6-5-2013

Behavioral Economics and Evidence Law

Fredrick E. Vars
University of Alabama - School of Law, fvars@law.ua.edu

Behavioral Economics and Evidence Law

Fredrick E. Vars

Forthcoming, OXFORD MANUAL ON BEHAVIORAL ECONOMICS AND LAW *(Eyal Zamir & Doron Teichman, eds.)*

Chapter 30

Behavioral Economics and Evidence Law

Fredrick E. Vars

1. Introduction

Psychology has long informed thinking on evidence law (Hutchins and Slesinger 1928; Smith 1942; Skolnick 1961; Woocher 1977; Mendez 1984; Tanford 1990; Wellborn 1991; Leo and Ofshe 1998; Robbennolt 2005). The first challenge of this Chapter is therefore to narrow the scope. My focus will be the analysis of psychological phenomena that squarely call into question assumptions of normative economic and rational choice theory. This includes primarily, but not exclusively, heuristics and biases and related findings (Tversky and Kahneman 1974). Heuristics and biases are cognitive processes that expedite rationalization and analytic reasoning. These short-cuts reduce the costs of decision-making, but sometimes produce errors. Behavioral economics attempts to identify and measure how cognitive short-cuts lead to predictable variances from traditional rational choice theory.

Given this definition, the applicability of behavioral economics to the field of evidence law should be apparent. Evidence law primarily concerns what pieces of information should be allowed to be considered in the fact-finder's decision process. Of key importance is the truth-seeking objective of trial. Given that rationality is predictably compromised by heuristics and biases, the relation between relevant evidence and the heuristics and biases triggered by such evidence can effectuate a better understanding of and a better formulation of evidence law in general.

In particular, behavioral economics has been applied to evidence law in at least four different ways: (1) to explain or justify current law; (2) to argue that current law counteracts or

fails to counteract the findings; (3) to advocate changes in practice under existing law; and (4) to argue for law reform. On the surface, the first two applications appear descriptive and the second two normative. Below the surface all four missions share common assumptions about the power and generalizability of the psychological findings. Even superficially explanatory assertions are premised on the normative view that psychology should inform our understanding of evidence law. Whether it should, and how law and practice should be changed, are incredibly complex questions. Illustrating the complexity and suggesting ways forward are the goals of this Chapter.

After the introduction, Section 2 of this Chapter discusses exemplars of each of the four descriptive and normative uses of behavioral findings outlined above. Several types of evidence and rules of evidence have received more sustained attention by behavioralists, cutting across the four types of arguments. Section 3 next takes up several of those types and rules of evidence: (1) eyewitness testimony; (2) experts; (3) standards of proof; and (4) subsequent remedial measures. The final and longest section, Section 4, is devoted to character evidence, surveying the literature and offering a new illustration. The example is the exclusion of past offense evidence from sex offense trials, but the broader goal is to caution against simplistic application of psychology to law and to suggest directions for richer and more nuanced future research. Moving from description to prescription presents serious challenges.

2. Descriptive and Normative Uses of Behavioral Insights in Evidence Law

2.1. Do biases explain or justify current evidence law?

Purely explanatory uses of behavioral findings are rare. One article argues that many guilty criminal defendants do not exercise their right to silence due to the "availability heuristic": they underestimate risks irrationally perceived as remote (Seidmann and Stein 2000, p. 448).

Being caught in a lie seems more remote than a certain appearance of guilt by remaining silent. Another example of the availability heuristic may be helpful: people read many news stories about particular families losing their homes to foreclosure. Because such instances are easy to recall, the newspaper readers overestimate the likelihood of foreclosure (Tversky 1973). The right-to-silence authors go on to make the normative claim that their game-theoretic model of the Fifth Amendment, including this feature, leads to more accurate determinations of guilt (Seidmann and Stein 2000, pp. 498-502).

The psychological phenomenon of overcorrection has been described as follows: "Subjects who are initially presented with a piece of information, but are later required to discount this information, generally tend to subtract too much" (Guttel 2004, p. 247). Slipping seamlessly from descriptive to normative, one commentator claims that avoiding overcorrection can help "explain" (p. 245) and "justify" (p. 261) "rules concerning discovery, the review powers of appellate courts, the exclusion of hearsay, and other evidentiary rules of the common law" (p. 245). The idea is to weed out weak evidence before its admission and later exclusion unduly influences jurors. Some authors jump straight to justification: for example, cognitive biases such as availability and representativeness indicate that juries will overreact to evidence of the sexual conduct of victims of sexual assault, thereby justifying suppression of it in most cases (Kessler 1992, pp. 94-96).

Sometimes the claim is that current law reflects rather than combats biases in decision-making. Starting with the bias: "A person who follows [the representativeness] heuristic evaluates the probability of an uncertain event . . . by the degree to which it is (i) similar in essential properties to its parent population and (ii) reflects the salient features of the process by which it is generated." (Kahneman 1972, p. 431) The classic example of where this strategy

fails is overestimating the likelihood that a person described in terms similar to a stereotypical engineer is in fact an engineer even though he is drawn from a population containing relatively few engineers (pp. 241-242). The representativeness heuristic has been linked to the legal doctrine of res ipsa loquitor: "According to the Restatement (Second) of Torts, when judging liability for an event that resulted in injury to a plaintiff, a fact-finder may infer that the defendant was negligent if 'the event is of a kind which ordinarily does not occur in the absence of negligence'" (Rachlinkski 2000, p. 90). Putting the two together, one author argues that, "[a]s it is ordinarily stated, res ipsa loquitur represents a profound misunderstanding of the laws of probability in precisely the way that the representativeness heuristic predicts" (Rachlinski 2000, p. 90; see also Kaye 1979). The obvious implication is that the doctrine should be changed.

