

# SEEING IS BELIEVING: THE ANTI-INFERENCE BIAS

Eyal Zamir, Ilana Ritov, and Doron Teichman\*

1. INTRODUCTION
2. DOCTRINAL AND THEORETICAL BACKGROUND
3. EXPERIMENTAL FINDINGS
  - 3.1. Experiment 1 (“Highway”): Conviction Based on Technological Evidence
  - 3.2. Experiment 2 (“Enforcement”): Ex ante Policy Making
  - 3.3. Experiment 3 (“Bus”): Conviction Based on Eyewitness Testimony
  - 3.4. Experiment 4 (“Antibiotics”): Forensic Evidence in Civil Litigation
4. DISCUSSION
  - 4.1. Explanations
  - 4.2. Normative Implications
5. CONCLUSION

## 1. Introduction

Sherlock Holmes, Hercules Poirot, Sir Henry Merrivale, and even Dr. Gregory House have long been admired for their ability to solve mysteries using logic, deductive reasoning, and inferences from circumstantial evidence. However, the creators of detective stories and TV dramas almost never rely solely on this type of evidence. The detective story or TV episode will typically include a climactic scene in which direct evidence, such as a confession by the culprit, is presented. This serves to reassure the readers and viewers that the inferential arguments were indeed correct. The implicit assumption seems to be that inferences are insufficient to determine the culpability of the suspect (an assumption which is reinforced by the technique of incorporating misleading circumstantial evidence into detective stories and TV dramas).

The same is plausibly true in the courtroom. There are grounds to believe that judicial factfinders treat direct and circumstantial evidence differently and are more reluctant to impose liability on the basis of circumstantial evidence alone (Heller 2006, pp. 247-55). In exhibiting this reluctance, factfinders sometimes follow legal norms that caution against relying on circumstantial evidence, but more often disregard legal norms that deny the relevance of this distinction (Greenstein 2009). Scholars largely

---

\* EYAL ZAMIR is Augusto Levi Professor of Commercial Law, Faculty of Law, Hebrew University of Jerusalem. ILANA RITOV is Professor at the School of Education and Center for Rationality, Hebrew University of Jerusalem. DORON TEICHMAN is Joseph H. and Belle R. Braun Senior Lecturer in Law, Faculty of Law, Hebrew University of Jerusalem. We would like to thank Netta Barak-Corren, Oren Gazal-Ayal, Ofer Malcai, Amos Schurr, Alex Stein, and Fredrick Vars for valuable comments on earlier drafts, and Michael Cohen, Meirav Furth, and Ori Katz for excellent research assistance. This research was supported by the Israel Science Foundation (grants No. 100/11 and 1283/11) and the Aharon Barak Center for Interdisciplinary Legal Research.

agree that this tendency is indefensible (Wigmore 1983, pp. 957-64; Heller 2006; Greenstein 2009). In fact, it is sometimes noted that circumstantial evidence is *more* reliable than direct evidence, and thus discounting its probative value is a “paradox” (Heller 2006, pp. 247-55; see also Greenstein 2009, p. 1803).

Several explanations have been offered for the tendency not to impose liability solely on the basis of circumstantial evidence, but most are incomplete at best. They are insufficient because too often they conflate the direct vs. circumstantial evidence distinction with related ones, such as between eyewitness testimonies and scientific data, between concrete and statistical proof, and between probabilistic and conclusive evidence. Once we recognize that circumstantial evidence may be either eyewitness or forensic, concrete or statistical, probabilistic or conclusive, and so forth, we can appreciate the limitations of theories that rest on these distinctions.

This article describes new experiments demonstrating that even when we have controlled for such factors as the eyewitness/forensic, concrete/statistical, and probabilistic/conclusive character of the evidence, the disinclination to impose liability based on non-direct evidence remains. While our findings do not necessarily refute the existing theories, they indicate that these theories are incomplete and point to the existence of a deep-seated bias against basing liability on inferences from circumstantial evidence—an anti-inference bias. Like other biases and heuristics, the anti-inference heuristic very often yields accurate judgments, yet it also gives rise to systematic errors.

The article is structured as follows. Section 2 defines circumstantial evidence and surveys the extant explanations for the persistent reluctance to base criminal and civil liability on such evidence. Section 3 presents four experiments designed to inquire whether this reluctance is unique to statistical, probabilistic, inconclusive, forensic, or non-case-specific evidence. Section 4 first analyzes the contribution of our results to a better understanding of the phenomenon and then discusses the potential normative implications of our findings.

## **2. DOCTRINAL AND THEORETICAL BACKGROUND**

Direct evidence, if accepted, proves a material fact without the mediation of a deductive process. Circumstantial evidence, in contrast, even if accepted, requires an additional mental step of inference in order to determine whether the material fact did or did not exist (Moore et al. 1996, pp. 2-3; McCormick 2006, p. 308; Murphy 2008, pp. 20-21). The very existence of this distinction is contested analytically. Some commentators argue that any determination of facts by judges or juries—including determination based on eyewitness testimony—inevitably involves inferences (Wigmore 1983, pp. 952-56); hence direct evidence is a myth (Greenstein 2009).<sup>1</sup> Indeed, elements of what is conventionally viewed as direct evidence are sometimes circumstantial. For instance, when a litigant takes the stand and testifies that she made an oral agreement with the defendant on a certain date, her testimony would conventionally be viewed as direct evidence. However, typically much of this testimony will refer to the prior relationships between the parties and their subsequent

---

<sup>1</sup> This position echoes the reductionist view in the philosophical debate over the epistemological problems of testimony (see generally Coady 1992; Adler 2006).

behavior, in order to corroborate the litigant's assertion that the alleged agreement was actually made. These parts of her testimony are analytically circumstantial evidence (Bergman 1996, pp. 988-89; Moore et al. 1996, pp. 3-4). One may thus distinguish between evidence that is perceived as wholly circumstantial and evidence that is not. In addition, to some extent the distinction between the two types of evidence is a matter of degree: Some evidence requires more inferential steps than others to draw a conclusion about the material fact. These complexities notwithstanding, the distinction between direct and circumstantial evidence is entrenched in popular and legal vocabulary. Whether or not one accepts the analytic critique of the distinction, this critique does not negate the impact of the distinction on people's perceptions and decisions. Inasmuch as people treat circumstantial evidence differently than direct evidence, this differentiation may have considerable ramifications in both criminal and civil litigation, as often there is no direct evidence bearing on the disputed facts.

As a doctrinal matter, some legal systems instruct, or used to instruct, factfinders to avoid convictions based exclusively on circumstantial evidence (on the rule in Jewish law, see, e.g., Rosenberg and Rosenberg 1995, pp. 1376-90). Other legal systems have allowed factfinders to rely on such evidence, yet have required them to exercise special caution in both criminal and civil cases (Wigmore 1983, pp. 957-61, 963; Rosenberg and Rosenberg 1995, pp. 1390-402; Greenstein 2009). At times, stricter limitations have been imposed on drawing an inference upon an inference, namely inferring the existence or non-existence of a material fact from another fact that was itself inferred from other evidence. According to one version of this limitation, even in civil litigation, an inference upon an inference is impermissible, and according to another version it is only permitted if the first inference meets the "beyond a reasonable doubt" standard (Cohen 1977, pp. 68-73; Eggleston 1983, pp. 35, 39-40, 237-40; Wigmore 1983, pp. 1106-19; Foster 1994).

