2

The Naturalized Epistemology Approach to Evidence

Gabriel Broughton and Brian Leiter

1. Introduction

At the core of epistemology, as the philosophical study of knowledge, is the critical examination of our methods of inquiry. The epistemologist wants to know how we should manage our beliefs about the world. What counts as good evidence for a conclusion? What can we reasonably infer from a given body of information? On a traditional view, these questions must be answered a priori—by the light of pure reason, as it were—without appeal to any supposed empirical knowledge. Naturalism rejects this view. A good method of inquiry must be *reliable*, in the sense that it consistently leads to the truth. But whether any given method is reliable depends on facts about the world that we can't discover from the armchair. Accordingly, naturalists insist that epistemology must be an a posteriori discipline, continuous with and dependent upon empirical science.

Naturalized epistemology has an individual and a social part. Individual epistemology concerns the intellectual practices of particular agents considered in isolation. Social epistemology takes up the social practices that inculcate belief. More specifically, it evaluates those practices instrumentally, in terms of how likely they are to yield true beliefs on important questions. Call this *veritistic* evaluation.[1] Since the effects of any given practice on the beliefs of relevant agents is an empirical question, amenable in principle to scientific investigation, naturalized social epistemology draws on the sciences. Still, its ultimate conclusions are *evaluative*. When the social epistemologist finds that a practice is more likely to lead to true beliefs than the alternatives, she identifies a reason to implement it. But practices that rate well on the veritistic dimension may perform so poorly on others that, all things considered, they should not be implemented. The social epistemologist focuses on the veritistic reasons without necessarily taking a stand on how they stack up against the rest.

In this chapter, we consider the law of evidence from the perspective of naturalized social epistemology, provisionally setting aside the other values implicated by rules of evidence and focusing on the promotion of accurate fact finding.[2] Of course, there are good reasons, all else being equal, to prefer accurate verdicts to inaccurate ones. For instance, a criminal prohibition's deterrence value depends on how reliably violators are convicted.[3] Indeed, many of the key aims of primary legal rules—from corrective justice

[1] Goldman 1999.
[2] cf. Allen and Leiter 2001.
[3] Kaplow 1994.

Gabriel Broughton and Brian Leiter, *The Naturalized Epistemology Approach to Evidence* In: *Philosophical Foundations of Evidence Law*. Edited by: Christian Dahlman, Alex Stein, and Giovanni Tuzet, Oxford University Press. DOI: 10.1093/oso/9780198859307.003.0003

to economic efficiency—can only be performed if the disputes arising under them are generally decided accurately. And there are other reasons too, having to do with justice in the individual case, with rule-of-law values, and with the perceived legitimacy of the law and its institutions.

The law of evidence matters to naturalized epistemology because it affects the accuracy of adjudication. But accurate adjudication depends on more than evidence law. It depends, for example, on civil and criminal procedure. It depends on the funding allocated to prosecutors and public defenders, on the contents of law school curricula, and on much else besides. The study of evidence law thus falls into place as one component of the broader project of studying *adjudication* as part of naturalized epistemology. This turns out to be important. For in the epistemological study of evidence law one sometimes identifies a persistent fact finding error that evidence law seems helpless to address. In such cases, one must remember that there may be other ways, outside of evidence law, to fix the problem.

2. Prospects

Studying evidence law as part of naturalized epistemology means (i) evaluating evidence law veritistically (ii) using the tools and results of the sciences. Start with (i). Rules of evidence can be evaluated along many dimensions. Do they produce outcomes that are generally accepted as fair or legitimate? Do they express respect for the parties? Do they keep administrative costs down? Studying evidence law as part of naturalized social epistemology means evaluating it in terms of the accuracy of the verdicts that it can be expected to produce. And this brings us to (ii). Rather than appeal to armchair intuition or the authority of tradition, naturalized epistemology examines evidence law with the help of empirical science.

How much can the sciences tell us about the veritistic value of different rules of evidence? Not all that much, according to some commentators.[4] These skeptics point out that we generally lack the sort of independent access to the facts of legal disputes that would allow us to reliably determine whether a given verdict is correct. And there are obvious practical, legal, and moral impediments to running actual legal trials under "experimental" rules of evidence. But how, then, are we supposed to empirically assess the reliability of alternative rules of evidence?

Many psychologists and social scientists interested in legal decision-making have focused on conducting controlled experiments known as *jury simulations*. These are studies in which participants acting as mock jurors observe a "trial" and then render a "verdict" or answer questions about the case. Typically, jury simulations involve randomly assigning participants to view different versions of the trial—one version might include an eyewitness or an expert, say, that another omits—in order to measure the effects of these differences. Critics claim that jury simulations tell us nothing about real juries.[5] They argue that

[4] See, e.g., Redmayne 2003, 866; Friedman 2001, 2034.
[5] See, e.g., Nunez et al. 2011. See also *Lockhart v. McCree*, 476 U.S. 162, 172 (1986).

simulation studies are unrealistic—they lack *ecological validity*—because (i) they use college students as mock jurors rather than drawing on a more representative sample; (ii) they use brief written case summaries instead of live testimony and real evidence presented over days or weeks; (iii) they omit jury deliberations and focus on the "verdicts" of individual jurors; and (iv) they use mock jurors who know that their decisions will have no real consequences. As a result, simulation studies don't generalize. And, the critics reason, since jury simulations are the best that social science—and, so, naturalized epistemology—can do, this means that naturalized epistemology has little to offer evidence law.

We think this overstates the case, for four reasons.

