ROGER G. KOPPL, ROBERT KURZBAN, AND LAWRENCE KOBILINSKY

EPISTEMICS FOR FORENSICS

ABSTRACT

Forensic science error rates are needlessly high. Applying the perspective of veritistic social epistemology to forensic science could produce new institutional designs that would lower forensic error rates. We make such an application through experiments in the laboratory with human subjects. Redundancy is the key to error prevention, discovery, and elimination. In the "monopoly epistemics" characterizing forensics today, one privileged actor is asked to identify the truth. In "democratic epistemics," several independent parties are asked. In an experiment contrasting them, democratic epistemics reduced the rate at which biased observers obscured the truth by two-thirds. These results highlight, first, the potential of "epistemic systems design," which employs the techniques of economic systems design to address issues of veracity rather than efficiency, and second, the value of "experimental epistemology," which employs experimental techniques in the study of science.

INTRODUCTION

Economists typically employ the normative criterion of efficiency, but have given less attention to truth. In many contexts, however, veracity may be more important than efficiency. Police forensics, pure science, espionage, auditing, clinical medical testing, drug screening, judicial proceedings, and government investigations are all social processes that generate, in one way or another, judgments of truth and falsity. In such processes it is worth knowing which institutional arrangements, or "epistemic systems" – "social processes viewed from the perspective of their

*Research for this paper was supported in part by the National Science Foundation under Grant Number 0622477. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation. Early stages of the research for this paper were funded jointly by two colleges of Fairleigh Dickinson University, namely, the Silberman College of Business and the Petrocelli College of Continuing Studies. For that funding, we thank David Steele, former Silberman dean, and Kenneth Vehrkens, the Petrocelli dean. For helpful comments we thank Alvin Goldman, Kathleen Haspel, Dan Houser, Jeffrey Hsu, Richard Langlois, Kelli Lanier, Christian List, Heather Loenser, Maria Minniti, Paul Rubin, John Schiemann, Laurie Treleven, and Xia "Lily" Wang. We thank Kelly Asao, Katelyn Sjogren, Bryan Bencomo, Alex Shaw, and Alicia Canzius for research assistance. We alone are responsible for any errors in the paper.

DOI: 10.3366/E1742360008000294

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

tendency to help or frustrate the production of truth" (Koppl 2005a, 91; cf. Goldman 1999) – tend to produce true beliefs and which tend to produce falsehood and error. Here the theory of epistemic systems is applied to the problem of reducing error rates, using the network structure of forensic science as a model.

The current institutional structure of forensics typically grants an individual lab a kind of monopoly on the analysis of any evidence sent to it: once a given lab has received and analyzed a body of evidence, it is unlikely that the evidence will be examined by any other lab. Periodically and randomly sending evidence to more than one lab might help break this monopoly (Koppl 2005b). The experiments discussed below provide an initial test of the possible consequences of implementing such "democratic epistemics."

The experiments described below are exercises in veritistic social epistemology (Goldman 1999). They address the problem of experts from a different perspective than that adopted by Goldman (2001). Goldman (2001) adopts the perspective of a novice and asks when a novice can be justified in preferring the testimony of one expert to that of another. We adopt a systems perspective and ask what social structures tend to induce more accurate judgments among novices who must evaluate expert testimony.

Our experiments might also be thought of as exercises in "epistemic engineering." (Our use of the term differs from that of Sterelny 2003, which concerns the co-evolution of artifacts and mental models.) The notion of epistemic engineering has been alluded to by Quine, who views "normative epistemology [as] a branch of engineering," making "free use of whatever scientific findings may suit its purpose" (1986, 664–5). In a similar spirit, Bartley (1987) identifies evolutionary epistemology as "a new social account of knowledge: an account dealing with the question of how to optimize the rules and practices of the community so as to diminish distortion" (447). Evolutionary epistemologists do not seem to have taken seriously, however, the engineering problem Bartley points to. Our experiments illustrate how one might do rigorous scientific research to address the engineering or design problem identified in the epistemological writings of Goldman, Quine, Bartley, and others. As such, they continue a trend of recent decades to examine epistemological questions using game theory and experiments with human subjects. A few salient examples are Akerlof (1970), Aumann (1976), Fudenberg and Kreps (1988), McCabe et al. (2001), Milgrom & Roberts (1986), and van Winden (1999). Goldman (1992) reflects this trend by choosing the subtitle "Philosophy Meets the Cognitive and Social Sciences." Peart and Levy (2008) discuss complementary work on social moral epistemology (Buchanan 2002, 2007) in relation to econometric replication (Dewald et al. 1986) and model uncertainty (Brock et al. 2007).

Section I discusses forensic science and its institutional structure. Section II reviews the theory of epistemic systems, applies the theory to forensic science, and draws out some testable implications of the analysis. Section III explains the experimental designs and results. Section IV is a concluding discussion.

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

I. FORENSIC SCIENCE AS THE PROBLEM CONTEXT

DNA exonerations in the US and other events have induced an extensive literature on the limited reliability of forensic testimony in court (Cole 2005; Giannelli 1997; State of Illinois 2002; Jonakait 1991; Kaufman 1998; Kelly and Wearne 1998; McDougall 2006; McRoberts et al. 2004; Moenssens 1993; Office of the Inspector General 1997, 2004, 2006; Risinger et al. 2002; Saks 1998; Saks and Koehler 2005; and Saks et al. 2001). This literature has brought all forensic disciplines into question, including DNA typing and fingerprint examination.

The case of Josiah Sutton illustrates problems with DNA. Sutton was jailed on a rape conviction that was based primarily on DNA evidence from the Houston Crime Lab that was later revealed to be false and mistaken (Koppl 2005b). Sutton's case was part of a scandal that resulted in the Houston Crime Lab's DNA/serology section being closed from December 2002 to July 2006 (Bromwich 2005, 2; Khanna 2006, Glenn 2006).