Craig Callen (1994) offers one of the most sensitive, non-empirical applications of psychological findings to evidence law, involving the hearsay rule and its exceptions. Hearsay is a "statement, other than one made by the declarant while testifying at the trial or hearing, offered in evidence to prove the truth of the matter asserted" (Fed. R. Evid. 801). Callen shows that norms of cooperation in communication better explain existing doctrine than prevailing theories, but also justify some changes to that doctrine.

The lesson of these examples is that legal scholars generally cite behavioral findings in the service of normative claims. This no doubt reflects both the nature of legal scholarship generally and the fact that these findings are dubbed "biases," "errors," and "illusions." Their very names would seem to justify correction.

2.2. Does the adversarial process counteract biases?

But perhaps the legal process already counteracts the effects of these biases. Some commentators think so, at least in certain areas (Froeb and Kobayashi 1996; Rachlinski 2000).

Specifically, one commentator argues that the adversary process combats the tendency to give more weight to evidence that confirms a belief than to comparable evidence that refutes it (Bersoff 1992). Another commentator argues that the pretrial discovery of evidence and the settlement process reduce the impact of the overcorrection bias (Guttel 2004). Others are less sure (Rachlinski 2011; Eisenberg et al. 2001). A third group thinks that the legal system does not counteract biases in at least some areas---in particular, the strategic manipulation of fact-finders using psychology (Gold 1987) and the pervasive belief that what actually happened was inevitable and should have been foreseeable (i.e., the "hindsight bias") (Hastie et al. 1999). One commentator more radically asserts that cognitive limitations affirmatively *promote* the truth-seeking process by making it more difficult to lie successfully (Sanchirico 2004). Memory is more internally consistent, and thus resistant to cross-examination, than imagination. This diversity of opinion reflects the complexity of drawing normative, or even descriptive, conclusions for the legal system based on behavioral findings.

2.3 Should practice under the current evidence rules be changed?

After demonstrating that judges are subject to a range of heuristics and biases, one group of commentators suggests that "judges might learn to educate themselves about cognitive illusions so that they can try to avoid the errors that these illusions tend to produce" (Guthrie et al. 2001). This modest proposal is followed by more aggressive ones, but it is important to note that simply learning about biases may help mitigate their effects (although the point is disputed). Not every behavioral finding justifies changes in the law.

2.4. Should the rules of evidence be changed?

Because so few authors limit themselves to description (i.e., how a bias impacts fact-finding), this category overlaps substantially with the first category above (i.e., whether a bias

justifies current evidence law). The only apparent difference is whether the commentator believes the status quo or a recommended alternative better counteracts a given bias. In addition to the other recommendations described in this Chapter, one commentator has called for eliminating the exclusionary rule for evidence obtained from an illegal search and seizure and replacing it with a cause of action for damages (Slobogin 1999). Under current law, only individuals found to possess incriminating evidence seek suppression. This selection bias, combined with the representativeness and availability heuristics (described above), will lead judges to systematically deny suppression motions. Exposing judges to damage actions brought by innocent individuals would mitigate the biased inference of guilt (Slobogin 1999). Others suggest changes without expressly advocating them (Guthrie et al. 2001; Jolls et al. 1998).

3. Central Applications of Behavioral Insights to Evidence Law

The following types and rules of evidence have received more sustained attention from commentators, implicating the same four descriptive and normative uses of behavioral evidence. Examining these uses one context at a time will hopefully solidify an understanding of these important pockets of evidence law and illustrate the evolution of thinking in each area.

3.1. Eyewitness testimony

Factfinders overvalue direct evidence, like eyewitness testimony, and undervalue circumstantial evidence (Heller 2006; see also Zamir et al. 2012). (Of course, this assumes that one knows the "correct" value of evidence, which usually will not be true. Controlled experiments can, however, provide good estimates in some domains.) The alleged culprits for erroneous weighting include the simulation (Heller 2006, pp. 259-64) and representativeness (Guthrie et al. 2001) heuristics. One version of the simulation heuristic is estimating the

probability of an event by trying to imagine a causal scenario; the easier it is to imagine such a scenario, the more probable the event appears. (Heller 2006, p. 260) Some have prescribed changes in practice: more and better studies on witness identifications (Schacter et al. 2008) and reformed jury instructions (Heller 2006, pp. 304-05). Others more radically propose requiring corroboration for eyewitness testimony (Barzun 2008, p. 1993). Of these, the call for more research has the soundest footing given the limitations of present knowledge and complexity of translating findings into practice.