(i) *Simulation studies don't exhaust the naturalized epistemology of evidence law*. When a rule of evidence excludes relevant information, the rationale is often that, while the evidence has *some* probative value, the jury may significantly misconstrue that value (or misuse the evidence in some other way). So the rule is based on two distinct claims, one about the actual value of the evidence and another about the value the jury will attribute to it. Jury studies address the second claim, but other empirical findings bear on the first. Consider, for instance, the reliability of various (so-called) forensic sciences. We don't need jury studies to find the error rates associated with spectographic voice identification, microscopic hair comparison, or bite mark analysis. This suggests that even if the sciences had nothing to tell us about juries, they would still have something to contribute to the veritistic evaluation of evidence law.

But even when it comes to jury psychology, we are not limited to simulations. We also have (a) surveys, (b) archival analyses, and (c) field experiments.[6] While none of these methodologies has been as popular among social scientists as simulations, each has contributed to our understanding of juries. Start with surveys. Many surveys in this area involve contacting jurors after a trial and asking why they decided as they did. But other sorts of surveys can also be useful. For example, courts have frequently insisted that jurors understand the factors affecting the reliability of eyewitness identifications. A properly conducted survey can put this empirical claim to the test.[7] Archival analyses use large datasets to discern relationships between case characteristics and outcomes. Such analyses can be used to study, for instance, whether juries are more likely to convict a defendant if she has a criminal record, or whether experienced jurors are more likely than inexperienced ones to find civil defendants liable. Field experiments involve randomly assigning actual juries to different procedural conditions and then measuring differences between experimental groups on various dependent variables. While understandably rare, field experiments have been used to study the effects of permitting juror discussion during civil trials,[8] among other procedural innovations.

(ii) *Low ecological validity is probably not as worrisome as the objection suggests, since what really counts is external validity*. An experiment's ecological validity is the

[6] See Bornstein 2017.
[7] More on this in the following section.
[8] Diamond et al. 2003.

extent to which it mimics salient surface features of a real-world target setting. Its external validity is the extent to which its results generalize. There is, of course, a relationship between the two. Increasing ecological validity sometimes increases external validity by eliminating differences between the experimental and the target context that affect important dependent variables. But sometimes there is no payoff in external validity, because the eliminated differences are causally irrelevant. The question is whether the differences between simulations and trials really matter.

Suppose that we are interested in the effects of admitting a certain kind of "bad character" evidence against criminal defendants. We might run a jury simulation experiment that sacrifices some level of ecological validity by using college students as mock jurors. And suppose that we find that the admission of the character evidence causes a rise in convictions. Now, what kind of differences between student jurors and *realistic jurors*, as we'll call them, should we be worried about? One possibility is that realistic jurors are simply more (or less) likely to convict than student jurors. In fact, this would *not* limit our ability to generalize the effect observed in our simulation, because it would involve student jurors and realistic jurors reacting to the character evidence similarly, qualitatively and quantitatively, even if they differ in other ways. The simulation's external validity would only be limited if there were some interaction between juror type and character evidence, for instance if admission had no effect on realistic jurors. Of course, whether student jurors and realistic jurors differ in these ways is an *empirical* question.

The verdicts of student jurors and realistic jurors have now been compared in dozens of experiments, involving a variety of civil and criminal cases and the manipulation of numerous independent variables beyond juror type, thus facilitating consideration of both main effects and interaction effects. Both types of effects are rare, and interaction effects are especially so. In one review of twenty-six studies, for instance, main effects of juror type were observed in just five, while interaction effects were only found in two. Thus, "interactions are the exception rather than the norm."[9] Subsequent research confirms this result.[10] Similar studies suggest that concerns about the medium in which trials are presented to mock jurors as well as the presence or absence of deliberation may be somewhat overblown.[11]

(iii) *Jury simulation studies are becoming increasingly realistic*. Even skeptics about the preceding can take solace in the fact that ecological validity in jury studies is improving. Critics conjure an image of college sophomores scanning a brief case summary and declaring a personal verdict, but this is no longer how most jury simulations are done. A review of every jury simulation reported over two recent years in *Law and Human Behavior* showed that the majority did *not* use student jurors or written case summaries.[12] Instead they used a more representative sample of community members and, in the usual case, a video presentation of a mock trial performed by professional actors. And while most of the studies still bypassed jury deliberation, a substantial minority (about 20%) did not.

9 Bornstein 1999, 80.
10 Penrod et al. 2011, 205
11 See, e.g., Kerr and Bray 2005.
12 Bornstein 2017.

(iv) *Triangulation and conceptual replication can help address reasonable concerns about external validity*. Even more realistic simulation studies differ from trials, of course. While a two-hour video is certainly better than a two-paragraph summary, it is still a far cry from two weeks or even two days in a courtroom. More important, none of these improvements in ecological validity addresses the problem of *consequentiality*.[13] No matter how representative the sample or how realistic the trial presentation, mock jurors still know that they are not real jurors. They still know that their decisions do not have real consequences. A handful of experimental studies have attempted to determine whether this difference limits the external validity of jury simulations. One study arranged for two separate juries, an actual jury and an experimental one, to hear a number of criminal cases.[14] The mock jury consisted of a random sample of potential jurors from the pool who were not selected or questioned by the attorneys. For each trial, the mock jury, like the actual one, was present in the courtroom during the proceedings and eventually deliberated in private before rendering a verdict. A few additional studies have attempted to test the effects of consequentiality in other ways.[15] Unfortunately, these isolated studies point in different directions, so that no firm conclusion can be drawn. This is clearly an area where further research would be useful.

