JOURNAL OF
EMPIRICAL LEGAL STUDIES

Journal of Empirical Legal Studies
Volume 6, Issue 4, 793–817, December 2009

Competition in the Courtroom: When Does Expert Testimony Improve Jurors' Decisions?

*Cheryl Boudreau and Mathew D. McCubbins**

Many scholars lament the increasing complexity of jury trials and question whether the testimony of competing experts helps unsophisticated jurors to make informed decisions. In this article, we analyze experimentally the effects that the testimony of competing experts has on (1) sophisticated versus unsophisticated subjects' decisions and (2) subjects' decisions on difficult versus easy problems. Our results demonstrate that competing expert testimony, by itself, does not help unsophisticated subjects to behave as though they are sophisticated, nor does it help subjects make comparable decisions on difficult and easy problems. When we impose additional institutions (such as penalties for lying or a threat of verification) on the competing experts, we observe such dramatic improvements in unsophisticated subjects' decisions that the gap between their decisions and those of sophisticated subjects closes. We find similar results when the competing experts exchange reasons for why their statements may be correct. However, additional institutions and the experts' exchange of reasons are less effective at closing the gap between subjects' decisions on difficult versus easy problems.

> The end result of the adversary process is often conflicting testimony from experts—the "battle of the experts."
>
> Vidmar and Diamond (2001:1134)

> In some types of modern trials—such as those involving complex scientific findings or esoteric economic or mathematical evidence—there probably is no adequate substitute for actually comprehending the evidence, the arguments of the parties, and the judge's instructions.
>
> Lilly (2001:70)

*Address correspondence to Cheryl Boudreau, University of California, Davis, Department of Political Science, One Shields Ave., Davis, CA 95616; email: clboudreau@udavis.edu. Boudreau is Assistant Professor of Political Science, University of California, Davis; McCubbins is Chancellor's Associates Chair of Political Science, University of California, San Diego and Visiting Professor of Law, University of Southern California.

This article was presented at the Conference on Empirical Legal Studies at Cornell University, 2008. We thank Valerie Hans and Ted Eisenberg for their generous invitation to present our work at that conference. We also thank the National Science Foundation (Grant SES-0616904) and the Kavli Institute for Brain and Mind for providing financial support for these experiments. We are also grateful to Craig Burnett, David Dunning, William Heller, Scott MacKenzie, Rebecca Morton, Jeff Rachlinski, Dan Rodriguez, Matthew Spitzer, Jeff Staton, Lydia Tiede, and members of the University of San Diego Law School for helpful comments on earlier drafts of this article.

Can expert testimony help unsophisticated jurors make informed decisions? In modern jury trials, jurors are frequently exposed to complex mathematical and scientific evidence that expert witnesses present. Upon hearing such evidence, jurors must then decide the guilt or innocence of the accused or the liability of parties to a civil suit. Given the increasing complexity of modern jury trials, as well as research showing declines in jurors' sophistication levels (see Cecil et al. 1987; Lilly 2001), many scholars question whether lay jurors can understand expert testimony and use it to make informed decisions. Specifically, some scholars fear that jurors are not sufficiently sophisticated to process the competing information and complex arguments that are presented during trials (Elwork et al. 1977; Hastie & Viscusi 1998; Mogin 1998; Fisher 2000–2001; Lilly 2001). On the other hand, many scholars argue that jurors, despite their lack of legal or scientific sophistication, can assess the value of particular pieces of evidence and make informed decisions (Kalven & Zeisel 1966; Hastie et al. 1983; Cecil et al. 1991; Cooper et al. 1996; Lupia & McCubbins 1998; Boudreau 2006; Boudreau & McCubbins 2008).

To address this debate about expert testimony, juror sophistication, and decision difficulty, we assess the effects that competition between experts has on: (1) sophisticated versus unsophisticated individuals' decisions and (2) individuals' decisions on easy versus difficult problems. Specifically, we ask whether and when the testimony of competing experts helps unsophisticated individuals make decisions that are comparable to those of sophisticated individuals. That is, when does expert testimony close the sophistication gap?[1] We also ask whether and when the testimony of competing experts helps individuals make comparable decisions on difficult and easy problems. That is, when does expert testimony close the difficulty gap?

To address these questions, we derive several theoretical predictions about the conditions under which the testimony of competing experts should close the sophistication gap and the difficulty gap. We then use data from laboratory experiments to test these predictions. Specifically, we assess whether the pattern of results that we observe in our experiments matches our pattern of theoretical predictions. This type of research design—known as pattern matching—is particularly strong in internal validity (Trochim 2001) and is well suited for our study, given the number of predictions that we derive and the number of treatment groups that we include. Further, although we admittedly sacrifice some external validity by conducting experiments, we gain the ability to objectively and reliably measure our main constructs of interest: sophistication, decision difficulty, and the quality of subjects' decisions.

Consistent with our pattern of predictions, our results demonstrate that when unsophisticated subjects are exposed to the testimony of competing experts, they consistently make worse decisions than do sophisticated subjects. Similarly, we find that subjects, in the aggregate, make significantly worse decisions on difficult problems than on easy problems. Thus, the testimony of competing experts, by itself, does not close the sophistication gap among subjects, nor does it close the difficulty gap in subjects' decisions.

[1]See Boudreau (2009) for a study of the conditions under which a single expert's statements can close the gap between sophisticated and unsophisticated citizens' decisions.

However, when we impose additional institutions (such as a penalty for lying or a threat of verification[2]) on the competing experts or when we allow the competing experts to exchange reasons for why their statements may be correct, both sophisticated and unsophisticated subjects improve their decisions significantly. Importantly, the additional institutions and the exchange of reasons induce such large improvements among unsophisticated subjects that the gap between their decisions and those of sophisticated subjects closes. Somewhat surprisingly, we find that for the range of penalties for lying and chances of verification we examine, smaller penalties and slimmer chances of verification can be just as effective as large ones at closing the sophistication gap among subjects.

With respect to decision difficulty, we find that although some additional institutions help subjects make comparable decisions on difficult and easy problems, others do not. Thus, the additional institutions we impose on the competing experts do not consistently close the difficulty gap in subjects' decisions. Similarly, when the competing experts exchange reasons with one another, subjects make significantly worse decisions on difficult problems than on easy problems. Taken together, these results suggest that the additional institutions and the experts' exchange of reasons are less effective at closing the difficulty gap than they are at closing the sophistication gap. That said, the additional institutions sometimes help subjects achieve significant improvements in their decisions—just not enough to help subjects make comparable decisions on difficult and easy problems.

This article proceeds as follows. We begin by describing our experiments, which expose sophisticated and unsophisticated subjects to the statements of two competing experts before they make both easy and difficult decisions. Then, we make several predictions about how the competing experts' statements will affect the quality of subjects' decisions (1) when subjects are sophisticated versus unsophisticated and (2) when the decisions they must make are difficult versus easy. Next, we present our results. We conclude with a discussion of the implications that our results have for debates about the competence of lay jurors and the efficacy of our adversarial legal system.

I. Can Competing Expert Testimony Improve Jurors' Decisions?

Much theoretical and empirical research suggests that competition between experts reveals truthful information and improves decision making, even among unsophisticated jurors (Milgrom & Roberts 1986; Lipman & Seppi 1995; Froeb & Kobayashi 1996; Kim 2001; Walpin 2003). For example, Milgrom and Roberts (1986) demonstrate that an unsophisticated decisionmaker will make a fully informed decision as long as the interests of the competing litigants are sufficiently opposed (see also Froeb & Kobayashi 1996). Similarly, many other scholars suggest that jurors, despite their lack of legal or scientific sophistica-

[2]In legal contexts, jurors know that witnesses for both the prosecution and the defense face penalties for perjury if they lie on the stand. Similarly, attorneys' cross-examinations (which are a form of verification) often reveal when witnesses have made false statements.

tion, can learn what they need to know during the course of a trial. These scholars recognize that jurors may not possess the formal education or training that is needed to process the complex, technical evidence presented during trials. However, these scholars suggest that jurors are able to use cues to assess the value of particular pieces of evidence and to make informed decisions (Hovland & Weiss 1951; Kalven & Zeisel 1966; Chaiken 1980; Hass 1981; Hastie et al. 1983; Petty & Cacioppo 1984; Cecil et al. 1991; Cooper et al. 1996; Shuman et al. 1996; Lupia & McCubbins 1998; Boudreau 2006; Boudreau & McCubbins 2008).[3]