3.2. Experts

Expert testimony is perhaps the area of evidence law where the calls for change are most diverse. Commentators have argued that heuristics and biases justify (1) more expert testimony, (2) more careful screening of putative expert testimony by judges, and (3) replacing expert opinion testimony with data. This divergence in prescriptions reflects a peeling back of the onion: (1) lay jurors are biased; (2) judges are less biased than jurors; and (3) experts themselves, including judges, are subject to bias.

An early article in this line suggested that expert testimony combats lay jurors' reliance on the representativeness heuristic and related neglect of base rates (recall the engineer example) (Saks and Kidd 1980, p. 134). It cited others' calls for additional base-rate experts, special masters, court advisors, and special courts (p. 134). The authors' own prescription was more modest: "experts ought to be permitted to offer their data, their algorithms, and their Bayesian [probability] theorems. The errors that may be introduced will be subject to adversarial cross-examination" (p. 148).

Notably, this article preceded the United States Supreme Court's adoption in *Daubert v. Merrell Dow Pharmaceuticals* (1993) of a more active judicial gatekeeping role for expert

testimony (Fed. R. Evid. 702). Behavioralists have largely embraced that role on the theory that judicial expertise and accountability can mitigate the ill effects of biases (Beecher-Monas 2003, p. 989; see also Gold 1987, pp. 510-12). This might appear like a flip from "more experts" to "fewer experts," but the new position can be described as an evolution toward "more and better experts" (for related suggestions to improve presentation of statistical evidence, see Lyon and Koehler 1996).

Perhaps not surprisingly, there have been new calls for additional experts, this time to help judges determine which expert testimony is reliable enough to be admitted. The rationale is that while judicial expertise is good, it is not good enough to evaluate scientific or other technical expert methodology. Such experts could avoid hindsight bias in particular, one set of commentators argue (Worthington et al. 2002).

Good experts are the solution to all problems, one might conclude from the foregoing. But experts are still human and thus prey to systematic bias (National Academy of Sciences 2009). Overconfidence and over-optomism are two particularly troubling tendencies for adjudication. Data may have other problems, but not these. One empirical study concludes that: "In many settings, the fallible opinions of isolated experts should be supplemented or replaced by statistical data" (Meadow and Sunstein 2001, p. 631). The authors support this conclusion with data showing that experts substantially and systematically underestimate time to treatment for a particular medical condition (pp. 637-38).

The evolution of recommendations with respect to expert testimony illustrates a critical and sometimes neglected step in the behavioral economics of evidence law. The easy part is identifying a bias that affects laypeople. The hard part is crafting a solution that fixes more problems than it creates. An obvious first response to juror bias is to throw more sophisticated,

and hopefully less biased, experts at the problem. But some of what they offer will be junk science and their opinions may be subject to the same or different biases that affect laypeople. Disqualifying bad experts (perhaps with the help of independent experts), or even replacing expert opinion with unbiased statistics, may end up being the best solution.

3.3. Standards of Proof

The standard of proof instructs factfinders how to decide cases given uncertainty. Rational choice theory has an elegant solution to the question of where to set the standard of proof: find for the plaintiff if and only if the probability that he or she meets the elements of his or her case is above an algebraic combination of the utilities of the four possible outcomes (Kaplan 1968; Cullison 1969). The details of the formula are not important for present purposes, but the four outcomes are, of course, true positive (i.e., a correct finding for the plaintiff), false positive, true negative, and false negative. Setting aside other criticisms (Tribe 1971; Guttel and Teichman 2012), one commentator argues from cognitive psychology results that standards of proof should be selected from a menu of no more than three, not from the continuum implied by rational choice theory (Clermont 1987, p. 1149). Selecting among standards is more complicated than the equation above (Vars 2010).

The two most studied existing standards are the criminal standard and the normal civil standard. In the common law world, a criminal defendant should be convicted only if guilt is established "beyond a reasonable doubt." In contrast, the plaintiff in a typical civil case should prevail if they make their case by a "preponderance of the evidence." The latter standard has been equated to a probability greater than 50%, which is generally believed to minimize the total number of trial errors. However, behavioral research suggests that the "omission bias" leads individuals to demand a significantly higher probability (66-75%) (Zamir and Ritov 2012, pp.

186-87). The omission bias is a tendency to favor inaction. In civil litigation, the researchers argue, accepting a claim is viewed as action, whereas rejecting a claim is inaction in that it preserves the legal status quo. The experimental finding of a higher threshold is important, but its normative implications are unclear (p. 197).

Along similar lines, studies have found that mock jurors' verdicts are not impacted by verbal statements of the standard of proof (Kagehiro 1990). Quantifying the standards as probability thresholds, on the other hand, generated smaller win rates as the proof standard was raised. This might appear to be a victory for traditional rational choice theory, but at least one study has found verbal formulations effective and offered a psychological explanation. Rather than waiting until the end of trial to combine all of the evidence in Bayesian fashion, jurors begin trying to make sense of each piece of evidence as it is presented. Information supportive of the initial conclusion is then overvalued and conflicting information undervalued in what has been dubbed "a coherence shift" (Simon 2004, pp. 511-86; Engel 2009, p. 456). The tendency to overlook evidence of acquittal was apparently mitigated in an experimental setting by an actual beyond-a-reasonable-doubt instruction emphasizing subjective emotion, not probability: "you must be convinced of [his] [her] guilt to the same degree you would be convinced about a matter of importance in your own life in which you would act with confidence and without restraint or hesitation" (Engel, p. 441, quoting Pennsylvania jury instruction). The study concludes that both the objectivists and subjectivists can learn from each other. The important point is that a behavioral understanding of juror decision-making may be essential in crafting standards of proof that accomplish their legal and social objectives (Korobkin and Ulen 2000, p. 1097; Jolls et al. 1998, pp. 1429-30).