Clearly, the modern trend (rooted in early modern England, Shapiro 1991, pp. 240-41), is to abandon rules limiting the use of circumstantial evidence, including an inference upon an inference. Instead, factfinders are instructed to determine the reliability and weight of any evidence, direct or indirect, without prejudice against the latter (Eggleston 1983, pp. 35, 39-40, 237-40; Wigmore 1983, pp. 957-64, 1120-38; Rosenberg and Rosenberg 1995, pp. 1400-02; Greenstein 2009, p. 1803; Phipson 2010, p. 5). Vestiges of the distrust of circumstantial evidence can nevertheless be found in jury instructions in some jurisdictions (e.g. Mississippi Model Jury Instructions – Criminal §1:18 (2011): "The court instructs the jury that if the State has relied on circumstantial evidence to establish its theory of guilt of the defendant, then the evidence for the State must be so strong as to establish the guilt of the defendant, not only beyond a reasonable doubt, but the evidence must be so strong as to exclude every other reasonable hypothesis other than that of guilt") and in case law (Hammarskjöld 2011, pp. 1120-22; see also Phipson 2010, p. 414). In the same vein, while the rule restricting an inference upon an inference has largely been abandoned, it is still in force in some jurisdictions (Foster 1994).

Commentators generally agree that the prejudice against circumstantial evidence is unwarranted (e.g. Wigmore 1983, pp. 957-64; Foster 1994; Greenstein 2009; but see Rosenberg and Rosenberg 1995, pp. 1402-22). Yet, empirical and experimental studies have long demonstrated that factfinders are far more inclined to base their conclusions

on direct evidence. Factfinders tend to undervalue the reliability and probative value of circumstantial evidence, including statistical and forensic (Goodman 1992; Nance and Morris 2005; Heller 2006, pp. 250-52), and to overvalue the trustworthiness and weight of direct evidence, such as eyewitness testimonies (Lindsay et al. 1981; Penrod and Cutler 1995; Heller 2006, pp. 247-50, 253-54).

Several explanations have been offered for this phenomenon. One is that unlike direct evidence, inferences are based on generalizations “which by definition are accurate less than 100% of the time” (Bergman 1996, p. 988). Even reliable circumstantial evidence may be compatible with numerous competing inferences; hence it is necessarily “polyvocal,” to use the term coined by Heller (Heller 2006, pp. 267-68; see also Moore et al. 1996, pp. 4-7). Another explanation does not assume that circumstantial evidence is objectively less conclusive than direct evidence, yet argues that it may be perceived as such. Focusing on people’s reluctance to use statistical evidence as a basis for imposing liability, it has been argued that factfinders make decisions on the basis of subjective probability estimates, which often differ from the objective probability (Tribe 1971; Thompson and Schumann 1987). Factfinders may believe that the reliability and probative weight of circumstantial evidence is considerably lower than it actually is, and therefore be reluctant to rely on it (the opposite would hold for direct evidence).

A third explanation focuses on the typical features of eyewitness testimonies, often atypical of circumstantial evidence. *Inter alia*, eyewitness testimonies provide “a verbal representation of the crime itself,” whereas indirect evidence is often abstract (Heller 2006, p. 265). Direct evidence is a story-like narrative, while arguments based on circumstantial evidence often resemble deductive reasoning (*id.*, pp. 265-67). The former are often vivid, concrete, and stirring, while the latter are pallid, general, and unexciting (*id.*, pp. 276-80). These characteristics make it easier for factfinders to form a coherent story of the events from direct evidence (Cf. Scurich and John 2011, pp. 58-59).

A fourth explanation, resonating with both the alleged generalization-based nature of circumstantial evidence (underlying the first explanation) and its non-concrete quality (part of the third explanation), has to do with the case-specificity of the evidence. A central justification for the reluctance to accept a claim on the basis of naked statistical evidence draws on the distinction between the probability that a certain fact is true and the weight or resiliency of the evidence supporting the claim. Decision-makers may sensibly reject a claim even if the probability that the plaintiff’s version is correct meets the controlling standard of proof, if assessment of this probability rests on too little information or on general, non-case-specific evidence (Cohen 1977; Kaye 1987; Stein 2005, pp. 40-56, 80-106).

Yet another possibility is that from the perspective of the factfinders, decisions based on testimonies and other direct evidence involve a smaller degree of responsibility. If it turns out that a testimony was inaccurate or deceptive, the factfinder can rationalize that the witness is to blame for the erroneous verdict. In contrast, if it transpires that liability was erroneously imposed due to a faulty inference from circumstantial evidence, the responsibility arguably lies with the factfinder (Heller 2006, p. 287). To avoid a feeling of regret, a factfinder would likely be disinclined to rely on circumstantial evidence. This explanation bears some similarity

to the epistemological notion that when a speaker makes testimonial assertions, she undertakes responsibility for the accuracy of what she says; and when a hearer takes a speaker's word for something, she ascribes her authority, a normative ascription (Adler 2006; Keren 2007).

The sixth and seventh theories were originally offered in the context of naked statistical evidence and later extended to circumstantial evidence more generally. Gary Wells (1992, p. 746) conjectured that “in order for evidence to have a significant impact on people's verdict preferences, one's hypothetical belief about the ultimate fact must affect one's belief about the evidence.” For example, the mere fact that 80% of the buses in a certain town belong to the blue bus company and 20% to the grey bus company is insufficient to find the former liable for an accident caused by an unidentified bus, because the determination of liability would not change one's belief about the accuracy of the statistical fact. In contrast, when a weigh-station attendant testifies that according to his records, a blue bus was weighed in the nearby station just before the accident—thus tying that bus company to the accident—the determination of liability would more likely bear on the reliability of this testimony, even if the defendant had already established that the records were wrong 20% of the time.

Finally, according to the ease-of-simulation explanation—offered by Keith Niedermeier and his coauthors (1999), inspired by Daniel Kahneman and Amos Tversky's simulation heuristic (1982), and later elaborated on by Kevin Heller (2006)—the willingness to ground liability on statistical/circumstantial evidence depends on how easily one can imagine an alternative scenario that would be compatible with the evidence.<sup>2</sup> Arguably, it is easier to imagine such a scenario when the only evidence is circumstantial, since by its very nature, such evidence does not prove the material fact itself. Hence, factfinders are considerably less willing to find defendants liable on the basis of such evidence. This explanation is associated with the third explanation discussed above, as direct evidence helps the factfinder to form a coherent story of the events, which then makes it harder to imagine an alternative story. In the same vein, Deanna Sykes and Joel Johnson (1999) argued that since comprehension of a witness's assertion entails an initial belief in that assertion, undoing the mental representation created by the testimony and imagining an alternative scenario require cognitive effort. Such an effort is not required when

---

<sup>2</sup> Following the experimental design of Wells (1992), Niedermeier et al. (1999, pp. 536-39) presented the participants in their study with various vignettes describing a legal suit for damages. The lawsuit was filed by a woman whose dog was killed by a bus. In all the vignettes, there was an objective probability of 80% that the defendant—one of the only two bus companies in the county—was responsible for the accident. In the “complete match” version, the evidence was based on a comparison between the tracks left by the bus that ran over the dog and the tracks of all the buses belonging to the two companies. Each company operated ten buses, and the track prints matched eight of the defendant's buses and two buses belonging to the other company. In the “partial match” version, there was a partial match with only one of the defendant's buses and one of the other company's buses, indicating that there was an 80% probability that the defendant's bus was the one that ran over the dog, and a 20% chance that it was a bus belonging to the other company. The responders were asked to indicate whether they would accept the claim and to assess the probability that the defendant's bus was involved in the accident. While the assessed probability was similar under the two conditions, the willingness to accept the claim was much lower in the complete-match condition, arguably because in this case it was easier to imagine that a bus belonging to the other company killed the dog. Cf. Scurich and John (2011, p. 58).

decision-makers are presented with probabilistic evidence that does not include a concrete assertion that needs to be undone.

All of these explanations can be—and have been—challenged. Thus, for example, the generality of the subjective probability argument (the second explanation), was called into question by Wells (1992), who instead offered the fact-to-evidence theory. Wells's theory was in turn challenged by Niedermeier et al. (1999) and Sykes and Johnson (1999), who proposed the ease-of-simulation explanation. The ease-of-simulation explanation was subsequently called into question by Arkes et al. (2011). As the experiments in the next section demonstrate, none of these arguments captures the full scope of the factfinders' bias against indirect evidence.