Ultimately, though, no single methodology can give us everything that we want in a study. Simulations offer the benefit of internal validity. In a laboratory setting, because we can randomly assign mock jurors to trial conditions distinguished only by some particular variable of interest, we can be confident that any differences in juror behavior between conditions is the result of the experimental manipulation. The disadvantage of simulations is that we have no a priori guarantee of external validity. Laboratories and courtrooms are different, and it's always possible that an effect observed in one will not be observed in the other. Archival analyses and other field studies, by examining the behavior of actual juries deciding actual cases, offer stronger assurances of external validity. Unfortunately, they do so at the expense of internal validity. Outside the laboratory, cases that differ in one way are liable to differ in many others as well. As a result, we can seldom draw causal conclusions with any confidence based on field studies alone. Every study is imperfect in some way.

The proper response is not to wash our hands of the investigation. It is to (i) *replicate* our simulation results in a variety of contexts, and (ii) *triangulate* those results with evidence from the field.[16] Suppose, for instance, that the admission of a certain kind of statistical testimony in a simulated products liability trial causes a massive increase in jurors' liability judgments. To establish the external validity of this result, we would next want to replicate it in another simulation using deliberating juries rather than individual jurors, or using a different products liability trial. We might also conduct an archival analysis to test for a correlation between such testimony and plaintiff judgments. If we continue to observe the same effects in many different simulation studies, and we find converging evidence from outside the laboratory, then we can be reasonably confident in generalizing our findings to actual trials.

13 Bornstein and McCabe 2005.
14 Diamond and Zeisel 1974.
15 See, e.g., Kaplan and Krupa 1986; Suggs and Berman 1979.
16 Saks 1997, 5.

3. Applications

For naturalized epistemology, the proof is in the pudding. Accordingly, we turn, in the remainder of this chapter, to two applications of the approach: eyewitness testimony and, more briefly, character evidence.

3.1 Eyewitness Testimony

An eyewitness identification likely represents the *principal* evidence in upwards of 100,000 criminal prosecutions in the United States each year.[17] And the law generally takes eyewitness identifications to be very *good* evidence, as appellate courts have consistently sustained guilty verdicts based on the testimony of a single eyewitness, even where that testimony is undermined by other evidence. Yet eyewitnesses are often wrong. Mistaken eyewitness identifications contributed to at least 258 of the 375 wrongful convictions (69%) in the US that have so far been overturned based on DNA evidence.[18] These exonerations suggest that mistaken eyewitness identification is the single leading cause of wrongful convictions in our criminal justice system today.

This raises, but does not settle, the question of the reliability of identification evidence. The science of eyewitness memory, developed over the last forty years, helps to provide an answer. Research psychologists most often study eyewitness memory by conducting laboratory experiments in which participants view a simulated crime and later attempt to pick the culprit out of a lineup. In one study, for example, participants signed up for an experiment on "complex information processing." In fact, the experimenters had arranged for someone to interrupt each session by bursting into the lab and stealing an expensive piece of electronic equipment.[19] Having become eyewitnesses to a crime, the participants were later shown either a target-present or a target-absent lineup and asked to identify the thief. Eyewitnesses perform surprisingly poorly in such studies. A review of ninety-four experiments found that among subjects shown target-present lineups, just 46 % correctly identified the culprit, while 21% incorrectly identified a filler.[20] (The remainder declined to make an identification.) Among subjects shown target-absent lineups, 48% incorrectly identified an innocent person. Experiments conducted outside the laboratory suggest that these results generalize.[21]

These results converge with the findings of field studies of lineups conducted in actual police investigations. Unlike in the experimental context, researchers observing eyewitness identifications at a police station don't know who actually committed the crime. The lineup consists of a suspect and several fillers known to be innocent. If an eyewitness identifies the suspect, then she may or may not have identified the actual culprit. We simply don't know whether her identification is accurate. If she identifies a filler, however, then we know that she has made a mistake. The published field studies include data from 6,734 lineups

[17] See Goldstein et al. 1989.
[18] Innocence Project 2020.
[19] Lindsay 1986.
[20] Clark et al. 2008.
[21] See, e.g., Pigott et al. 1990, in which forty-seven unwitting bank tellers were confronted by, and later asked to identify, a man trying to cash a crudely forged money order.

conducted in a variety of jurisdictions.[22] All told, eyewitnesses identified the suspect 2,746 times (40.8%), identified a known-innocent filler 1,599 times (23.7%), and declined to make an identification 2,389 times (35.5%). Thus, nearly one in every four eyewitnesses (23.7%) identified someone known to be innocent. Setting aside the cases in which no identification was made, 36.8% of the choosers pointed to someone known to be innocent. Since some suspect identifications were surely mistaken as to the actual culprit, the true rate of mistaken identification is even higher.

Of course, eyewitnesses perform better in some circumstances than others. In the research literature, factors related to eyewitness accuracy are divided into *system variables* and *estimator variables*.[23] A system variable is a variable that is potentially under the control of the criminal justice system, such as the size of the lineup shown to an eyewitness. An estimator variable is one that is outside the system's control, such as the age of the eyewitness. Some research on estimator variables has produced predictable results, for instance that eyewitness accuracy falls when the culprit is seen only briefly, or only in the dark, or only from a great distance. But the literature includes some startling findings as well. For instance, eyewitness accuracy falls when the culprit's race differs from the eyewitness's,[24] when the eyewitness only sees the culprit under highly stressful conditions,[25] and when the culprit openly displays a weapon,[26] among other conditions.