3.4. Subsequent Remedial Measures

People overestimate the predictability of past events. This tendency is known as "hindsight bias" (Fischhoff 1975). One set of commentators argue that excluding evidence of remedial measures taken after an injury to prove negligence before the injury (Federal Rules of Evidence 407) "represents an adaptation to the effects of the hindsight bias" (Guthrie et al. 2001). It appears that the commentators are making both descriptive and normative claims: this is how the law came to be, and it is a good thing. Logic provides some support for both claims. The hindsight bias is a robust psychological phenomenon and would predictably lead factfinders sometimes to conclude that a defendant should have predicted an injury because the defendant took measures after it to prevent its recurrence. The defendant should have seen it coming sooner, the biased factfinder might reason. To be sure, updating probability estimates based on new data is perfectly rational, if done correctly (Sanchirico 2003, pp. 1197-1200).

So far so good, but there's a missing step, as one of the authors previously observed: "suppressing subsequent remedial measures in accident cases slightly reduces the pool of relevant evidence that the fact-finder has available to decide the case. If the effect of the hindsight bias is small, these adaptations might be overreactions to the problem of judging liability in hindsight" (Rachlinski 2000, p. 73). Dan Kahan picked up on this important observation, concluding that the excluding or admitting subsequent remedial measures on a case-specific basis may be optimal (Kahan 2010).

Kahan's argument is significant on its own terms, but also as an exemplar of careful application of psychological findings to law. The critical insight is that evidence of subsequent remedial measures makes negligence somewhat more likely, even if not by as much as the hindsight bias leads factfinders to believe, so a per se rule of exclusion will generate erroneous outcomes. Kahan, while expressly eschewing explanation, offers the following prescription: "If

a judge excludes [subsequent remedial measures] proofs in cases in which the plaintiff's case is otherwise weak, and is receptive to admission of such proofs in cases in which the issue of the defendant's substandard conduct is otherwise close, she will minimize the sum total of erroneous outcomes---either mistaken findings of liability or mistaken findings of nonliability---relative to a rule that categorically excludes or admits such proofs" (p. 1637).

But Kahan cannot resist the gravitational pull of current practice when he argues that this is what judges do all the time in balancing probative value and prejudicial effect under Federal Rule of Evidence 403, which basically requires that the value of evidence be net neutral or positive to be admissible (pp. 1638-39). Here Kahan falls short. He effectively assumes that the subsequent remedial measures evidence is the last piece of evidence offered so that "the full evidentiary context" is otherwise complete (p. 1639). This may not always be true. Relatedly, Kahan also assumes that judges measure probative value in terms of how likely evidence is to change the outcome ("contribution . . . to accurate factfinding" (p. 1649)). More plausibly, judges ask how far new evidence moves the meter. If hindsight bias is strong enough and probative value fixed, then unfair prejudice may substantially outweigh probative value in every case: weak, close, or overwhelming. Kahan may or may not be correct in all of his conclusions, but at least he frames the question well.

4. Character Evidence

In the behavioral economics literature, perhaps the most discussed rule of evidence is the exclusion of character evidence. For example, under New York law, "[t]he prosecutor in a criminal case may not introduce, during her case in chief, evidence of the defendant's general character or propensity for criminal behavior . . . [or] specific incidents of defendant's bad

conduct" (Barker and Alexander 2012). In many jurisdictions, numerous exceptions may in practice overwhelm the rule, but for simplicity this illustration concerns only the general rule.

Past behavior is the best predictor of future behavior, so this rule on its face would seem accuracy-defeating (Tillers 1998, p. 792). Scholars have offered numerous justifications for the exclusion of character evidence (Sanchirico 2003, pp. 1239-59). Only one type of justification is relevant here: for psychological reasons juries will systematically overweight character evidence. The fundamental attribution error and representativeness heuristic are the most commonly cited and persuasive psychological culprits (Sanchirico 2003, pp. 1242-43; see also Saks and Kidd 1980, p. 136; Gold 1983, pp. 525-30; Park 1996; Korobkin and Ulen 2000, pp. 1086-87). Representativeness has been discussed above. The fundamental attribution error is the tendency to over-value personality-based explanations for behavior and under-value the importance of the situation. But does the undue weight overwhelm the probative or other positive values?

To assess this question a framework is needed. Accuracy in trial outcomes is an important purpose of evidence law. Minimizing the total number of errors is perhaps a sufficient definition of accuracy in the civil context, but avoiding false convictions (FP) is generally considered more important than avoiding false acquittals (FN). This justifies the requirement that guilt be proved beyond a reasonable doubt, which in one large survey judges most often defined as 90% certainty (McCauliff 1982, p. 1325 tbl. 2; but see Simon and Mahan 1971).