### 3. EXPERIMENTAL FINDINGS

#### 3.1. Experiment 1 (“Highway”): Conviction Based on Technological Evidence

The first experiment was designed to examine how the type of evidence impacts on the willingness to impose criminal liability, namely convicting a driver of a traffic violation. Specifically, the experiment examined the difference between evidence that directly establishes the defendant's fault and evidence that establishes fault indirectly, requiring an additional inferential step. Both the direct and the inference evidence were technological.

*3.1.1. Participants.* Eighty-eight law students (57 women, 31 men), who had finished their first year of studies, participated in this study. They ranged in age from 19 to 40, with a mean of 24.

*3.1.2. Experimental Design.* We used a between-subject design, presenting each participant with one of two versions of a short scenario, followed by three questions. The “direct-evidence” condition read as follows:

Speed cameras were installed on a toll road. The probability of an error in the camera system is 2%. The speed limit on this road is 100 KPH. According to the camera, a driver drove his car at a speed of 125 KPH late at night.

The “inference” condition was:

Cameras that document the exact time at which each vehicle passes by them were installed on a toll road. The cameras do not document the speed of the passing vehicle, but from the distance between the cameras and the time that elapses between the points they document, it is possible to infer the driver's speed in that section of the road. The probability of an error in the camera system is 2%. The speed limit on this road is 100 KPH. According to the time elapsed and the distance between the two cameras, a driver drove his car at a speed of 125 KPH late at night.

Responders were reminded that the standard of proof in such cases is “beyond a reasonable doubt” and asked to indicate whether they would convict the driver of violating the speed limit. Those who decided affirmatively were then asked to indicate the penalty they would impose on the driver. The penalty could range from 50 NIS to 1000 NIS (one NIS roughly equals 0.25 US\$). Finally, all responders were asked to

rate the fairness of conviction on the basis of the evidence presented, using a scale of 0 to 100, where 0 indicates “absolutely unfair” and 100 “absolutely fair.”

The two questionnaires were posted on a website designed for controlled judgment and decision-making experiments. Responders were recruited by an e-mail message sent to students at the Faculty of Law of the Hebrew University of Jerusalem. To encourage participation, one out of ten participants was randomly selected to win a prize of 100 NIS. Responders were randomly assigned to one of the two conditions.

*3.1.3. Results.* The rate of convictions differed by condition: 81.4% of the responders in the direct-evidence condition decided to convict the driver, while only 60% of the responders in the inference condition did (chi-square=4.835,  $p=.028$ ). The average fine, counting acquittals as a fine of zero, was not significantly affected by condition (300.3 and 286.2 for direct-evidence and inference, respectively;  $p=.832$ ).<sup>3</sup> The mean fairness rating of convicting the driver in the direct-evidence condition (74.8 on a scale of 0 to 100, where 0 is absolutely unfair and 100 absolutely fair) was higher than in the inference condition (66.1 on the same scale), albeit marginally significant ( $t(86)=1.64$ ,  $p=.1$ ).

Interestingly, the effect of the condition, direct evidence vs. inference, was possibly mediated by the effect of this factor on the judgment of fairness. Predicting conviction decision by a logistic regression analysis that includes both the condition (direct evidence vs. inference as a dummy variable, coded as 1 for direct evidence and 0 for inference condition) and fairness judgments as independent variables renders the effect of the condition non-significant ( $B=.782$ ,  $Wald=1.422$ ,  $p=.233$  and  $B=.072$ ,  $Wald=20.696$ ,  $p<.001$  for condition and fairness, respectively).

## **3.2. Experiment 2 (“Enforcement”): Ex ante Policy Making**

The second experiment examined whether the results obtained in experiment 1, which used a between-subject design mimicking an ex post judicial decision, would be replicated in a within-subject design mimicking an ex ante choice of an enforcement system. It also tested the strength of the responders’ preference for any system by exploring whether they would favor it even if it were less accurate than the alternative system.

*3.2.1. Participants.* Forty-nine first-year law students participated in the study (31 women, 18 men). Their ages ranged from 18 to 32, with a mean of 23.

*3.2.2. Experimental Design.* Participants were asked to imagine that they were members of a team of experts whose job is to set policy regarding the enforcement of traffic laws. They were asked to choose between two speed-enforcement systems. One system was based on speed-enforcement cameras, as in the direct-evidence condition of experiment 1. The other system was based on two cameras documenting the exact time at which a car passes by, from which it is possible to infer the driver’s speed, as in the inference condition of experiment 1. It was further indicated that the cost of

---

<sup>3</sup> Similarly, the average fine excluding the acquittals did not significantly differ between the two conditions (368.9 vs. 477.0 for direct evidence and inference, respectively,  $t(60)=1.472$ ,  $p=.15$ ).

purchasing, installing, and operating the two systems was identical and that the accuracy of both systems was 99%. The order of presentation of the two systems, labeled A and B, was varied between subjects. The students answered the questionnaires voluntarily in one of their law classes.

Responders were given three options: preference for system A, preference for system B, and indifference. Responders with a preference for one of the two systems were then asked whether they would stick with their choice even if it was 98% accurate and the other system remained 99% accurate. Finally, they were asked to indicate the lowest level of accuracy at which they would still choose their preferred system over the alternative, assuming the other remained 99% accurate.

*3.2.3. Results.* Order of presentation did not significantly affect preference (chi-square=1.414,  $p=.493$ ). Of the 49 responders, 53.1% preferred the “direct-evidence system,” 24.5% preferred the “inference system,” and 22.4% were indifferent. The percent of responders who preferred the direct-evidence system was significantly higher than the percent preferring the inference one ( $z=2.953$ ,  $p<.01$ ).

Of the 26 responders who preferred the direct-evidence system, 11 would have preferred it even if it were 98% accurate and 13 would not (two did not answer this question). The mean lowest accuracy they would accept was 95.2%. Of the 12 responders who preferred the inference system, eight would have preferred it even if it were 98% accurate and four would not. The mean lowest accuracy they would accept was 92.2%. This seems to suggest that the subjects who preferred the indirect-evidence enforcement regime exhibited a stronger preference for this technology relative to the preference exhibited by subjects who preferred the direct-evidence regime. The difference, however, was not statistically significant.

### **3.3. Experiment 3 (“Bus”): Conviction Based on Eyewitness Testimony**

Experiment 3 was designed to examine how general the findings of experiments 1 and 2 were. First, we wanted to investigate whether the results were due to any unique feature of the pertinent enforcement mechanisms. For example, one could conjecture that responders were reluctant to convict in the inference condition because the two-camera system might have been perceived as more intrusive and a greater threat to people’s privacy. Second, we wanted to see whether the results of experiment 1 were a function of the technological nature of the evidence or whether they would hold for eyewitness testimony as well. Third, we wanted to test whether the results are mediated by the responders’ subjective probability estimates, which may differ from the objective probability. Finally, we were interested in whether the difference between the two conditions in experiment 1 had to do with the perceived responsibility-taking by the decision-maker. Thus, instead of asking participants how they would decide the case, we asked them how a judge should decide the case, to see whether the difference would disappear if it were not the responder’s decision but rather somebody else’s.

*3.3.1. Participants.* One-hundred-and-seventeen first-year law students (59 women, 58 men) answered the questionnaire in the first week of their studies. They ranged in age from 18 to 36, with a mean of 23.

