(\text{stat}[99 : 1]|G)}{P(\text{stat}[99 : 1]|I)} \times k. \quad (7)$$

Notice that (7) says: convict iff the inequality holds. Yet, this is not the rule that Di Bello uses. According to him, in the prisoners scenario, we acquit not because we know the inequality does not hold. We acquit because we do not know if it does. Instead of $P(\text{stat}[99 : 1]|G)/P(\text{stat}[99 : 1]|I)$, let us write $LR(\text{stat}[99 : 1])$ (it is still a conditional likelihood ratio, but we suppress G and I in the background). To match Di Bello's move, the decision rule should be changed as follows:

$$\text{convict iff it is known that } \frac{\text{cost}(CI)}{\text{cost}(AG)} < LR(\text{stat}[99 : 1]) \times k. \quad (8)$$

This rule, however, does not follow from purely decision-theoretic considerations. Perhaps, it can be motivated by some considerations regarding the presumption of innocence, but a more explicit argument would be useful.

Let us concede this point and let us assume that we are to follow (8). Even if we agree with Di Bello that the exact value of the likelihood ratio cannot be established, does it follow that we should not convict? Well, $\text{cost}(CI)/\text{cost}(AG)$ is just some constant, say c , so the question is: do we need to know the exact value of $LR(\text{stat}[99 : 1])$ to know that $LR(\text{stat}[99 : 1]) > c$? Again, this seems too hasty. Say I meet Marcello Di Bello on the street. I can tell that his height is >1.5 m without knowing how tall exactly he is. In general, from the fact that I do not know the exact value of a random variable X it does not follow that I do not know whether $X > t$ for some threshold t . So even if we accept (8) and we agree that the precise value of $LR(\text{stat}[99 : 1])$ is unknown, it still does not follow that we do not know whether $LR(\text{stat}[99 : 1]) > k$ in the prisoners scenario. What needs to be excluded is that in this scenario (and, in real-life scenarios employing statistical evidence) no imprecise estimate of the likelihood ratio can be made to the effect that it is greater than the relevant threshold.

Moreover, the prisoners scenario, especially on Di Bello's proposal, is rather unusual. Normally, the reason why we would like to know the likelihood of evidence is because we already have some priors and want to find out what the posterior probabilities are. In the prisoners scenario it is agreed that we know what the posterior odds are: the 99:1, and the question is whether this information (based on statistics) is sufficient for conviction. Di Bello disposes with the prior probability, proposing to replace it with some small constant k , and insists that even if we know the posterior probability, we need to find out the likelihood ratio to estimate the risk of a mistaken conviction. Let us play along, say k is $1/1000$. Is it really the case that we cannot estimate $LR(\text{stat}[99 : 1])$ in this setting?

Let us rewrite (1), replacing the reciprocal prior odds with the reciprocal of k , and the posterior odds with the ones which are assumed in this thought experiment. Since we are using the reciprocal of k instead of the actual prior, let us call this ' LR' ' rather than ' LR '.

$$LR'(\text{stat}[99 : 1]) = \frac{99}{1} \times \frac{1000}{1} = 99\,000. \quad (9)$$

So, if we are to take the proposal seriously and use k instead of prior odds, if we agree that the posterior probability in the prisoners scenario is as described in the thought experiment, we not only can find the likelihood ratio, but also we should think that the evidence is very highly probative.

One could try to reply that given that we disposed of the real priors, this is not the 'real' likelihood ratio, and that even though we should use k in our decision criterion, we should still use real priors when thinking about the actual likelihood ratio. Fair enough. Then, by (1) we get:

$$LR(\text{stat}[99 : 1]) = \frac{99}{1} \times \frac{1}{\text{Odds}(G)} \quad (10)$$

This makes $LR(\text{stat}[99 : 1])$ a function of $P(G)$. Let us concede we do not have a legitimate point estimate of $P(G)$. Still, it is quite clear that high prior probability of guilt would be unreasonable. If $P(G) \leq 0.99$, it follows that $LR(\text{stat}[99 : 1]) \geq 1$. Here are some examples of how $P(G)$ is connected with the likelihood ratio:

$P(G)$	$LR(stat[99 : 1])$
0.01	9801.00
0.10	891.00
0.25	297.00
0.50	99.00
0.75	33.00
0.95	5.21
0.99	1.00

The function is illustrated in Figs. 1 and 2.