Concerning system variables, we note three significant findings. First, research shows that accuracy is significantly affected by whether the eyewitness has been warned that the culprit might or might not be in the lineup. Pre-lineup instructions that lack this warning—what are called *biased instructions*—produce a negligible increase in correct identifications and a massive increase in false positives as compared to unbiased instructions.[27] Second, research shows that *filler selection* is extremely important. In particular, using fillers that make the suspect stand out—for instance, by failing to match eyewitness descriptions of the culprit—severely undermines the reliability of an identification.[28] Third, if the lineup administrator knows who the suspect is, then she can influence the identification, and undermine its reliability, in a variety of ways.[29]

In fact, an unblinded lineup administrator can cause problems even after an identification has been made. For instance, she can raise the confidence a mistaken eyewitness has in her identification by providing positive feedback. Research has repeatedly shown that even modest encouragement ("Good, you picked out the suspect") significantly inflates both the eyewitness's confidence that her identification was correct and the confidence that she *recalls* having when she originally made the identification.[30]

The courts have often touted eyewitness confidence as an important index of accuracy.[31] Are they right about this? Are confident eyewitnesses reliable? We need to distinguish between (i) an eyewitness's confidence in her identification at the time she makes it, (ii) her

22 Wells et al. 2020.
23 Wells 1978.
24 Meissner and Brigham 2001.
25 Deffenbacher et al. 2004.
26 Steblay 1992.
27 Steblay 1997.
28 Fitzgerald et al. 2013.
29 See, e.g., Greathouse and Kovera 2009.
30 Semmler et al. 2004; Wells and Bradfield 1998.
31 See, e.g., *Neil v. Biggers*, 409 U.S. 188, 199–200 (1972).

confidence at the time of trial, and (iii) her belief at the time of trial about her confidence at the time of identification. It is quite clear that because of the malleability of eyewitness confidence and retrospective assessments of confidence, neither (ii) nor (iii) is a remotely reliable index of accuracy.[32] The real question is whether initial eyewitness confidence, if accurately recorded by a blinded administrator, reliably tracks accuracy.

A consensus has emerged that, under certain conditions, it does.[33] But the boundary conditions of this phenomenon are not yet clear. Some argue that initial confidence reliably predicts accuracy so long as the testing conditions are *pristine*—so long, that is, as (i) the lineup includes only one suspect; (ii) the lineup is fair, in the sense that the suspect does not stand out; (iii) the pre-lineup instructions are unbiased; and (iv) the lineup administrator does not know who the suspect is (i.e., double-blind administration).[34] Others argue, with some empirical support, that pristine testing conditions are not good enough.[35] In addition, they claim, the *witnessing conditions* must be favorable: the eyewitness must have a relatively long look at the culprit's face, from a relatively close distance, in relatively good light, and so on. This remains an open question. Even if pristine testing conditions turned out to be sufficient to ensure a strong confidence-accuracy relationship, however, the problem would remain that, although the situation is improving, most eyewitness identifications in the United States are not conducted under pristine conditions.[36]

Even if eyewitness identifications are not especially reliable, an identification will typically still be relevant. It will typically still make the defendant's guilt more probable than it otherwise would have been. The research goes to the probative value of eyewitness identifications, and we only have a problem if juries take them to be more or less probative than they actually are. Unfortunately, the evidence suggests that jurors do indeed *overbelieve* eyewitnesses. Consider an experiment in which a number of factors in a staged crime were manipulated to yield low (33%), moderate (50%), or high (74%) proportions of correct eyewitness identifications.[37] Jurors then watched as defense counsel cross-examined an eyewitness drawn from one of these conditions. Under every condition, jurors' belief rates were higher than witnesses' accuracy rates. The disparity was especially severe when witnessing conditions were poor. Under those conditions—in which only 33% of eyewitnesses were accurate—jurors believed eyewitnesses 62% of the time.[38]

A number of US courts have suggested that jurors are at least sensitive to the factors affecting eyewitness accuracy.[39] They insist that the results just canvassed are simply common sense, perfectly familiar to the average juror. Survey research shows that this is wrong. In one study, experimenters asked jury-eligible adults to judge the truth or falsity of thirty statements concerning various issues affecting eyewitness accuracy (e.g., "The presence of a weapon impairs an eyewitness's ability to accurately identify the perpetrator's face").[40] They then compared these responses to those of research psychologists in the field. The

[32] Bradfield et al. 2002; Wells et al. 1981.
[33] See, e.g., Wixted and Wells 2017; Palmer et al. 2013.
[34] Wixted and Wells 2017.
[35] See, e.g., Lockamyeir et al. 2020.
[36] See, e.g., McNabb et al. 2017; Greene and Evelo 2015; Police Executive Research Forum 2013.
[37] Lindsay et al. 1981.
[38] These results were replicated in Wells et al. 1980. See also Brigham and Bothwell 1983.
[39] See, e.g., *United States v. Rodriguez-Felix*, 450 F.3d 1117, 1125 (10th Cir. 2006); *State v. Coley*, 32 S.W. 3d 831, 833–834 (Tenn. 2000).
[40] Benton et al. 2006.

results showed that jurors generally agreed with the experts on just four out of the thirty statements. For instance, while 98% of experts said that police instructions can affect an eyewitness's willingness to make an identification, just 40% of jurors agreed. And while 90% of experts said that eyewitnesses are more accurate when identifying members of their own race, just 50% of jurors agreed.