Notwithstanding this high standard of proof, sex crime conviction rates for people brought to trial in two recent years (2009 and 2011) in New York City were 86 and 89 percent (Doll 2012). As alluded to above, evidence of past offenses is frequently admitted despite the general bar on character evidence of this kind. But assume for illustration that it isn't. What would be the impact of adding prior offense evidence (as Federal Rule of Evidence 413 now

allows)? The probability of conviction for defendants with prior sex offenses admitted would be given by the following formula, a straightforward application of Bayes's Rule:

$$P(Conviction|Prior) = \frac{P(Prior|Conviction) \times P(Conviction)}{P(Prior)}$$

One study found that roughly 12% of sex offenders in prison had a prior sex offense, which serves as a decent proxy for the probability of a prior given conviction (P(Prior|Conviction)) (Greenfeld 1997, p. 22 fig. 23). Assume next that a particular sex offense defendant has the same likelihood of conviction as the group of sex offenders who went to trial in New York City in 2011 (P(Conviction)=89%). The third term, P(Prior), is the toughest one to estimate. Sex offenders make up a tiny fraction of the population (p. 2), but the pool of individuals on trial for a sex offense is obviously not a random cross-section of the population. It is helpful to divide the pool by conviction or acquittal.

$$P(Prior) = P(Conviction) \times P(Prior|Conviction) + P(Acquittal) \times P(Prior|Acquittal)$$

The first two terms have already been defined at 89% and 12%, respectively. The probability of acquittal is 11% (1 minus 89%). We have successfully pushed back the problem, but we are still left with an unknown: the probability that a defendant who will be acquitted has a prior sex offense (P(Prior|Acquittal)). Given that the recidivism rate for sex offenders is orders of magnitude higher than the first offense rate (Langan et al. 2003, pp. 1, 2), it is almost certain that the unknown value is less than 12%. But by how much? Three possibilities will frame the issue.

First, assume that acquitted individuals are just as likely as convicts to have a prior offense, viz., 12%. Introducing prior sex offense evidence would have no impact on the probability of conviction. The evidence has no probative value. At the other extreme, assume that acquitted individuals have the same very small likelihood of a prior sex offense conviction

as the general population. Admitting evidence of a prior sex offense in this second scenario would all but ensure conviction (p>99%). Assuming a history of priors halfway between these extremes (6%) and evidence thereof would push the probability of conviction from 89% to around 94%. (Notably, the effect is not symmetric. In this third scenario, for example, introducing evidence of *no* priors would reduce the probability of conviction by less than a percentage point.)

Suppose this third, midpoint, scenario accurately reflects reality. If juries were good Bayesians and correctly gauged the probative value of past offenses---the latter of which here is to say, assigned the correct values to P(Prior|Conviction) and P(Prior|Acquittal)---then past offense evidence would increase correct convictions for past offenders and slightly increase correct acquittals for past non-offenders. The net result would be an essentially costless improvement in trial accuracy. The problem is that juries are not good Bayesians in general and systematically overweight prior sex offense information in particular.

As alluded to above, people tend to neglect base rates (like the percentage of engineers in the population)---or, more precisely, to give base rates less weight than Bayes's Rule dictates. In the present context, that would mean assigning too little weight to the past offense evidence. The probative value of the past offense evidence turns on two base rates: P(Prior|Conviction) and P(Prior|Acquittal). In contrast, the other variable in the Equation 1, P(Conviction), is based on event-specific evidence like eyewitness identification, forensics, and victim testimony. In the extreme case juries would give no weight to past offense evidence (scenario one above) and its exclusion or admission would make no difference in outcomes.

A more likely and more serious error in this context is overweight. As noted above, the fundamental attribution error and the availability and representativeness heuristics point in that

direction. In fact, people vastly overestimate recividism among sex offenders: according to one survey, the public believes 74% of sex offenders will commit another sex offense (Levenson et al. 2007, p. 13 tbl. 2); in contrast, only 5.3% of released sex offenders were rearrested within 3 years for a sex crime (Langan et al. 2003, pp. 1, 2). Of course, that means people believe sex offenders are much more likely than non-offenders to have a prior offense (assuming that convictions are closely correlated with guilt, this implies that P(Prior|Conviction) >> P(Prior|Acquittal)). Again taken to the extreme, this would ensure conviction at trial of almost every sex offender with a prior offense where the other evidence justified only an 89% conviction rate (scenario two above).

This is the danger that leads most commentators to defend the exclusion of prior offense evidence (Saks and Kidd 1980; Korobkin and Ulen 2000, Gold 1983). If overweight is really this extreme, then the commentators are probably right. In the present illustration, admitting priors leads to over 99% convictions when the correct conviction rate is 94%. However, by not allowing priors, the actual conviction rate remains a sub-optimal 89%. Adding 5% correct convictions and 5% incorrect convictions would be justified only if false convictions and false acquittals were weighted equally (Friedman 2003, p. 969). They are not, as the beyond-a-reasonable doubt standard implies. If, however, the degree of overweight were slight, the trade-off might make sense. The important point is that the mere fact of overweight does not decide the question (Friedman 2003; Sanchirico 2001, p. 1246; Kahan 2010, p. 1634), even in this very stylized example.