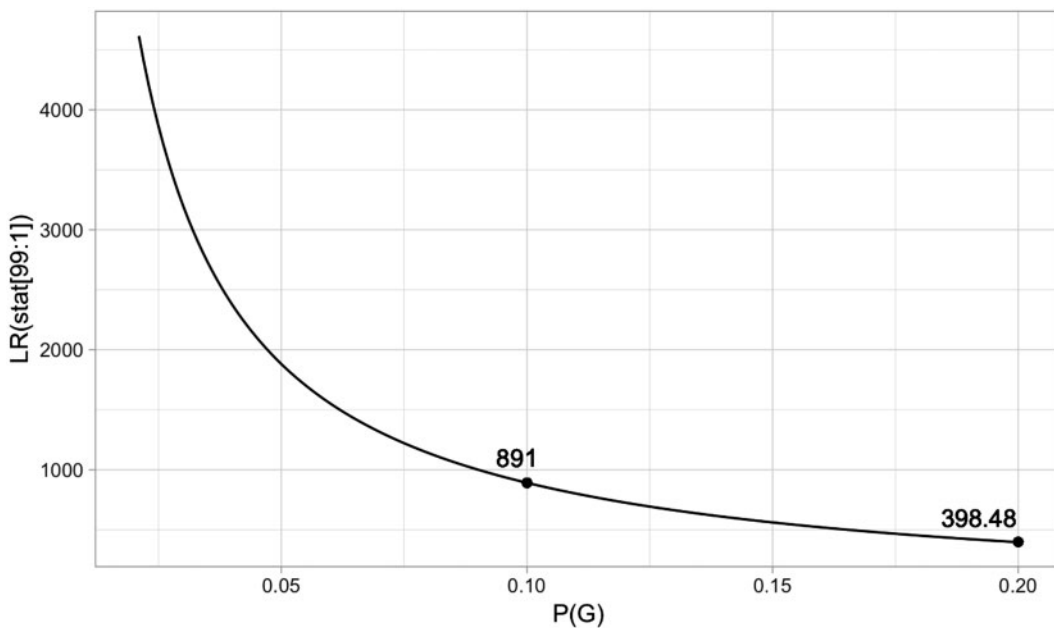


FIG. 1. LR in *prisoners* as a function of $P(G)$ for cases where $5000 > LR > 200$.

So, on one hand, if one accepts the set-up in which the posterior probability of guilt is 0.99, any low value of the prior will make the evidence highly probative. Perhaps, we cannot know for sure the exact likelihood ratio, but the considerations above indicate that we at least should expect it to be very high.

On the other hand, if we want both $LR(stat[99 : 1]) \leq 1$ and $P(G)$ to be low, we cannot consistently claim that the posterior odds are high. For instance, if $LR(stat[99 : 1]) = 1$ and $Odds(G) = 1/1000$, the posterior is 0.001. And in general, by Bayes' theorem, if $LR(stat[99 : 1]) \leq 1$, the posterior odds cannot exceed the prior odds.