*3.3.2. Experimental Design.* As in experiment 1, this experiment used a between-subject design, employing two versions of a short scenario, followed by three questions. The two versions described a tourist bus that was stuck in a sparsely populated area late in the evening. A policeman who arrived on the scene assisted the driver in arranging for minibuses to bring the tourists to their destination. In the direct-evidence condition, the policeman, who got on the bus, noticed that there were 54 tourists on it, despite the fact that its permit allowed it to carry only 50 passengers. In the inference condition, two minibuses arrived empty, were filled up, and drove away. It then turned out that four tourists were left and an additional vehicle was summoned for them. Since each minibus had seats for 25 passengers, the policeman inferred that the bus driver had violated the terms of the permit, which allowed only 50 passengers on the bus. In both scenarios, based on the policeman's report, the driver was accused of carrying too many passengers. The driver pleaded not guilty. The judge had the impression that the policeman was a reliable person and assessed that the likelihood that he had miscounted the passengers was very low: 1 in 20.

After being reminded of the standard of proof in criminal proceedings ("beyond a reasonable doubt"), participants were asked whether in their opinion the judge should find the driver guilty. They were then asked to assess the probability that the driver had actually transported more passengers than permitted. Finally, the participants were asked to rate the fairness of convicting the driver on a scale of 1 to 9, where 1 indicates that a conviction is absolutely unfair and 9 that it is absolutely fair. The students answered the questionnaires voluntarily in one of their law classes. They were randomly assigned to one of the two conditions.

*3.3.3. Results.* The rate of convictions differed by condition: 64.8% of the participants in the direct-evidence condition determined that the judge should convict the driver, while only 38.1% in the inference condition were of this opinion (chi-square=8.304,  $p=.004$ ).

The participants estimated the probability of the driver actually violating the terms of the permit as considerably lower than the probability arguably implied by the vignette. While the chance of miscounting according to the Judge's impression was described as 1 in 20, implying a 95% reliability of the testimony, the mean estimated probability of the driver actually committing the violation was 80.1% across the two conditions. The mean probability estimates were not significantly different for the two conditions: 82.2% and 78.3% for direct-evidence and inference conditions, respectively ( $p=.29$ ).

Importantly, although the probability estimate affected the conviction decision, it did not mediate the effect of the condition. A logistic regression analysis predicting the conviction decision by condition (direct evidence vs. inference, coded as in Experiment 1) and subjectively estimated probability yielded significant effects of both probability and condition ( $B=1.091$ ,  $Wald=7.092$ ,  $p<.01$  and  $B=.038$ ,  $Wald=11.097$ ,  $p<.01$  for the effects of the condition and the subjective probability, respectively). The difference between the two conditions remains highly significant even when controlling for the estimated probability.

Convicting the driver was perceived as fairer in the direct-evidence condition than in the inference condition: mean ratings of 6.09 and 4.70 for direct evidence and inference, respectively, on a scale of 1 to 9, where 1 is absolutely unfair and 9 absolutely fair ( $t(115)=3.005$ ,  $p=.003$ ). Across the two conditions, probability estimates were significantly correlated with fairness judgment ( $r=.279$ ,  $p<.01$ ). The effect of condition (type of evidence) on fairness judgment was not mediated by probability estimates. A regression analysis predicting fairness judgment by condition (direct evidence vs. inference) and subjectively estimated probability yielded significant effects of both probability and condition ( $F(114)=7.829$ ,  $p<.01$  and  $F(114)=8.508$ ,  $p<.01$  for the effects of the condition and the subjective probability, respectively).

The effect of the condition, direct evidence vs. inference, on the conviction decisions was possibly mediated by the effect of this factor on judged fairness. Predicting conviction decision by a logistic regression analysis that includes both condition (direct evidence vs. inference) and fairness judgment as independent variables renders the effect of condition non-significant ( $B=1.278$ ,  $Wald=30.291$ ,  $p<.001$  and  $B=.719$ ,  $Wald=1.090$ ,  $p=.297$  for the effects of fairness and condition, respectively). Thus the difference between the two conditions with respect to conviction rate is no longer significant when we control for judged fairness. It is also possible, however, that the similar effects of the condition on the decision to convict and on the perceived fairness of the conviction may reflect two outcomes of the same cognitive process.

### **3.4. Experiment 4 (“Antibiotics”): Forensic Evidence in Civil Litigation**

The fourth experiment had several purposes. First, it broadened our focus to include civil litigation. Second, it extended the inquiry to a case where the question was not whether a wrong had been committed, but rather who had committed it. Third, whereas the previous experiments related to technological and eyewitness evidence, this one dealt with laboratory findings. Fourth, since some jurisdictions set particularly strict limits on drawing an inference upon an inference, the present experiment compared not only direct evidence and inference, but also single and double inference. Finally, while the participants in the previous experiments were first-year law students (or students who completed their first year of legal studies), participants in the present experiment were advanced law students, including master’s candidates (many of whom are practicing lawyers).

*3.4.1. Participants.* One-hundred and ninety three advanced LL.B. and LL.M. students responded to the questionnaire (100 women, 93 men). They ranged in age from 19 to 58, with a mean of 26.

*3.4.2. Experimental Design.* We used a between-subject design, employing three versions of a vignette describing a civil action, followed by three questions. Responders were asked to imagine that they were serving as a judge in a suit for damages filed by a small dairy against a dairy farmer. The farmer sells the dairy the milk he produces. According to the contract, the farmer should make sure that there are no antibiotics residues in the milk, because such residues obstruct the production of

various products. The milk is delivered to the dairy by a tank truck that transports the milk of two farmers. Since the milk of the two farmers is mixed in the tank, a sample is taken from each farmer's milk before pumping it into the tanker, and the samples are delivered to a laboratory, where they will be examined should a problem arise. The dairy claims that the farmer provided milk containing antibiotics residues, and as a consequence yogurt production failed and the dairy had to discard all of the raw materials it had used in this process. As a result, the dairy suffered a loss of 30,000 NIS.

In the direct-evidence condition, when the yogurt-production process failed, the milk samples of the two farmers were examined. This examination revealed that there were antibiotics residues in the defendant's milk and no residues in the other farmer's milk. The probability that the results of the laboratory examinations are correct (that is, that there were antibiotics residues in the defendant's milk but not in the other farmer's milk) is 85%. It is undisputed that the source of the antibiotics residues could only be milk from one of the farmers.

In the single-inference condition, when the yogurt-production process failed, the milk samples from the two farmers were about to be examined (here too, it was undisputed that the source of the antibiotics residues could only be milk from one of them), but it turned out that the sample of the defendant's milk had been lost in the laboratory, and therefore it was only possible to examine the other farmer's sample. This examination showed no antibiotics residues. The probability that the results of the laboratory examination are correct (that is, that there were no antibiotics residues in the other farmer's milk) is 85%.

In the double-inference condition, when the yogurt-production process failed, the first possibility examined was that the failure was due to other substances added to the milk during the process. This examination revealed that none of the other substances were tainted. When the milk samples from the two farmers were then about to be examined (here again, it was undisputed that the source of the antibiotics residues could only be milk from one of the farmers), it turned out that the sample of the defendant's milk had been lost in the laboratory, and therefore it was only possible to examine the other farmer's sample. This examination showed no antibiotics residues. The probability that the results of the laboratory examinations are correct (that is, that the production failure was not caused by other substances or by antibiotics residues in the other farmer's milk) is 85%. Based on the test results of the other substances, the dairy claims that one should infer that the milk caused the failure of the yogurt-production process, and based on the results relating to the other farmer's sample, one should infer that the defendant was the one responsible.

Responders were then asked (1) whether they would accept the claim, (2) what was their assessment of the probability that there were antibiotics residues in the defendant's milk, and (3) how they would rate the fairness of accepting the claim. The three questionnaires were posted on the same website and participants were recruited in the same way as in experiment 1. To encourage participation, four participants were randomly selected to win a prize of 150 NIS each. Responders were randomly assigned to one of the three conditions.