Several studies of fingerprint proficiency tests illustrate problems in that field. One study suggests a false positive rate for fingerprints of at least 2% (Peterson and Markham 1995a, 1995b). A more recent test produced a 20% rate of false positives (Grieve 1996). A third study considered a larger set of test results having an overall false positive error rate of 0.8%, which "may represent only a lower bound" to the rate existing in practice (Cole 2005, 1030, 1034). Whether the rate of false identifications for fingerprints is closer to 0.8%, 2%, or 20%, it is more than zero, the rate sometimes claimed by fingerprint examiners (Stiles et al. 1999, 22).

A recent court decision suggests that American jurists may be growing more sensitive to the apparent limits of forensic science as it is practiced in the US today. Citing Cole and others, the circuit court for Baltimore County has recently excluded fingerprint evidence in a capital case. The court found it "more likely so, than not so," that standard fingerprint identification technology is "a subjective, untested, unverifiable identification procedure that purports to be infallible." The court ruled that the State of Maryland, which wished to introduce such evidence, "did not show by a preponderance of evidence that a fingerprint examiner can reliably identify a fingerprint to an individual to the exclusion of all others using" the standard method (State of Maryland v. Bryan Rose 2007).

In short, forensic sciences, including DNA typing, are less reliable than generally supposed. Many scholars, journalists, activists, and others have recognized the need to improve forensic science (Pyrek 2007). Former FBI agent Jim Fisher, for example, says "forensic science is a failed promise" because it did not "significantly increase crime solution rates" as it was expected to do. "Since forensic science is mainly in service to criminal investigation, scientific crime detection is not being fully utilized and so hasn't lived up to its full criminal justice potential." (2008, ix)

There are a number of remediable features of the current institutional structure. One important remedy is "rivalrous redundancy," which would produce several competing forensic labs in any jurisdiction, eliminating the monopoly position now

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

enjoyed by most forensic labs (Koppl 2005b). For example, some DNA evidence, chosen at random, might be sent to more than one lab for analysis. The forensic worker need not know whether the evidence is examined by another lab, only that there *could* be another lab and sometimes there is.

Crucially, rivalrous redundancy would not increase the costs of forensics, because the cost savings from reducing errors swamp the costs of the extra tests that would be required. One of us (Koppl forthcoming) examines the cost consequences of independent triplicate fingerprint examinations in all felony cases in the US that have fingerprint evidence and go to trial. He shows that in 2002 it would have cost about $100 to add two redundant fingerprint examinations (for a total of three examinations) to each of the felony cases in the US that went to trial and used fingerprint evidence, and that the average cost of incarceration for a false felony conviction was about $100,000 in 2002. He concludes that triplicate fingerprint examinations would not only eliminate most of the false felony convictions resulting from false positive errors of fingerprint examination, but also reduce the taxpayer costs of administering the criminal justice system.

We now turn to a more formal presentation of the problem and the suggested solution.

II. EPISTEMIC SYSTEMS AS THE THEORETICAL FRAMEWORK

An epistemic system is an ordered triple, $\langle \mathfrak{S}, \mathfrak{R}, \mathfrak{M} \rangle$. The set $\mathfrak{S}$ (of Senders) is indexed by $i \in I^*$. A member of $\mathfrak{S}$ is represented by $s_i \in \mathfrak{S}$. The set $\mathfrak{R}$ (of Receivers) is indexed by $j \in J^*$. A member of $\mathfrak{R}$ is represented by $r_j \in \mathfrak{R}$. The set $\mathfrak{M}$ (of messages) is indexed by $h \in H^*$. A member of $\mathfrak{M}$ is represented by $m_h \in \mathfrak{M}$. Presumably, one would typically model the number of Senders and Receivers as finite. It may often be convenient, however, to assume the message space is infinite. If I^* has a largest element, denote that element I. Define J and H similarly.

Senders and Receivers have value functions over messages. These might be utility functions or payoff functions. For Receivers, $V_{r_j} = f_{r_j}(m_{s_1}, m_{s_2}, \ldots, m_{s_I})$. For Senders, $V_{s_i} = f_{s_i}(V_{r_1}, V_{r_2}, \ldots, V_{r_J}; m_{s_1}, m_{s_2}, \ldots, m_{s_I})$. For example, suppose we have a lawsuit in which the plaintiff's attorney has hired an expert witness to estimate the money value of the claimed harm. Here, the Sender is an expert witness hired by the Receiver, the plaintiff's lawyer. Up to some limit of plausibility, the Receiver prefers higher estimates to lower estimates. This may induce a similar preference in the Sender, who wants the plaintiff's lawyer to become a repeat customer. If there are two or more Senders, they are in a position of strategic interdependence with respect to the messages they send.

In the experiments reported here, we are not in a position to make any general claims about how or whether Sender values are influenced by expectations of Receiver values because Senders were given no specific inducement to consider Receiver values.

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

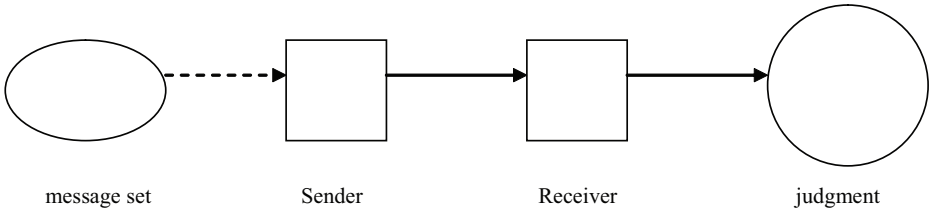


Figure 1. Monopoly: One crime lab reports on the evidence.

Application to Forensics

Under current institutions in the common-law countries, crime labs are typically organized under the police. This mode of organization tends to create a pro-police bias (Risinger et al. 2002, Giannelli 2007). The bias need not be conscious. The police generally ask for a test when they believe a match will identify their suspect. Thus, some forensic workers tend to have a bias in favor of finding matches. For this analysis, therefore, we will assume that forensic labs prefer (*ceteris paribus*) to send the message 1, "match."