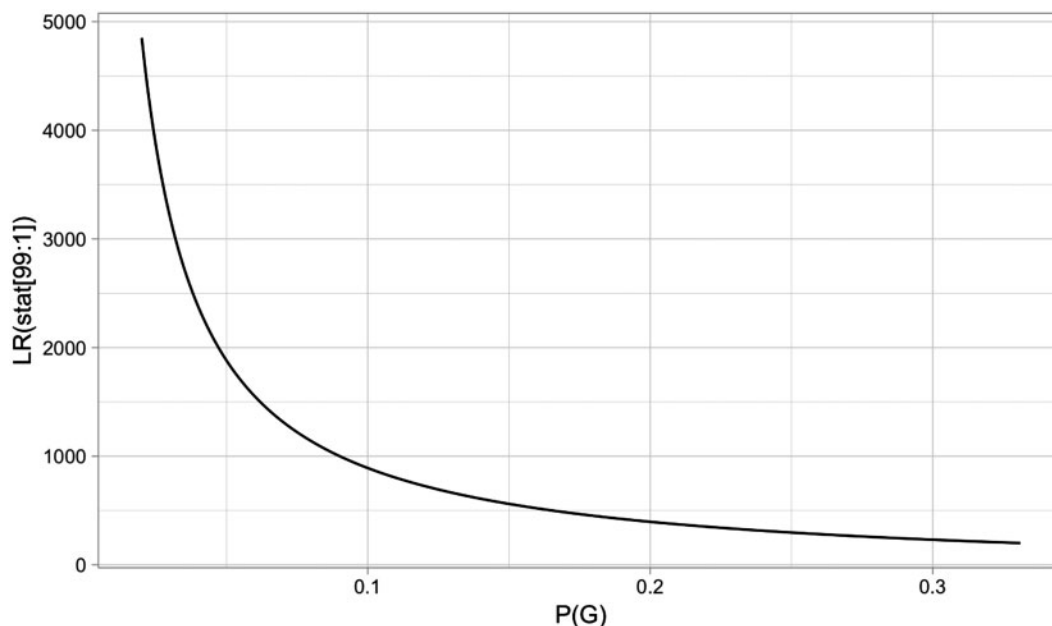


Fig. 2. LR in *prisoners* as a function of $P(G)$ for $0.2 > P(G) > 0.02$.

7. Eyewitness testimony

One may say that the prisoners scenario is a very unusual case. Let us turn to a more mundane example—we will look at how this approach handles eyewitness evidence. Di Bello suggests that in many occasions, eyewitness testimony, after a cross-examination which succeeds at making the probability of a misleading testimony ($P(witness_G|I)$) sufficiently low, might suffice for conviction:⁹

Once the low constant k and the high cost ratio $cost(CI)/cost(AG)$ are fixed, a conviction based on eyewitness testimony alone is acceptable provided the likelihood ratio $P(witness_G|G)/P(witness_G|I)$ is suitably high. My claim is that unlike the 99:1 statistics, whose probative value is unknown, the probative value of eyewitness evidence (as measured by the likelihood ratio) is not unknown, and is in fact relatively high in a large number of circumstances. [...] Given this assumption, there is still some variability in the likelihood ratio, but in many cases the ratio will be fairly high [...]. The important point is that this variability is not left entirely unchecked by the trial system, because the process of cross-examination is meant to distinguish good from bad eyewitness testimonies, those with low susceptibility to a false positive identification from those with high susceptibility to error. This is where eyewitness evidence differs from the 99:1 statistics.

⁹ Here, $witness_G$ reads: ‘the eyewitness testified to the effect that the defendant is guilty’.

Let us think about this claim working with a particular choice of numbers. According to Di Bello (2019, p. 11), on some occasions in criminal trials it is fair to assume that $\frac{\text{cost}(CI)}{\text{cost}(AG)} = 9$ and that, in general, high ratio of costs ‘encodes the moral demand that the system should avoid mistaken convictions even at the cost of allowing more mistaken acquittals’. The possible values that Di Bello (2019, p. 29) considers for $P(\text{witness}_G|I)$ are between 0.01 and 0.04. We will take their average and assume $P(\text{witness}_G|I) = 0.025$. Then, Di Bello assumes that $P(\text{witness}_G|G) > 0.5$. Again, we will take an average and assume $P(\text{witness}_G|G) = 0.75$. About the value of k we only know that it should be low ($0 < k < 1$). For the sake of example, we will make an assumption that $k = 1/9$, which corresponds to $P(I) = 0.9$. With all of those assumptions in mind if we substitute the values for variables in (6), we obtain:

$$\begin{aligned} 9 &< \frac{75}{2.5} \times 1/9 \\ 9 &< 3.3333 \dots \end{aligned} \tag{11}$$