On the other hand, many scholars question whether competition between experts actually produces these positive outcomes. For example, much legal and social science research suggests that jurors are not sufficiently sophisticated to process the competing information and arguments that are presented during trials (Elwork et al. 1977; Selvin & Picus 1987; Faigman & Baglioni 1988; Thompson 1989; Hastie & Viscusi 1998; Mogin 1998; Fisher 2000–2001; Lilly 2001). Similarly, many scholars suggest that the increasing complexity of modern jury trials hinders lay jurors' ability to make informed decisions (Strier 1994; Fisher 2000–2001; Lilly 2001). Indeed, these scholars lament the fact that complex mathematical and technical issues are now commonplace in ordinary criminal and civil trials. Such concerns over jurors' levels of sophistication and the increasing complexity of trials have led many scholars to question the efficacy of the jury system and to advocate various reforms (Sutton 1990; Strier 1994; Broyles 1996; Mogin 1998; Fisher 2000–2001).[4]

It is this research on juror sophistication and decision difficulty that we build on in our experiments. In a previous article (Boudreau & McCubbins 2008), we reported the results of experiments that were designed to test whether and when competition between experts helps jurors, in the aggregate, to improve their decisions. There, we found that when additional institutions were imposed on competing experts or when competing experts exchanged reasons with one another, subjects achieved large improvements in their decisions. However, the experimental design originally described in Boudreau and McCubbins (2008) also allows us to address other important questions.[5] Thus, here we use our experiments to answer two new questions: (1) When does competition between two experts help unsophisticated jurors to make decisions that are comparable to those of sophisticated jurors? and (2) When does competition between two experts help jurors to make comparable decisions on difficult and easy problems?

Although there are many theoretical, experimental, and empirical studies of competing expert testimony, our experimental design makes a number of new contributions to the literature and has several important advantages. One advantage of our design stems

[3]For an interesting discussion of trial innovations that may help jurors understand scientific evidence, see Hans (2007). For an analysis of how judges might help jurors understand expert testimony, see Diamond (2007).

[4]There is also much interesting research on whether and when certain types of expert testimony should not be permitted (see Lyon 2000a), as well as research on the accuracy of eyewitness testimony (see Dunning & Stern 1994; Dunning & Perretta 2002) and testimony from children (Lyon 2000b, 2002).

[5]Taken together, Boudreau and McCubbins's (2008, 2009) results are, in some important respects, different from what economic signaling models would predict.

from the nature of the choices that subjects make in our experiments. Specifically, instead of asking subjects to make decisions about fictional court cases or to vote for fictional candidates (as psychologists, legal scholars, and political scientists often do when running experiments), we ask subjects to make choices about math problems after they hear two expert subjects make statements about whether answer "a" or answer "b" is the correct answer.[6] One reason that this type of decision is advantageous is that solving math problems provides a straightforward way to identify correct decisions and to assess whether and when the experts' statements induce an improvement in decision making. Stated differently, although it is often difficult to identify when subjects have chosen the "correct" legal decision,[7] it is very easy to tell when they have chosen the correct answer to a math problem. Thus, although our experiments are low in mundane realism (i.e., on the surface, math problems do not resemble the decisions that jurors make in courtroom settings; see Aronson et al. 1998), we gain the ability to objectively measure whether and when competing expert testimony helps subjects make better decisions.

By using math problems in our experiments, we also gain an objective, valid, and reliable measure of subjects' levels of sophistication. Indeed, although an agreed upon measure of legal sophistication does not exist, there does exist an agreed upon, widely used, and straightforward measure of mathematical sophistication—namely, SAT math scores. Further, subjects' SAT math scores provide a measure of sophistication that is directly related to the task that subjects perform in our experiment (i.e., solving math problems). For these reasons, we collect subjects' SAT math scores prior to the experiment, which enables us to assess whether sophisticated and unsophisticated subjects are equally likely to benefit from the competing experts' statements.

Another advantage of our design is that math problems, even though they do not *look* like legal decisions on the surface, capture many key characteristics of the information that jurors receive and the decisions that jurors make in real-world courtroom settings. Thus, they can tell us a great deal about how jurors in the real world make choices. For example, in the real world, jurors are not blank slates when they listen to the statements of competing experts. That is, they often have preexisting knowledge or beliefs about the topics that the experts discuss. Similarly, subjects in our experiments are not blank slates when they hear the experts' statements about whether answer "a" or answer "b" is the correct choice because they have preexisting knowledge about how to solve math problems.

That said, jurors in the real world might be uncertain about their decisions; that is, they may not know which choice (e.g., guilty vs. not guilty) is the correct one. Similarly, subjects in our experiments may be uncertain about whether "a" or "b" is the correct choice. As in the real world, the uncertainty that subjects experience depends, in part, on their

---

[6]Note that the two subjects who are chosen to be the experts are shown the correct answer to a particular math problem before they make their statements; thus, they are made expert by the experimenter. This is common knowledge to all participants in the experiment.

[7]See Diamond for a discussion of the difficulties associated with identifying whether and when jurors make "correct" decisions. Indeed, she states: "To assess how the jury operates as a decision-maker, we cannot compare the jury's verdict with some gold standard of truth because no such dependable standard exists . . . [I]n the end we cannot be certain that the correct conclusions have been drawn" (2003:150–51).

levels of sophistication. Indeed, just as unsophisticated jurors in the real world may be more uncertain about which choice is correct, so, too, may unsophisticated subjects in our experiments be more uncertain about whether "a" or "b" is the best choice. And, just as jurors in the real world vary in their levels of sophistication, so, too, do our subjects, as their SAT math scores range from 450 (the 27th percentile) to 800 (a perfect score).

Further, in real-world courtroom settings, there is something at stake for jurors when they make their decisions, but the stakes may not be very large. This is especially true in low-profile, run-of-the-mill court cases. Similarly, there is something at stake for subjects in our experiments because they earn money if they make a correct choice and lose money if they make an incorrect choice. As is often the case in the real world, the stakes in our experiments are not very large, as subjects either earn or lose 50 cents for each decision they make.

Further, jurors in real-world contexts often receive factual information, and they must then make decisions that are objectively correct or incorrect. For example, in courtroom settings, jurors listen to factual information that competing witnesses provide and then make a decision about whether the accused is guilty or innocent or whether a party to a civil suit is liable or not. Interestingly, in both criminal and civil trials, jurors are often exposed to mathematical information that is not unlike the information that subjects in our experiments receive (Lilly 2001; Fisher 2000–2001).[8] Like jurors in the real world, subjects in our experiments receive factual, mathematical information and then make decisions that are either correct or incorrect. The difference is that we are able to identify when subjects make correct choices, whereas in the real world, it is more difficult (if not impossible) to do this. Given the many similarities between real-world legal decisions and decisions about math problems, there is a close mapping between the psychological processes of subjects in our experiments and the psychological processes of jurors in real-world courtrooms (stated differently, our experiments have a great deal of psychological realism; see Aronson et al. 1998).

II. Experimental Design

To shed light on debates about juror sophistication and decision difficulty, we conduct laboratory experiments in which two competing experts make statements to jurors. In these experiments, we obtain a pretest measure of subjects' levels of sophistication (i.e., their SAT math scores) prior to the experiment. We then randomly assign subjects to either a control group or to one of several different treatment groups. We next ask subjects in all groups to answer binary-choice math problems that are drawn from an SAT math test and consist of several different types of problems and various levels of difficulty. We tell subjects in our treatment and control groups that they have 60 seconds to answer each math problem and

[8]Lilly (2001:71) states: "[W]ith increasing frequency, contemporary juries are faced with sophisticated, highly technical evidence drawn from such diverse fields as economics, mathematics, statistics, psychiatry, engineering, epidemiology, toxicology, serology, and genetics."

that they will earn 50 cents for each problem they answer correctly, lose 50 cents for each problem they answer incorrectly, and neither earn nor lose 50 cents if they leave a problem blank.

The main difference between the treatment and control groups has to do with the conditions under which subjects answer the math problems. In the control group, subjects answer the math problems on their own, one at a time. The purpose of the control group is to establish a baseline for how well sophisticated and unsophisticated subjects perform on difficult and easy math problems when they must make their choices on their own (i.e., without an opportunity to learn from two competing experts' statements).

In our various treatment groups, subjects also answer these math problems one at a time; however, subjects in the treatment groups receive the statements of two competing experts before they make their decisions. Specifically, before each math problem, the experimenter selects two subjects at random to act as the experts for that particular math problem. The experts are then shown the correct answer to a particular math problem (i.e., the experts are given knowledge about the correct choice), and they then choose what statement they would like to make to subjects (statements take the form of answer "a" or answer "b").[9] Once the experts choose their statements, the "testimony" of the two experts is passed on to the remaining subjects who must decide within 60 seconds whether to answer the problem, and if they choose to answer, whether to pick "a" or "b."

