Appalachian STATE UNIVERSITY
BOONE, NORTH CAROLINA

An Explicit Test Of Plea Bargaining In The "Shadow Of The Trial"

By: Shawn D. Bushway, Allison D. Redlich, And **Robert J. Norris**

Abstract

Bargaining in the "shadow of the trial," which hinges on the expectations of trial outcomes, is the primary theory used by noncriminologists to explain variation in the plea discount given to defendants who plead guilty. This study develops a formal mathematical representation of the theory and then presents an empirical test of the theory using an innovative online survey with responses to a hypothetical case from 1,585 prosecutors, defense attorneys, and judges. The key outcomes are the probability that the defendant will be convicted at trial, the sentence for the defendant if convicted, and the best plea that the respondent would accept or offer. Variation in the outcomes is created through experimental variation in the information presented to the respondents. Structural regression models are estimated to fit the formal theoretical models, and the instrumental variables method is used to correct for measurement error in the estimate for probability of conviction. The data support the basic shadow model, with minor modifications, for only prosecutors and defense attorneys. Controlling for the characteristics of the individual actors and their jurisdictions adds explanatory value to the model, although these control variables did not affect the key coefficients from the shadow model.

Shawn D. Bushway, Allison D. Redlich, And **Robert J. Norris** (2014) "An Explicit Test Of Plea Bargaining In The "Shadow Of The Trial". *CRIMINOLOGY* Volume 52 Number 4 723–754 [DOI: 10.1111/1745-9125.12054] Version Of Record Available At www.onlinelibrary.wiley.com

Introduction

Defendants who plead guilty usually receive substantially shorter sentences than observably equivalent people convicted at trial. The implication is that defendants re-ceive a discount for pleading guilty or, alternatively, a penalty for going to trial.

The size of this plea discount varies across individual cases (Bradley-Engen et al., 2012; Bush-way and Redlich, 2012; Smith, 1986; Ulmer and Bradley, 2006; Ulmer, Eisenstein, and Johnson, 2010). Criminological scholars have used perspectives such as focal concerns to argue that this variation can be explained by case-level factors, such as race, criminal history, crime type, and the probability of conviction, and institutional factors, such as workgroup norms and workload. Empirical tests with standard sentencing data sets do a reasonable job of "explaining" the size of the trial penalty with these explanatory fac-tors included in multilevel models (Bradley-Engen et al., 2012; Ulmer and Bradley, 2006; Ulmer, Eisenstein, and Johnson, 2010).

Outside of criminology, "bargaining in the shadow of the trial" (hereafter the "shadow" model) is the primary model used to explain variation in the plea discount (Bibas, 2004; Landes, 1971; Mnookin and Kornhauser, 1979; Nagel and Neef, 1979). In this model, a defendant pleads guilty if the offered sentence is less than or equal to his or her expected value of the trial. For example, if the expected sentence for a conviction at trial is 20 years and the defendant believes his or her probability of conviction at trial is .8, then a plea to a sentence of no more than 16 years (80 percent of 20) represents a rational choice for a risk-neutral defendant. A risk-neutral defendant is not affected by the degree of uncertainty in a choice, and therefore, this individual is indifferent to the choice between accepting the plea bargain and going to trial as long as he or she has the same expected outcome. In contrast, a risk-averse defendant would prefer the certainty of a plea bargain to a choice to proceed to trial even when the expected sentence if the person goes to trial is lower than the expected value of the plea bargain.

Despite the prevalence of this theory in the legal literature, there were no explicit tests of the theory prior to 2012. Legal scholars such as Bibas (2004) also have begun to raise important questions about the face validity of the model. In its simplest incarnation, the theory assumes that actors in the criminal justices system act rationally; yet behavioral economists and psychologists now routinely show that people, including professionals in the criminal justice system, do not act in strictly rational ways (Guthrie, Rachlinski, and Wistrich, 2001; Plous, 1993; Tversky and Kahneman, 1974). No systematic attempt has been made to integrate these ideas into the basic "shadow" model. More generally, Bibas (2004) questioned whether the complex institutional choices involved in bargaining are congruent with the simple shadow model.

Bushway and Redlich (2012) formally introduced the shadow model into criminology and provided the first-known attempt to test the shadow model using individual case record data. To conduct such a test, the researcher needs the following three pieces of information: the probability of conviction at trial, the sentence at trial (which together produce the expected value of the trial), and the value of the plea bargain. Obviously, these data are not available for each person because each defendant either pleads guilty or goes to trial. Bushway and Redlich (2012) overcame this problem by using statistical models to create predicted counterfactuals for those who pled guilty based on the data from people who actually went to trial. At the aggregate level, their initial result provided support for the shadow model—the average plea sentence for the sample was equivalent to the average sentence at trial discounted by the probability of conviction for the sample. A paper by Abrams (2011) with different administrative data found a contrasting result: the average plea value was actually higher than the average expected value of the trial (the sentence at trial discounted by the probability of conviction).

At the individual level, the results from Bushway and Redlich (2012) did not support the shadow model. In many cases, a defendant's actual plea value was not at all similar to the estimate of that defendant's discounted probability of a sentence at trial. Individual estimates of the probability of conviction at trial for those who pled guilty were either uncorrelated with key pieces of evidence known to increase the probability of conviction at trial, such as confessions, or correlated in the opposite direction than the one predicted by the shadow theory. This unexplained noise at the individual level leaves open the possibility that other theories, including criminological theories that focus on interjurisdictional differences in plea outcomes, may be needed to explain individual-level variation in the size of plea discounts.

Bushway and Redlich (2012) cautioned against making too many conclusions based on their study and called for replication. They used 35-year-old data that were considerably less detailed than the norm in modern criminological data sets. In addition, the econometric methods used to create the counterfactuals for those who pled guilty required a nontestable assumption that those who pled guilty are observationally equivalent to those who went to trial after controlling for case characteristics.

In this article, we attempt to advance understanding of plea bargaining by resurrecting and revitalizing a survey approach with a hypothetical case first used by Miller, McDonald, and Cramer (1978). Attorneys and judges were asked to review the case and provide their best estimates for the three main components of the shadow of the trial: the probability of conviction at trial, the outcome at trial, and the plea deal they would advocate for in this situation (all within their own jurisdiction). Because the study uses hypothetical rather than real case data, this approach does not require estimation to test the shadow of the trial model—the key parameters are obtained directly from the respondents. More broadly, there is no need to assume that cases convicted at trial and by guilty plea are observationally equivalent, as in traditional empirical studies using case data. The trade-off is that concerns about external validity become more prominent. Actors may respond to questions in a hypothetical survey differently from how they would act when faced with the constraints of a real case, constraints that include the need to negotiate with other actors. We start the next section by comparing the criminological approach with the shadow model. After explaining the model, we describe the data more fully and then present the study results.

THEORETICAL FRAMEWORK

The criminological approach to the study of sentencing starts with an organizational paradigm that emphasizes that judges, prosecutors, and defense attorneys within a jurisdiction are part of an interdependent workgroup working together toward shared goals such as disposing of cases efficiently and minimizing uncertainty (Eisenstein and Jacob, 1977). Within this framework, criminologists predict that the workgroup will establish norms, or going rates, that are less than the punishment at trial for defendants who are willing to plead guilty. These plea discounts save time and resources while reducing uncertainty. Over time, theorists have extended this basic framework. Dixon (1995) and Engen and Steen (2000) developed the organizational efficiency perspective that emphasizes the efficient disposition of cases. Albonetti (1986, 1987) developed theories of uncertainty avoidance and causal attributions in punishment (see also Albonetti, 1991) to identify the possible use of heuristics and other shortcuts by judges to reduce uncertainty in the sentencing process. Steffensmeier, Ulmer, and Kramer (1998) integrated these perspectives into the focal concerns perspective, which explains how rational actors use heuristics to make sentencing decisions that try to incorporate concerns about 1) offender blameworthiness and the harm caused by the crime, 2) protection of the community, and 3) pragmatic implications such as resource constraints. For example, prosecutors in jurisdictions with higher caseloads might exert greater pressure on defendants to plead guilty than prosecutors in jurisdictions with lower caseloads in response to cost pressures.