*3.4.3. Results.* Since the single-inference and double-inference conditions did not differ with respect to any of the measures, we collapsed the data across the two inference conditions and compared the direct-evidence results to those of the combined-inference conditions. The rate of accepting the claim differed by condition: 81.7% of the participants in the direct-evidence condition accepted the claim, while only 40.2% in the inference conditions did (chi-square=31.328,  $p<.001$ ).<sup>4</sup>

The estimated probability of the existence of antibiotics residues in the defendant's milk significantly differed in the direct-evidence and the inference conditions, with means of 81.7% and 64.8% for direct evidence and inference, respectively ( $t(191)=6.663$ ,  $p<.001$ ). Not surprisingly, across all conditions, participants who indicated they would accept the claim evaluated the probability of the presence of antibiotics residues in the defendant's milk higher than participants who did not accept the claim (80.6% and 69.1%, respectively,  $t(191)=8.119$ ,  $p<.001$ ). Although the subjective probability estimates are closely related to the decision to accept the claim, the difference between the conditions with respect to the tendency to accept the claim was not mediated by the difference in probability estimates. A logistic regression predicting the decision to accept the claim from probability estimate and condition (direct evidence vs. inference, coded as in Experiment 1) yielded significant effects of both factors ( $B=.054$ ,  $Wald=25.666$ , and  $B=1.296$ ,  $Wald=10.976$ , for probability estimate and condition, respectively, with  $p<.01$  for both factors).<sup>5</sup>

A similar pattern emerges from the analysis of fairness judgment. The mean perceived fairness of accepting the claim in the direct-evidence condition was significantly higher than in the inference conditions: 4.992 vs. 7.014 on a scale of 1 to 9 (where 1 indicates "absolutely unfair" and 9 "absolutely fair") for the direct-evidence and the inference conditions, respectively ( $t(191)=7.047$ ,  $p<.001$ ). Although the fairness judgments significantly correlated with the probability estimates ( $r=.527$ ,  $p<.01$ ), the effect of condition (type of evidence) on the fairness judgment was not mediated by probability estimates. A regression analysis predicting fairness judgment by condition (direct evidence vs. inference) and subjectively estimated probability yielded significant effects of both condition and probability ( $F(190)=21.491$ ,  $p<.01$  and  $F(190)=42.398$ ,  $p<.01$  for the effects of the condition and the subjective probability, respectively).

As in experiment 3, the effect of condition, direct evidence vs. inference, was possibly mediated by the effect of this factor on judged fairness. Predicting a decision to accept the claim by a logistic regression analysis that includes both condition (direct evidence vs. inference) and fairness judgment as independent variables renders the

---

<sup>4</sup> The percentage of acceptance in the single- and double-inference conditions was 44% and 37%, respectively ( $p=.4$ ).

<sup>5</sup> These results differ from the results of Experiment 2 reported by Arkes et al. (2011). In that experiment, Arkes and his co-authors used vignettes similar to the tire-tracks scenario of Wells's experiment, yet while in one condition the witness found that the tire-tracks matched the Blue Bus company vehicles, in the other the witness found that the tracks did not match the Grey Bus company vehicles. In both conditions, the technique used was correct 80% of the time. Responders' willingness to impose liability on the Blue Bus company was more than twice greater in the first condition, and this difference was mediated by their different probability assessments. In other experiments Arkes and his coauthors describe, verdict was affected by various factors without being mediated by the assessed probability.

effect of condition non-significant ( $B=1.509$ ,  $Wald=54.372$ ,  $p<.001$  and  $B=.445$ ,  $Wald=.551$ ,  $p=.458$  for the effects of fairness and condition, respectively).

#### 4. DISCUSSION

All four experiments demonstrated an anti-inference bias: The respondents were far more reluctant to impose liability based solely on circumstantial evidence than on direct evidence. This reluctance was manifest across different scenarios; was true for both criminal and civil liability; was common to technological, scientific, and eyewitness evidence; and was evident both when the disputed question was whether a wrong had been committed and when the question was who had committed the wrong. In all four experiments, the anti-inference bias characterized case-specific circumstantial evidence, rather than merely statistical evidence. We found a strong disinclination to base liability on circumstantial evidence in both between- and within-subject experimental designs; in both ex post judicial decision-making and ex ante policy-making; and both when responders made the decision themselves and when they specified how another person should decide. It was true of students at the beginning of their first year of law school and of advanced LL.B. and LL.M. students.

While the type of evidence, direct or circumstantial, sometimes affected the subjectively assessed probability of the defendant's fault, in none of the experiments did it mediate the effect of the type of evidence on the decision of whether to impose liability. The effect of the type of evidence on this decision was possibly mediated by the effect of this factor on the judgment of fairness.

This section examines the contribution of these findings to a better understanding of the prevalent dissimilar treatment of direct and circumstantial evidence and explores the potential normative implications of our findings.

##### 4.1. Explanations

Our findings extend the empirical results reported by Wells (1992) and Niedermeier et al. (1999), and hence expose gaps in existing explanations for the differential treatment accorded to circumstantial evidence (e.g. Heller 2006).

*Probability and Probability Assessment.* Like Wells (1992), Niedermeier et al. (1999), and Arkest et al. (2011), we found that neither differences in objective probabilities nor differences in subjective probability assessments can fully account for people's reluctance to impose legal liability based on indirect evidence. In Highway, Bus, and Antibiotics, the objective probability that the defendant had committed the act attributed to him was identical in the direct-evidence and the inference conditions, yet the responders' willingness to impose legal liability differed markedly. In the same vein, despite the identical accurateness of the two enforcement systems in Enforcement, the responders strongly preferred the direct-evidence system over the inference one. The assumption that circumstantial evidence is inherently less conclusive than direct evidence thus cannot explain the prevalent disinclination to impose liability based on such evidence.

As for subjective probability assessments, in Bus, despite the fact that there was no significant difference between responders' mean probability assessments in the two conditions, about 65% of the responders found the driver guilty of violating the terms

of the permit in the direct-evidence condition, but only 38% did so in the inference condition. In Antibiotics, while there were significant differences in the subjective probability assessments between the two conditions, these differences did not mediate the difference in the responders' willingness to accept the claim under the two conditions, as this difference remained significant even when we controlled for the differences in probability assessments. In this regard too, our findings are in line with those of Wells (1992), Niedermeier et al. (1999), and Arkes et al. (2011), yet they generalize their findings and expose a wide-ranging bias against circumstantial evidence that is not limited to the context of statistical and forensic evidence. In all of our experiments, conditions differed with regard to the directness of the evidence, but not with regard to whether the evidence was statistical, scientific, or eyewitness. Our findings also extend the previous experimental findings in covering criminal cases (Highway, Enforcement, and Bus) in addition to civil ones; an ex ante choice between enforcement systems (Enforcement) in addition to ex post judicial decision (Highway, Bus, and Antibiotics); and within-subject (Enforcement) in addition to between-subject design (Highway, Bus, and Antibiotics).

*Complexity.* The incompatibility of our findings with the subjective probability explanation indicates that the greater complexity of circumstantial evidence cannot fully account for the responders' disinclination to rely on this type of evidence. Had the factfinders been reluctant to impose liability based on circumstantial evidence because of their difficulty in understanding and drawing conclusions from such evidence, one would expect this difficulty to affect, first and foremost, the judged probability of the defendant's fault, and that we would find a correlation between the subjective probability assessment and the responders' willingness to impose liability. Moreover, while in Antibiotics the circumstantial evidence (particularly in the double-inference condition) was indeed more complex than in the direct-evidence condition, this can hardly be said of Highway, Enforcement, and Bus. Finally, in Antibiotics there was a considerable difference between the single-inference and double-inference conditions in terms of their complexity, but no significant difference in terms of the responders' willingness to impose liability. To be sure, at times the complexity and clarity of the evidence certainly influence factfinders' willingness to assign liability. Nonetheless, these features cannot account for the general phenomenon we identified in our experiments.