Figure 1 represents the current situation in the United States and elsewhere, though in a kind of simplified cartoon caricature. The message space is given by $\mathfrak{M} = \{0, 1\}$, where 0 represents "no match" and 1 represents "match." The Sender is a forensics lab and the Receiver is a judge or jury. Figure 1 reflects the current monopoly situation in forensics whereby evidence going to any one lab will probably not be examined by any other lab.

In the super-simplified situation of Figure 1, the lab always sends the message 1 and the crime lab adds no new information to the system. In this case, forensic science does not increase the epistemic reliability of the criminal justice system. Reality is more complicated, but the monopolistic position of crime labs and their organization under law enforcement agencies probably increases the rate of false positive errors, thus reducing the contribution of forensics to the epistemic reliability of the criminal justice system.

Notice that we are counting all false reports as "errors" whether they are lies or honest mistakes. There are no bright lines as we pass from willful falsehood to conscious bias to unconscious bias to perfect objectivity. Thus, all false reports are lumped together for present purposes.

Now assume that multiple independent forensic scientists examine the evidence. Figure 2 illustrates the case in which there are three independent forensic examiners. This redundancy will not help if it is mere redundancy, that is, if more than one Sender examines the evidence but there are no incentives for any one Sender to discover the errors of the other(s). If each Sender sends the same message, the Receiver's judgment is likely to be the common message. In this model, redundancy does not alter the behavior of the Senders. Each of the Senders in this model sends 1. The epistemic efficiency of the system is

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

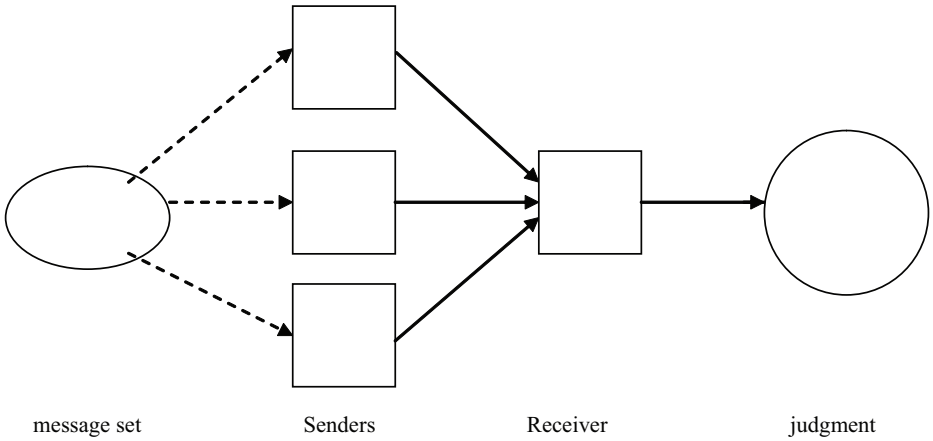


Figure 2. Alternative organization: Three crime labs report on the Evidence.

low because the system produces the right judgment only when the message "1" happens, coincidentally, to be true.

Now suppose, however, that there are at least three Senders and that each Sender's payoff depends on the choices of the other Senders and of the Receiver. Each Sender receives a positive payoff if and only if his selection is nominated by the Receiver. The Receiver is not obliged to accept the majority opinion among the Senders. If Senders expect him to, however, they will not wish to be in the minority. If a Sender loses, he receives nothing or perhaps is fined. This payoff structure creates "rivalrous redundancy." If there are two Senders, no one can be in a clear minority, which gives us reason to doubt that two Senders are better than one.

We expect rivalrous redundancy to increase the epistemic reliability of the system if the number of Senders is at least three. This gain in epistemic reliability may come about for either of two reasons. First, the independence of Sender reports increases the chance that majority opinion will be truthful. The relatively simple situation we are examining approximates the conditions of elementary versions of the Condorcet jury theorem. If Senders are independent and each Sender has at least a 50% chance of sending a correct message, then majority opinions will contain fewer errors than individual decisions. If Receivers accept majority opinion, then democratic epistemics will improve system performance. Second, if Senders wish to send reports that express the majority opinion and if the truth is a Schelling point, then the accuracy of Sender reports will be higher under rivalrous redundancy. Senders who expect Receivers to accept majority opinion will want to be in the majority. When there are exactly three Senders, for example, there are precisely two Sender strategy profiles in which all Senders are in the majority, namely, (0,0,0) and (1,1,1). There is an obvious symmetry between these two profiles. It is at least possible, however, that the truth may

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

be more salient (Schelling 1960) than other messages. In that case, the accuracy of Sender reports will be higher under rivalrous redundancy. For both reasons just reviewed, we predict rivalrous redundancy to increase the epistemic efficiency of the system. The experimental design of this study was meant to test this prediction.

Before turning to our experiments, it may be worth exploring more fully the above argument. (We thank Christian List for help with the logic and exposition of the rest of this section.) First, even if there is no conflict of interest between Senders because they all prefer Receiver choices to be correct, Senders may still have an incentive to send untruthful messages. Even when all Senders prefer truthful messages, each Sender's decision as to which message to send will be influenced by his sense of whether he is pivotal, and the Sender may then decide to send either always a positive or always a negative message, irrespective of his own true judgment.

This effect may undermine the applicability of the Condorcet jury theorem, which requires truthfulness on the part of the individuals. This point was first observed by Austen-Smith and Banks (1996) who showed that in many cases sincere voting "precludes Nash equilibrium behavior" (43). Brennan and Buchanan (1984) made a similar point, without reference to Condorcet or Nash. Feddersen and Pesendorfer (1999) give a succinct and helpful overview of many of the issues involved. This literature suggests that making truthfulness incentive compatible may sometimes require a departure from simple majority rule, depending on the details of the situation.