In terms of an equation, it is useful to think about the abstract situation where we can observe both the plea sentence Y_i and the trial sentence X_i for the same defendant. This ideal, although not realistic, allows us to focus on the underlying model and not on the identification strategy. In the ideal case, criminological models identify the proportional difference between the trial sentence and the plea sentence as the dependent variable $(Y_i - X_i)/X_i$ and then explain variation with case and defendant characteristics and jurisdictional differences that attempt to approximate or explain workgroup norms, as in equation 1:

$$\begin{aligned}\frac{(Y_i - X_i)}{X_i} = {} & \beta_0 + \beta_1 \, \text{Case Factors}_i \; + \beta_2 \, \text{Defendant Factors}_i \\ & + \beta_3 \, \text{Jurisdictional Factors}_i + \varepsilon_{\mathrm{i}} \end{aligned} \tag{1}$$

Criminological theory makes predictions about which factors to include and the expected sign on the coefficient, although for some factors, such as criminal history, the expected sign may be theoretically ambiguous (Ulmer, Eisenstein, and Johnson, 2010: 19).

No claim is made that knowing the value of any given set of factors will allow the reader to predict, a priori, the size of the plea discount. Although the model parameters could generate predicted values for the plea discount, the theory itself does not predict the value of a given coefficient, only the sign, and it does not model the underlying decision making of any given actor. Rather, the goal of the theory is to identify the factors driving the plea sentence.

The shadow model, in contrast, is a formal theoretical (i.e., mathematical) model that starts from the perspective of the defendant, but ultimately can be used to predict the behavior of other actors interacting with the defendant. The outcome of the model is a clear prediction about the maximum sentence the defendant would willingly accept as part of a plea bargain. We start with the simplest possible version of the shadow model.

The defendant faces a decision between a relatively certain outcome (the sentence after a guilty plea) and the outcome if the defendant decides not to plead guilty, which represents an uncertain mixture of two outcomes: the outcome if the person is convicted at trial and the outcome if the person is acquitted.[4] In probability theory, the average of all

4. In the criminal justice system, the outcome of accepting a plea is not necessarily certain because judges can, and sometimes do, change the terms of the plea deal. The impact of this added

possible outcomes, weighted by the probability of each outcome occurring, is known as the expected value. Therefore, the choice facing the defendant is between the expected value of the plea, which is the value of the plea sentence in the simplest case, and the expected value of the trial (equation 2a). The expected value of the trial is the probability of being convicted at trial × the outcome at trial plus the probability of being acquitted at trial × the outcome at trial if acquitted. Given that the punishment if acquitted is zero, the expected value of the trial reduces to the sentence at trial discounted by the probability of being convicted. For example, if the sentence at trial is 10 years and the probability of conviction is .7, then the expected value of the trial is 7 years. This description is formalized in equation (2b), where Y is the sentence at plea, P is the probability of conviction at trial, and X is the sentence if convicted at trial:

$$\text{Plea value} \approx \text{Expected value of trial} \tag{2a}$$

$$Y_i \approx P_i X_i \tag{2b}$$

$$Y_i = \gamma_1 P_i X_i + \varepsilon_i \text{ (regression format)} \tag{2c}$$

Equivalence is not used in equations (2a) and (2b) because the exact nature of the relationship will be driven by the attitude of the defendant toward risk. A defendant who is willing to accept any offer that has the same expected value regardless of the amount of uncertainty or risk is called "risk neutral." In terms of the preceding example, the risk-neutral defendant is indifferent between a plea sentence of 7 years and a trial with an expected value of 7 years. In contrast, a risk-seeking defendant would actually prefer to take the gamble of the trial. In practical terms, risk seeking creates a situation in which a plea deal would have to be less than the expected value of the trial to convince a defendant to take the plea. In contrast, a risk-averse individual prefers the certainty of the plea deal to the outcome of a trial with an uncertain outcome. This person would therefore be willing to accept a plea deal that might be higher than the expected value of the trial. In the preceding example, a risk-seeking defendant might need a deal of 6 years before he or she pleads guilty, whereas a risk-adverse individual might plead guilty once the deal reaches 8 years or less. We will start with the assumption of risk neutrality, which means the expected value of the trial becomes the upper bound of any plea deal that a defendant may accept. In the example, the defendant would readily accept any offer less than 7 years and is indifferent between a plea deal of 7 years and a trial with an expected value of 7 years. As a result, the shadow model predicts that the coefficient γ_1 will be a maximum of one. Values over one would be consistent with a defendant's risk aversion, and values of less than one would be consistent with a defendant's risk-seeking preferences.

This simple model leads to some straightforward predictions. For example, the average probability of conviction at trial in the largest 75 counties in the United States is 75 percent (Cohen and Kyckelhahn, 2010). The shadow of the trial would therefore predict that the

---

uncertainty can be added to the model, but for the sake of simplicity, we will stay with the simplest model. In the same way, the expected value of going to trial could include a much more complicated set of outcomes, including the possibility that the case is dismissed or additional crimes are investigated. Again, we stay with the simplest case for this initial model.

average plea value would be 75 percent of the trial sentence. This estimate is close to some of the best current estimates of the plea discount in the literature. For example, Ulmer, Eisenstein, and Johnson (2010) estimated the plea sentence to be 69 percent of the trial sentence.

Although the discrepancy between 69 percent and 75 percent might be driven by many factors including risk-seeking behavior on the part of defendants, Alschuler (2002) anticipated that this pattern is a result of prosecutor behavior. In his view, prosecutors trying to avoid the cost of a trial will lower the plea discount below the break-even point to coerce defendants into pleading guilty (see also Alschuler, 1979; Brereton and Casper, 1982; Halberstam, 1982; Nardulli, Eisenstein, and Fleming, 1988). The prosecutor might want to offer a sentence above the upper bound of the plea decision, but the shadow model says that the defendant will not accept that offer. Therefore, the upper bound in the defendant's frame of reference becomes the upper bound for the offer from a rational prosecutor.

The issue is less clear for judges. Judges often are not directly involved except in cases where a plea bargain is not reached, that is, a set of cases that is highly selected. For example, Priest and Klein (1984) argued in the civil context that cases go to trial only when the probability of conviction is near 50 percent. Although Priest and Klein acknowledged that the criminal court context might be different than the civil context (Priest and Klein, 1984: 46), if the basic insight holds, judges might not be as sensitive to the probability of conviction as prosecutors and defense attorneys.[5] In this case, a better model would claim that those who plead simply receive a fixed discount, regardless of the probability of conviction. This discount would reflect the higher costs of going to trial versus accepting a plea. This insight is captured by equation (3):

$$Y_i = \gamma_1 X_i + \varepsilon_i \tag{3}$$

where γ_1 is less than or equal to one. The difference between equation (3) (the fixed discount model) and equation (2c) (the shadow model) is that the discount in equation (3) is constant across all defendants and does not vary with the individual probability of conviction. For example, suppose that all defendants are automatically offered a 30 percent discount for pleading guilty. In this example, the γ_1 will be equal to .7.