This is obviously false, so contrary to the claim, Di Bello’s approach seems to entail that the eyewitness testimony does not suffice for a conviction. If the value of k was lower, it would be even more difficult to satisfy the inequality.¹⁰

In the argument, we followed Di Bello in assuming that $P(\text{witness}_G|I) < 0.04$. The assumption seems to be supported by a feature that distinguishes eyewitness testimony from naked statistical evidence—the possibility of cross-examination:

Cross-examination elicits information about the lighting and environmental conditions; it tests the memory of the witness; it attempts to detect biases the witness might have against the defendant; and so on [...]. Cross-examination can be viewed as a method of ensuring that the probability $P(\text{witness}_G|I)$ is kept suitably low for testimonies that survive cross-examination [...]. This prompts the question whether cross-examination is actually an effective method. This is an empirical question. Suppose, for the sake of argument, it turned out that cross-examination was no better than chance at testing the susceptibility to error of eyewitness testimonies. If that were true, we should reconsider our intuitions. (Di Bello, 2019, pp. 26–29)

Unfortunately, both the extremely low value of $P(\text{witness}_G|I) < 0.04$ and the claim that cross-examination works as advertised are unrealistic. Field studies indicate filler identification rates of 20–24% (Klobuchar *et al.*, 2006). In the USA, 4.1%, which is close to the upper limit Di Bello used for eyewitness identification, is a conservative estimate of false death sentence convictions (and those are based on much stronger and multiple pieces of evidence, not a single eyewitness’ testimony [Gross *et al.*, 2014]). A study of 340 exonerations in years 1989–2003 indicates that around 90% of false convictions for rape (and pretty much all inter-racial rape mistaken convictions) are based on eyewitness misidentification, and in 43% of false convictions for murder the defendant was misidentified by one or more eyewitnesses. How often exactly in real cases errors are made by eyewitnesses is very difficult to evaluate, but the abundance of such exonerations should make us suspicious of eyewitness testimony, even though in real cases these witnesses have been cross-examined.

¹⁰ Even if the probability that a witness would testify the way they did if the accused was guilty was 0.95 (and we will discuss some reasons to think it is that high in certain contexts later on), and the probability that a witness would testify against the accused if in fact they are not with probability 0.05, the approach would still require $9 < 0.95/0.05 \times 1/9 = 2.[1]$.

A study of lineups in 1561 witnesses and 616 suspects in real cases in Greater London (Wright and McDaid, 1996) and of 689 identification attempts in 271 real identification cases in Sacramento (Behrman and Davey, 2001) suggest false positive rate of around 20%. Also in experimental setting (in which the main difference is that the witnesses are less emotionally affected by the crime), eyewitnesses identify a known wrong person (a ‘filler’ or ‘foil’) in approximately 20% of all real criminal lineups (Thompson, 2007). Moreover, studies of cross-examination failed to show that it improves accuracy and that there is a clear relation between witness’ confidence and accuracy. In a series of experiments, subjects were asked to cross-examine eyewitnesses to determine whether witnesses made accurate or mistaken identifications. Subjects have shown little or no ability to make such discriminations (Wells and Olson, 2003). For instance, in an experiment by Lindsay *et al.* (1981), a representative sample of 48 witnesses was cross-examined. Subjects ($n = 96$) viewing the cross-examinations showed no ability to accurately distinguish false-identification witnesses within conditions as measured by subjects’ belief of witnesses.

But what about the additional information that, according to Di Bello, cross-examination can elicit? Isn’t it useful? Well, to some extent. Several of the eyewitness quality issues have been studied by means of experimental methods. Some of them are systemic variables (simultaneous/sequential lineups, showups,¹¹ presence of prior identification, culprit present/culprit absent, frequency of witnesses per suspect), some of them—estimator variables—are not (the effect of delay, cross versus own-race effects, weapon focus effects, presence of violence)—see Behrman and Davey (2001) and Wells and Olson (2003) for an interesting discussion of such factors. Obtaining information about these factors might help us exclude some sources of error. But once this is done, the research suggests, further cross-examination does pretty much nothing to improve accuracy.