Once subjects make their decisions, we move on to the next math problem. At this point, we randomly select two new subjects to act as the experts.[10] We select two new subjects for each math problem to ensure that the competing experts do not develop reputations from one problem to the next.[11] Indeed, avoiding repeat play effects is important because such effects could confound the treatments we impose. Further, because jurors' interactions with competing experts typically do not repeat over time (e.g., they are typically limited to the course of a trial), it makes sense to have subjects make "one-shot" decisions in our experiments.[12]

That said, the key to our experimental design is threefold. First, one of the experts knows that he or she shares common interests with subjects (i.e., he or she benefits when

---

[9]Note that the experimenter reads both experts' statements aloud to the subjects in order to prevent subjects from learning anything about the experts from the sound of their voices. Also, all subjects (including the experts) sit behind large partitions so that their identities are anonymous.

[10]In some experiments, we randomly select two new subjects to act as the experts by pulling two subject numbers out of a hat before each math problem. In other experiments, we randomly select four or more subjects at the beginning of the experiment to act as a panel from which we draw two competing experts on any particular math problem. Then, before distributing each math problem, we draw two numbers out of a hat to determine which two subjects on the panel will act as the experts on that problem. We repeat this procedure for each math problem.

[11]The total number of math problems that any one subject (acting as an expert) makes statements about depends on which procedure is used to select the experts, as well as the number of math problems used in a particular experiment. Across all experiments, the total number of math problems that an expert makes statements about ranges from 14 problems to 1 problem. Also, subjects answer between 5 and 24 math problems in each experiment. We control for this in our statistical analyses, and it does not affect our results.

[12]For an example of a repeated communication game, see Sobel (1985).

the remaining subjects answer the math problem correctly), and the other expert knows that he or she has conflicting interests with subjects (i.e., he or she benefits when the remaining subjects answer the math problem incorrectly). We refer to these as the common-interest and conflicting-interest experts, respectively. Second, it is common knowledge to all subjects that one expert shares common interests with them and that one expert has conflicting interests with them, but they are not told which expert shares common interests with them on any particular problem. Stated differently, subjects know that the experts are adversaries, but *they do not know which expert's interests are aligned with their own.*[13] Third, both experts and the subjects know that the experts can lie about the correct answer to the math problem or tell the truth; it is entirely up to them. The experts' ability to lie or tell the truth is constant throughout this experiment and is designed to be an analogy to Crawford and Sobel's (1982) and Lupia and McCubbins's (1998) models, as well as to many real-world competitive settings.

So how do we induce competition between the two experts within the context of our experiments? In short, we induce competition by manipulating the ways that the experts and the subjects earn money. Specifically, the common-interest expert is paid 50 cents for each subject who answers a particular math problem correctly. The conflicting-interest expert is paid 50 cents for each subject who answers a particular math problem *incorrectly*. So, for example, if 11 subjects (the typical number used in our experiments) answer the math problem correctly, they earn 50 cents each, the common-interest expert earns $5.50 (i.e., 50 cents for each of the 11 subjects who answer the problem correctly), and the conflicting-interest expert loses $5.50 (i.e., 50 cents for each of the 11 subjects who answer the problem correctly). Similarly, if 11 subjects answer the math problem incorrectly, then the common-interest expert loses $5.50, and the conflicting-interest expert earns $5.50. In this way, the interests of the two competing experts are strictly adversarial, or zero sum.

In addition to the competition treatment condition described above, treatment group subjects are exposed to several other treatment conditions in each experiment, which creates a within-group design. In these other treatment conditions, we alter the basic competition condition in one of two ways: (1) following Lupia and McCubbins (1998), we vary the institutional context in which the competing experts make their statements or (2)

[13]We design our experiments in this way because we are interested in analyzing the conditions under which citizens can learn from competing experts *who they do not know.* This aspect of our model corresponds to many legal and political settings. Indeed, when learning from competing experts in a courtroom, jurors may not know which expert's interests are aligned with their own. Similarly, in primary elections, citizens may not know which candidate (of the many candidates who share their party label) is more likely to have interests that are aligned with their own.

That said, we could design our experiments differently. Indeed, if we were interested in analyzing the conditions under which citizens can learn from competing experts *who they know something about,* we could tell subjects that there is a 70 percent chance that the second expert has conflicting interests with them. Because the two experts in our experiments are adversaries, this means that subjects would also know that there is a 70 percent chance that the first expert has common interests with them. Knowing this, subjects should ignore the second expert's statement, pay attention to the first expert's statement, and base their choices on it. However, because this design simply turns our experiments into a test of how perceived common interests between an expert and a citizen affect trust, persuasion, and learning, and because the effects of common interests are well understood (see Lupia & McCubbins 1998 for a game theoretic model and experiments demonstrating the effects of common interests), we focus instead on situations in which subjects do not know which expert is more likely to share common interests with them on any particular problem.

we allow the competing experts to exchange reasons for why "a" or "b" may be the correct choice for subjects. Each of these experimental variations is common knowledge. The details of how we implement these other treatment conditions are described below.

A. Additional Institutions

In these other treatment conditions, we alter the basic competition condition by imposing one of two additional institutions on the competing experts: namely, a *penalty for lying* or a *threat of verification*. Both these additional institutions have analogues in real-world legal contexts. Specifically, jurors know that witnesses for both the prosecution and the defense face penalties for perjury if they lie on the witness stand. Or, an expert witness may incur a penalty for lying (in the form of a loss of reputation) if he or she is caught lying while testifying during a trial. Similarly, attorneys' cross-examinations are a form of verification that may reveal when witnesses have made false statements.

To impose a penalty for lying in our experiments, we manipulate the way that the experts earn money. Specifically, we maintain the competition between the two experts, and we then impose a penalty for lying on both experts. So, in these treatment conditions, the experts are engaged in competition as before, but we announce to the experts and the subjects that both experts will incur a penalty (either a large \$15 penalty or a smaller \$5 or \$1 penalty, depending on the treatment condition) if they lie about the correct answer to the math problem. We vary the size of the penalty because penalties for lying in real-world courtroom settings also vary in how large they are, relative to what is at stake for the experts. For example, the \$15 penalty for lying is a very large penalty, relative to what is at stake for the experts in our experiments. The \$1 penalty, however, is small relative to the money the experts can gain in our experiments. Indeed, the conflicting-interest expert can gain as much as \$5.50 if 11 subjects choose the wrong answer to a problem and lose only \$1 for lying about the correct answer. Similarly, the penalties for lying that experts in real-world courtroom settings face can be large or small, depending on the nature of the trial, the stakes involved, and the value that the expert places on his or her reputation.

For the other additional institution—namely, verification—we again maintain competition between the two experts, but this time we verify both experts' statements with some probability to make sure that they are true statements before they are read to subjects. In the high (i.e., 100 percent) chance of verification condition, if either expert chooses to make a false statement about the correct answer to the math problem, we do not read the false statement(s) aloud to subjects, but replace it (them) with the correct answer when we announce the experts' statements. If either expert makes a true statement, then we simply read that expert's statement aloud to subjects.

However, because a 100 percent chance of verification is unlikely to occur in real-world courtroom settings, we also examine the effects of a smaller, 50 percent chance of verification. In the 50 percent chance of verification condition, we roll a six-sided die before the experts' statements are read aloud to subjects. If the die lands on 1 through 3, then we silently verify both experts' statements and replace any false statements with truthful statements when we read the experts' statements aloud to subjects. If the die lands on 4 through 6, however, then we simply announce the answers that the experts choose to report, regardless of whether they are correct or incorrect. In this way, subjects know that there is a 50 percent chance that the experts will be verified, but they do not know whether the experts have been verified on any particular problem.

To make our verification and penalty for lying conditions even more realistic, we add a conditional penalty for lying to the threat of verification in other treatment conditions. We do this because, in the real world, experts must often pay a cost if they are verified and caught lying. Further, experts' false statements are not likely to be detected or punished with certainty in real-world courtroom settings. Thus, in these treatment conditions, the competing experts must pay a cost (of either $1 or $2, depending on the treatment condition) if they make a false statement *and* are verified by the experimenter. In this way, the penalty is conditional (and is different from the penalty for lying treatment conditions described above) because *it is only imposed if verification occurs*. For example, in the 50 percent chance of verification plus a $1 penalty condition, both experts lose $1 of their experimental earnings if the die lands on 1 through 3 (i.e., verification occurs) *and* if they had chosen to make a false statement. However, if the die lands on 4 through 6 (i.e., verification does not occur) and if both experts had chosen to make false statements, then we do not subtract $1 from the experts' experimental earnings, and we read the experts' false statements aloud to subjects. The other treatment conditions in which we impose a chance of verification plus a conditional penalty proceed in a similar manner.