Also, it is possible to incorporate the differential costs of trials and pleas for defendants directly into the shadow model. These nonsentence costs include the costs of other sanctions, the personal costs of the court involvement, and the monetary costs of representation. The costs of a plea to an individual i can be modeled as C_i^P, and the costs of going to trial can be modeled as C_i^T. Then a more general form of equation (2b) can be written as equation (4a), where the costs of each decision are added to the decision framework. Equation (4a) reduces to equation (4b), where equations (4b) and (2b) are now analogous, except we have now added a positive cost term on the right-hand side (under the reasonable assumption that the costs of going to trial are higher than the costs of pleading guilty, $C_i = C_i^T - C_i^P$.) In regression terms, this could be tested by including a constant in equation (2c), which is now rendered as equation (2c), if we are willing to assume the cost function in equation (4b) is constant across individuals and $\gamma_1 = 1$. In practical terms, equation (4b) and (4c) imply that, in the face of nonsentence costs associated with going

5. We thank an anonymous reviewer for this important insight.

to trial, a risk-neutral defendant might in fact plead guilty to a sentence Y that is *higher* than what is predicted in the simplest form of the shadow model (equation (2c)) to avoid the nonsentence costs of a trial. Abrams (2011) presented empirical evidence supporting this possibility. A simple test of this idea involves evaluating whether the average plea sentence is higher than the average expected value of the sentence at trial.

$$Y_i + C_i^P \approx (1 - P_i)\, C_i^T + P_i \left(X_i + C_i^T\right) \tag{4a}$$

$$Y_i \approx \left(C_i^T - C_i^P\right) + P_i \,(X_i) \tag{4b}$$

$$Y_i = C + P_i X_i + \varepsilon_i \text{ (regression)} \tag{4c}$$

One key feature of the two versions of the shadow model is that the plea value Y is explained using at most two independent variables: the probability of conviction P and sentence at trial X. This is a striking difference from the criminological model as described in equation (1). This difference in the number of variables represents a difference in goals rather than a claim by the shadow model that the factors identified in criminological theory do not matter. In fact, the shadow model recognizes that P and X are a function of the factors identified in the focal concerns model. For example, workload can affect the probability of conviction, and workgroup norms can affect the average sentence at trial. But, the shadow theory is not trying to identify factors that can explain the size of the plea discount. Instead, tests of the shadow model are attempting to identify whether the plea value is a function of the expected value of the trial, as predicted by the shadow model. The two approaches are best viewed as complementary rather than as competing; the shadow approach is focused on the decision to accept a plea bargain, whereas the criminological approach is interested in identifying the features of the institutional landscape that shape the outcome.

The models are competing, however, when the claim is made that the factors identified by the criminological model have an impact through a mechanism other than their impact on P or X. For example, Ulmer and Bradley (2006: 658) claimed that the observed discount is larger than what could be accounted for even if the probability of conviction was taken into account. Ulmer, Eisenstein, and Johnson (2010) also made clear that, in their view, the institutional factors are separate factors that work independently of the probability of conviction. In this way, criminologists share much in common with legal scholars like Bibas (2004), who has argued that unexplained individual and jurisdictional level variation after controls for the shadow of the trial model are included are evidence of problems with the shadow model.

LITERATURE REVIEW OF EMPIRICAL MODELS OF PLEA BARGAINING

The focal concerns approach to plea bargaining has been tested with standard sentencing models using data on populations of convicted felons. Logged sentence length is the dependent variable, and standard legal controls for criminal history and crime severity are included along with a dummy variable indicating whether the case went to trial (Ulmer and Bradley, 2006; Ulmer, Eisenstein, and Johnson, 2010). This dummy variable is a measure of the "trial penalty" in proportional terms, under the assumption that by controlling for the legal factors, the cases are now observationally equivalent. Researchers have then attempted to explain the size of the trial penalty by adding in additional variables and then measuring the decline in the size of the coefficient. For example, Ulmer, Eisenstein, and Johnson (2010) found that in the federal system, controlling for substantial assistance departure, acceptance of responsibility, and obstruction of justice reduces the size of the "trial penalty" by 64 percent. Ulmer and Bradley (2006) and Ulmer, Eisenstein, and Johnson (2010) also looked for (and found) variation across jurisdictions in the size of the trial penalty, although factors such as caseload have only a modest ability to explain this variation. Finally, Ulmer, Eisenstein, and Johnson (2010) interacted the trial dummy with other individual characteristics like sex, race, and criminal history to test whether substantial variation existed in the size of the penalty across groups. These interactions typically have found only modest differences across groups, and the directions have not always corresponded with the predictions of the model.

A main problem with this approach is that the data sets used do not contain information on cases that go to trial and end in acquittal. As a result, the authors cannot calculate the expected value of the trial but instead compare convicted cases that are resolved through trial with those resolved through plea bargains (Abrams, 2011: 203). In addition, there are important concerns about inherent differences in the cases of defendants who are convicted by plea versus those convicted at trial.

According to Abrams (2011), the best empirical work comparing sentences after trial and after plea bargain is by Smith (1986). Smith (1986) used data collected by Miller, McDonald, and Cramer (1978) on the evidentiary and case characteristics, including sentencing outcomes, for 3,397 felony cases in six sites: New Orleans, Norfolk, Seattle, El Paso, Tucson, and Delaware County, Pennsylvania. Smith focused on 1,533 pled and 387 tried robbery and burglary cases from five sites, excluding El Paso because of missing data. The main advantages of the data set are twofold: It includes cases of individuals acquitted at trial, and it contains a rich set of characteristics, including evidentiary information, not included in sentencing data sets.

Using these data, Smith (1986) generated an estimate of the unconditional expected sentence at trial using observable characteristics to control for differences between cases that are resolved through pleas versus those resolved through trial. Smith found that the expected probability of going to prison after a trial is equal to the probability of incarceration after a plea bargain for cases that are otherwise identical based on observable characteristics. In other words, his estimate validated the shadow model as written in equation (2b) on average. Bushway and Redlich (2012) replicated Smith (1986) using the same data, but then also showed that although equation (2b) holds for the sample at equivalence (as if there are risk-neutral preferences), it does not hold for individuals, as discussed earlier. This is problematic because the shadow of the trial is an individual theory and should be true at the individual level as well as at the aggregate level.

Abrams (2011) followed Smith (1986) by using recent data from Cook County, Illinois, but Abrams found a different result. He found that on average the plea sentence is actually higher than the expected sentence at trial, suggesting that defendants 1) are acting irrationally (they should be going to trial), 2) are risk adverse, or 3) are facing significant nonsentencing costs for going to trial (e.g., equation (4b)). The main problem with all three papers (i.e., Abrams, 2011; Bushway and Redlich, 2012; Smith, 1986) is that they cannot control for unobserved differences across individuals whose cases are resolved through various modes of adjudication.[6]

Although scholars have been making new attempts to secure the kinds of rich data that will allow for more fully specified models of the sentencing process (e.g., Kutateladze, Andiloro, and Johnson, 2014; Kutateladze et al., 2014), detailed data on preconviction processes have been difficult to collect. As an alternative to administrative data, scholars have made use of hypothetical case scenarios to explore plea decision making.

As part of the same ground-breaking project that collected the sentencing data used by Smith (1986) and Bushway and Redlich (2012), Rossman, McDonald, and Cramer (1980) conducted a novel experimental analysis of two hypothetical cases using responses from prosecutors ($n = 136$) and defense attorneys ($n = 140$). They manipulated strength of the evidence (weak or strong), and the defendant's criminal history (long or short), thus creating four conditions to which attorneys were randomly assigned. Attorneys were briefly informed about the case, including that the defendant was charged with armed robbery (or burglary) and was willing to plead guilty for a consideration, and that the law in the hypothetical jurisdiction allowed for a maximum 30-year penalty. Then, attorneys were shown an opened folder with the tops of 43 index cards showing with written titles, such as "defendant's sex" and "trial judges' reputation for leniency." Attorneys were instructed to select and view the cards that they would normally consider and to stop viewing cards when they believed that had enough information to make a decision and recommend a sentence. The attorneys were then asked to provide their best estimates of the probability of conviction, likelihood of engaging in a plea bargain, and recommended sentence.

Rossman, McDonald, and Cramer (1980) did not use these data to test the shadow model directly but instead focused on the attorneys' willingness to plea. Their main finding was that for robberies (but not burglaries), the presence of strong evidence and a lengthy prior record increased both prosecutors' and defense attorneys' willingness to plea. They also found that defense attorneys and prosecutors did not respond to the same questions in the same way, suggesting that there may be a need to understand these actors separately.

There have been two partial replications of the Rossman, McDonald, and Cramer (1980) hypothetical plea simulation design (Kramer, Wolbransky, and Heilbrun, 2007; McAllister and Bregman, 1986).[7] Both studies found that defense attorneys were more likely to advise a plea when cases were stronger, and McAllister and Bregman (1986) found that prosecutors were less willing to plea bargain when cases were strong.