In the real world, a whole host of other considerations is relevant in deciding whether to admit character evidence. The rules of evidence serve goals beyond accuracy at trial and operate as part of a complex system. To take two examples: (1) if past offense evidence were admitted

in every case, prosecutors would more aggressively prosecute alleged repeat offenders, but, (2) due to case selection, this may or may not mean more of them would go to trial. Presumably, more individuals with past offenses would be charged but more would accept plea bargains. The other evidence of guilt for individuals with clean records would likely need to be stronger to justify indictment and prosecution. All of these responses to a change in the rules would affect the variables in the illustration above, with uncertain impact on net social welfare.

The rules of evidence also affect primary behavior (Kaplow 2012). Increasing the likelihood of sanction for individuals with prior offenses would presumably deter some from committing additional sex offenses. How strong is this effect likely to be? And how many sex offenses must be prevented to justify one false conviction? Christopher Sanchirico argues that allowing character evidence would have the opposite effect and would actually *increase* crime. This conclusion is apparently premised on the notion that evidence of a prior offense, standing alone, would be sufficient for a conviction (Sanchirico 2001, p. 1266). Theory alone cannot answer the question.

The important lesson of this example is that identifying a bias is merely the first step. The second step is estimating the magnitude of the effect. To determine whether the bias justifies current law or a change thereof requires consideration of how that effect impacts a complex fact-finding process. How many outcomes change? How do we value each type of outcome? How would a change in practice or the law affect primary behavior? The values question is normative, but the other questions are, or should be, empirical. Progress in the behavioral analysis of evidence law will be achieved through careful natural and controlled experimentation.

5. Conclusion

Other scholars have concluded that empiricism is the way forward for behavioral economics and law (Goldman 2003, p. 224). Sensitivity to context is critical. As one commentator colorfully urges, "empirical legal research should fashion itself after cartography, seeking to create detailed maps of legal behavior in context" (Mitchell 2003, p. 1147). This is perhaps nowhere truer than in the law of evidence. First, one cannot be sure without testing that a particular bias affects decision-makers in a particular situation. Second, any assessment of impact or proposed reform must recognize that the introduction of a piece of evidence is just one part of a complex process. "[B]ecause the rules of evidence and procedure are merely parts of the larger legal system, the collateral consequences of tinkering with some parts while leaving others untouched must be taken into account" (Wistrich et al. 2005, p. 1330).

Even with careful methodology, success is by no means assured. When psychological findings conflict with fundamental principles of the legal system, the system may reject the findings, no matter how robust (Rachlinski 2011, pp. 1690-96). A final word of caution about legal empiricism, which is not unique to behavioral economics or to evidence law: in the legal setting, measuring outcomes is difficult or sometimes even impossible. A primary goal of the law of evidence is accurate decision-making, but how does one measure accuracy? One commentator claims that "one reason that truth is problematic as a legal goal is that the fit between the truth and a jury verdict is impossible to measure" (Moreno 2003, p. 1186). This may be true in some cases, but DNA exonerations and related research (Risinger 2007) demonstrate that assessing the accuracy of trial outcomes is at times attainable.

References

BARKER, ROBERT A. and VINCENT C. ALEXANDER (2012), N.Y. Practice Series, Evidence in New York State and Federal Courts **5**, § 4:13.

BARZUN, CHARLES L. (2008). "Rules of Weight." Notre Dame Law Review **83**(5): 1957-2017.

BEECHER-MONAS, ERICA. "Heuristics, Biases, and the Importance of Gatekeeping." Michigan State Law Review **2003**(4): 987-1022.

BERSOFF, DONALD N. (1992). "Judicial Deference to Nonlegal Decisionmakers: Imposing Simplistic Solutions on Problems of Cognitive Complexity in Mental Disability Law." SMU Law Review **46**(2): 329-371.

CALLEN, CRAIG R. (1994). "Hearsay and Informal Reasoning." Vanderbilt Law Review **47**(1): 43-113.

CLERMONT, KEVIN M. (1987). "Procedure's Magical Number Three: Psychological Bases for Standards of Decision." Cornell Law Review **72**(6): 1115-1156.

CULLISON, ALAN D. (1969). "Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of the Subjective Approach." University of Toledo Law Review **1**(3): 538-598.

Daubert v. Merrell Dow Pharmaceuticals (1993), 509 U.S. 579.

DOLL, JEN (2012). "New York Does Convict Rapists, But They Don't Make the Front Page." The Atlantic Wire. http://www.theatlanticwire.com/national/2012/03/new-york-does-convict-rapists-they-dont-make-front-page/50516/ (visited Oct. 30, 2012).

EISENBERG, THEODORE, JEFFERY J. RACHLINSKI, and MARTIN T. WELLS (2002). "Reconciling Experimental Incoherence with Real-World Punitive Damages." Stanford Law Review **54**(6): 1239-1271.

ENGEL, CHRISTOPH (2009). "Preponderance of the Evidence versus Intime Conviction: A Behavioral Perspective on a Conflict between American and Continental European Law." Vermont Law Review **33**(3): 435-467.

FISCHHOFF, BARUCH (1975). "Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty." Journal of Experimental Psychology: Human Perceptions & Performance **1**(3): 288-299.

FRIEDMAN, RICHARD D. (2003). "Minimizing the Jury Over-Valuation Concern." Michigan State Law Review **2003**(4): 967-986.