These data converge with the results of experimental investigations of the factors that actually influence verdicts. For instance, one study involved mock jurors watching a video of an armed robbery trial in which an eyewitness identification was the key prosecution evidence.[41] Ten factors relevant to eyewitness accuracy were systematically manipulated, including culprit disguise, retention interval, weapon visibility, instruction bias, and lineup fairness. The *only* factor that significantly affected verdicts was eyewitness confidence *at the time of trial*, which, as we saw, is effectively worthless. The remaining factors had at most trivial effects, often in the wrong direction. This basic result—that jurors are highly sensitive to eyewitness confidence at trial but insensitive to important factors bearing on eyewitness accuracy—has been replicated repeatedly.[42]

A variety of possible reforms could improve the accuracy of the system. These reforms fall into three distinct categories. The first involves attempts to improve the reliability of the eyewitness identification evidence that law enforcement collects along the lines discussed earlier. If law enforcement agencies were forced to use best practices, the resulting identifications would be more reliable, and more in line with jurors' expectations.[43] This might be accomplished through legislation requiring their use or, indirectly, through a rule of evidence excluding identification evidence resulting from dubious procedures. (Such a rule of evidence might itself be adopted by legislation or by judicial interpretation of constitutional due process. The US Supreme Court has rejected this approach,[44] but a number of states have adopted something like it.[45])

The second category of reforms involves using the existing rules of evidence to exclude the least reliable eyewitness identification evidence. This could overlap with the previous reforms in jurisdictions that use evidence law to deter shoddy identification procedures, but courts should also exclude some identification evidence whose unreliability is not due to procedural defects. For instance, an eyewitness should not be allowed to testify at trial (or in pretrial hearings) about her confidence that the defendant (i.e., the person she has identified) is in fact the culprit. We have seen that eyewitness confidence can be inflated by dubious procedures. But a variety of additional variables have also been shown to inflate eyewitness confidence even in the absence of procedural defects.[46] As a result, an eyewitness's confidence at trial is liable to reflect factors unrelated to memory strength. In many cases, eyewitness confidence at trial may actually be strictly irrelevant, but even if not, given its likely influence on the jury, courts should exclude it as "prejudicial" or "confusing" under Rule 403 of the Federal

41 Cutler et al. 1988.

42 See, e.g., Jones et al. 2020; 2008.

43 For an extended discussion of best practices, see Wells et al. 2020.

44 See *Neil v. Biggers*, 409 U.S. 188 (1972); *Manson v. Brathwaite*, 432 U.S. 98 (1977). Compare *Stovall v. Denno*, 388 U.S. 293 (1967); *Manson v. Brathwaite*, 432 U.S. 98, 118–124 (1977) (Marshall J, dissenting). The empirical evidence shows the Court's current due process approach to identification evidence to be irredeemably flawed. See, e.g., Wells and Quinlivan 2009.

45 See, e.g., *Com. v. Johnson*, 420 Mass. 458, 472 (1995); *People v. Adams*, 53 N.Y.2d 241, 251 (1981).

46 See, e.g., Odinot et al. 2009; Shaw 1996; Wells et al. 1981.

Rules of Evidence (FRE) or the state equivalent. Other plausible candidates for exclusion include evidence from eyewitnesses whose initial confidence in an identification was low,[47] as well as courthouse identifications generally.[48]

The final category of possible reforms involves trying to improve jurors' ability to evaluate the eyewitness identification evidence that they do see. If jurors fail to appropriately evaluate identification evidence because their knowledge of the factors affecting identification accuracy is limited, perhaps their performance would improve if this information were provided to them. This might be done, for instance, through expert testimony from a qualified research psychologist.[49] Historically, eyewitness expert testimony was excluded as invading the province of the jury.[50] Under FRE 702, it's admissible if (i) it is based on reliable scientific knowledge, and (ii) it will help the jury evaluate the eyewitness testimony at issue. While courts generally allow that eyewitness expert testimony is based on reliable scientific knowledge, many still routinely exclude it as unhelpful, claiming that eyewitness psychology is just common sense for the average juror.[51] As we have seen, this is false.

The question remains, though, whether expert testimony will actually improve jurors' evaluations of identification evidence. If it has any effects at all, it might produce (i) *confusion*, leading to perverse evaluations of eyewitness identifications; (ii) *skepticism*, leading to fewer guilty verdicts regardless of the strength of the evidence; or (iii) *sensitivity*, leading to verdicts that track the strength of the evidence.[52] Given that jurors tend to over-believe eyewitnesses, some form of skepticism effect would arguably be salutary. But accuracy would hardly recommend that jurors be led to dismiss identifications made quickly and confidently in pristine testing conditions based on memories formed in ideal witnessing conditions. Such identifications remain, by all accounts, very strong evidence. Thus, sensitivity is ultimately the more desirable result.

Research shows that expert testimony can sometimes improve juror sensitivity to the factors that affect eyewitness accuracy.[53] The more common result, however, has been increased skepticism of eyewitness testimony in general.[54] This, at least, is the more common result when discrimination is measured at the level of *verdicts*. Interestingly, though, in many cases where expert testimony does not produce verdicts that are sensitive to the quality of the identification, it nevertheless improves jurors' general understanding of eyewitness factors. The problem, in other words, seems to be that of applying this knowledge to the particular case. Recent experiments with expert testimony modeled on the interview-identification-eyewitness (I-I-Eye) teaching aid[55] suggest that this problem can be solved.[56]

[47] See Wixted and Wells 2017.
[48] See Garrett 2012.
[49] Alternatively, it might be done through jury instructions. We ignore jury instructions for reasons of space.
[50] See, e.g., *Criglow v. State*, 183 Ark. 407 (1931).
[51] See, e.g., *State v. Young*, 35 So.3d 1042, 1050 (La. 2010).
[52] Cutler et al. 1989.
[53] See, e.g., Wise and Kehn 2020; Cutler et al. 1989.
[54] See, e.g., Jones et al. 2017; Lindsay 1994.
[55] Pawlenko et al. 2013.
[56] See, e.g., Wise and Kehn 2020.