The current study extends the approach developed in these three studies by creating exogenous variation in the probability of conviction. In the latter two replication studies, brief case facts were described and the subjects were told that the there was a "20 percent probability of conviction" or that it was a "strong case." Our goal is to elicit the assessments of the plea value, probability of conviction, and sentence at trial directly from the subjects so we can assess whether the basic shadow of the trial model is valid

6. Abrams (2011) used an instrumental variable approach to control for unobserved heterogeneity, but the instrument is weak and his estimates are imprecise.
7. These replications occur in the field of psychology, not criminology. The use of hypothetical cases to motivate experiments (e.g., mock trials) is a common paradigm in psychology and law. We believe that the field of sentencing in criminology could benefit from the adoption of methods from psychology and law (Baumer, 2013).

for our respondents. We also use different types and combinations of evidence to create exogenous variation in the expected value of the trial so we can eliminate concerns about endogeneity that may bias our estimates, and we implement the survey using the Internet, which allows us to implement in full the "information-capture" approach originally modeled by Rossman, McDonald, and Cramer (1980). To develop the online survey, we partnered with a software design company that designs customized Web-based training courses. To ensure that the tool was easily navigable and as seamless as possible prior to data collection, user testing was conducted, during which several volunteer lawyers worked through the survey in the presence of both the researchers and a programmer. Once a complete version of the tool was developed, pilot surveys were sent to approximately 20 volunteers to ensure the program was functional and intuitive for users on their own computers.

DATA

To recruit our sample, we sought out and received collaboration from several national organizations, including the American Bar Association, the American Judges Association, the National Judicial College, the Association of Prosecuting Attorneys, and the National Legal Aid and Defender Association. We also approached approximately 40 state-level organizations to assist us in recruiting. We were successful in gaining cooperation from the State Bars of South Carolina, Wisconsin, and Montana; the Pennsylvania District Attorneys Association; the New York State Defenders Association; and others.[8] These national and state organizations facilitated contact with their members by forwarding the survey on our behalf, providing e-mail lists for our use, or placing a blurb about the study and link on their website. In addition to national and state organizations, we obtained collaboration from more than 30 district attorney and public defender offices, and we contacted nearly 3,000 individual prosecuting and defense attorneys and judges through public e-mail lists available online. As we will describe, by using these methods, we were successful in recruiting nearly 2,600 attorneys and judges to log onto our survey, which to our knowledge is one of the largest samples to date of its kind. Data collection ran from June 2010 through March 2011.

After viewing and agreeing to an informed consent form, participants were asked to indicate their current role (i.e., defense attorney, prosecutor or state attorney, or judge), and were asked questions pertaining to their demographics (e.g., gender, age, and race), career (e.g., length of experience and number of cases per year), and jurisdiction (e.g., state and size).

Participants were then asked to imagine that a less-experienced junior colleague had come to them for advice about a case in his or her office, and they were presented with the following basic case facts about a hypothetical robbery:

> At 2:30 in the afternoon on a Saturday in a mixed residential commercial area, a male, age 19, was robbed at knifepoint. The robber demanded the man's wallet and watch. The victim handed over his watch and his wallet, which contained a $100 bill,

8. Determining precisely which state agencies chose to assist us was difficult. Specifically, some state agencies explicitly let us know that they were happy to help and would forward along our request. Others did not reply and may or may not have forwarded our request to their members.

> his driver's license, and two credit cards. The watch was originally $1,000 retail. The victim's arm was slashed by the robber without provocation, and he was pushed to the ground. After the robber ran off, the victim called the police, who arrived to the location approximately 5 minutes later. The victim gave the description of his attacker as a "White male in his twenties, wearing black sweat pants and a blue Yankee's cap." The victim also noted that the attacker did not have any distinguishing marks, such as a scar or tattoo. The victim was then taken to the hospital and received five stitches for the laceration he received. After interviewing the victim, the police searched the area. About three blocks away, they found the man's empty wallet in a dumpster. Around the corner from the crime scene, the police found a blue Yankee's cap similar to the one worn by the robber. The police apprehended a suspect 2 hours after the alleged incident several blocks away from the crime scene. The suspect (the defendant in this case) matched the physical description provided by the victim.

Participants were then presented with a set of 31 fact files, each containing a different piece of information pertaining to the hypothetical case (e.g., defendant's race and alibi), which they could choose to view at their discretion along with the basic facts of the case described. (See appendix A in the online supporting information for a complete list of facts provided to prosecutors.[9]) As in the original study (Miller, McDonald, and Cramer, 1978), participants were instructed to select and view the facts that they would normally consider, and to stop viewing the files when they had enough information to make a decision. The average respondent viewed 21 folders (standard deviation [SD] = 9.26). The most viewed fact was the confession statement (94 percent), and the least viewed fact was the backlog of the judge's docket (43 percent). The order of the facts was randomized for each participant to avoid primacy and other order effects. This randomization was successful, as the average position of all checked folders for all participants (both for the entire sample and within roles) did not vary significantly from the expected median position (15.5), and the average position in which a folder was selected did not differ significantly across the 31 randomized folders. The 32nd folder, the basic facts of the case, was always listed last.

The presence and absence of three specific types of evidence—eyewitness identification, confession, and DNA match—and the length of the defendant's criminal history were manipulated (see appendix A in the online supporting information for details). This study conforms to a 2 (eyewitness present or absent) × 2 (confession present or absent) × 2 (DNA match present or absent) × 2 (long or short history) design, creating a total of 16 between-subject conditions. In all conditions, additional identical evidence was provided about the circumstances surrounding the crime (see appendix A in the online supporting information). Participants were randomly assigned to 1 of the 16 conditions. Although the participants were not required to select the manipulated conditions, these facts were considered important by most participants. The confession statement was viewed by 94 percent of participants, the prior record folder was viewed by 92 percent, the eyewitness identification was viewed by 90 percent, and the DNA evidence was viewed by 84 percent.

9. Additional supporting information can be found in the listing for this article in the Wiley Online Library at http://onlinelibrary.wiley.com/doi/10.1111/crim.2014.52.issue-4/issuetoc.

After reading the case and viewing the facts they wished to consider, participants were asked several main questions. First, they were asked to estimate, on a scale from 0 to 100 percent, the probability of conviction if the case were to go to trial. This information was provided by moving a slider along a line that ranged from 0 to 100. This method was used in an attempt to minimize bunching at common points like 50 percent (Spetzler and Von Holstein, 1975; Wallsten and Budescu, 1983). Then, participants were asked to assess the average sentence for a trial conviction, as well as the least and most severe sentences the defendant in the hypothetical case could receive if convicted at trial. Respondents were told that the maximum penalty was 25 years in state prison. Finally, respondents were asked to indicate the least severe sentence that would be acceptable for a plea deal, and what their likely course of action would be in this case (i.e., dismiss case, plea under certain conditions, plea regardless of deal, or go to trial).[10] Respondents were asked four manipulation-check questions to assess whether they could recall relevant information about each of the four conditions that we manipulated. The complete questionnaire is provided in appendix B in the online supporting information.

The survey was sent to approximately 3,000 individual e-mail addresses, but it also was distributed indirectly by organizations that did not wish to share their e-mail lists. As a result, we do not have the precise number of people who received the e-mail invitation. We do know that 2,593 people completed at least the first page of the survey and that 1,664 completed the survey in its entirety, for a completion rate among those who started of 64.2 percent. Survey completion differed across roles, with judges completing the survey at a rate of 54.5 percent, compared with completion rates for defense and prosecuting attorneys of 69.8 percent and 64.6 percent, respectively. We tested whether significant differences were found in age, race, sex, or years of practice between those who completed and those who did not complete the surveys. Among defense attorneys, there were no significant differences between those who did and did not complete the survey on age, race, sex, or years of practice. Among judges, those who completed the survey were significantly more likely to be White and male than those who did not. Finally, among prosecutors, completers were significantly younger than noncompleters.