B. Reason Giving

In our reason-giving treatment conditions, we alter the basic competition condition by allowing the competing experts to exchange reasons for why "a" or "b" may be the correct choice. These experiments are identical to the basic competition condition (i.e., competition without additional institutions) with four exceptions. First, the competing experts are no longer required to make statements about whether "a" or "b" is the correct choice; that is, both experts can choose to remain silent. Second, if either expert chooses to make a statement, he or she must not only state whether "a" or "b" is the correct choice, but also give a reason for why the answer the expert chooses to report may be correct. Third, the experts have an opportunity to respond to one another's statements and reasons; that is, the experts have three opportunities to give reasons and counter-reasons. Fourth, in some reason-giving treatment conditions, the competing experts must pay a small cost ($1) each time they wish to recommend an answer and give a reason. Thus, competition between experts in these reason-giving conditions is analogous to the testimony given to juries, where witnesses may be called to rebut one another's statements and attorneys may counter the arguments made by opposing counsel.[14]

[14]As Vidmar and Diamond (2001:1133–34) note: "The adversary system relies on the opposing side to cross-examine and deconstruct the testimony of the expert to expose its weakness or irrelevance to the dispute. Then, the first party is allowed to reexamine its experts to rehabilitate them. When the opposing party begins its evidence presentation, it may call its own experts in an effort to refute the other party's experts. The end result of the adversary process is often conflicting testimony from experts—the 'battle of the experts.' "

Specifically, our reason-giving experiments proceed as follows. After both experts are shown the correct answer to a math problem, the first expert chooses whether to make a statement to the other subjects about whether "a" or "b" is the correct answer. If the first expert chooses to make a statement, then he or she must decide whether to state "a" or "b." After the first expert selects which answer to report, then he or she must also select a reason for why the answer that he or she chose may be correct. When selecting a reason to support his or her statement of "a" or "b," the first expert may choose from a brief menu of correct and incorrect reasons that we provide or write down his or her own reason. As before, the first expert's statement and reason may truthfully reflect the correct answer or falsely indicate a different answer. After the first expert selects a statement and reason, we distribute that statement and reason to the second expert. If the first expert chooses not to make a statement, then we tell that to the second expert and see whether the second expert wishes to make a statement and send a reason.

The second expert then chooses whether to make a statement. If the second expert chooses to make a statement, then he or she must decide whether to recommend "a" or "b" to subjects. After choosing his or her statement, the second expert must then provide a reason for the answer that he or she chose. When selecting a reason to support his or her statement of "a" or "b," the second expert may also choose from the list of correct and incorrect reasons that we provide or write down his or her own reason. We then distribute this statement and reason to the first expert. If the second expert chooses not to make a statement, then we tell that to the first expert and see whether the first expert would like to make another statement.

We continue this process until both experts have had three opportunities to provide reasons for their recommended answer. Although both experts can choose up to three reasons for why their recommendation of "a" or "b" may be correct, the experts cannot change their statement on a particular math problem. That said, they can repeat the same reason over and over, or they can choose a new reason each time. In this way, we allow the competing experts to go back and forth about whether "a" or "b" is the correct choice for subjects and about why "a" or "b" may be correct or why the other expert may be wrong. At the end of the trial, we allow subjects to see a transcript of the experts' exchange of statements and reasons. That is, we have the two experts debate first, and we then distribute the transcript of the debate in order to save time and reduce subject boredom. It also allows us to maintain the anonymity of the experts, as subjects do not hear them, see their writing (except after we transcribe it), or have any other contact or communication with the experts. We then give subjects 60 seconds to choose an answer to the math problem.

III. Pattern of Predictions

We now derive a pattern of predictions about how the testimony of competing experts in each treatment condition should affect the quality of subjects' decisions when subjects are sophisticated versus unsophisticated and when the decisions are difficult versus easy. We measure the quality of subjects' decisions by observing (1) the relative amounts of money that sophisticated and unsophisticated subjects earn in our treatment and control groups and (2) the relative amounts of money that subjects earn on difficult versus easy problems in our treatment and control groups. Indeed, because subjects earn money for answering problems correctly and lose money for answering problems incorrectly in all treatment conditions and in the control group, the amount of money that subjects earn is a straightforward measure of whether and when competition between experts (with or without additional institutions or the exchange of reasons) enables subjects to learn from the experts' statements and improve their decisions.

Specifically, in the basic competition condition, we make the following predictions.

- *Competition-Sophistication Hypothesis*: Sophisticated subjects who are exposed to the competing experts' statements will earn more money than unsophisticated subjects who are exposed to the competing experts' statements.
- *Competition-Difficulty Hypothesis*: Subjects who are exposed to the competing experts' statements will earn more money on easy problems than on difficult problems.

The logic behind these predictions is best understood by considering the nature of our experiments. Recall that subjects are not told which expert is more likely to share common interests with them. Thus, if subjects hear the statement "a" from one expert, followed by the statement "b" from the other expert, they cannot necessarily learn anything from these statements because they do not know whether the first expert (who stated "a") or the second expert (who stated "b") has common interests with them. However, subjects may be able to infer which expert made a truthful statement (1) if they are sophisticated (i.e., they can determine whether "a" or "b" is the correct choice on their own, thereby verifying the experts' statements) or (2) if the problem is easy enough that even unsophisticated subjects can determine the correct answer on their own. For this reason, we expect sophisticated subjects to earn more money than unsophisticated subjects in the competition condition. We also expect subjects to earn more money on easy problems than on difficult problems in this condition.

In the $15 penalty for lying condition, we expect different results. Specifically, when both competing experts are subject to a $15 penalty for lying, we make the following predictions.

- *Large Penalty-Sophistication Hypothesis*: There will be no difference in the amounts of money that sophisticated and unsophisticated subjects earn.
- *Large Penalty-Difficulty Hypothesis*: There will be no difference in the amounts of money that subjects earn on difficult versus easy problems.

The logic behind these predictions stems from Lupia and McCubbins's (1998) model, which demonstrates that when a penalty for lying is sufficiently large, then, in equilibrium, a speaker never has an incentive to lie, and the subjects trust and learn from the speaker's statements. In the context of our experiments, a $15 penalty is "sufficiently large"—that is, it is large enough to ensure that both experts have a dominant strategy to tell the truth and that all subjects (regardless of their levels of sophistication) know this.[15] Indeed, to ensure that subjects understand that the experts always have an incentive to tell the truth in this treatment condition, we give them a quiz at the beginning of the experiment that asks them to state how much money the experts earn under various circumstances. Subjects, by and large, answer the quiz questions correctly. If a subject answers a quiz question incorrectly, we explain to that subject why the answer that he or she chose is incorrect and reveal the correct answer to that subject. In this way, we are certain that all subjects understand the experts' incentives in this condition. Thus, we expect subjects to trust the experts' statements and base their decisions on them, regardless of their levels of sophistication and the difficulty of the problem.

When a 100 percent chance of verification (with or without a conditional penalty for lying) is imposed on the competing experts, we expect results that are similar to those in the $15 penalty for lying condition. Specifically, when both competing experts are subject to a 100 percent chance of verification, we make the following predictions.

- *Verification-Sophistication Hypothesis*: There will be no difference in the amounts of money that sophisticated and unsophisticated subjects earn.
- *Verification-Difficulty Hypothesis*: There will be no difference in the amounts of money that subjects earn on difficult versus easy problems.

These predictions also stem from Lupia and McCubbins's (1998) model, which demonstrates that increasing the probability of verification decreases the probability that a speaker can benefit from making a false statement. Thus, when there is a 100 percent chance that both experts' statements will be verified, subjects are certain to receive two truthful statements about the correct choice. Thus, subjects should trust the experts' statements, base their choices on them, and make better decisions as a result—again, regardless of their levels of sophistication and the difficulty of the problem.[16]

A. *Smaller Penalties and Slimmer Chances of Verification*

As for the smaller penalties for lying and slimmer chances of verification that we impose in our experiments (i.e., a $5 penalty, a $1 penalty, and a 50 percent chance of verification),

---

[15]To see why this is the case, consider the way that the conflicting-interest expert earns money in this condition: given that there are conflicting interests between this expert and the subjects, this expert earns $5.50 if each subject answers a problem incorrectly. Although at first blush this might seem to give this expert an incentive to lie, note that the $15 penalty for lying will reduce this expert's gain of $5.50 down to a loss of $9.50. Further, if this expert lies and all the subjects happen to answer the problem correctly, then this expert will lose $20.50 (i.e., a $15 loss because of the penalty for lying and a $5.50 loss because 11 subjects answered the problem correctly). If the conflicting-interest expert tells the truth, however, then the worst he or she can do is to lose $5.50 (which will happen if each subject answers the problem correctly), and the best that he or she can do is to earn $5.50 (which will happen if each subject answers the problem incorrectly). As these payoffs make clear, this expert is always better off if he or she tells the truth about the correct answer to the math problem. The same is, of course, true for the common-interest expert.