*Abstract, General, Rhetorical, and Pallid Nature of Circumstantial Evidence.* The evidence presented in all four experiments in both conditions was either eyewitness testimony or scientific/technological; none of the scenarios involved naked statistical evidence. This would indicate that the explanations based on the relative representational vs. abstract, narrative vs. rhetorical, unconditional vs. probabilistic, vivid vs. pallid, concrete vs. general, and emotional vs. unexciting nature of the evidence—highlighted by Heller (2006) as the defining differences between direct and circumstantial evidence—are incomplete. For instance, in Highway and Enforcement the evidence produced by the one-camera and two-camera systems was equally abstract, probabilistic, pallid, concrete, and unexciting. While Heller's differences are important in some cases, they do not provide a comprehensive explanation for the reluctance to impose liability based on non-direct evidence.

*Case Specificity.* Similar observations apply to the case-specificity argument, according to which naked statistical evidence is insufficient for imposing liability even when the probability is very high, because the weight or resiliency of the evidence is equally important. If the assessed probability rests on too little information or on general, non-case-specific evidence, it should not be relied upon (Cohen 1977; Stein 2005, pp. 40-56, 80-106; but see Pundik 2008, pp. 474-87). Unlike Heller's (2006) discussion of circumstantial evidence, these explanations were never meant to apply outside of the naked-statistical-evidence debate. And indeed, they do not seem relevant to any of our experiments, none of which involved the imposition of liability based on non-case-specific, statistical evidence.

*Responsibility Taking.* The responsibility-taking explanation does not account for our findings either. According to this theory, when relying on direct evidence, the factfinder feels less accountable for erroneous judgments because it is not *her* inference, but rather the witness's fallibility or dishonesty that brought about the regrettable judicial error. As a result, the factfinder is less hesitant to impose liability based on direct evidence. In line with the findings of Niedermeier et al. (1999) and Sykes and Johnson (1999), our results do not lend support to the responsibility-taking explanation. There was no real difference between the two conditions used in the Highway, Bus, and Antibiotics scenarios in this regard. Certainly no issue of responsibility-taking was involved in the Enforcement experiment, where responders were asked to choose between two speed-enforcement systems *ex ante*, in the abstract, rather than to impose liability on any particular person.

Furthermore, in Bus responders were not asked to make the decision themselves, but rather to indicate how the judge should decide. This did not counteract the responders' reluctance to rely on circumstantial circumstances. To be sure, our experiments—whether considered on their own or together—do not refute the responsibility-taking explanation. Nonetheless, they do establish that this explanation is incomplete.<sup>6</sup>

*The Fact-to-Evidence Hypothesis.* Gary Wells's (1992) fact-to-evidence theory posits that people are far more inclined to rely on evidence when the decision of whether to impose liability bears on the reliability of the evidence. Much like Niedermeier et al. (1999, pp. 536-39), our findings do not support this conjecture. For instance, in both conditions of the Bus experiment, convicting the driver implied that the policeman's testimony was very reliable, and acquitting the driver implied that it was insufficiently reliable; yet there was a huge difference between the two conditions in the willingness to convict.<sup>7</sup>

---

<sup>6</sup> Our vignettes varied also in terms of the identity of the person who drew the inference. In Highway and Enforcement, it was "possible to infer..." from the indirect evidence; in Bus, "the policeman inferred"; and in Antibiotics the plaintiff urged the court (that is, the responder) to infer. Arguably, responders should have experienced greater personal responsibility in Antibiotics and less in Bus. In fact, the disinclination to impose inference-based liability was common to all four scenarios.

<sup>7</sup> In Enforcement, it could hardly be argued that choosing one system necessarily implied that it was more accurate than the alternative. In that experiment, 38 of 49 responders preferred one of the systems (one-camera/direct system or two-camera/inference system) to the other. Of those 38, 19 (50%) would stick with their preferred choice even if its accuracy were lower than that of the alternative.

*Ease of Simulation.* The theory favored by Niedermeier et al. (1999), Sykes and Johnson (1999), and Heller (2006) is the ease of simulation. As originally conceived by Kahneman and Tversky (1982) and subsequently studied in various contexts (e.g. Levi and Pryor 1987; Raune et al. 2005), the simulation heuristic pertains to how likely one thinks it is that an outcome will occur or an event has occurred (or could have occurred). The easier it is to imagine a scenario (and the harder it is to imagine a competing one), the more probable it appears to be. The ease of simulation may depend on the ease of simulating other similar events (Miller et al. 1989) and counterfactuals (Hirt and Markman 1995), as well as on the perceived normality of different events (Kahneman and Miller 1986; Miller et al. 1989). The ease of simulation is to some extent manipulable (e.g. Niedermeier et al. 1999, pp. 539-41; Roese et al. 2006).

The ease of simulation may thus explain factfinders' reluctance to base their decisions on circumstantial evidence if such evidence makes it easier to imagine alternative scenarios in which the defendant is not liable, hence decreasing the assessed probability that the defendant is liable. The ease-of-simulation heuristic cannot explain factfinders' disinclination to impose liability based on circumstantial evidence when the type of evidence, direct or circumstantial, does not significantly affect the subjective probability assessment (as in *Bus* and the experiments conducted by Wells (1992) and Niedermeier et al. (1999)), or when this disinclination remains significant even when one controls for the differences in subjective probability (as in *Antibiotics*). Explaining the reluctance to base liability on purely circumstantial evidence by the ease-of-simulation heuristic thus does not fall in line with the conventional meaning of this heuristic. Niedermeier et al. (1999) and Heller (2006) apparently use the notion of ease of simulation in an idiosyncratic fashion.

Moreover, when these authors seek to demonstrate differences in ease of simulation, they refer to concrete alternative scenarios. For example, Niedermeier et al. (1999, p. 540) ask the responders in the third condition of their third experiment to imagine that a second bus obscured the tracks of the bus that ran over the dog. Similarly, Heller (2006, pp. 296-97) explains that the presence of the defendant's skin under the fingernails of the alleged victim does not rule out a scenario in which the victim scratched the defendant accidentally when he tried to help her. Our vignettes were designed to eliminate such alternative explanations. For instance, in *Highway*, given the distance between the two cameras and the time that elapsed between the documenting of the car by each camera, there is simply no way the driver didn't violate the speed limit. True, one could imagine an error in the system, but the phrasing of the two conditions was identical in this respect: "The probability of an error in the camera system is 2%." Similarly, in the inference condition of *Bus*, the only explanation for the fact that four tourists were left behind after the two minibuses that arrived empty were filled up and drove away, is that there were 54 passengers on the tourist bus. In fact, in this experiment there was no statistically significant difference between the subjective probability assessments in the two conditions.

The ease-of-simulation theory fails to provide a complete explanation for the different treatment of direct and circumstantial evidence even in Niedermeier and his coauthors' own study. Niedermeier et al. (1999, pp. 536-41) directly tested the mediating effect of the ease-of-simulation variable on verdicts in two of their

experiments. While in both experiments the ease of simulation affected the verdicts,<sup>8</sup> in neither did it fully mediate the effect of type of evidence. Although controlling for ease of simulation considerably reduced the proportion of variance in verdicts, there remained a statistically significant difference between the verdicts in the different conditions (id., pp. 538, 541). Even according to Niedermeier and his coauthors, the ease of simulation thus provides an incomplete explanation for factfinders' behavior. Direct and more elaborate examination of the ease-of-simulation explanation was also conducted by Arkes and his coauthors (2011). They found that sometimes this factor (which they dub "imaginability") possibly mediates the verdict, yet sometimes it is unrelated to the verdict or even related in a manner completely contrary to the ease-of-simulation hypothesis.

*Fairness.* Last we consider the fairness explanation arguably supported by our findings. In addition to deciding whether to convict the accused or accept the claim, the responders in Highway, Bus, and Antibiotics were asked to judge the fairness of conviction/acceptance of the claim based on the evidence presented to them. In all three experiments, there was a statistically significant positive correlation between the answers to the two questions: The fairer the responders judged the imposition of liability, the likelier they were to convict the defendant or accept the claim.