List and Pettit (forthcoming) analyze the logic of majority rule. Strictly speaking, our experiment does not assume or require majority rule because the Receiver is not required to follow the majority opinion of Senders. It may help us to understand the strategic issues facing Senders, however, to momentarily imagine that Receivers always accept majority opinion among Senders. With that assumption, it can be shown that if senders expect certain benefits from being in the majority, perhaps because of the incentives that forensic labs are given, then (0,0,0) and (1,1,1) may both be Nash equilibria *irrespective of the Senders' truthful judgments*. Suppose you are one of the Senders. Assuming that both of the other Senders send the message 0, then the majority verdict will be 0, irrespective of what message you send. It is therefore your preference to be in the majority that drives your decision as to which message to send. Given your preference to be in the majority, the assumption that the other two Senders send the message 0 will thus lead you to send the same message too. Since this reasoning is symmetrical across Senders, (0,0,0) is a Nash equilibrium. By an analogous argument, (1,1,1) is also a Nash equilibrium. Truthfulness can be a Nash equilibrium as well, but arguably it is a more fragile one. Our experiments are part of an effort to discover institutional designs that will make truth-telling robust. We now drop our temporary assumption that the Receiver is bound to accept majority opinion and turn to our experiments.

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

III. THE EXPERIMENTS

Objectives

Our experiments examined four questions:

(1) Everything else equal, will people transmit false information to a decision-maker if they have an incentive to do so? (Can a disposition toward truthfulness be countered by financial incentives?)
(2) Can a simple institutional mechanism, democratic epistemics, be implemented to reduce or eliminate the distorting effects of the bias in (1)?
(3) How do factors such as the size of the bias in (1) interact with the mechanism in (2)?
(4) Does system performance vary monotonically with the number of Senders?

Answering question 2 means testing democratic epistemics as an institutional mechanism, tying together the research question with the experimental method. Questions 3 and 4 address robustness. Is the mechanism of rivalrous redundancy sensitive to factors such as the size of the bias and the number of Senders?

The experiment gives Senders an incentive to produce biased reports on an observed event. The Receivers guess what event happened after reading the report(s) of one or more Senders. In the Baseline, "monopoly," condition, each Receiver gets one report from one Sender. In the "democratic" condition, each Receiver gets multiple reports from multiple Senders. In the Baseline condition, the Sender has an incentive to lie whenever the truth does not conform to the Sender's bias and the Sender's bias is sufficiently large. In the democratic condition, this incentive competes with the incentive to issue a report that is consistent with majority opinion. Thus, the Senders have a coordination problem. The chance that Senders will choose truth-telling depends on several variables such as the size of the bias and number of Senders per Receiver. These dimensions are explored in different experimental treatments as described below.

First Experiment

Experimental Design

A new task, dubbed "The Science Game" is introduced to mirror the structure of forensic science. In this game, Senders represent forensic labs and Receivers represent a judge or jury. There are N Senders and one Receiver, each describable, in the behavioral game theory tradition, as having a set of information, a set of decisions, and payoffs associated with all possible outcomes. In each round, Nature (i.e., a randomizing device) chooses a "correct" Object from a set of 3 Objects: circle, triangle, square. This information is known to Senders but not the Receivers. Senders send a message consisting of one of the Objects to Receivers, who then must submit a decision about what they believe the correct Object to be. (In the context of forensic science, this message corresponds to a lab indicating

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

"match," "no match," or "inconclusive.") The Receivers' information consists only of the communications they will receive from Senders, their own payoffs for correct versus incorrect guesses, and limited information about Senders' payoffs (see below).

Senders get a payment, P_C, only if the Receiver indicates the correct Object. Senders, however, also get an additional payment depending on which Object the Receiver chooses *independent of the correct Object*. This payment corresponds to the bias on the part of the lab personnel. This payoff information is private (i.e., unknown to the Receivers).

All experimental sessions were run in the Penn Laboratory for Experimental Evolutionary Psychology (PLEEP) using pen and paper data collection. The PLEEP lab has 16 subject stations with partitions, which prevent players from seeing one another or each others' actions or materials. Anonymity was preserved by assigning participants identification numbers at the start of the experiment. Participants were paid by taking an envelope labeled with their unique numeric identifier filled with their earnings. Participants were recruited through the web-based "experiments @ Penn" database, which includes members of the general University of Pennsylvania community, including staff and students.

Experimental Treatment I: High & Low Bias

In the Baseline condition, there was one Sender and one Receiver. Senders received a payment of $3 for leading the Receiver to the correct Object (a "hit") if the Sender sent the correct Object and the Receiver chose this Object. Senders may have received a separate payment, depending on the treatment, of either $1.00 or $5.00 for a "supplemental" Object. This payment induced a bias in the preferences of Senders. The Sender received this payment as a function of the Object chosen by the Receiver independent of what the correct Object was. Thus, if the correct Object was square and the supplemental Object was circle, and the Receiver choose circle, the Sender got a payoff of either $1.00 or $5. A Sender with a $5 supplemental payoff thus had an incentive to report the supplemental Object (circle in our example) regardless of what the correct Object was. This incentive is the Sender's bias.

The size of the bias was manipulated as a within-subjects treatment. In the "High Bias" condition, the bias was greater than the payoff to the Sender for the Receiver making the correct decision. If the Receiver chose the correct Object, the Receiver received $5, otherwise $2.

Senders were informed that (1) the only information that Receivers had was the information from Senders, (2) Receivers earn more money if they guess the correct Object, and (3) this is the *only* means by which Receivers earn money. Senders were not given quantitative information about Receiver's payoffs. Not informing Senders how much they can benefit the Receiver by sending correct information minimizes problems that might be associated with other-regarding preferences

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

(Rabin 1993, Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Charness and Rabin 2002, Hoffman et al. 1996).

In each experimental session, 10 rounds of this game were played. The correct Object and supplemental shape for each round were determined randomly before the experiment and given to Senders at the beginning of each round. Five High and Five Low bias rounds were played, with their order being randomly pre-set. No feedback was given to players regarding the Receiver's guess about the Object in each individual round.