Last, but not least, the situation is even more complicated than the received picture in the psychological literature would suggest, as a fascinating meta-analysis by Wixted and Wells (2017) indicates. Let us go over the main points they bring up.

- In light of the empirical results that we have discussed and results similar to them, the justice systems have grown more suspicious of eyewitness evidence over the last 20 years or so, especially doubting whether the eyewitness confidence is predictive of accuracy.
- Isolating cases in which the identification has been made ‘in pristine conditions’ with high ‘initial’ confidence focusing on cases which would go to trial, that is, in which the suspect indeed was identified with high confidence, Wixted and Wells (2017) argue that the data show that in such circumstances, the eyewitness confidence indeed is highly predictive of accuracy.
- Pristine conditions involve a double-blind experiment containing only one suspect, at least five fillers, with no fillers who do not look like the perpetrator at all, from among which the suspect does not stand out as obviously fitting the description which the eyewitness is familiar with, with no extreme coincidental resemblance of a filler to the suspect. The witness has to be cautioned that the offender might not be in the lineup and understand that they are not failing if they do not indicate anyone, and the confidence statement needs to be collected at the time of first identification. Very few police departments are known to run their lineups in pristine conditions.
- Whether the conditions were pristine and whether the eyewitness initial confidence was high are factors that trump the role of the estimator variables.

¹¹ Two basic types of identification procedures can be found in the literature, lineups, and field showups. A showup refers to the observation of a single suspect by a witness in the field, typically at the crime scene, whereas a lineup refers to the presentation of the suspect and several foils, either live or via photographs.

- The extent to which initial high confidence in pristine conditions is predictive of identification accuracy depends on the base rate of target-present lineups. In lab studies it is usually fixed at around 50%, but it is expected to vary widely in real circumstances. The best estimate is around 35%, and it seems that at that base rate high-confidence witnesses (initial confidence, pristine conditions) are around 90% accurate.
- With time, and especially in the contexts in which a witness expects a cross-examination, witness confidence loses its predictive power. Preparations for the trial are known to inflate witness confidence, especially if they received a positive feedback from anyone after the initial identification.

With this in mind, let us try to apply (6) again. This time, instead of 0.025, we will use the more realistic (and still fairly optimistic, because now we assume the lineup took place in optimal conditions and resulted in high initial confidence) 0.1. As for $P(witness_G|G)$, real cases surveys, such as those in Sacramento, indicate that the ratio with which witnesses correctly point out the suspects in lineups, is actually below 0.5, but let us follow [Wixted and Wells \(2017, p. 53\)](#), and use 0.9. What we get is that in optimal lineup conditions we should convict just in case (it is known that)

$$\frac{cost(CI)}{cost(AG)} < \frac{9}{1} \times 1/9 = 1 \quad (12)$$

$$cost(CI) < cost(AG) \quad (13)$$

This means that if the lineup and its result are optimal (in the sense already described), and the prior probability of guilt is 10% (which seems quite far from the presumption of innocence!), if convicting an innocent is slightly better than acquitting a guilty person, the eyewitness testimony will be enough to convict.

This, however, we submit, still fails to prove the superiority of eyewitness testimony as such. For one thing, this is highly sensitive to the choice of the prior. If the prior is 1% instead, what we need is now $cost(CI)/cost(AG) < \frac{9}{1} \times 1/99 \approx 0.0909$, so convicting an innocent now would have to be almost 10 times better than acquitting a guilty suspect.

Secondly, whatever power eyewitness testimony obtained through such considerations was through experiment, data analysis, and statistical evaluation, and not some extra-statistical power of cross-examination, which seems to either not help at all, or actually incorrectly inflates the witness' confidence.