[16]As in the $15 penalty for lying condition, we give subjects a quiz at the beginning of the experiment to ensure that they understand how this treatment condition will proceed. Thus, we are certain that subjects understand that the experts will be verified with certainty in this condition.

we predict that we will observe differences in the amounts of money that subjects earn, depending on their levels of sophistication and the difficulty of the problem. Specifically, we make the following predictions.

- *Smaller Penalty/Verification-Sophistication Hypothesis*: Sophisticated subjects will earn more money than unsophisticated subjects.
- *Smaller Penalty/Verification-Difficulty Hypothesis*: Subjects will earn more money on easy problems than on difficult problems.

The logic behind these predictions is that, under each one of these conditions, the conflicting-interest expert does not necessarily have an incentive to make truthful statements, in equilibrium. For example, if the conflicting-interest expert believes that a particular math problem is difficult, then he or she may have an incentive to lie about the correct answer to the problem. The conflicting-interest expert has this incentive on the difficult problems because only the most sophisticated subjects will be able to solve the problem on their own, thereby verifying the experts' statements and recognizing which expert is lying. The unsophisticated subjects, however, may be fooled by the conflicting-interest expert's lie or not know which choice to make if they receive conflicting statements from the two experts. On the easy problems, however, the conflicting-interest expert should tell the truth about the correct answer whenever there is a penalty for lying in place. The reason for this is that on the easy problems, subjects do not need expert testimony to answer the problem correctly. Thus, the expert's lie will fool few (if any) subjects, and the expert will suffer a penalty for lying nonetheless.

What the above discussion boils down to is an expectation that sophisticated subjects should earn more money than unsophisticated subjects in these treatment conditions because they are less dependent on the experts, who may have an incentive to lie. Further, subjects should earn more money on the easy problems both because they are more likely to be able to determine the correct choice on their own and because they are more likely to receive truthful statements from the experts. Subjects should earn less money on the difficult problems because they are less likely to be able to determine the correct choice on their own; thus, they are more dependent on the experts, who may lie about the correct answer to these problems.

B. Exchange of Reasons

Allowing both competing experts to exchange reasons for why their statements may be correct does not alter the basic competition hypotheses that we state above. That is, even though the competing experts in this treatment condition offer reasons for why their statements may be correct and have the opportunity to go back and forth, we still predict that we will observe differences in the amounts of money that subjects earn, depending on their levels of sophistication and the difficulty of the problem. Specifically, we make the following predictions.

- *Reasoning-Sophistication Hypothesis*: Sophisticated subjects will earn more money than unsophisticated subjects.

- *Reasoning-Difficulty Hypothesis*: Subjects will earn more money on easy problems than on difficult problems.

We make these predictions because, as in the basic competition condition, subjects are not told which expert is more likely to share common interests with them. Thus, subjects cannot necessarily learn anything from the experts' statements and reasons because they do not know which expert has common interests with them. However, subjects may be able to infer which expert made a truthful statement from the statements and reasons that the experts give. For example, subjects may be able to identify a flaw in one expert's logic or an incorrect reason, even if they do not immediately know the correct answer to the math problem. And, subjects should be better able to identify faulty reasoning if they are sophisticated or if the problem is easy enough so that the exchange of reasons is clear and comprehensible to even unsophisticated subjects.

IV. Data Analysis

When analyzing the data gleaned from our experiments, we first define our main variables of interest: sophistication and decision difficulty. When classifying subjects as sophisticated or unsophisticated, we use subjects' SAT math scores, as well as the nationwide SAT math percentile rankings that the Educational Testing Service releases. Specifically, subjects whose SAT math scores fall above the median score for our sample are considered sophisticated, while subjects whose SAT math scores fall below the median are considered unsophisticated. In terms of the scores associated with these classifications, sophisticated subjects' scores range from 680 to 800 (the 91st percentile and higher), while unsophisticated subjects' scores range from 450 to 660 (the 27th percentile through the 87th percentile).[17]

When classifying decisions as difficult or easy, we use the results from subjects in our control group, who answer the math problems on their own. That is, we classify a math problem as difficult if less than 50 percent of subjects in the control group answer it correctly. We classify a math problem as easy if more than 50 percent of subjects in the control group answer it correctly.[18] In terms of the ranges associated with these classifications, the difficult problems in our experiment have between 6 percent and 46 percent of control group subjects answering them correctly. The easy problems have between 54 percent and 90 percent of control group subjects answering them correctly.

We then estimate an ordinary least squares (OLS) regression[19] in which we regress a variable that reflects the amount of money that subjects earn on: (1) a dummy variable for

---

[17]Our results are robust to different specifications of sophistication. Specifically, our conclusions do not change if we alter the high and low SAT math scores used to classify subjects as sophisticated or unsophisticated.

[18]Our results are also robust to different specifications of decision difficulty.

[19]As a robustness check, we also estimated a random-effects GLS regression, which produced results that are similar to the ones we report in this article. A random-effects model was used to account for unobserved subject heterogeneity.

each treatment condition (i.e., the $15 PENALTY FOR LYING variable is coded 1 for the $15 penalty for lying condition and 0 otherwise), (2) a SOPHISTICATION dummy variable (coded 1 if a subject is considered sophisticated and 0 if a subject is considered unsophisticated), (3) a DIFFICULTY dummy variable (coded 1 if a problem is considered difficult and 0 if a problem is considered easy), (4) interactions between SOPHISTICATION and the dummy variables for each treatment condition, (5) interactions between DIFFICULTY and the dummy variables for each treatment condition, and (6) variables that control for the order in which subjects solve the problems, as well as characteristics of the subjects.[20] The omitted category in this regression is the control group.[21]

The regression results reported in Table 1 estimate the effects of each variable, assuming that all other variables are held constant at their mean values. Because most of the variables in the model are dummy variables that represent participation in a particular treatment condition, these coefficients do not provide estimates that correspond to real subjects or groups of subjects. Thus, we use simulations to estimate the expected amount of money that subjects earn in each treatment condition and in the control group (King et al. 2000; Tomz et al. 2003). Specifically, we estimate the expected amount of money that sophisticated and unsophisticated subjects earn in each experimental condition, holding the difficulty of the decision and the control variables constant at their mean values. We then analyze the expected amount of money that subjects earn on easy versus difficult problems in each experimental condition, holding subjects' levels of sophistication and the control variables constant at their mean values. These analyses allow us to compare the decisions of sophisticated and unsophisticated subjects, as well as subjects' decisions on easy versus difficult problems.

V. Results

A. Sophisticated Versus Unsophisticated Subjects

Our results for the competition, $15 penalty for lying, and 100 percent chance of verification conditions are consistent with our predictions.[22] As expected, sophisticated subjects earn significantly more money than unsophisticated subjects in the competition condition

---

[20]Specifically, the ORDER variable captures the number of math problems used in a particular experiment, as well as the order in which they were presented to subjects. As for the characteristics of subjects, we control for whether they have taken a college math class and their year in school, as well as whether their college major is in a math-related discipline. We control for these subject characteristics because there were small differences in them across our various treatment groups. To save space, we do not report the results of these control variables in Table 1, but they are available from the authors upon request.

[21]We also estimate an ordered probit model in which our dependent variable takes on a value of 0 if subjects answer the problem incorrectly, 1 if subjects leave the problem blank, and 2 if subjects answer the problem correctly. The results of this model are consistent with the conclusions presented in this article.

[22]In our previous article, we analyze the truthfulness of the experts' statements in each treatment condition (Boudreau & McCubbins 2008). We find that the experts' propensity to make truthful statements in each treatment condition is largely consistent with our predictions and with the logic of Lupia and McCubbins's (1998) model.