Participants were told that one round of play would be selected randomly, and that the outcome for that round would determine their monetary payoff for the experiment.

Experimental Treatment II: Number of Senders

A second experimental treatment was run with three Senders. Everything was identical, including the High and Low Bias treatments, except that there were three Senders per Receiver. Each Sender was told that the Receiver would have access to the information (shape) sent by each of the three Senders, and then would guess the correct shape based on this information. In each round all three Senders had the same supplemental shape, but they did not necessarily have the same bias level (High or Low). Crucially, Senders were not informed of the biases of the other Senders.

Hypotheses

The hypotheses for this experiment were that

(1) the biases of the Baseline condition will induce inaccurate reporting on observed events,
(2) larger biases will produce larger number of inaccurate reports, and
(3) democratic epistemics will reduce these inaccuracies.

The first two hypotheses are principally a matter of confirming the effect of the method and should be considered to be manipulation checks.

If the third hypothesis is sustained, then it is likely that decision makers can understand and respond to the incentives of democratic epistemics in lab-based social processes. In this case we will have evidence that the epistemic performance of the system depends on its institutional structure as well as the skill or integrity of the individuals in the system. Such a result would strengthen the view that institutional change might be an effective method of reducing error rates in forensic science and other areas.

Results

4 Sessions in the 1-Sender condition (N = 30) and 3 Sessions in the 3-Sender condition (N = 24) were run. To test the statistical significance of each result we conducted standard Pearson's chi-square tests. The basic idea of the test is that deviations between the observed and theoretical distribution of "hits"

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

across several categorical alternatives should be "small" if one's theory or hypothesis is right. Given one's theory or hypothesis, the sum of the squared deviations over expected values, $\sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$, should be chi-square distributed. The computed value is "large" if it is sufficiently improbable given your theory, i.e., if the chi-square distribution would generate so large a computed value only, say, 5% of the time or 1% of the time.

Collapsing the 1- and 3-Sender conditions, participants sent the correct information 96% of the time with low bias, and 50% of the time with high bias ($\chi^2 = 89$, d.f. $= 1$, $p < 0.0005$). Such a disparity would happen less than 0.05% of the time under the hypothesis that chance alone is responsible for the difference. Since this is well below the customary threshold of 1%, we reject the hypothesis and conclude that the difference in performance is correlated with the difference in experimental treatment. In other words, the size of the bias does matter. In high-bias cases when the bias disagreed with the truth, the rate of errors in messages was lower in the 3-Sender condition, 67%, than in the 1-Sender condition, 86% ($\chi^2 = 5$, d.f. $= 1$, $p = 0.02$). Receivers' guesses were more often correct in the 3-Sender condition. Across conditions (1- & 3-Senders and high & low Sender bias), there was a systemic error rate of 36% in the 1-Sender condition and 15% in the 3-Sender condition ($\chi^2 = 10$, d.f. $= 1$, $p = 0.003$). When Sender bias disagreed with the truth, the systemic error rate was 47% in the 1-Sender condition and only 21% in the 3-Sender condition ($\chi^2 = 26$, d.f. $= 1$, $p < 0.0005$). Further restricting attention to cases in which at least one Sender was in the high-bias condition reveals a systemic error rate of 75% in the 1-Sender condition and only 25% in the 3-Sender condition ($\chi^2 = 20$, d.f. $= 1$, $p < 0.0005$).

The logic of our results on systemic error rates can be seen by considering the case in which each Receiver always accepts the majority opinion and in each group of three Senders, one sends a false message and the other two send truthful messages. The error rate in messages would be 33%, but the systemic error rate would be zero. Through strategic redundancy, democratic epistemics reduces systemic error rates by more than the reductions in error rates in messages.

These results show that (1) experimental participants can be induced to send true information when the induced bias is small, (2) participants can be induced to send biased information when the induced bias is large, (3) democratic epistemics may reduce the error rate in messages, and (4) *democratic epistemics can produce an even larger reduction in systemic error rates.* As we shall see in our second experiment, conclusion (3), that democratic epistemics may reduce the error rate in messages, seems to be sensitive to the details of our experimental design. For institutional design, conclusion (4), that democratic epistemics reduces the systemic error rate, is our most important result. We believe it worth emphasizing that error rate in messages understates the benefits of democratic epistemics, which derive mostly from the still larger reductions produced in systemic error rates.

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

Second Experiment

Experimental Design

We conducted a second experiment with a nearly identical design. In the multiple-Sender treatments of the second experiment, however, all Senders reporting to a given Receiver had the same bias level, but not necessarily the same supplementary shape. Again, Senders were not informed of the biases of the other Senders. In each experimental session of the second experiment, 12 rounds were played. The hypotheses for this experiment were the same as for our first experiment.

For the second experiment, all experimental sessions were again run in the Penn Laboratory for Experimental Evolutionary Psychology (PLEEP) using pen and paper data collection.

Experimental Treatment I: High & Low Bias

This treatment was identical to experimental treatment 1 of the first experiment except that the number of rounds per session was 12 rather than 10. The correct Object and supplemental shape for each round were determined randomly before the experiment and given to Senders at the beginning of each round. Six High and six Low bias rounds were played, with their order being randomly pre-set. No feedback was given to players regarding the Receiver's guess about the Object in each individual round. Participants were told that one round of play would be selected randomly, and that the outcome for that round would determine their monetary payoff for the experiment.

Experimental Treatment II: Number of Senders

A second experimental treatment was run with multiple Senders. Everything was identical, including the High and Low Bias treatments, except that there were two, three, four, or five Senders per Receiver. Each Sender was told that the Receiver would have access to the information (shape) sent by each of the Senders, and then would guess the correct shape based on this information. In each round, all Senders reporting to a given Receiver had the same bias level, but not necessarily the same supplemental shape. Senders were not informed of the biases of the other Senders.