Of the 1,664 participants who completed the survey, 4 were dropped from the final sample because they spent less than 5 minutes on the survey and 24 were dropped for answering less than half of the manipulation-check questions correctly.[11] Our final sample consisted of 1,585 participants representing all 50 states and the District of Columbia. On

10. These options varied slightly depending on whether they were a judge, defense attorney, or prosecutor. For example, judges were given the options of refusing to accept any plea deal, accepting the deal only if it matched what they perceived to be an appropriate sentence, or accept the plea regardless of the deal. Defense attorneys were provided with the options to refuse to accept any offer and go to trial, accept a plea offer only if it matched what they perceived to be a fair deal, or accept the plea regardless of the offer. Finally, prosecutors were given the options of not offering a plea deal and taking the case to trial, offering a plea only if it matched what they perceived to be a fair deal, pleading out the case in any way they could, or dismissing the case.

11. In addition, several respondents provided answers that were inconsistent with or illogical given their responses to other questions. Thus, some minor data adjustments had to be made. First, participants who responded with a zero to the average sentence question were assigned the midpoint between their given least and most severe possible sentences. This midpoint also was used for respondents who provided an average sentence that was lower than their given least severe possible sentence and those whose average sentence was greater than their most severe sentence. An additional 51 participants were excluded from the final sample because they entered a plea sentence of

Table 1. Descriptive Statistics of Survey Respondents

	Variable	Judges (SE)	Defense (SE)	Prosecutors (SE)	Total (SE)
1	Age*	56.30	46.05	45.03	48.21
		(7.64)	(12.35)	(10.53)	(11.86)
2	Male*	80.65%	64.31%	68.25%	69.09%
3	Non-White	8.87%	9.58%	8.73%	9.21%
4	Years Practice*	11.40	14.99	12.94	13.66
		(8.56)	(11.09)	(9.07)	(10.19)
5	Percent Urban*	29.38%	43.86%	21.51%	35.16%
6	County Size*				
	Very small	10.75%	5.99%	13.49%	8.90%
	Small to midsize	28.23%	25.15%	33.86%	27.95%
	Midsize to large	31.72%	33.41%	30.42%	32.30%
	Very large	29.03%	34.37%	21.69%	30.09%
7	Yearly Caseload*				
	$X \leq 100$ cases	9.41%	40.72%	20.37%	28.52%
	$100 < X \leq 250$ cases	12.37%	27.78%	18.52%	21.96%
	$250 < X \leq 600$ cases	25.81%	25.63%	23.81%	25.24%
	$X > 600$ cases	52.42%	5.87%	37.30%	24.29%
	N	372	835	378	1,585

NOTE. About 1 percent of the cases in "county size" are missing, so the columns do not add to 100 percent.
$^{*}p < .01$.

average, they spent 22.2 minutes on the survey and got 3.7 out of 4 manipulation check questions correct. The final sample is described in table 1.

METHOD

The goal of this article is to assess whether the variation in plea sentence, probability of conviction, and the expected trial outcome that we induced through our experiment is consistent with any of the simple formal models that we outlined in the Theoretical Framework section. As such, this exercise is different from the typical empirical test in criminology. We are not trying to explain all of the variation in the plea sentence identified by the respondent, nor are we trying to get the best, unbiased estimate of the causal relationship between the trial or the expected trial outcome and the plea sentence. Instead, we estimate a structural model that represents the equations from the Theoretical Framework section and compare how well they can explain variation in the observed plea sentence reported by the respondents. A finding that the model fits the data well does not prove that the model is true, but it does suggest that the theory has some merit empirically.

We start by demonstrating that our experimental manipulation of evidence quality and the defendant's prior record created substantial and meaningful variation in our three outcome variables (plea sentence, probability of conviction, and sentence at trial). Next, we calculate the expected sentence of the trial for the total sample and each set of respondents—prosecutors, judges, and defense attorneys. Furthermore, we compare this

zero or still had a zero or other illogical average sentence value after making the adjustments described earlier, thus making our final sample 1,585. Dropped participants were statistically equally likely to be from each of the three roles.

average with the average plea sentence reported by each respondent group. This simple comparison is a direct analog to the tests reported by Smith (1987) and Abrams (2011), both of whom used empirical models to generate estimates of the average plea sentence and the expected trial outcome. This approach also represents a test of the cost model outlined in equation (4c). Note the absence of any coefficients in equation (4c)—the key value is the constant, which is just the average difference between the plea sentence and the expected trial. A positive value is consistent with the cost model discussed previously.

In the Theoretical Framework section, we showed that, under the simplest assumptions, plea sentences designed to overcome the larger costs of trial will be higher than the expected value of the trial by a fixed amount (Abrams, 2011). A finding of near equality between the plea sentence and the expected value of the trial (Smith, 1986), in contrast, would suggest that the simplest version of the bargaining in the shadow of the model has promise. Bushway and Redlich (2012) argued that the shadow theory should be further evaluated by considering its predictive capacity across individual-level decision makers.[12]

Therefore, we will test the simple bargaining in the shadow model at the individual level, as described in equations (2a) and (2b). A simple way to start would be to run the regression specified in equation (2c) and to test whether the coefficient on PX, γ_1, is equal to one. But equation (2c) is a highly restrictive equation. The coefficient on the probability of conviction P and the trial sentence X are forced to be the same. As discussed, it is possible that the probability of conviction does not impact the decision (equation (3)). In addition, the relationship between PX and Y is forced to be linear. Fortunately, there is an easy way to estimate a more general version of equation (2c) that loosens both of those restrictions. If we take the log transformation of both sides of equation (2c), then equation (2c) becomes equation (5a). This log transformation is not the typical log transformation done in the sentencing literature, where sentencing outcomes are logged to deal with skew. Here, we are taking the log to transform the restrictive equation (2c) into an identical but more general form. However, to the extent to which there are concerns that the data from the experiment are skewed, the logging will have the additional benefit of transforming the data into a form that is more consistent with the assumptions of the ordinary least-squares (OLS) regression model.

Equations (2c) and (5a) are mathematically equivalent. In equation (5a), the log of the plea value becomes the new dependent variable, and the constant will be the log of the coefficient in equation (2c). P and X are now separated and are allowed to have their own coefficients. In the simplest case of the shadow model with risk-neutral defendants, we expect to find the constant would be zero, and the constants on the logged values of P and X, which are the exponents on P and X in equation (2c) (as shown again here), will be equal to one. These exponents are the elasticities of Y with P and Y with X, respectively:

$$Y_i = \gamma_1 P_i X_i + \varepsilon_i \tag{2c}$$

12. In response to reviewer comments, we considered adding control variables for context like the state of practice for each respondent. These variables are highly relevant because in terms of both statute and standard practice, there is wide variation in the average sentence for robbery crimes across states. However, this variation should be captured by the trial sentence X, which also will vary across states. Therefore, we chose to stay with the simplest model without control variables for most of our models, but we will test whether jurisdiction and decision-maker characteristics add anything to the most general model.

$$\ln Y_i = \ln \gamma_1 + \ln P_i + \ln X_i + \mu_i$$

(5a)

$$Y_i = \gamma_1 P_i^{\delta_P} X_i^{\delta_X} + \varepsilon_i \tag{5b}$$

$$\ln Y_i = \ln \gamma_1 + \delta_P \ln P_i + \delta_X \ln X_i + \mu_i \tag{5c}$$

The elasticities need not be restricted to be one. In equation (5b), we rewrite equation (2c) to include the possibility that the exponents on P and X, δ_P and δ_X, are different from one another, and are different from one.[13] The relationship between Y and the pair of variables P and X is already nonlinear because P and X are multiplied by each other. However, the elasticity provides for additional flexibility. The simplest way to see this is to imagine a situation in which $\gamma_1 = 1$ and $\delta_P = 0$. If δ_X is unit elastic, or equal to one, then the plea value Y is equal to the trial sentence X, a straight line on the diagonal.[14] However, if $\delta_X = .8$, then the actor is inelastic, and a 10 percent increase in trial sentence, for example, is met with only an 8 percent increase in plea sentence. The implication here is that the plea discount in real terms is bigger when the trial sentences are large. If $\delta_X = 1.2$, then the actor is elastic, and a 10 percent increase in trial sentence is met with a 12 percent increase in plea sentence. Here, there is actually a plea penalty, and that penalty gets larger as the trial sentence gets bigger. The former situation is more probable than the latter, but nothing in the equation will force the elasticity to take on certain values.