How should we understand the notion that it is less fair to impose liability based on circumstantial evidence alone? One possibility is that such an imposition of liability is unfair because it is prone to error due to the inherent inconclusiveness of such evidence, its limited weight, or its lesser resiliency. However, this does not seem to be the notion of fairness envisioned by the responders in our experiments. While in all three experiments the imposition of liability correlated with, and was possibly mediated by, the judgment of fairness, in none was the imposition of liability mediated by the assessed probability that the defendant actually committed the act.<sup>9</sup> The issue of weight or resiliency was similarly irrelevant because all evidence was case-specific.

A different notion of fairness might have to do with people's expectations and the "rules of the game." Perhaps people believe that they should only be penalized if they are caught in the act. Imposing an inference-based liability is thus seen as unfair, regardless of the reliability of the circumstantial evidence.<sup>10</sup> While such a notion of fairness possibly plays a role in some contexts, we doubt that it can serve as a general explanation for the reluctance to impose inference-based liability. First, it is unclear that in deciding whether to impose criminal or civil liability, judges and juries—or the

---

<sup>8</sup> In experiment 3 (Niedermeier et al. 1999, p. 540-41), only one of the two measures of the ease of simulation yielded the expected result, with the other producing an opposite (though not statistically significant) result.

<sup>9</sup> It may also be argued that in Highway, the speed-camera system was perceived as fairer because it is more forgiving: It does not penalize drivers who violate the speed limit in the parts of the highway not covered by a speed camera, i.e. most of the road. However, the two-camera system is more forgiving in the sense that it only captures the driver's *average* speed in the pertinent section of the road, which is typically lower than the highest speed she drove in that section.

<sup>10</sup> The rules of the game may change over time. Thus, while speed cameras have come to be regarded as a commonplace in Israel (where the experiments were conducted), two-camera systems of the kind described in Highway and Enforcement are not ordinarily used in Israel and can thus be perceived as unfair. Even if this conjecture helps to explain the results of Highway and Enforcement, it does not seem relevant to Bus or Antibiotics.

participants in our experiments—consider the wrongdoer’s expectations as to the type of evidence that might be used against her. Second, when one’s behavior harm another person, as in Antibiotics and in Well’s (1992) blue-bus scenario, not compensating the injured party because the evidence supporting her claim is not the type expected by the defendant, seems considerably more unfair than using such evidence. Finally, the present notion of fairness does not account for the disinclination to rely on circumstantial evidence even when the pertinent evidence, such as forensic evidence found at the scene of the crime, has long been used by the police and is now part of the “rules of the game.”

Yet another possibility is that the responders in our experiments understood “fair” as roughly equivalent to “appropriate” or “acceptable.” The decision to impose criminal or civil liability and the characterization of the imposition of liability as fair echoed one another, possibly reflecting the same cognitive process. Thus, the answer to the fairness question was possibly a post hoc rationalization of responders’ answer to the verdict question.

To sum up, our findings show that the reluctance of factfinders to impose liability based solely on indirect, circumstantial evidence is stronger and farther reaching than demonstrated in previous studies. None of the explanations heretofore proposed for this reluctance is completely satisfactory.

Much like other simplifying heuristics, the anti-inference heuristic assists people to make judgments under uncertainty. It functions as a substitute for extensive algorithmic processing and yields judgments that are usually accurate. Ordinarily, when we see something with our own eyes, or when someone tells us that she saw something herself, that event actually happened; this is not necessarily true of conclusions drawn from circumstantial evidence. However—like other biases and heuristics—the anti-inference heuristic also gives rise to systematic error. It strongly affects decisions even when the objective and subjective probabilities of the pertinent occurrence are equal according to circumstantial and direct evidence.<sup>11</sup>

## 4.2. Normative Implications

While our study sheds new light on factfinders’ behavior, it leaves the normative question associated with the findings largely unanswered and awaiting future research. On the one hand, similar to other irrational biases, the anti-inference bias might drive judicial decision-making astray. Thus, legal systems might be able to improve the quality of factfinding by circumventing the bias through proper regulation. On the other hand, if the bias reflects deeply held epistemological and moral intuitions, circumventing it might adversely affect the desirable correspondence between peoples’ prevalent perceptions and the outcomes of adjudication. This disconnect might in turn undermine the legitimacy and efficiency of the legal system (Robinson and Darley 1997).

Furthermore, even if one concludes that circumventing the bias is desirable, it is unclear whether this is feasible. While jury instructions often stress that direct and

---

<sup>11</sup> There is an intriguing correspondence between our findings and the findings of experiments studying young children’s overly rigid application of the “seeing = knowing” rule. See, e.g., Ruffman 1996; Friedman et al. 2003).

circumstantial evidence are in principle equally reliable (e.g. Ohio Jury Instructions – Civil §207.07(2) (2010): “Evidence may be direct, circumstantial, or both. Direct and circumstantial evidence are of equal weight”), the effectiveness of such instructions is not self-evident (e.g. Tiersma and Curtis 2008). The research on instructions relating to other areas (e.g. inadmissible evidence, the exercise of the right to remain silent) suggests that instructions are often ineffective since juries tend to disregard them (Lieberman and Arndt 2000; Shafer 1985).

Another factor that bears on the need for regulation, and that may warrant future research, is the way in which lawyers respond to the bias. Arguably, repeat players such as lawyers are well aware of the precise scope of the bias. Thus, defense lawyers might attempt to strengthen the bias in order to make a favorable outcome more probable. Framing evidence as *merely* circumstantial seems to be a routine tactic of the defense (Bergman 1996). They might exacerbate the anti-inference bias not only by pointing to the existence of alternative explanations for the circumstantial evidence (as demonstrated by Niedermeier et al. (1999, pp. 539-41)), but also by making those alternatives as vivid as possible, for instance by using computer simulations (see e.g. Kassin and Dunn 1997; Dunn et al. 2006). Prosecutors and plaintiff lawyers are expected to try to counteract such maneuvers.

Finally, special attention should be devoted to examining the way in which judges are influenced by the bias. While there is no reason to assume that the psychology of legally trained subjects, including judges, differs fundamentally from that of other people, and some studies have indeed demonstrated that it does not (e.g. Guthrie, Rachlinski, and Wistrich 2001), some studies do show that professionals can learn to employ debiasing mechanisms (cf. Guthrie and Rachlinski 2006). Further research with judges would enhance our understanding of judicial behavior and the need for regulation in this context.

## 5. CONCLUSION

The experiments described in this study demonstrate that the disinclination of factfinders to convict defendants or impose civil liability based solely on circumstantial evidence is more powerful than heretofore realized. It characterizes not only statistical, probabilistic, inconclusive, forensic, and non-case-specific evidence, but also non-statistical, non-probabilistic, conclusive, eyewitness, and case-specific evidence. None of the theories proposed as explanations for this phenomenon provide a satisfactory account, indicating the presence of a robust, deep-rooted anti-inference bias.

Sherlock Holmes famously says that “when you have eliminated the impossible, whatever remains, however improbable, must be the truth.”<sup>12</sup> It seems, however, that most people—including judicial factfinders—adopt the maxim used as the title for one of Carter Dixon’s Sir Henry Merrivale mysteries, “Seeing is believing.”<sup>13</sup>

---

<sup>12</sup> See, e.g., Doyle, Arthur Conan. 2001 (1927). The Adventure of the Blanched Soldier in *The Case-Book of Sherlock Holmes* 33, 54. Cornwall: House of Stratus.

<sup>13</sup> Carter Dickson [John Dickson Carr]. 1941. *Seeing Is Believing*. United States: William Morrow and Company.