Results

We ran 9 sessions in the 1-Sender condition (N = 106), 7 Sessions in the 2-Sender condition (N = 66), 8 sessions in the 3-Sender condition (N = 80), 8 sessions (N = 90) in the 4-Sender condition, and 7 sessions in the 5-Sender condition (N = 84). In this experiment, Senders again almost always sent correct information in the low-bias condition, but did so only about half the time in the high-bias condition. In these experiments, however, increasing the number of Senders did not seem to reduce the Sender error rates. Instead, Sender error rates went up, although not always by a statistically significant degree. Overall, our second

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

experiment produced weak evidence that democratic epistemics may reduce the reliability of Senders.

Receivers' guesses were more often correct in the 3-, 4-, and 5-Sender conditions than in the 1- or 2-Sender conditions. For the cases in which Senders had a high bias, for example, the Receiver error rate was 51% in the 1-Sender condition, 55% in the 2-Sender condition and lower in the 3-, 4-, and 5-Sender conditions, namely 33%, 36%, and 43%, respectively.

The results of our second experiment show that (1) experimental participants can be induced to send accurate information when the induced bias is small, (2) participants can be induced to send inaccurate information when the induced bias is large, (3) *democratic epistemics can improve system performance* in the sense of lowering Receiver error rates, and (4) the benefits of democratic epistemics derive not from making the experts in the system better, but from the structural redundancy that lowers Receiver error rates.

Result (4) suggests the possibility that improved system performance may come at the cost of degraded performance of the Senders within the system. The logic of this result can be seen by considering an imaginary case. Imagine each Receiver always accepts the majority opinion and in each group of, say, three Senders, one sends an inaccurate message and the other two send accurate messages. Imagine further that in the 1-Sender treatment, 15% of Sender reports are inaccurate. Under the 1-Sender treatment, the error rate for Senders and Receivers alike would then be 15%. But under the 3-Sender treatment in this imaginary example, the error rate for Senders will be 33%, while that for Receivers will be zero. The more "democratic" network has a higher rate of errors for the Senders, but a lower rate of errors for the Receivers. Thus, the overall system can be made more reliable by the very alteration that reduces the reliability of an important set of actors within the system.

Our results suggest that three is the "magic number." They suggest that democratic epistemics will improve system performance if the number of Senders is three or higher, but may not improve performance if the number of Senders is two. Moreover, further increasing the number of Senders beyond three does not seem to improve system performance.

We do not know why it does not help to increase the number of Senders above three, which is a topic for future work. We suspect, however, that the more Senders there are the less a Sender's "vote" counts. The less your vote counts, the less pull truth has on your decision (Brennan and Buchanan 1984, Brennan and Lomasky 1993, 19–31). In the extreme case of a democratic election, voting ends up a purely esthetic act. You remain rationally ignorant of the issues because the chance your wisdom will influence the result sinks arbitrarily close to zero. When your vote counts for nothing, the pleasure of expressing, say, your goodness through voting completely trumps your interest in forming and expressing a truthful judgment. More generally, the importance of non-epistemic motives grows in comparison to truth as the importance of your vote shrinks. Since Brennan and Buchanan (1984) and Brennan and Lomasky (1993),

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

such considerations have become standard public-choice theory. We don't know of any experimental support for this explanation, however, beyond our result reported above.

IV. DISCUSSION

In both experiments, democratic epistemics improved system performance. In the first experiment, democratic epistemics lowered the error rate in messages; in the second experiment, it raised the error rate in messages. Whether democratic epistemics does, in fact, reduce Sender reliability should be viewed as an open question. Its effect on Sender performance seems to depend sensitively on factors we have not controlled for. We hope future research will discover experimental manipulations that reliably predict the effect of democratic epistemics on Sender performance.

In any event, to improve system reliability it is not necessary to improve the reliability of the individual units with the system (Senders). A chain is only as strong as its weakest link; a net(work) is stronger than its individual knots (nodes). A net may be stronger than a chain even though the average knot in the net may be weaker than the average link in the chain. This result is important in part because it will be counterintuitive for many, including participants in public discussions on how to improve forensic science. Our second experiment suggests it is not necessary to improve the reliability of individual crime labs to improve the reliability of the *system*, i.e., the network of crime labs. Establishing this fact may increase the chance that reform measures in forensic science will improve, not reduce, the performance of the system. This result may also be an important input to policy decisions in other areas such as medical testing, drug screening, and science policy.

While the current study supports the claim that democratic epistemics should replace monopoly epistemics, more experimental work is needed. Future work should, for example, vary the relative size of the Senders' bias and randomize the number of Senders. There are, moreover, many similar lines of experimentation in the same general area, wherein the normative criterion applied is not efficiency, but veracity. Forensics is one of many areas of application. All such experiments would form part of the field of "epistemic systems design," which is the application of the techniques of economic systems design to issues of veracity, rather than efficiency.

This study has adapted the techniques of economic systems design to aid in discovering institutional changes that will improve not the efficiency, but the veracity of forensic science and other social processes, including other lab-based social processes. The change in normative criterion from efficiency to veracity creates epistemic systems design. Economic systems design uses "the lab as a test bed to examine the performance of proposed new institutions, and modifies their rules and implementation features in the light of the test results." (Smith 2003) It has produced a major change in how researchers design economic institutions.

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

Epistemic systems design may have a similar potential to change how researchers design institutions.

Epistemic systems design is possible because we *construct* the truth in an experimental economics laboratory. We are in the godlike position of saying unambiguously what the truth is and how close to it our experimental subjects come. (On our godlike position, compare Schuetz 1943, 144-5.) We construct the truth, the preferences, and the institutional environment of choice. We construct, in other words, the world in which we place our subjects. From this godlike perspective we are in a position to compare the epistemic properties of different institutional arrangements. When we return from our constructed world to the real world, we lose our privileged access to the truth and return to the normal uncertainty common to all. But we carry with us a knowledge of which institutional structures promote the discovery and elimination of error and which institutional structures promote error and ignorance. This knowledge can be carried from the constructed world of the laboratory to the natural world of social life because of the common element in both worlds, namely, the human mind. The one vital element of the experimental world that is *not* constructed is the human mind, which makes choices within the institutional context of the laboratory experiment. It is this same element that makes choices in the institutional structures of the natural world of social life. Thus, the sort of laboratory experiment described in this study cannot tell us which particular expert judgments are correct and which incorrect, but they can tell us that the monopoly structure of forensics today produces a needlessly high error rate.