After estimating equation (5c), the new test for the shadow of the trial model will be whether the constant $\ln \gamma_1$ is different from zero because the natural log of one is zero. The two independent variables will be the natural log of the two component parts of the expected value of the trial: the probability of conviction P and the expected sentence at trial X. The coefficients on these two terms, δ_X and δ_P, will then allow for direct tests of whether, for each set of actors, the probability of conviction and the trial at conviction are given equal weight in the setting of a plea value. In particular, we are interested in testing whether the coefficient on P, δ_P, is significantly different from zero; if it is zero, then the model reduces to a more general form of equation (3), the constant discount model.

An added benefit of the log transformation is that it facilitates a solution to the issue of clumping in the estimates of P, the probability of conviction, which occurs in our data despite the use of the sliding scale. One interpretation of the clumping is that the clumping reflects reality. If so, the clumping will affect the error terms but not the coefficient estimates. A second, more problematic interpretation is that the clumping is the product of measurement error, where the observed values of P are a mixture of the true value of P, P^*, plus measurement error:

$$P_i = P_i^* + \varepsilon_i \tag{6}$$

13. δ_P = percentΔ Y/percentΔ P and δ_X = percentΔ Y/percentΔ X. Because these elasticities are allowed to vary, we can capture nonlinearities in the relationship between "X and Y" and "P and Y" without the use of polynomial terms.

14. An elasticity of one does not always imply linearity because it refers to the relative movement in percent change rather than to the relative movement in absolute changes. This example is provided for simplicity, not because it demonstrates a general rule.

For example, suppose prosecutor Sam's true probability of conviction is .71, but Sam chooses to move the line to the three-quarter mark (.75) for convenience. Random measurement error of this type on the independent variable will bias the coefficient on P, δ_P, toward zero, and it will produce an invalid inference about the nature of the relationship between P and Y.

The instrumental variable or two stage least-squares method can create consistent estimates of the desired independent variable even in the presence of measurement error (Bushway and Apel, 2010; Tita, Petras, and Greenbaum, 2006). In the current context, an instrumental variable for the probability of conviction is anything that affects the true probability of conviction P^* but is uncorrelated with the measurement error in the probability of conviction and is only indirectly correlated with the dependent variable (the plea value) through the independent variable, the probability of conviction.

Mechanically, the instrumental variable model works in two stages, as shown in equation 7. In the first stage (equation (7a)), the independent variable is predicted by the instrumental variable(s), shown as Z, plus the exogenous independent variables ($\ln X$) from the main equation of interest (equation (5c)). In the experiment, we varied three evidence conditions—DNA match, confession, and positive eyewitness ID—that should affect probability of conviction but not average sentence. We will use these three main effects as instrumental variables.[15] Although we only need one instrument, we exploit the fact that we have three main effects to maximize power.

In the second stage (equation (7b)), the dependent variable is regressed on the predicted value of the dependent variable ($\widehat{\ln P}$) from the first stage. The use of the predicted value from the first stage is the only difference between equations (7b) and (5c). Because the coefficient in the second stage is identified only by the exogenous variation in the probability of conviction created by the instrumental variables, the result will be less susceptible to measurement bias. This benefit comes at the cost of a decrease in precision. Depending on the power of the first stage, the standard errors in the second stage can be up to an order of magnitude larger than the standard errors in the standard OLS. If the clumping reflects the actual preferences of individuals and is not caused by measurement error, then the instrumental variable model will not lead to any change in the estimates of the underlying coefficients in the second stage.

$$\ln P_i = \alpha + \beta_1 Z_i + \beta_2 \ln X_i + \varepsilon_i \qquad \text{First stage (7a)}$$

$$\ln Y_i = \ln \gamma_1 + \delta_P \widehat{\ln P}_i + \delta_X \ln X_i + \mu_i \qquad \text{Second stage (7b)}$$

In our final model, we add standard criminological variables such as jurisdiction size, caseload, and practitioner characteristics into the model (equation (5c)) (Ulmer and Bradley, 2006). We add dummy variables for state, county size, and caseload level as well as practitioner gender, race, age, age squared, and experience. We are looking for two things: first, whether these variables affect the coefficients estimated from equation 5, and

15. Evidence should drive probability of conviction and not average sentence. We omitted the prior history manipulation as an independent variable because it should affect both the probability of conviction and the average sentence (Eisenberg and Hans, 2009). Because the average sentence is also included in the model, any relationship that the evidence conditions have with this variable will not affect the estimates of interest.

Table 2. Plea Bargaining Outcomes, By Actor

	Variable	Judges (SE)	Defense (SE)	Prosecutors (SE)	Total (SE)
1	Probability of conviction	64.03%	63.69%	66.26%	64.38%
		(27.23)	(24.44)	(25.36)	(25.34)
2	Sentence at trial (years)	9.99	10.52	9.73	10.21
		(5.69)	(6.18)	(6.19)	(6.08)
3	Expected value of a trial (1x2)	6.58	6.87	6.65	6.75
		(4.99)	(5.22)	(5.24)	(5.17)
4	Acceptable plea (years)	6.26	6.07	6.05	6.11
		(4.03)	(4.54)	(4.60)	(4.44)
	N	372	835	378	1,585

second, whether these variables have an independent ability to affect the plea variable more than the variables in the shadow of the trial model. Although this is not the main thrust of this article, this last model will allow for a preliminary test of the ideas from Bibas (2004) and Ulmer and Bradley (2006) that factors such as these should influence plea sentences over and above their influence on the probability of conviction and going rates of trial sentences.

RESULTS

As shown in table 1, the sample is roughly half defense attorneys, with the remainder split equally between prosecutors and judges. The participants were nearly 70 percent male and more than 90 percent White. The respondents were, on average, 48 years of age, with an average of 13.66 years of experience in their current role. Judges were significantly older and more likely to be male than both prosecutors and defense attorneys. Defense attorneys had significantly more years of experience and were more likely to work in urban districts than the other two legal actor types. Across roles, respondents were distributed evenly across very small to very large counties, with prosecutors more likely to report being from very small and small counties, and defense attorneys the most likely to report being from large and very large counties. Annual caseloads, perhaps not surprisingly, varied considerably across actors. Judges reported the highest caseloads, with more than half reporting that they had more than 600 cases. In contrast, only 6 percent of the defense attorneys reported that level of work. Prosecutors fell in the middle, with the median prosecutor reporting between 250 and 600 cases. The median number of cases was 150 for defense lawyers, 400 for prosecutors, and 750 for judges.

Rows 1, 2, and 4 of table 2 provide the average responses for the sample to the three key questions in the survey regarding the case. The probability of conviction ranged from a low of 63.7 percent for the defense lawyers to a high of 66.3 percent for the prosecutors, and the sentence at trial ranged from 9.73 years for the prosecutors to 10.52 years for judges. Finally, the acceptable plea sentence ranged from 6.05 years for the prosecutors and from 6.26 years for the judges. Across all measures, none of the differences across roles were statistically significant, suggesting (but not proving) a similar underlying process of decision making.