Table 1: Determinants of the Amount of Money Subjects Earn

Independent Variables	*Main Effects*	*Sophistication × Treatment Condition*	*Difficulty × Treatment Condition*
Competition	–0.035	0.042	0.142*
	(0.045)	(0.050)	(0.053)
$15 penalty for lying	0.237*	–0.063	0.141*
	(0.054)	(0.062)	(0.063)
$5 penalty for lying	0.209*	–0.070	0.111
	(0.056)	(0.065)	(0.067)
$1 penalty for lying	0.155*	–0.022	0.187*
	(0.052)	(0.062)	(0.064)
100% verification	0.161*	–0.009	0.198*
	(0.054)	(0.072)	(0.065)
100% verification + $1 penalty	0.240*	–0.086	0.162*
	(0.043)	(0.060)	(0.057)
100% verification + $2 penalty	0.258*	–0.080	0.256*
	(0.055)	(0.062)	(0.064)
50% verification	0.211*	–0.096	–0.326*
	(0.060)	(0.072)	(0.073)
50% verification + $1 penalty	0.062	–0.097	0.344*
	(0.050)	(0.064)	(0.062)
50% verification + $2 penalty	0.163*	0.037	–0.046
	(0.064)	(0.069)	(0.071)
Free reasoning	0.160*	0.002	–0.068
	(0.070)	(0.071)	(0.076)
Costly reasoning	0.124	0.042	–0.029
	(0.069)	(0.071)	(0.076)
Difficulty	–0.269*	—	—
	(0.022)		
Sophistication	0.076*	—	—
	(0.016)		
Constant	0.229	—	—
	(0.030)		
N	2754	—	—
R^2	0.272		

NOTE: Table 1 displays the coefficients from the OLS regression we estimated. The control group is the omitted category. The results for the additional control variables are not shown in this table. These results are available from the authors upon request. Dependent variable = Amount of money subjects earn on each problem; standard errors in parentheses; *$p < 0.05$.

(just as they do in the control group). As shown in Table 2, sophisticated subjects in the competition condition earn an estimated $0.19 per problem, while unsophisticated subjects in the competition condition earn only $0.07 per problem. This difference between sophisticated and unsophisticated subjects is statistically significant ($p < 0.05$).

Our sophistication results for the $15 penalty for lying and 100 percent chance of verification conditions are also consistent with our predictions. That is, we find that in each one of these conditions, there is no difference in the amount of money that sophisticated and unsophisticated subjects earn. As shown in Table 2, sophisticated subjects in the $15

Table 2: Expected Amounts of Money that Unsophisticated and Sophisticated Subjects Earn in Each Experimental Condition

Experimental Condition	*Unsophisticated: Predicted Amount of Money Earned*	*Sophisticated: Predicted Amount of Money Earned*	*Size and Significance of Difference Between Unsophisticated and Sophisticated*
Control group	\$0.09	\$0.16	**\$0.07** (0.04, 0.11)
Competition	\$0.07	\$0.19	**\$0.12** (0.04, 0.22)
\$15 penalty for lying	\$0.35	\$0.36	\$0.01 (−0.10, 0.12)
\$5 penalty for lying	\$0.34	\$0.34	\$0 (−0.12, 0.13)
\$1 penalty for lying	\$0.28	\$0.32	\$0.04 (−0.07, 0.16)
100% verification	\$0.26	\$0.32	\$0.06 (−0.07, 0.20)
100% verification + \$1 penalty	\$0.37	\$0.35	\$0.02 (−0.12, 0.10)
100% verification + \$2 penalty	\$0.38	\$0.37	\$0.01 (−0.12, 0.11)
50% verification	\$0.32	\$0.29	\$0.03 (−0.16, 0.11)
50% verification + \$1 penalty	\$0.20	\$0.16	\$0.04 (−0.16, 0.10)
50% verification + \$2 penalty	\$0.29	\$0.39	\$0.10 (−0.02, 0.24)
Free reasoning	\$0.27	\$0.34	\$0.07 (−0.06, 0.22)
Costly reasoning	\$0.24	\$0.35	\$0.11 (−0.02, 0.24)

NOTE: Table 2 converts the coefficients from the regression in Table 1 to expected values for sophisticated and unsophisticated subjects. These results reflect first differences with all other treatment variables set to zero and the control variables held constant at their mean values (King et al. 2000; Tomz et al. 2003). Confidence intervals in parentheses; boldface indicates that the 95 percent confidence interval does not contain zero, signifying a statistically significant difference between sophisticated and unsophisticated subjects in a particular experimental condition.

penalty for lying condition earn an estimated \$0.36 per problem, and unsophisticated subjects in the \$15 penalty for lying condition earn an estimated \$0.35 per problem. This one-cent difference between sophisticated and unsophisticated subjects is not statistically significant. Similarly, in the 100 percent chance of verification condition, sophisticated subjects earn an estimated \$0.32 per problem, and unsophisticated subjects earn an estimated \$0.26 per problem. In the 100 percent chance of verification plus a \$1 penalty condition, sophisticated subjects earn an estimated \$0.35 per problem, and unsophisticated subjects earn an estimated \$0.37 per problem. In the 100 percent chance of verification plus a \$2 penalty condition, sophisticated subjects earn an estimated \$0.37 per problem and unsophisticated subjects earn an estimated \$0.38 per problem. None of these differences between sophisticated and unsophisticated subjects in the 100 percent chance of verification conditions are statistically significant.

Contrary to our expectations, we find that smaller penalties for lying and slimmer chances of verification also quite effectively close the gap between the performance of sophisticated and unsophisticated subjects. As shown in Table 2, both sophisticated and unsophisticated subjects in the $5 penalty for lying condition earn an estimated $0.34 per problem. Similarly, sophisticated subjects in the $1 penalty for lying condition earn an estimated $0.32 per problem, while unsophisticated subjects earn an estimated $0.28 per problem This difference between sophisticated and unsophisticated subjects is not statistically significant. Further, in the 50 percent chance of verification condition, sophisticated subjects earn an estimated $0.29 per problem and unsophisticated subjects earn an estimated $0.32 per problem. In the 50 percent chance of verification plus a $1 penalty condition, sophisticated subjects earn an estimated $0.16 per problem, while unsophisticated subjects earn an estimated $0.20 per problem. In the 50 percent chance of verification plus a $2 penalty condition, sophisticated subjects earn an estimated $0.39 per problem, and unsophisticated subjects earn an estimated $0.29 per problem. None of these differences between sophisticated and unsophisticated subjects in the 50 percent chance of verification conditions are statistically significant.

Interestingly, we also find that the competing experts' exchange of reasons closes the gap between the performance of sophisticated and unsophisticated subjects. As shown in Table 2, sophisticated subjects in the free reasoning condition (i.e., where the experts may exchange reasons without paying a cost) earn an estimated $0.34 per problem, and unsophisticated subjects earn an estimated $0.27 per problem. Similarly, in the costly reasoning condition (i.e., where the experts must pay $1 to exchange reasons), sophisticated subjects earn an estimated $0.35 per problem and unsophisticated subjects earn an estimated $0.24 per problem. Contrary to our expectations, neither of these differences between sophisticated and unsophisticated subjects is statistically significant.

Taken together, these results demonstrate that even small penalties for lying, slimmer chances of verification, and the exchange of reasons close the sophistication gap among subjects. Importantly, these additional institutions and the exchange of reasons do more than simply boost the performance of unsophisticated subjects up to the levels of sophisticated subjects; rather, the additional institutions and the exchange of reasons help *both* sophisticated and unsophisticated subjects to make significantly better decisions than their sophisticated and unsophisticated counterparts in the control group.[23] The even larger improvements that unsophisticated subjects achieve are what close the gap between their decisions and those of sophisticated subjects.

B. Difficult Versus Easy Decisions

Our results for decision difficulty are also largely consistent with our predictions. As expected, subjects earn significantly more money on easy problems than on difficult

---

[23]The one exception to this statement is sophisticated subjects in the 50 percent chance of verification plus a $1 penalty condition (see Table 2).