When applied to pure science, the techniques of epistemic systems design give us an experimental approach to science studies, which might be called experimental epistemology. Epistemic systems design as here conceived falls squarely within the broadly Mertonian tradition of science studies (Merton 1937/1957, Kitcher 1993, Goldman 1978, 1999, 2007). In the past, disputes in this field could be addressed only through historical research and field studies. It now seems possible to address a significant fraction of them with the tools of epistemic systems design. In particular, the network structure relating one lab to another can be manipulated in the epistemics laboratory. Thus, it seems possible to address the role of the network structure of pure science in producing reliable knowledge. Indeed, one may describe the proposed move from monopoly to democratic epistemics in forensic science as a move to a network structure for forensic science that more nearly resembles the network structure of pure science.

Epistemic systems design might help us to understand which social institutions produce truth and which do not. The related strategy for the discovery of truth and the elimination of error is indirect. Rather than attempting to instruct people in how to form true opinions, we might reform our social institutions in ways that tend to induce people to find and speak the truth. Comparing the epistemic properties of alternative social institutions is "comparative institutional epistemics." At the margin it may be more effective to give people an interest in discovering the truth

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

than to invoke the value of honesty or teach people the fallacies they should avoid. When we rely on experts such as forensic scientists to tell us the truth, it seems especially likely that institutional reforms will have a higher marginal value than exhortations to be good or rational. If virtue and rationality are scarce goods, we should design our epistemic institutions to economize on them.

REFERENCES

Akerlof, G. A. 1970. "The Market for 'Lemons': Quality, Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84(3): 488–500.

Aumann, R. J. 1976. "Agreeing to Disagree." *Annals of Statistics* 4: 1236–9.

Austen-Smith, D. and J. S. Banks. 1996. "Information Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review* 90(1): 34–45.

Bartley, W. W., III. 1987. "Alienation Alienated: The Economics of Knowledge versus the Psychology and Sociology of Knowledge." In G. Radnitzsky and W. W. Bartley, III (eds.), *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*. La Salle, IL: Open Court.

Bolton, G. and A. Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90: 166–93.

Brennan, G. and J. Buchanan. 1984. "Voter Choice: Evaluating Political Alternatives." *The American Behavioral Scientist* 28(2): 185–201.

Brennan, G. and L. Lomasky. 1993. *Democracy and Decision: The Pure Theory of Electoral Preference*. Cambridge: Cambridge University Press.

Brock, W. A., S. N. Durlauf, and K. D. West. 2007. "Model Uncertainty and Policy Evaluation: Some Theory and Empirics." *Journal of Econometrics* 136: 629–64.

Bromwich, M. R. 2005. *Second Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room*. Retrieved May 24, 2007, from http://www.hpdlabinvestigation.org/reports/050531report.pdf

Buchanan, A. 2002. "Social Moral Epistemology." *Social Philosophy and Policy* 19: 126–51.

Buchanan, A. 2007. "Institutions, Beliefs and Ethics: Eugenics as a Case Study." *Journal of Political Philosophy* 15: 22–45.

Charness, G. and M. Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117: 817–69.

Cole, S. 2005. "More Than Zero: Accounting for Error in Latent Fingerprint Identification." *The Journal of Criminal Law & Criminology* 95(3): 985–1078.

Dewald, W. G., J. G. Thursby and R. G. Anderson. 1986. "Replication in Empirical Economics: *The Journal of Money, Credit and Banking* Project." *American Economic Review* 76: 587–603.

Feddersen, T. J. and W. Pesendorfer. 1999. "Elections, Information Aggregation, and Strategic Voting." *Proceedings of the National Academy of Sciences* 96: 10572–4.

Fehr, E. and K. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114: 817–68.

Fisher, J. 2008. *Forensics Under Fire: Are Bad Science and Dueling Experts Corrupting Criminal Justice?* New Brunswick, NJ: Rutgers University Press.

Fudenberg, D. and D. M. Kreps. 1988. "A Theory of Learning, Experimentation, and Equilibrium in Games." Unpublished manuscript.

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

Giannelli, P. C. 1997. "The Abuse of Evidence in Criminal Cases: The Need for Independent Crime Laboratories." *Virginia Journal of Social Policy & the Law* 4: 439–78.

Giannelli, P. C. 2007. "Confirmation Bias." *Criminal Justice*, 22(3): 60–1.

Glenn, M. 2006. "Under the Microscope: HPD's Evidence Closely Watched This Time Around." *Houston Chronicle*, July 11. Retrieved May 24, 2007 from http://www.chron.com/disp/story.mpl/special/crimelab/4038116.html

Goldman, A. I. 1978. "Epistemics: The Regulative Theory of Cognition." *The Journal of Philosophy* 75: 509–23.

Goldman, A. I. 1992. *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: MIT Press.

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

Goldman, A. I. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63: 85–109.

Goldman, A. I. 2007. "Social Epistemology," *The Stanford Encyclopedia of Philosophy,* Spring 2007 edition. E. N. Zalta (ed.). Retrieved from http://plato.stanford.edu/archives/spr2007/entries/epistemology-social/

Grieve, D. 1996. "Possession of Truth." *Journal of Forensic Identification*. 46(5): 521–8.

Hoffman, E., K. McCabe, and V. Smith. 1996. "Social Distance and Other-regarding Behavior in Dictator Games." *American Economic Review* 86: 653–60.

Jonakait, R. N. 1991. "Forensic Science: The Need for Regulation." *Harvard Journal of Law and Technology* 4: 109–91.

Kaufman, F. 1998. *Commission on Proceedings Involving Guy Paul Morin*. Queen's Printer for Ontario.