8. The risk of wrongful conviction argument

Di Bello offers another type of argument against the use of purely statistical evidence in criminal trials. The argument is again centered around the idea that we do not know the value of the likelihood ratio. In this case, because of the unknown likelihood, we do not know the risk at which we place the defendant. In other words, the worry is not that we put a defendant at a significant risk, but that we place her at an unknown risk. Let us take a look at the argument.

Di Bello takes the risk to be the conditional probability that a defendant, if innocent, would be mistakenly convicted:

$$P(C|I \wedge R_s). \quad (14)$$

In the formula, C stands for ‘convicted’, I stands for ‘innocent’, and R_s is the claim that a statistic-based decision rule is used that says, e.g. the following:

R_s : If, given statistical evidence, the probability of the defendant’s guilt is high-for example, it meets a probability threshold of 99% - convict; otherwise, acquit.

Even though it seems difficult to estimate the value of (14), Di Bello argues that we can approximate it; at least in the prisoners scenario. Asking about the probability of an innocent person being convicted by applying the statistic-based rule seems to boil down to asking about the probability of an innocent defendant facing the 99 : 1 statistics. So, instead of estimating $P(C|I \wedge R_s)$, we would like to estimate $P(\text{stat}[99 : 1]|I)$. Let us grant that the probability of obtaining the statistical evidence if the suspect is guilty is quite high, say, close to 1. Now, if we were able to estimate the risk of incorrect conviction, $P(\text{stat}[99 : 1]|I)$, we would be able to estimate the likelihood ratio, $P(\text{stat}[99 : 1]|G)/P(\text{stat}[99 : 1]|I)$. But, if we follow Di Bello in thinking that the likelihood in such cases cannot be estimated, we also should not think that the risk of innocent conviction can be estimated. Our ignorance, Di Bello concludes, should lead us to abstain from using purely statistical evidence because we want to protect innocent defendants against mistaken convictions. Decisions for which we do not know the risk seem to violate this basic juridical norm.

Di Bello considers the relevant risk to be the conditional probability $P(C|I \wedge R_s)$ which, he argues, is difficult to estimate. But it is not clear that this is the conditional probability we should be relying on. Suppose for instance that we want to evaluate the risk for the prisoners scenario. We are interested not only in the conditional probability that the defendant is convicted given the statistic-based rule and her innocence. The condition should also include the additional fact that the known statistic is 99 : 1. Otherwise, we throw away free information and fail to conditionalize on our total evidence.

If we add the statistic into the information to be conditionalized on, it becomes easy to evaluate the risk in the prisoners case. The risk simply is 100%. The reason is that any defendant will be convicted, because, given the known statistic, the probabilistic threshold in R_s is met. If the statistic is anything lower than 99 : 1, then the risk is 0%, because nobody will be convicted.

Di Bello insists that we need to ignore the evidence in our estimation of the risk of wrongful conviction:

The risk whose value we want to know is the conditional probability that a mistaken conviction would occur as a result of applying the statistics-based rule in the prisoner scenario without already assuming that the statistics have taken a specific form.
(Di Bello, 2019, p. 20)

But what would be the reason for ignoring available evidence in our risk estimation? One could argue that including the statistical facts, we trivialize the risk-based argument because a defendant can face either risk 0% or 100%. Be that as it may, what exactly is wrong with a simple outcome of considerations of a rather simplistic thought experiment?

In real-life cases, strictly speaking, the risk of incorrect conviction based on a certain type of evidence cannot be known, because we cannot really find out which convictions were correct. But no one takes this to be the reason to refuse to rely on all sorts of evidence. The risk of wrongful conviction given some types of evidence can to some extent be estimated *post hoc* by looking at

exonerations and known errors involved in trials. And here, it seems, statistical evidence (such as DNA match) fares much better than eyewitness evidence.

9. Discussion

Does all this undermine the risk-based approach to evidence evaluation? Not at all: the risk of mistaken conviction is an important factor that should be considered in legal decisions. After all, requiring that innocent defendants should not be exposed to unequal risks of mistaken convictions sounds at least like a reasonable candidate for a norm of fairness. However, deploying the concept of such a risk, we argued, has consequences different from those that one might have in mind when initially approaching the so-called ‘paradoxes of naked statistical evidence’. Whether we should find those consequences acceptable, is a separate question.