FROEB, LUKE M. and BRUCE H. KOBAYASHI (1996). "Naïve, Biased, Yet Bayesian: Can Juries Interpret Selectively Produced Evidence?" Journal of Law, Economics & Organization **12** (1): 257-276.

GOLD, VICTOR (1987). "Covert Advocacy: Reflections on the Use of Psychological Persuasion Techniques in the Courtroom." North Carolina Law Review **65**(3): 481-515.

GOLD, VICTOR J. (1983). "Federal Rule of Evidence 403: Observations on the Nature of Unfairly Prejudicial Evidence." Washington Law Review **58**(3): 497-533.

GOLDMAN, ALVIN I. (2003). "Simple Heuristics and Legal Evidence." Law, Probability & Risk **2**: 215-226.

GREENFELD, LAWRENCE A. (1997). Bureau of Justice Statistics, Sex Offenses and Offenders: An Analysis of Data on Rape and Sexual Assault, (Feb. 1997).

GUTHRIE, CHRIS, JEFFREY J. RACHLINSKI and ANDREW J. WISTRICH (2001). "Inside the Judicial Mind." Cornell Law Review **86**(4): 777-830.

GUTTEL, EHUD (2004). "Overcorrection." Georgetown Law Journal **93**(1): 241-283.

GUTTEL, EHUD and DORON TEICHMAN (2012). "Criminal Sanctions in the Defense of the Innocent." Michigan Law Review **110**(4): 597-645.

HASTIE, REID, DAVID A. SCHKADE, and JOHN W. PAYNE (1999). "Juror Judgments in Civil Cases: Hindsight Effects on Judgments of Liability for Punitive Damages." Law & Human Behavior **23**(5): 597-614.

HELLER, KEVIN JON (2006). "The Cognitive Psychology of Circumstantial Evidence." Michigan Law Review **105**(2): 241-305.

HUTCHINS, ROBERT M. and DONALD SLESINGER (1928). "Some Observations on the Law of Evidence." Columbia Law Review **28**(4): 432-440.

JOLLS, CHRISTINE, CASS R. SUNSTEIN, and RICHARD THALER (1998). "A Behavioral Approach to Law and Economics." Stanford Law Review **50**(5): 1471-1550.

KAGEHIRO, DOROTHY K. (1990). "Defining the Standard of Proof in Jury Instructions." Psychological Science **1**(3): 194-200.

KAHAN, DAN M. (2010). "The Economics--Conventional, Behavioral, and Political--of 'Subsequent Remedial Measures' Evidence." Columbia Law Review **110**(6): 1616-1653.

KAPLAN, JOHN (1968). "Decision Theory and the Factfinding Process." Stanford Law Review **20**(6): 1065-1092.

KAPLOW, LOUIS (2012). "Burdens of Proof." Yale Law Journal **121**(4):738-859.

KAYE, DAVID (1979). "Probability Theory Meets Res Ipsa Loquitur." Michigan Law Review **77**(6): 1456-1484.

KESSLER, ELIZABETH (1992). "Pattern of Sexual Conduct Evidence and Present Consent: Limiting the Admissibility of Sexual History Evidence in Rape Prosecutions." Women's Rights Law Reporter **14**(1): 79-96.

KOROBKIN, RUSSELL B. and THOMAS S. ULEN (2000). "Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics." California Law Review **88**(4): 1051-1144.

LANGAN, PATRICK A., ERICA L. SCHMITT, and MATTHEW R. DUROSE (2003). Bureau of Justice Statistics, Recidivism of Sex Offenders Released from Prison in 1994.

LEO, RICHARD A. and RICHARD J. OFSHE (1998). "The Consequences of False Confessions: Deprivations of Liberty and Miscarriages of Justice in the Age of Psychological Interrogation." Journal of Criminal Law and Criminology **88**(2): 429-496.

LEVENSON, JILL, YOLANDA N. BRANNON, TIMOTHY FORTNEY and JUANITA BAKER (2007). "Public Perceptions About Sex Offenders and Community Protect Policies." Analyses of Social Issues and Public Policy **7**(1): 137-161.

LYON, THOMAS D. and JONATHAN J. KOEHLER (1996). "The Relevance Ratio: Evaluating the Probative Value of Expert Testimony in Child Sexual Abuse Cases." Cornell Law Review **82**(1): 43-78.

MCCAULIFF, C.M.A. (1982). "Burdens of Proof: Degrees of Belief, Quanta of Evidence or Constitutional Guarantees?" Vanderbilt Law Review **35**(6): 1293-1335.

MEADOW, WILLIAM and CASS R. SUNSTEIN (2001). "Statistics, Not Experts." Duke Law Journal **51**(2): 629-646.

MENDEZ, MIGUEL ANGEL (1984). "California's New Law on Character Evidence: Evidence Code Section 352 and the Impact of Recent Psychological Studies." UCLA Law Review **31**(5): 1003-1060.

MITCHELL, GREGORY (2003). "Mapping Evidence Law." Michigan State Law Review **2003**(4): 1065-1148.

MORENO, JOËLLE ANNE (2003). "Translating 'Visions of Rationality' into Specific Legal Reforms." Michigan State Law Review **2003**(4): 1175-1188.