Kelly, J. F. and P. Wearne. 1998. *Tainting Evidence: Inside the Scandals at the FBI Crime Lab*. New York: The Free Press.

Khanna, R. 2006. "Crime Lab's DNA Section, Shut in '02, Will Reopen," *Houston Chronicle*, June 21. Retrieved May 24, 2007, from http://www.chron.com/disp/story.mpl/special/crimelab/3988198.html

Kitcher, P. 1993. *The Advancement of Science*. Oxford: Oxford University Press.

Koppl, R. 2005a. "Epistemic Systems." *Episteme, A Journal of Social Epistemology* 2(2): 91–106.

Koppl, R. 2005b. "How to Improve Forensic Science." *European Journal of Law and Economics* 20(3): 255–86.

Koppl, R. Forthcoming. "Romancing Forensics: Legal Failure in Forensic Science Administration." In E. López (ed.), *Government Failure in the Legal System: A Public Choice Review of the Law*. Independent Institute.

List, C. and P. Pettit. Forthcoming. *Group Agency: The Possibility, Design and Status of Corporate Agents*.

McCabe, K., D. Houser, L. Ryan, V. Smith, and T. Trouard. 2001. "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange." *Proceedings of the National Academy of Sciences* 98(4): 11832–5.

McDougall, L. 2006. "Shirley McKie: Was it Really an Honest Mistake?" *Sunday Herald,* February 12.

McRoberts, F., S. Mills, and M. Possley. 2004. "Forensics Under the Microscope: Unproven Techniques Sway Courts, Erode Justice." *Chicago Tribune,* October 17.

Merton, R. K. 1937/1957. *Science and the Social Order*. Reproduced in his *Social Theory and Social Structure*. Rev. ed. New York: The Free Press.

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

Milgrom, P. and J. Roberts. 1986. "Relying on Information of Interested Parties." *RAND Journal of Economics* 17(1): 18–32.

Moenssens, A. A. 1993. "Novel Scientific Evidence in Criminal Cases: Some Words of Caution." *Journal of Criminal Law and Criminology* 84: 1–21.

Office of the Inspector General, United States Department of Justice. 1997. "The FBI Laboratory: An Investigation into Laboratory Practices and Alleged Misconduct in Explosives-Related and Other Cases." Retrieved from http://www.usdoj.gov/oig/special/97-04a/index.htm

Office of the Inspector General, United States Department of Justice. 2004. "The FBI DNA Laboratory: A Review of Protocol and Practice."

Office of the Inspector General, United States Department of Justice. 2006. "A Review of the FBI's Handling of the Brandon Mayfield Case: Unclassified Executive Summary."

Peart, S. and D. Levy. 2008 "Introduction to the Symposium on Ethics." *Eastern Economic Journal* 34(1): 101–2.

Peterson, J. L. and P. N. Markham. 1995a. "Crime Lab Proficiency Testing Results, 1978–1991, I: Identification and Classification of Physical Evidence." *Journal of Forensic Sciences* 40(6): 994–1008.

Peterson, J. L. and P. N. Markham. 1995b. "Crime Lab Proficiency Testing Results, 1978–1991, II: Resolving Questions of Common Origin." *Journal of Forensic Sciences* 40(6): 1009–29.

Pyrek, K. M. 2007. *Forensic Science Under Siege: The Challenges of Forensic Laboratories and the Medico-Legal Death Investigation System*. Amsterdam: Academic Press.

Quine, W. V. O. 1986. "Reply to Morton White." In L. E. Hahn & P. A. Schilpp (eds.), *The Philosophy of W. V. Quine,* pp. 663-5. La Salle, IL: Open Court.

Rabin, M. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83: 1281–1302.

Risinger, M., M. J. Saks, W. C. Thompson, and R. Rosenthal. 2002. "The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion." *California Law Review* 90: 1–56.

Saks, M. J. 1998. "Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science." *Hastings Law Journal* 49: 1069–141.

Saks, M. J. et al. 2001. "Model Prevention and Remedy of Erroneous Convictions Act." *Arizona State Law Journal* 33: 665–718.

Saks, M. J. and J. J. Koehler. 2005. "The Coming Paradigm Shift in Forensic Identification Science." *Science* 309: 892–5.

Schelling, T. C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schuetz, A. 1943. "The Problem of Rationality in the Social World." *Economica*, N.S. 10(38): 130–49.

Smith, V. L. 2003. "Constructivist and Ecological Rationality in Economics." *American Economic Review* 93: 465–508.

State of Illinois. 2002. *Report of the Governor's Commission on Capital Punishment* April 15.

State of Maryland v. Bryan Rose. 2007. Case K06-0545. Memorandum Decision, October 19.

Sterelny, K. 2003. *Thought in a Hostile World*. Oxford: Blackwell.

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press

Stiles, M. R., R. H. Levine, and P. A. Sarmousakis. Government's Combined Report to the Court and Motions in Limine Concerning Fingerprint Evidence, United States v. Mitchell, No. 96-00407, United States District Court for the Eastern District of Pennsylvania. Retrieved August 2006, from http://www.clpex.com

van Winden, F. 1999. "Experimental Studies of Signaling Games." In L. Luini (ed.), *Uncertain Decisions, Bridging Theory and Experiments*, pp. 147–73. Boston: Kluwer.

Roger Koppl is a Professor of Economics and Finance at Fairleigh Dickinson University. He studies how error rates in forensic science may be affected by institutional and organizational structures. He founded the Institute for Forensic Science Administration at Fairleigh Dickinson University in 2006. Robert Kurzban is an Assistant Professor of Psychology at the University of Pennsylvania. His research focuses on evolutionary approaches to understanding human social behavior. Lawrence Kobilinsky is a forensic scientist and Science Advisor to the President of John Jay College of Criminal Justice, where he is also Professor of Biology and Immunology.

https://doi.org/10.3366/E1742360008000294 Published online by Cambridge University Press