3.2 Character

FRE 404(a) prohibits the use of evidence of someone's "character"—evidence that she is careless, or that she is aggressive—to prove that she acted accordingly on a particular occasion (subject to some exceptions). This basic prohibition extends to evidence of previous *actions* when offered to show character in order to prove action in accordance. Yet FRE 404(b) permits evidence of past actions to be used for other purposes, for instance, to show motive, or opportunity, or absence of mistake. When specific acts evidence is admitted under FRE 404(b), the court may instruct the jury to consider it only for a specified permitted purpose and not as the basis for the prohibited character inference. FRE 404(a) does contemplate an exception for impeachment in accordance with FRE 608, which permits evidence of a dishonest character to impeach a witness.

There is a large social psychology literature examining the explanatory power of the notion of character. Do people act in keeping with stable personality traits across a diverse range of situations? Or is behavior so situation-specific that personality traits lack predictive value? *Situationism* holds that, in fact, people's actions are primarily the result of situational factors—often factors operating outside conscious awareness—rather than reflecting stable dispositions constitutive of character. Thus, situationism repudiates the core premise underlying the most obvious use of character evidence—namely, that character can be used to predict behavior on a particular occasion. As Ross and Nisbett note, "standard correlation coefficients determined in well-controlled research settings" show that "personality traits" lack substantial "explanatory and predictive power."[57] If situationism is correct, then the FRE 404(a) bar on character evidence is sound.

But consider the FRE 608 exception for impeachment by evidence of dishonest character. If "manipulations of the immediate social situation can overwhelm in importance the type of individual differences in personal traits or dispositions that people normally think of as being determinative of social behavior,"[58] then why think a witness's dishonest behavior at work or in her personal life bears on her truth-telling *in court*, *under oath*, *before a judge*, *under threat of perjury*? The situation giving rise to the impeachment evidence and the situation in which the witness testifies are usually nothing alike.

Many also worry that permitting evidence of past actions under FRE 404(b) will lead juries to draw officially forbidden inferences about the "bad character" of criminal defendants. In fact, there is substantial experimental evidence to suggest that any limiting instruction associated with the admission of such evidence is likely to be ineffective.[59] Given situationism, the danger of "unfair prejudice" under FRE 403 appears substantial: if the jury draws (forbidden) inferences from putative traits of character, the jury will be misled and prejudiced, since situationism teaches that character traits have relatively little predictive power. Should such evidence generally be excluded?

Caution is required here. Consider the famous situationist study of Good Samaritan behavior,[60] which found that "[i]f the subjects were in a hurry . . . , only about 10 percent helped [the person needing assistance]. By contrast, if they were not in a hurry . . . about 63 percent

[57] Ross and Nisbett 1991, 91.
[58] Ross and Nisbett 1991, xiv.
[59] See, e.g., Lieberman and Arndt 2000.
[60] Darley and Baton 1973.

of them helped."[61] But what about that 10%? Would it not be reasonable to invoke their good character in explaining their behavior, unlike the majority? Indeed, other researchers have argued that character traits can have quite large impacts on behavior. Suppose we want to know whether a trait of "honesty" can be used to predict the degree to which children will engage in a broad array of related behaviors. If we try to predict just *one* such behavior on the basis of one other behavior, we obtain a correlation that explains only 5% of the behavioral variance. However, if we look at the overall honesty that a child shows across a whole battery of tests and then try to predict the honesty that the same child will show in another battery of tests, we obtain a much higher correlation—this time, explaining a full 81% of the variance.[62] This suggests that it is the *quality* of the evidence that matters: we need *better proof* of character.

Existing rules, however, permit rather weak evidence of character: the opinion of those who know the person or her reputation, or evidence that she committed a serious crime, or committed any crime involving dishonesty. Moreover, at trial, we are concerned with a single instance of conduct—Did the defendant act in accordance with character *on the occasion that resulted in charges?* Did the witness tell the truth *today*?—which is precisely where situationism counsels skepticism about the predictive value of character evidence. In this context, the case for excluding evidence that supports inferences about character deserves more serious consideration from courts, although sometimes the probative value of prior bad acts for permissible purposes will outweigh the danger of unfair prejudice.

References

Allen, R. J., and Leiter, B. (2001). "Naturalized Epistemology and the Law of Evidence," *Virginia Law Review*, 87: 1491–550.

Benton, T. R., Ross, D. E., Bradshaw, E., Thomas, W. N., and Bradshaw, G. S. (2006). Eyewitness Memory is Still Not Common Sense: Comparing Jurors, Judges and Law Enforcement to Eyewitness Experts. *Applied Cognitive Psychology*, 20: 115–29.

Bornstein, B. H. (1999). "The Ecological Validity of Jury Simulations: Is the Jury Still Out?" *Law and Human Behavior*, 23(1): 75–91.

Bornstein, B. H. (2017). "Jury Simulation Research: Pros, Cons, Trends, and Alternatives," in M. B. Kovera (ed.), *The Psychology of Juries*, 207–26. Washington, DC: American Psychological Association.

Bornstein, B. H., and McCabe, S. G. (2005). "Jurors of the Absurd: The Role of Consequentiality in Jury Simulation Research," *Florida State University Law Review*, 32(2): 443–68.