Table 3: Expected Amounts of Money that Subjects Earn on Easy Versus Difficult Problems in Each Experimental Condition

Experimental Condition	*Easy Decisions: Predicted Amount of Money Earned*	*Difficult Decisions: Predicted Amount of Money Earned*	*Size and Significance of Difference Between Easy and Difficult Problems*
Control group	\$0.29	\$0.02	**\$0.27** (0.23, 0.31)
Competition	\$0.25	\$0.13	**\$0.12** (0.04, 0.22)
\$15 penalty for lying	\$0.53	\$0.40	**\$0.13** (0.02, 0.25)
\$5 penalty for lying	\$0.50	\$0.34	**\$0.16** (0.03, 0.29)
\$1 penalty for lying	\$0.44	\$0.36	\$0.08 (–0.02, 0.19)
100% verification	\$0.45	\$0.38	\$0.07 (–0.05, 0.19)
100% verification + \$1 penalty	\$0.53	\$0.42	\$0.11 (–0.002, 0.21)
100% verification + \$2 penalty	\$0.55	\$0.53	\$0.02 (–0.10, 0.14)
50% verification	\$0.50	\$–0.09	**\$0.59** (0.46, 0.73)
50% verification + \$1 penalty	\$0.35	\$0.43	\$0.08 (–0.03, 0.19)
50% verification + \$2 penalty	\$0.45	\$0.14	**\$0.31** (0.18, 0.45)
Free reasoning	\$0.45	\$0.12	**\$0.33** (0.20, 0.48)
Costly reasoning	\$0.41	\$0.12	**\$0.29** (0.15, 0.43)

NOTE: Table 3 converts the coefficients from the regression reported in Table 1 to expected values for easy and difficult problems. These results reflect first differences with all other treatment variables set to zero and the control variables held constant at their mean values (King et al. 2000; Tomz et al. 2003). Confidence intervals in parentheses; boldface indicates that the 95 percent confidence interval does not contain zero, signifying a statistically significant difference between easy and difficult decisions in a particular experimental condition.

problems in the competition condition (just as they do in the control group). Specifically, Table 3 shows that subjects earn an estimated \$0.25 on the easy problems, while they earn only \$0.13 on the difficult problems. This difference in the amounts of money that subjects earn on easy versus difficult problems in the competition condition is statistically significant ($p < 0.05$). As expected, the 100 percent chance of verification conditions effectively close the gap between subjects' decisions on easy and difficult problems. As shown in Table 3, subjects in the 100 percent chance of verification condition earn an estimated \$0.45 on easy problems, and they earn an estimated \$0.38 on difficult problems. In the 100 percent chance of verification plus a \$1 penalty condition, subjects earn an estimated \$0.53 on easy problems and an estimated \$0.42 on difficult problems. In the 100 percent chance of verification plus a $2 penalty condition, subjects earn an estimated $0.55 on easy problems and an estimated $0.53 on difficult problems.[24] As expected, none of these differences between easy and difficult decisions are statistically significant.

Unexpectedly, the $15 penalty for lying failed to close the difficulty gap in subjects' decisions. That is, subjects in the $15 penalty for lying condition earn significantly more money on easy problems than on difficult problems. As shown in Table 3, subjects in the $15 penalty for lying condition earn an estimated $0.53 on the easy problems and only $0.40 on the difficult problems. This difference is statistically significant ($p < 0.05$). That said, although the $15 penalty for lying fails to close the difficulty gap, it does help subjects to achieve significant improvements in their decisions, relative to subjects in the control group. These improvements on the difficult decisions are simply not large enough to make them comparable to the easy decisions, where subjects make virtually perfect decisions.

Our results for decision difficulty in the smaller penalty for lying and slimmer chance of verification conditions are mixed. As predicted (and as shown in Table 3), we find that subjects earn significantly more money on easy problems than on difficult problems in the $5 penalty for lying, 50 percent chance of verification, and 50 percent chance of verification plus a $2 penalty conditions. That said, most of these additional institutions still help subjects to achieve significant improvements in their decisions, relative to subjects in the control group. Again, the improvements are just not large enough on the difficult problems to help subjects make comparable decisions on difficult and easy problems.[25]

Interestingly, when a $1 penalty for lying or a 50 percent chance of verification plus a $1 penalty is imposed on the competing experts, the gap between the amounts of money that subjects earn on easy versus difficult problems closes. As shown in Table 3, subjects in the $1 penalty for lying condition earn an estimated $0.44 on easy problems and an estimated $0.36 on difficult problems. Similarly, subjects in the 50 percent chance of verification plus a $1 penalty condition earn an estimated $0.35 on easy problems and $0.43 on difficult problems. Neither of these differences in the amounts of money that subjects earn on easy versus difficult problems is statistically significant.

Our results for decision difficulty in the reason-giving conditions, however, are perfectly consistent with our predictions. Specifically, we find that, as predicted, subjects earn significantly more money on easy problems than on difficult problems in both the free reasoning and costly reasoning conditions. As shown in Table 3, subjects in the free reasoning condition earn an estimated $0.45 on easy problems, while they earn only $0.12 on difficult problems ($p < 0.05$). Similarly, in the costly reasoning condition, subjects earn an estimated $0.41 on easy problems, while they earn only $0.12 on difficult problems ($p < 0.05$). These results likely stem from the fact that the experts' exchange of reasons on difficult problems is either too hard for most subjects to understand or devolves from an exchange of reasons into an uninformative exchange of accusations that the other expert is lying.

---

[24]These results are estimated first differences of the amounts of money that subjects earn on difficult and easy problems in each experimental condition. Thus, in a few instances, the estimated amount of money earned exceeds the actual amount of money that subjects can earn on each problem in the experiment (i.e., 50 cents).

[25]The one exception to this statement is the 50 percent chance of verification on the difficult problems (see Table 3).

VI. Conclusion

In this article, we derived a pattern of predictions regarding whether and when the testimony of competing experts helps jurors make informed decisions when they are unsophisticated versus sophisticated and when the decisions they must make are difficult versus easy. In a previous article (Boudreau & McCubbins 2008), we found that competition between experts tended not to help subjects, in the aggregate, improve their decisions.[26] However, when additional institutions were imposed on the competing experts or when the experts exchanged reasons with one another, subjects, in the aggregate, achieved large improvements in their decisions. Because our previous article analyzed only the aggregate decisions of all subjects, it left open the questions of (1) whether and when the testimony of competing experts helps unsophisticated subjects make decisions that are comparable to those of sophisticated subjects and (2) whether and when the testimony of competing experts helps subjects make comparable decisions on difficult and easy problems.

Our results here address both these open questions. Consistent with our predictions, our results demonstrate that competition between experts, by itself, does not help unsophisticated subjects to make decisions that are comparable to those of sophisticated subjects; that is, it does not close the sophistication gap among subjects. Indeed, we find that unsophisticated subjects who are exposed to competing experts make significantly worse decisions than do sophisticated subjects. Similarly, we find that competition between experts does not close the difficulty gap. That is, subjects, in the aggregate, make significantly worse decisions on difficult problems than on easy problems. Thus, the testimony of competing experts, by itself, is not particularly helpful to unsophisticated subjects, nor does it help subjects make equally good decisions on difficult and easy problems.

Once competition between experts is coupled with additional institutions (institutions that are common in many legal contexts), unsophisticated subjects achieve such large improvements in their decisions that the gap between their decisions and those of sophisticated subjects closes. Surprisingly, we find that smaller penalties for lying and slimmer chances of verification can be just as effective as large ones at closing the sophistication gap. We also find that, contrary to our expectations, the competing experts' exchange of reasons quite effectively closes the gap between the decisions of sophisticated and unsophisticated subjects. Importantly, the sophistication gap closes because the additional institutions and the exchange of reasons help *both* sophisticated and unsophisticated subjects improve their decisions, with even larger improvements occurring among unsophisticated subjects.

With respect to decision difficulty, we find that some institutions help subjects make equally good decisions on difficult versus easy problems (e.g., a 100 percent chance of verification), while others do not. Thus, the additional institutions we impose on the competing experts do not consistently close the difficulty gap in subjects' decisions. Similarly, when the competing experts exchange reasons with one another, subjects make

[26]The exception to this statement occurs when both competing experts make truthful statements in the competition condition. When both experts make truthful statements, subjects appear to assume that both experts are telling the truth. This assumption largely turns out to be correct and it therefore helps subjects improve their decisions in the competition condition (Boudreau & McCubbins 2008).

significantly worse decisions on difficult problems than on easy problems. These results suggest that the additional institutions and the exchange of reasons are less effective at closing the difficulty gap than they are at closing the sophistication gap. That said, these additional institutions often help subjects achieve significant improvements in their decisions, even on the difficult problems. It is just not always enough of an improvement to help subjects make equally good decisions on difficult and easy problems.

Taken together, these results indicate that legal institutions, as well as the back and forth that occurs between witnesses and lawyers during trials, are beneficial not only because they close the sophistication gap, but also because of the way this closing of the sophistication gap occurs. That is, rather than simply boosting the performance of unsophisticated individuals up to the level of sophisticated individuals, institutions and the experts' exchange of reasons help *both* sophisticated and unsophisticated individuals improve their decisions, with even larger improvements occurring among unsophisticated individuals. That unsophisticated individuals achieve such large improvements when even relatively small penalties for lying or slim chances of verification are in place suggests that citizen juries may not be as problematic as some scholars suggest. Indeed, our results demonstrate that unsophisticated individuals can make decisions that are comparable to those of sophisticated individuals when (1) additional institutions are imposed on the experts, (2) individuals are aware of the effects of these institutions (i.e., they are common knowledge), and (3) individuals have at least a small incentive to make correct decisions.