NATIONAL ACADEMY OF SCIENCES (2009). Strengthening Forensic Science in the United States: A Path Forward. http://www.nap.edu/catalog/12589.html.

PARK, ROGER C. (1996). "Character Evidence Issues in the O.J. Simpson Case." University of Colorado Law Review **67**(4): 747-776.

RACHLINSKI, JEFFERY J. (2011). "The Psychological Foundations of Behavioral Law and Economics." University of Illinois Law Review **2011**(5): 1675-1696.

RACHLINSKI, JEFFREY J. (2000). "Heuristics and Biases in the Courts: Ignorance or Adaptation?" Oregon Law Review **79**(1): 61-102.

RISINGER, D. MICHAEL (2007). "Innocents Convicted: An Empirically Justified Factual Wrongful Conviction Rate." Journal of Criminal Law and Criminology **97**(3): 761-806.

ROBBENNOLT, JENNIFER K. (2005). "Evaluating Juries by Comparison to Judges: A Benchmark for Judging?" Florida State University Law Review **32**(2): 469-509.

SAKS, MICHAEL J. and ROBERT F. KIDD (1980). "Human Information Processing and Adjudication: Trial by Heuristics." Law & Society Review **15**(1): 123-160.

SANCHIRICO, CHRIS WILLIAM (2003). "Finding Error." Michigan State Law Review **2003**(4): 1189-1203.

SANCHIRICO, CHRIS WILLIAM (2004). "Evidence, Procedure, and the Upside of Cognitive Error." Stanford Law Review **57**(2): 291-365.

SANCHIRICO, CHRISTOPHER WILLIAM (2001). "Character Evidence and the Object of Trial." Columbia Law Review **101**(6): 1227-1311.

SCHACTER, DANIEL L., ROBYN DAWES, LARRY L. JACOBY, DANIEL KAHNEMAN, RICHARD LEMPERT, HENRY L. ROEDIGER, and ROBERT ROSENTHAL (2008). "Policy Forum: Studying Eyewitness Investigations in the Field." Law & Human Behavior **32**(1): 3-5.

SEIDMANN, DANIEL J. and ALEX STEIN (2000). "The Right to Silence Helps the Innocent: A Game-Theoretic Analysis of the Fifth Amendment Privilege." Harvard Law Review **114**(2): 430-510.

SIMON, DAN (2004). "A Third View of the Black Box: Cognitive Coherence in Legal Decision Making." University of Chicago Law Review **71**(2): 511-86.

SIMON, RITA JAMES, and LINDA MAHAN (1971). "Quantifying Burdens of Proof: A View from the Bench, the Jury and the Classroom." Law & Society Review **5**(3):319-30.

SKOLNICK, JEROME H. (1961). "Scientific Theory and Scientific Evidence: An Analysis of Lie-Detection." Yale Law Journal **70**(5): 694-728.

SLOBOGIN, CHRISTOPHER (1999). "Why Liberals Should Chuck the Exclusionary Rule." University of Illinois Law Review **1999**(2): 363-446.

SMITH, HUBERT W. (1942). "Components of Proof in Legal Proceedings." Yale Law Journal **51**(4): 537-581.

TANFORD, J. ALEXANDER (1990). "The Law and Psychology of Jury Instruction." Nebraska Law Review **69**(1): 71-111.

TILLERS, PETER (1998). "What Is Wrong with Character Evidence?" Hastings Law Journal **49**(3): 781-834.

TRIBE, LAURENCE H. (1971). "Trial by Mathematics: Precision and Ritual in the Legal Process." Harvard Law Review **84**(6): 1329-1393.

TVERSKY, AMOS and DANIEL KAHNEMAN (1973). "Availability: A heuristic for judging frequency and probability." Cognitive Psychology **5**(1): 207-233.

TVERSKY, AMOS and DANIEL KAHNEMAN (1974). "Judgment Under Uncertainty: Heuristics and Biases." Science **185**(4157): 1124-1131.

VARS, FREDRICK E. (2010). "Toward a General Theory of Standards of Proof." Catholic University Law Review **60**(1): 1-45.

WELLBORN, OLIN GUY III (1991). "Demeanor." Cornell Law Review **76**(5): 1075-1105.

WISTRICH, ANDREW J., CHRIS GUTHRIE, and JEFFREY J. RACHLINSKI (2005). "Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding." University of Pennsylvania Law Review **153**(4): 1251-1345.

WOOCHER, FREDIRIC D. (1977). "Did Your Eyes Deceive You? Expert Psychological Testimony on the Unreliability of Eyewitness Identification." Stanford Law Review **29**(5): 969-1030.

WORTHINGTON, DEBRA L. MERRIE JO STALLARD, JOSEPH M. PRICE, and PETER J. GOSS (2002). "Hindsight Bias, *Daubert*, and the Silicone Breast Implant Litigation." Psychology Public Policy & Law **8**(2): 154-179.

ZAMIR, EYAL and ILANA RITOV (2012). "Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation." Journal of Legal Studies **41**(1): 165-207.

ZAMIR, EYAL, ILLANA RITOV, and DORON TEICHMAN (2012). "Seeing Is Believing: The Anti-Inference Bias." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1989561 (Apr. 21, 2012).