## REFERENCES

- Adler, Jonathan. 2006. Epistemological Problems of Testimony. *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/testimony-episprob>).
- Arkes, Hal R., Brittany Shoots-Reihard, and Ryan S. Mayers. 2011. Disjunction Between Probability and Verdict in Juror Decision Making. *Journal of Behavioral Decision Making* XX:XX-XX.
- Bergman, Paul. 1996. A Bunch of Circumstantial Evidence. *University of San Francisco Law Review* 30:985-90.
- Coady, C. A. J. 1992. *Testimony: A Philosophical Study*. Oxford: Oxford University Press.
- Cohen, L. Jonathan. 1977. *The Probable and the Provable*. Oxford: Oxford University Press.
- Dunn, Meghan A., Peter Salovey, and Neal Feigenson. 2006. The Jury Persuaded (and Not): Computer Animation in the Courtroom. *Law and Policy* 28:228-48.
- Eggleston, Richard. 1983. *Evidence, Proof, and Probability* (2d ed.). London: Weidnefeld and Nicolson.
- Foster, Michael. 1994. A Review and Reconsideration of Florida's Rule against Basing an Inference on an Inference in Civil Cases. *Stetson Law Review* 23:743-803.
- Friedman Ori, Richard Griffin, Hiram Brownell, and Ellen Winner. 2003. Problems with the Seeing = Knowing Rule. *Developmental Science* 6:505-13.
- Goodman, Jane. 1992. Jurors' Comprehension and Assessment of Probabilistic Evidence. *American Journal of Trial Advocacy* 16:361-89.
- Greenstein, Richard K. 2009. Determining Facts: The Myth of Direct Evidence. *Houston Law Review* 45:1801-30.
- Guthrie, Chris, and Jeffrey J. Rachlinski. 2006. Insurers, Illusions of Judgment & Litigation. *Vanderbilt Law Review* 59: 2017-49.
- Guthrie, Chris, Jeffrey J. Rachlinski, and Andrew J. Wistrich. 2001. Inside the Judicial Mind. *Cornell Law Review* 86:777-830.
- Hammarskjold, Carl N. 2011. Smokes, Candy, and the Bloody Sword: How Classifying Jailhouse Snitch Testimony as Direct, Rather than Circumstantial, Evidence Contributes to Wrongful Convictions. *University of San Francisco Law Review* 45:1103-43.
- Heller, Kevin Jon. 2006. The Cognitive Psychology of Circumstantial Evidence. *Michigan Law Review* 105:241-305.
- Hirt, Edward R., and Keith D. Markman. 1995. Multiple Explanation: A Consider-an-Alternative Strategy for Debiasing Judgments. *Journal of Personality and Social Psychology* 69:1069-86.
- Kahneman, Daniel, and Dale T. Miller. 1986. Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review* 93:136-53.
- Kahneman, Daniel, and Amos Tversky. 1982. The Simulation Heuristic. *Judgment Under Uncertainty: Heuristics and Biases* 201-08 (edited by Daniel Kahneman, Paul Slovic, and Amos Tversky). Cambridge: Cambridge University Press.
- Kassin, Saul M., and Meghan A. Dunn. 1997. Computer-Animated Displays and the Jury: Facilitative and Prejudicial Effects. *Law and Human Behavior* 21:269-81.

- Kaye, David. 1987. Apples and Oranges: Confidence Coefficient and the Burden of Persuasion. *Cornell Law Review* 73:54-77.
- Keren, Arnon. 2007. Epistemic Authority, Testimony and the Transmission of Knowledge. *Episteme* 4:368-381.
- Levi, Ariel S., and John B. Pryor. 1987. Use of the Availability Heuristic in Probability Estimates of Future Events: The Effects of Imagining Outcomes versus Imagining Reasons. *Organizational Behavior and Human Decision Process* 40:219-34.
- Lieberman Joel D., and Jamie Arndt. 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* 6:677-711.
- Lindsay, R.C.L., Gary L. Wells, and Carolyn M. Rumpel. 1981. Can People Detect Eyewitness-Identification Accuracy Within and Across Situations? *Journal of Applied Psychology* 66:79-89.
- McCormick, Charles T. 2006. *McCormick on Evidence* (6th, one-volume ed. by Kenneth S. Broun et al.). St. Paul, Minn.: Thompson/West.
- Miller, Dale T., William Turnbull, and Cathy McFarland. 1989. When a Coincidence is Suspicious: The Role of Mental Simulation. *Journal of Personality and Social Psychology* 57:581-89.
- Moore, Albert J., Paul Bergman, and David A. Binder. 1996. *Trial Advocacy: Inferences, Arguments, and Techniques*. St. Paul, Minn.: West Pub. Co.
- Murphy Peter. 2008. *Murphy on Evidence* (10th ed.). New York: Oxford University Press.
- Nance, Dale A., and Scott B. Morris. 2005. Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random-Match Probability. *Journal of Legal Studies* 34:395-444.
- Niedermeier, Keith E., Norbert L. Kerr, and Lawrence A. Messé. 1999. Jurors' Use of Naked Statistical Evidence: Exploring Bases and Implications of the Wells Effect. *Journal of Personality and Social Psychology* 76: 533-42.
- Penrod, Steven and Brian Cutler. 1995. Witness Confidence and Witness Accuracy: Assessing Their Forensic Relation. *Psychology, Public Policy, and Law* 1:817-45.
- Phipson, Sidney L. 2010. *Phipson on Evidence* (17th ed., Hodge M. Malek et al. eds.) London: Thomson Reuters (Legal) Limited.
- Robinson, Paul H., and John M. Darley. 1997. The Utility of Desert. *Northwestern University Law Review* 91:453-99.
- Roese, Neal J., Florian Fessel, Amy Summerville, Justin Kruger J, and Michael A. Dilich. 2006. The Propensity Effect: When Foresight Trumps Hindsight. *Psychological Science* 17:305-10.
- Rosenberg, Irene Merker, and Yale L. Rosenberg. 1995. "Perhaps What Ye Say is Based Only on Conjecture"—Circumstantial Evidence, Then and Now. *Houston Law Review* 31:1371-427.
- Ruffman, Ted. 1996. Do Children Understand the Mind by Means of Simulation or a Theory? Evidence From Their Understanding of Inference. *Mind & Language* 11:388-414.

- Raune, David, Andrew MacLeod, and Emily A. Holmes. 2005. The Simulation Heuristic and Visual Imagery in Pessimism for Future Negative Events in Anxiety. *Clinical Psychology and Psychotherapy* 12:313–325.
- Scurich, Nicholas, and Richard S. John. 2011. Trawling Genetic Databases: When a DNA Match is Just a Naked Statistic. *Journal of Empirical Legal Studies* 8:49-71.
- Shaffer, David R. 1985. The Defendant's Testimony. *The Psychology of Evidence and Trial Procedure* 124-49 (edited by Saul M. Kassin and Lawrence S. Wrightsman). Beverly Hills, CA: Sage.
- Shapiro, Barbara J. 1991. *Beyond Reasonable Doubt and Probable Cause: Historical Perspectives on the Anglo-American Law of Evidence*. Berkeley: University of California Press.
- Stein, Alex. 2005. *Foundations of Evidence Law*. New York: Oxford University Press.
- Sykes, Deanna L. and Joel T. Johnson. 1999. Probabilistic Evidence versus the Representation of an Event: The Curious Case of Mrs. Prob's Dog. *Basic and Applied Social Psychology* 21:199-212.
- Thompson, William C., and Edward L. Schumann. 1987. Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy. *Law and Human Behavior* 11:167-87.
- Tiersma, Peter, and Mathew Curtis. 2008. Testing the Comprehensibility of Jury Instructions: California's Old and New Instructions on Circumstantial Evidence. *Journal of Court Innovation* 1:231-61.
- Tribe, Laurence H. 1971, Trial by Mathematics: Precision and Ritual in the Legal Process. *Harvard Law Review* 84:1329-93.
- Wells, Gary L. 1992. Naked Statistical Evidence of Liability: Is Subjective Probability Enough? *Journal of Personality and Social Psychology* 62:739-52.
- Wigmore, John Henry. 1983. *Evidence in Trials at Common Law*, vol. 1A (revised by Peter Tillers). Boston: Little, Brown and Company.