Acknowledgements

The authors would like to express gratitude to Marcello Di Bello for many discussions related to the topic and for his feedback on earlier drafts of this paper.

Funding

This research has been funded by the Polish National Science Centre [grant number 2016/22/E/HS1/00304].

REFERENCES

- AITKEN, C., ROBERTS, P., and JACKSON, G. (2010). Fundamentals of probability and statistical evidence in criminal proceedings (Practitioner Guide No. 1), Guidance for judges, lawyers, forensic scientists and expert witnesses. *Royal Statistical Society's Working Group on Statistics and the Law*.
- BASU, R. (2018). The wrongs of racist beliefs. *Philosophical Studies*. [forthcoming].
- BEHRMAN, B. W. and DAVEY, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior*, **25**(5):475–491.
- COHEN, J. (1977). *The Probable and the Provable*. Oxford: Oxford University Press.
- DI BELLO, M. (2019). Trial by statistics: Is a high probability of guilt enough to convict? *Mind*, **128**(512): 1045–1084.
- EDDINGTON, A. (1939). *The Philosophy of Physical Science*. Cambridge: Cambridge University Press.
- ENOCH, D. and FISHER, T. (2015). Sense and sensitivity: Epistemic and instrumental approaches to statistical evidence. *Stanford Law Review*, **67**:557–611.
- FENTON, N., LAGNADO, D., DAHLMAN, C., and NEIL, M. (2019). The opportunity prior: A proof-based prior for criminal cases. *Law, Probability and Risk*, page [online first].
- FINKELSTEIN, M. (2009). *Basic Concepts of Probability and Statistics in the Law*. New York: Springer.
- GROSS, S. R., O'BRIEN, B., HU, C., and KENNEDY, E. H. (2014). Rate of false conviction of criminal defendants who are sentenced to death. *Proceedings of the National Academy of Sciences*, **111**(20):7230–7235.
- KLOBUCHAR, A., STEBLAY, N. K. M., and CALIGIURI, H. L. (2006). Improving eyewitness identifications: Hennepin county's blind sequential lineup pilot project. *Cardozo Public Law, Policy & Ethics Journal*, **4**: 381–413.
- KOTZEN, M. (2012). Selection biases in likelihood arguments. *British Journal for the Philosophy of Science*, **63**:825–839.
- LINDSAY, R. C. L., WELLS, G. L., and RUMPEL, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations? *Journal of Applied Psychology*, **66**(1):79–89.

- LUCY, D. (2013). *Introduction to Statistics for Forensic Scientists*. Chichester: John Wiley & Sons.
- NESSON, C. R. (1979). Reasonable doubt and permissive inferences: The value of complexity. *Harvard Law Review*, **92**(6):1187–1225.
- REDMAYNE, M. (2008). Exploring the proof paradoxes. *Legal Theory*, **14**(4):281–309.
- SMITH, M. (2017). When does evidence suffice for conviction? *Mind*. DOI: 10.1093/mind/fzx026.
- SOTTAS, P.-E., ROBINSON, N., SAUGY, M., and NIGGLI, O. (2008). A forensic approach to the interpretation of blood doping markers. *Law, Probability and Risk*, **7**(3):191–210.
- THOMPSON, S. G. (2007). Beyond a reasonable doubt-reconsidering uncorroborated eyewitness identification testimony. *UC Davis Law Review*, **41**:1487–1545.
- WEISBERG, J. (2005). Firing squads and fine-tuning: Sober on the design argument. *British Journal for the Philosophy of Science*, **56**:809–821.
- WELLS, G. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology*, **62**(5):739–752.
- WELLS, G. L. and OLSON, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, **54**(1):277–295.
- WIXTED, J. T. and WELLS, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, **18**(1):10–65.
- WRIGHT, D. and McDAID, A. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology*, **10**:75–84.