Our results also suggest that it is harder to close the difficulty gap. Indeed, only under certain conditions (typically when the institutions are very strong) do individuals make comparable decisions on difficult and easy problems. That said, even though the difficulty gap does not completely close under a variety of conditions, we frequently observe large improvements in decision making even on the difficult decisions. These improvements are simply not large enough to make them comparable to easy decisions, where individuals make near perfect decisions when additional institutions are imposed on the experts or when the experts exchange reasons. That individuals can achieve large improvements in their decisions even on the difficult problems again suggests that improving our jury system need not involve the replacement of citizen juries with juries composed of experts. Rather, our results suggest that strengthening our existing institutions can dramatically improve decision making even when jurors are unsophisticated and even when the decisions they must make are difficult.

References

Aronson, Eliot, Timothy D. Wilson, & Marilynn B. Brewer (1998) "Experimentation in Social Psychology," in D. T. Gilbert, S. T. Fiske, & G. Lindzey, eds., *The Handbook of Social Psychology*, 4th ed., vol. 1, pp. 99–142. New York: Oxford Univ. Press.

Boudreau, Cheryl (2006) "Jurors are Competent Cue-Takers: How Institutions Substitute for Legal Sophistication," 2(3) *International J. of Law in Context* 293.

——— (2009) "Closing the Gap: When Do Cues Eliminate Differences Between Sophisticated and Unsophisticated Citizens?" 71(3) *J. of Politics* 964.

Boudreau, Cheryl, & Mathew D. McCubbins (2008) "Nothing But the Truth? Experiments on Adversarial Competition, Expert Testimony, and Decision Making," 5(4) *J. of Empirical Legal Studies* 751.

——— (2009) "Competition in Courtroom: When Does Expert Testimary Improve Jurors' Decisions?" 6(4) *J. of Empirical Legal Studies* 835.

Broyles, Keith (1996) "Taking the Courtroom into the Classroom: A Proposal for Educating the Lay Juror in Complex Litigation Cases," 64 *George Washington Law Rev.* 714.

Cecil, Joe S., Valerie P. Hans, & Elizabeth C. Wiggins (1991) "Citizen Comprehension of Difficult Issues: Lessons from Civil Jury Trials," 40 *American Univ. Law Rev.* 727.

Cecil, Joe S., E. Allan Lind, & Gordon Bermant (1987) *Jury Service in Lengthy Civil Trials*. Washington, DC: Federal Judicial Center.

Chaiken, Shelly (1980) "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion," 39 *J. of Personality & Social Psychology* 752.

Cooper, Joel, Elizabeth A. Bennett, & Holly L. Sukel (1996) "Complex Scientific Testimony: How Do Jurors Make Decisions?" 20(4) *Law & Human Behavior* 379.

Crawford, Vincent, & Joel Sobel (1982) "Strategic Information Transmission," 50 *Econometrica* 1431.

Diamond, Shari Seidman (2003) "Truth, Justice, and the Jury," 26 *Harvard J. of Law & Public Policy* 143.

——— (2007) "How Jurors Deal with Expert Testimony and How Judges Can Help," 16 *J. of Law & Policy* 47.

Dunning, David, & Scott Perretta (2002) "Automaticity and Eyewitness Accuracy: A 10- to 12-Second Rule for Distinguishing Accurate from Inaccurate Positive Identifications," 87(5) *J. of Applied Psychology* 951.

Dunning, David, & Lisa B. Stern (1994) "Distinguishing Accurate from Inaccurate Eyewitness Identifications via Inquiries About Decision Processes," 67 *J. of Personality & Social Psychology* 818.

Elwork, Amiram, Bruce D. Sales, & James J. Alfini (1977) "Juridic Decisions: In Ignorance of the Law or in Light of It?" 1 *Law & Human Behavior* 163.

Faigman, David L., & A. J. Baglioni, Jr. (1988) "Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence," 12(1) *Law & Human Behavior* 1.

Fisher, Michael A. (2000–2001) "Going for the Blue Ribbon: The Legality of Expert Juries in Patent Litigation," 2 *Columbia Science & Technical Law Rev.* 1.

Froeb, Luke M., & Bruce H. Kobayashi (1996) "Naïve, Biased, Yet Bayesian: Can Juries Interpret Selectively Produced Evidence?" 12(1) *J. of Law, Economics, & Organization* 257.

Hans, Valerie P. (2007) "Judges, Juries, and Scientific Evidence," 16 *J. of Law & Policy* 19.

Hass, R. Glen (1981) "Effects of Source Characteristics on Cognitive Responses and Persuasion," in R. E. Petty, T. M. Ostrom, & T. C. Brock, eds., *Cognitive Responses in Persuasion*, pp. 141–72. Hillsdale: Erlbaum.

Hastie, Reid, Steven D. Penrod, & Nancy Pennington (1983) *Inside the Jury*. Cambridge: Harvard University Press.

Hastie, Reid, & W. Kip Viscusi (1998) "What Juries Can't Do Well: The Jury's Performance as a Risk Manager," 40 *Arizona Law Rev.* 901.

Hovland, Carl I., & Walter Weiss (1951) "The Influence of Source Credibility on Communication Effectiveness," 15 *Public Opinion Q.* 635.

Kalven, Harry, Jr., & Hans Zeisel (1966) *The American Jury*. Boston, MA: Little, Brown.

Kim, Hyongsoon (2001) "Adversarialism Defended: *Daubert* and the Judge's Role in Evaluating Expert Evidence," 34 *Columbia J. of Law & Social Problems* 223.

King, Gary, Michael Tomz, & Jason Wittenberg (2000) "Making the Most of Statistical Analyses: Improving Interpretation and Presentation," 44(2) *American J. of Political Science* 347.

Lilly, Graham C. (2001) "The Decline of the American Jury," 72 *Univ. of Colorado Law Rev.* 53.

Lipman, Barton L., & Duane J. Seppi (1995) "Robust Inference in Communication Games with Partial Provability," 66 *J. of Economic Theory* 370.

Lupia, Arthur, & Mathew D. McCubbins (1998) *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.

Lyon, Thomas D. (2000a) "Expert Testimony on the Suggestibility of Children: Does it Fit?" in B. L. Bottoms, M. B. Kovera, & B. D. McAuliff, eds., *Children, Social Science, and the Law*, pp. 378–411, New York: Cambridge Univ. Press.

——— (2000b) "Child Witnesses and the Oath: Empirical Evidence," 73 *Univ. of Southern California Law Rev.* 1017.

——— (2002) "Applying Suggestibility Research to the Real World: The Case of Repeated Questions," 65 *Law & Contemporary Problems* 97.

Milgrom, Paul, & John Roberts (1986) "Relying on the Information of Interested Parties," 17 *Rand J. of Economics* 18.

Mogin, Paul (1998) "Why Judges, Not Juries, Should Set Punitive Damages," 65 *Univ. of Chicago Law Rev.* 179.

Petty, Richard E., & John T. Cacioppo (1984) "The Effects of Involvement on Responses to Argument Quantity and Quality: Central and Peripheral Routes to Persuasion," 46 *J. of Personality & Social Psychology* 69.

Selvin, Molly, & Larry Picus (1987) *The Debate Over Jury Performance: Observations from a Recent Asbestos Case.* Santa Monica, CA: RAND Corporation.

Shuman, Daniel W., Anthony Champagne, & Elizabeth Whitaker (1996) "Assessing the Believability of Expert Witnesses: Science in the Jury Box," 37 *Jurimetrics J.* 23.

Sobel, Joel (1985) "A Theory of Credibility," 52 *Rev. of Economic Studies* 557.

Strier, Franklin (1994) *Reconstructing Justice: An Agenda for Trial Reform.* Westport: Quorum Books.

Sutton, Rita (1990) "A More Rational Approach to Complex Civil Litigation in the Federal Courts: The Special Jury," 1990 *Univ. of Chicago Legal Forum* 575.

Thompson, William (1989) "Are Juries Competent to Evaluate Statistical Evidence?" 52(4) *Law & Contemporary Problems* 9.

Tomz, Michael, Jason Wittenberg, & Gary King (2003) *CLARIFY: Software for Interpreting and Presenting Statistical Results,* v. 2.1. Stanford Univ., Univ. of Wisconsin, and Harvard Univ. Available at ⟨http://gking.harvard.edu/⟩.

Trochim, William M. K. (2001) *The Research Methods Knowledge Base,* 2d ed. Cincinnati, OH: Atomic Dog Publishing.

Vidmar, Neil, & Shari Seidman Diamond (2001) "Juries and Expert Evidence," 66 *Brooklyn Law Rev.* 1121.

Walpin, Gerald (2003) "America's Adversarial and Jury Systems: More Likely to Do Justice," 26 *Harvard J. of Law & Public Policy* 175.