Table 3. Plea Bargain Outcomes Across Experimental Situations, Total Sample

	Condition*	*N* (1,585)	Probability of Conviction	Sentence at Trial	Expected Value of Trial	Acceptable Plea
1	Confess, ID, DNA, history	102	82.08% (17.98)	12.56 (6.39)	10.52 (6.16)	8.68 (4.47)
2	Confess, ID, DNA	95	75.86% (20.80)	8.76 (5.36)	6.92 (4.87)	5.71 (3.62)
3	Confess, DNA, history	101	68.82% (23.60)	11.71 (6.59)	8.01 (5.42)	7.96 (5.23)
4	Confess, ID, history	102	75.48% (19.85)	11.91 (6.46)	8.99 (5.66)	7.38 (4.43)
5	ID, DNA, history	100	77.49% (16.60)	12.20 (6.29)	9.65 (5.88)	7.68 (4.36)
6	ID, DNA	127†	72.31% (20.53)	8.86 (5.74)	6.64 (5.04)	5.90 (4.64)
7	DNA, history	96	61.21% (24.70)	11.37 (6.23)	7.22 (5.34)	6.34 (4.43)
8	ID, history	98	67.80% (21.53)	11.50 (5.46)	7.90 (4.98)	6.93 (4.00)
9	Confess, DNA	94	68.48% (22.98)	8.87 (5.70)	6.15 (4.73)	5.42 (4.29)
10	Confess	104	56.90% (24.30)	8.28 (5.43)	4.77 (3.63)	4.52 (4.48)
11	Confess, history	95	57.82% (23.61)	11.83 (5.61)	6.96 (4.31)	6.88 (4.18)
12	Confess, ID	94	71.56% (20.25)	8.01 (5.31)	5.95 (4.47)	5.32 (4.18)
13	DNA	88	54.99% (24.62)	9.00 (5.26)	4.97 (3.93)	4.85 (3.59)
14	ID	86	58.90% (21.49)	7.74 (4.72)	4.60 (3.29)	4.88 (4.01)
15	History	99	47.19% (24.04)	11.79 (6.48)	5.70 (4.47)	5.84 (3.96)
16	None	104	30.91% (23.61)	8.73 (6.13)	2.65 (3.06)	3.26 (3.21)

*This column lists which of the four experimental conditions is toggled on in each case. The exact wording of each condition is given in appendix A in the online supporting information. History refers to the condition in which the defendant has a long criminal history.
†Condition 14 was accidentally written as condition 6 in the initial code. This was caught in the first 2 weeks during which the survey was live and fixed. This explains why condition 6 has more cases than the other conditions.

By multiplying the probability of conviction (row 1) by the sentence at trial (row 2) for each individual, we estimated the expected value of the trial to be between 6.58 (for judges) and 6.87 years (for defense lawyers), with a standard deviation of approximately 5 years. The shadow model predicts that we would find plea values equal to the expected value of the trial, and the estimates are close, with pleas that are a 40 percent discount off the average sentence at trial and 90 percent of the expected value of trial. Like Smith (1986) and Bushway and Redlich (2012), we find that the shadow model has at least face validity in the aggregate sample.

Table 3 provides the same four variables for the entire sample across the 16 experimental conditions. The experiment did an excellent job of inducing variation in the key outcome variables. For example, compare the values across row 1, which has all three evidence conditions and the longer criminal history, with the values from row 16, which has none of the three evidence conditions and the shorter criminal history: In row 1, the average probability of conviction was 82 percent, the average sentence at trial was 12.56 years, and the acceptable plea was 8.68 years, whereas in row 16, the values plummet to 31 percent, 8.73 years, and 3.26 years, respectively.

Evidence is particularly effective at moving probability of conviction and not very important for the sentence at trial. For example, compare rows 10, 13, and 14 (cases with just a single piece of evidence) with row 16 (a case with no evidence and the same shorter criminal history condition): In each case, the probability of conviction rises substantially (from 30.9 percent to either 56.9 percent, 55.0 percent, or 58.9 percent), whereas the sentence at trial is basically the same—8.7 in row 16, 8.3 in row 10, 9.0 in row 13, and 7.7 in row 14. In terms of main effects for the total sample, eyewitness identification (ID) had the largest impact on the probability of conviction at 17 percentage points, followed by DNA match at 12 percentage points, confession at 11 percentage points, and criminal history at 6 percentage points. The finding that criminal history has a significant, albeit small, impact on the probability of conviction replicates the important finding from Rossman, McDonald, and Cramer (1980). For more on the experimental variation of each factor, see Redlich, Bushway, and Norris (2013).

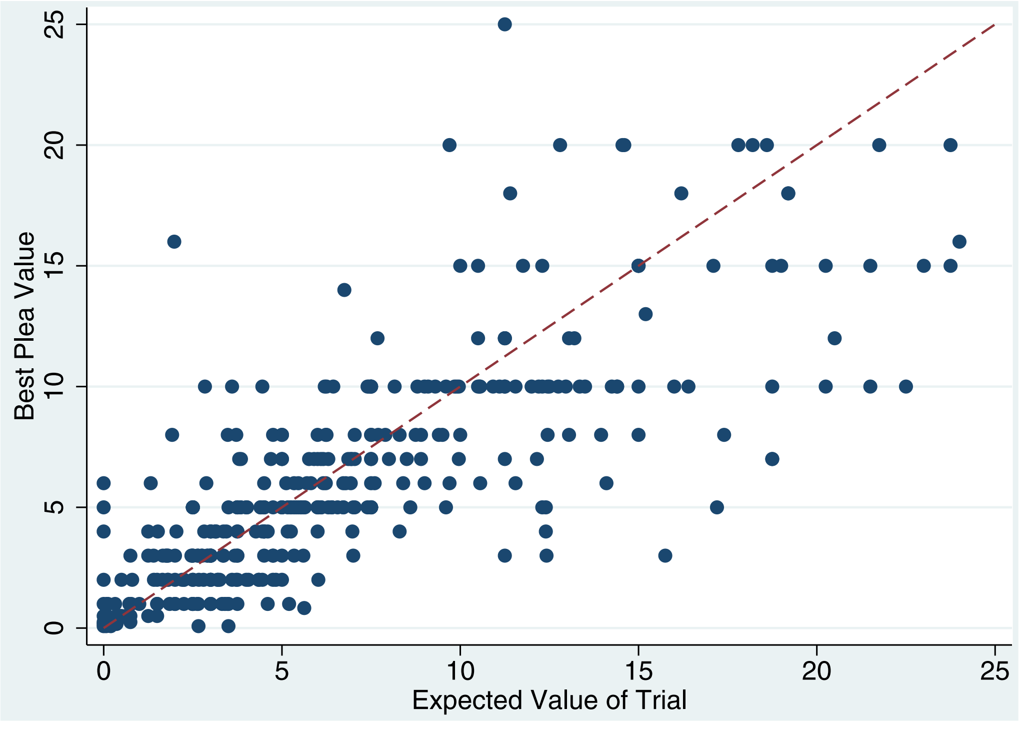
In general, the sentence at trial seems to respond only to changes in the information about the criminal history. For example, compare the expected average trial sentence in row 1 (12.56 years) and row 2 (8.76 years): In both situations, the case had all of the evidence, but in row 1, the case had the longer criminal history, and in row 2, the case had the shorter criminal history. In contrast, the average trial sentences in rows 3, 4, and 5—scenarios with the longer criminal history and one less piece of evidence—all have average trial sentences that are similar to row 1 (11.71, 11.91, and 12.20, respectively), suggesting, as expected, that evidence does not induce substantial variation in the average trial sentence.

Table 3 demonstrates that the basic shadow of the trial theory seems to hold across the experimental manipulations, with the acceptable plea value varying across each condition in a manner that is, at least, strongly correlated with, if not identical to, the expected value of the trial. For example, the case with the highest expected value of trial (row 1) has the highest acceptable plea, and the case with the lowest expected value of trial (row 16) has the lowest acceptable plea. In addition, in all but three conditions (rows 14–16), the acceptable plea is below the expected value of the trial.

These descriptive statistics tell an interesting story. We experimentally induced substantial and statistically significant variation in the expected value of the trial through variation in the evidence and criminal history. The responses do not seem to be consistent with the simple cost model outlined in equation (4c), which predicts that the plea value would be greater than the expected value of the trial (Abrams, 2011). Instead, the responses seem to be consistent with the shadow model, with acceptable plea values that are similar to the expected value of the trial on average.