Bradfield, A. L., Wells, G. L., and Olson, E. A. (2002). "The Damaging Effects of Confirming Feedback on the Relation between Eyewitness Certainty and Identification Accuracy," *Journal of Applied Psychology*, 87(1): 112–20.

Brigham, J. C. and Bothwell, R. K. (1983). "The Ability of Prospective Jurors to Estimate the Accuracy of Eyewitness Identifications," *Law and Human Behavior*, 7(1): 19–30.

Clark, S. E., Howell, R. T., and Davey, S. L. (2008). "Regularities in Eyewitness Identification: The Role of System and Estimator Variables," *Law and Human Behavior*, 32(3): 233–58.

Cutler, B. L., Dexter, H. R., and Penrod, S. D. (1989). "Expert Testimony and Jury Decision Making: An Empirical Analysis," *Behavioral Sciences and the Law*, 7: 215–25.

Cutler, B. L., Penrod, S. D., and Stuve, T. E. (1988). "Juror Decision Making in Eyewitness Identification Cases," *Law and Human Behavior*, 12(1): 41–55.

[61] Ross and Nisbett 1991, 4.
[62] Hartshorne and May 1928.

Darley, J. M. and Baton, C. D. (1973). “From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior,” *Journal of Personality & Social Psychology*, 27: 100–8.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., and McGorty, E. K. (2004). “A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory,” *Law and Human Behavior*, 28(6): 687–706.

Diamond, S. S., Vidmar, N., Rose, M., Ellis, L., and Murphy, B. (2003). “Inside the Jury Room: Evaluating Juror Discussions during Trial,” *Judicature*, 87(2): 54–9.

Diamond, S. S., and Zeisel, H. (1974). “A Courtroom Experiment on Juror Selection and Decision-Making,” *Personality and Social Psychology Bulletin*, 1(1): 276–7.

Fitzgerald, R. J., Price, H. L., Oriet, C., and Charman, S. D. (2013). “The Effects of Suspect-Filler Similarity on Eyewitness Identification Decisions: A Meta-Analysis,” *Psychology, Public Policy, and Law*, 19(2): 151–64.

Friedman, R. D. (2001). “E is for Eclectic: Multiple Perspectives on Evidence,” *Virginia Law Review*, 87: 2029–54.

Garrett, B. L. (2012). “Eyewitnesses and Exclusion,” *Vanderbilt Law Review*, 65(2): 449–506.

Goldman, A. (1999). *Knowledge in a Social World*. Oxford: Oxford University Press.

Goldstein, A. G., Chance, J. E., and Schneller, G. R. (1989). “Frequency of Eyewitness Identification in Criminal Cases: A Survey of Prosecutors,” *Bulletin of the Psychonomic Society*, 27(1): 71–4.

Greathouse, S. M., and Kovera, M. B. (2009). “Instruction Bias and Lineup Presentation Moderate the Effects of Administrator Knowledge on Eyewitness Identification,” *Law and Human Behavior*, 33(1): 70–82.

Greene, E., and Evelo, A. J. (2015). “Cops and Robbers (and Eyewitnesses): A Comparison of Lineup Administration by Robbery Detectives in the USA and Canada,” *Psychology, Crime & Law*, 21(3): 297–313.

Hartshorne, H., and May, M. A. (1928). *Studies in Deceit*. New York: Macmillan.

Innocence Project (2020). *DNA Exonerations in the United States*. Available at https://www.innocenceproject.org/dna-exonerations-in-the-united-states/ (Accessed: March 29, 2021).

Jones, A. M., Bergold, A. N., Dillon, M. K., and Penrod, S. D. (2017). “Comparing the Effectiveness of *Henderson* Instructions and Expert Testimony: Which Safeguard Improves Jurors’ Evaluations of Eyewitness Evidence?,” *Journal of Experimental Criminology*, 13: 29–52.

Jones, A. M., Bergold, A. N., and Penrod, S. (2020). “Improving juror sensitivity to Specific Eyewitness Factors: Judicial Instructions Fail the Test,” *Psychiatry, Psychology and Law*, 27(3): 366–85.

Jones, E. E., Williams, K. D., and Brewer, N. (2008). “‘I had a Confidence Epiphany!’: Obstacles to Combating Post-Identification Confidence Inflation,” *Law and Human Behavior*, 32: 164–76.

Kaplan, M. F., and Krupa, S. (1986). “Severe Penalties under the Control of Others can Reduce Guilty Verdicts,” *Law & Psychology Review*, 10: 1–18.

Kaplow, L. (1994). “The Value of Accuracy in Adjudication: An Economic Analysis,” *Journal of Legal Studies*, 23(1): 307–401.

Kerr, N. L., and Bray, R. M. (2005). “Simulation, Realism, and the Study of the Jury,” in N. Brewer and K. D. Williams (eds.), *Psychology and Law: An Empirical* Perspective, 322–64. New York: Guilford Press.

Lieberman, J. D., and Arndt, J. (2000). “Understanding the Limits of Limiting Instructions: Social Psychological Explanations for the Failures of Instructions to Disregard Pretrial Publicity and other Inadmissible Evidence,” *Psychology, Public Policy, and Law*, 6(3): 677–711.

Lindsay, R. C. L. (1986). “Confidence and Accuracy of Eyewitness Identification from Lineups,” *Law and Human Behavior*, 10(3): 229–39.

Lindsay, R. C. L. (1994). “Expectations of Eyewitness Performance: Jurors’ Verdicts do not Follow from their Beliefs,” in D. F. Ross, J. D. Read, and M. P. Toglia (eds.), *Adult Eyewitness Testimony: Current Trends and Developments*, 362–84. New York: Cambridge University Press.

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): 75–89.

Lockamyeir, R. F., Carlson, C. A., Jones, A. R., Carlson, M. A., and Weatherford, D. R. (2020). “The Effect of Viewing Distance on Empirical Discriminability and the Confidence-Accuracy

Relationship for Eyewitness Identification," *Applied Cognitive Psychology*, 34(5):1047–60. doi: 10.1002/acp.3683.

McNabb, N. S., Farrell, B. R., and Brown, C. R. (2017). "Voluntary Adoption of Evidence-Based Practices by Local Law Enforcement: Eyewitness Identification Procedures in Arkansas, Iowa, Kansas, Missouri, and Nebraska," *Journal of Gender, Race, and Justice*, 20(3): 509–42.

Meissner, C. A., and Brigham, J. C. (2001). "Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review," *Psychology, Public Policy, and Law*, 7(1): 3–35.

Nunez, N., McCrea, S. M., and Culhane, S. E. (2011). "Jury Decision Making Research: Are Researchers Focusing on the Mouse and not the Elephant in the Room?," *Behavioral Sciences and the Law*, 29: 439–51.

Odinot, G., Wolters, G., and Lavender, T. (2009). "Repeated Partial Eyewitness Questioning causes Confidence Inflation but not Retrieval-Induced Forgetting," *Applied Cognitive Psychology*, 23: 90–7.

Palmer, M. A., Brewer, N., Weber, N., and Nagesh, A. (2013). "The Confidence-Accuracy Relationship for Eyewitness Identification Decisions: Effects of Exposure Duration, Retention Interval, and Divided Attention," *Journal of Experimental Psychology: Applied*, 19(1): 55–71.

Pawlenko, N. B., Safer, M. A., Wise, R. A., and Holfeld, B. (2013). "A Teaching Aid for Improving Jurors' Assessments of Eyewitness Accuracy," *Applied Cognitive Psychology*, 27(2): 190–7.

Penrod, S. D., Kovera, M. B., and Groscup, J. (2011). "Jury Research Methods," in B. Rosenfeld and S. D. Penrod (eds.), *Research Methods in Forensic Psychology*, 191–214. Hoboken, NJ: Wiley.

Pigott, M. A., Brigham, J. C., and Bothwell, R. K. (1990). "A Field Study on the Relationship Between Eyewitnesses' Descriptions and Identification Accuracy," *Journal of Police Science and Administration*, 17(2): 84–8.

Police Executive Research Forum (2013). "A National Survey of Eyewitness Identification Processes in Law Enforcement Agencies" (March 8).

Redmayne, M. (2003). "Rationality, Naturalism, and Evidence Law," *Michigan State Law Review*, 4: 849–83.

Ross, L., and Nisbett, R.E. (1991). *The Person and the Situation: Perspectives of Social Psychology*. Philadelphia, PA: Temple University Press.

Saks, M. J. (1997). "What do Jury Experiments tell us About how Juries (should) make Decisions?," *Southern California Interdisciplinary Law Journal*, 6(1): 1–53.

Semmler, C., Brewer, N., and Wells, G. L. (2004). "Effects of Postidentification Feedback on Eyewitness Identification and Nonidentification Confidence," *Journal of Applied Psychology*, 89(2): 334–46.

Shaw, III, J. S. (1996). "Increases in Eyewitness Confidence Resulting from Postevent Questioning," *Journal of Experimental Psychology: Applied*, 2(2): 126–46.

Steblay, N. M. (1992). "A Meta-Analytic Review of the Weapon Focus Effect," *Law and Human Behavior*, 16(4): 413–24.

Steblay, N. M. (1997). "Social Influence in Eyewitness Recall: A Meta-Analytic Review of Lineup Instruction Effects," *Law and Human Behavior*, 21(3): 283–97.

Suggs, D., and Berman, J. J. (1979). "Factors affecting testimony About Mitigating Circumstances and the Fixing of Punishment," *Law and Human Behavior*, 3(4): 251–60.

Wells, G. L. (1978). "Applied Eyewitness Testimony Research: System Variables and Estimator Variables," *Journal of Personality and Social Psychology*, 36: 1546–57.

Wells, G. L., and Bradfield, A. L. (1998). "'Good, you Identified the Suspect': Feedback to Eyewitnesses Distorts their Reports of the Witnessing Experience," *Journal of Applied Psychology*, 83(3): 360–76.

Wells, G. L., Ferguson, T. J., and Lindsay, R. C. L. (1981). "The Tractability of Eyewitness Confidence and its Implications for Triers of Fact," *Journal of Applied Psychology*, 66(6): 688–96.

Wells, G. L., Kovera, M. B., Douglas, A. B., Brewer, N., Meissner, C. A., and Wixted, J. T. (2020). "Policy and Procedure Recommendations for the Collection and Preservation of Eyewitness Identification Evidence," *Law and Human Behavior*, 44(1): 3–36.

Wells, G. L., Lindsay, R. C. L., and Tousignant, J. P. (1980). "Effects of Expert Psychological Advice on Human Performance in Judging the Validity of Eyewitness Testimony," *Law and Human Behavior*, 4(4): 275–85.

Wells, G. L., and Quinlivan, D. S. (2009). "Suggestive Eyewitness Identification Procedures and the Supreme Court's Reliability Test in Light of Eyewitness Science: 30 Years Later," *Law and Human Behavior*, 33(1): 1–24.

Wise, R. A., and Kehn, A. (2020). "Can the Effectiveness of Eyewitness Expert Testimony be Improved?," *Psychiatry, Psychology and Law*, 27(2): 315–30.

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.