Although the theory has implications for what will be observed at this aggregate level, the shadow model can be evaluated most rigorously by considering its predictive capacity for individual-level decision makers: The acceptable plea value for a given respondent should equal the expected value of the trial for that individual. As a preliminary test of the insight behind the shadow model, we tested the correlation between the estimated expected value of the trial for each respondent and the acceptable plea for each respondent. The resulting correlation coefficients are .72 for the total sample, .66 for the judges, .71 for the defense attorneys, and .78 for the prosecutors. At least three

Figure 1. Scatter Plot of Plea Value and Expected Value of Trial for Prosecutors

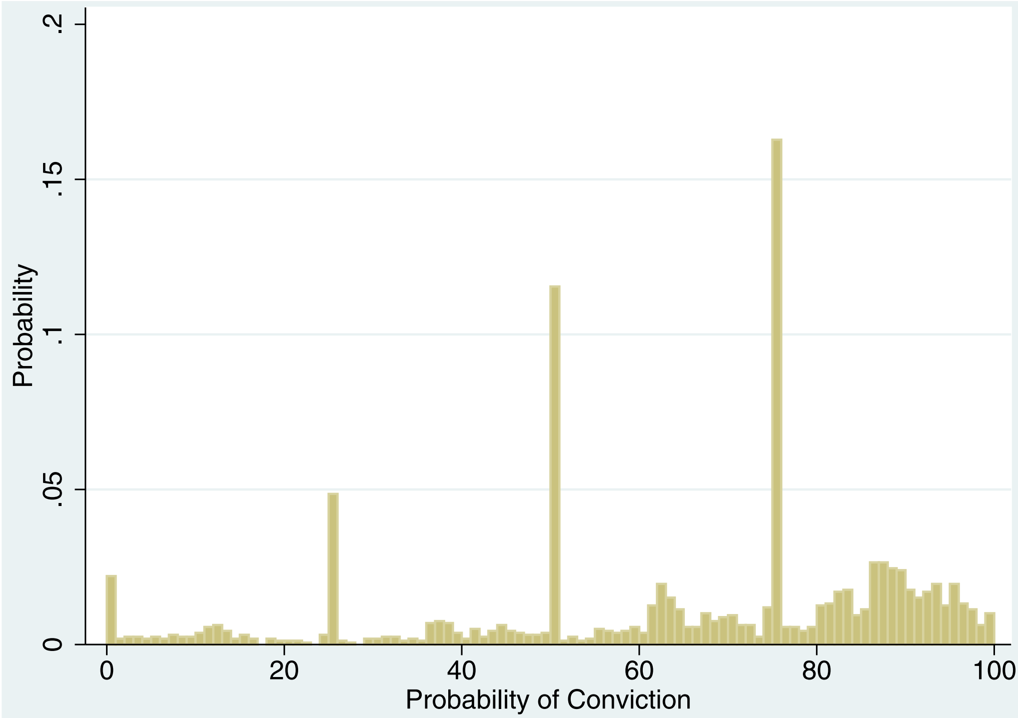


messages can be found here. First, the expected value of the trial is highly correlated with the plea sentence and can explain between 44 percent and 61 percent of the variation in plea values. Second, interesting variation might occur across actors, despite the evidence from table 2 that the main variables are statistically similar across actors. Third, the correlation is not perfect, and substantial variation is found in the plea value not captured by the expected value of the trial.

To make this latter point more concrete, figure 1 presents a scatterplot of the relationship between the expected value of the trial and the best plea deal offered by prosecutors. The diagonal line represents the slope expected by the shadow of the trial model. Although the relationship is clearly positive and strong, considerable variation is found around the line. Some prosecutors offer deals that are below the expected value of the trial (below the diagonal line), whereas other prosecutors offer deals that are above the expected value of the trial (above the diagonal line).

This noise may be partially an artifact of clumping in the probability of conviction variable. Although we used a method designed to minimize clumping, the probability of conviction has substantial bunching at 25 percent, 50 percent, and 75 percent (see figure 2). Although it is possible that that these numbers reflect the true probability of conviction for the respondents, it also is possible that the true probability of conviction induced by

” **Figure 2. Histogram of Probability of Conviction**



the experiment is smoother than this observed distribution. In other words, our measure of probability of conviction includes measurement error, which could bias our regression estimates toward zero.

This variation could be random noise or measurement error induced by the clumping of the responses in the probability of conviction, but it also could reflect other causal mechanisms. The standard deviations reported in table 3 display substantial variation in the responses within each experimental condition, suggesting that other factors are at play besides evidence and criminal history. In addition, figure 2 shows that the relationship is likely to be nonlinear, with a slope that seems to flatten as the expected value of the trial increases.

The next step is to capture this relationship in an econometric form. Table 4 reports the results from our estimate of equation (5c), with a basic ordinary least-squares model with corrections for heteroskedasticity and possible clustering by state. The dependent variable is the logged plea value. The R^2 values for these models are impressive, ranging from .49 for defense lawyers to .62 for judges.

The first test is whether the coefficient on the expected value of the trial γ_1 is equal to one, which, because of the log transformation, is the test for whether the constant $\ln(\gamma_1)$ is different than zero. In the total sample, we can reject the null hypothesis. However, despite the nearly identical descriptive statistics in table 2, the estimates for the theoretical model differ across actors. In the case of both the defense lawyers and prosecutors, we

Table 4. OLS Model Estimates of General Shadow of Trial Model

Variable	Judges (SE)	Defense (SE)	Prosecutors (SE)	Total (SE)
Ln (γ_1) (constant)	$-.35^{**}$	$-.12$	$-.13$	$-.22^{**}$
	(.12)	(.11)	(.10)	(.07)
δ_P, Ln (probability of conviction at trial)	.04	$.50^{**}$	$.48^{**}$	$.31^{**}$
	(.02)	(.10)	(.08)	(.06)
δ_X, Ln (sentence at trial)	$.92^{**}$	$.85^{**}$	$.89^{**}$	$.88^{**}$
	(.05)	(.05)	(.04)	(.03)
F test for $\delta_P = 1$	1559.24^{**}	25.27^{**}	37.88^{**}	140.73^{**}
F test for $\delta_X = 1$	2.13	9.58^{**}	7.11^{*}	14.96^{**}
R^2	.62	.49	.60	.50
N	372	835	378	1,585

$^{*}p < .05$; $^{**}p < .01$.

do not reject the null hypothesis that the coefficient γ_1 is one, whereas we reject the hypothesis for judges. As shown in table 4, this is not just a question of power. The total sample has a value of γ_1 equal to .8 ($\exp^{-.22} = .80$), whereas the same value is .89 for defense lawyers, .88 for prosecutors, and .71 for judges. The shadow model fit is worse for judges, who are offering higher discounts relative to defense lawyers and prosecutors.

The next test is for the coefficients on the logged probability of conviction δ_P and logged trial sentence length δ_X, which are the elasticities between the plea values and these explanatory variables. We provide results for tests of whether these are significantly different from zero and whether they are significantly different from one, which is the prediction of the most basic shadow model. The results for the total sample are equivocal. The elasticity between the probability of conviction and plea sentence δ_P is significantly different from zero, but it is nowhere near one ($\delta_P = .31$). In contrast, the elasticity between the trial sentence and plea value δ_X is .88, which is close to the predicted value of one (although it is significantly different from one). The relationship in both cases is concave, suggesting that the plea value does not increase at the same rate as the probability of conviction and the trial length. The probability of conviction has a much weaker impact on the plea sentence than the trial sentence length. Overall, these results suggest that the respondents behave in a way that is more consistent with the fixed discount model (equation 3) than the shadow model.

The story becomes a bit clearer when the models are run separately by role, also shown in table 4. Prosecutors and defense attorneys look like the full sample, with elasticities for probability of conviction ($\delta_P = .50$ and .48, respectively) that are significantly different from zero but far from one, and the elasticities for sentence length are close to one, although significantly different from one ($\delta_X = .85$ and .89). However, the story for judges is different. For judges, the elasticity between the probability of conviction and plea sentence δ_P is not significantly different from zero, indicating that the sentence chosen by the judge does not depend on the probability of conviction. Moreover, the elasticity between the sentence length at trial and plea value δ_X is .92 and is not significantly different from one. The fact that the coefficient on probability of conviction is close to zero suggests that the best model for the judges is the fixed discount model (equation (3)).

It is possible, however, that measurement error in the probability of conviction (figure 2) has biased the coefficient on the probability of conviction toward zero. This

Table 5. Instrumental Variable Estimates of the General Shadow of Trial Model

First Stage DV = ln*P*	Judges (SE)	Defense (SE)	Prosecutors (SE)	Total (SE)
DNA match	.33**	.20**	.28**	.24**
	(.12)	(.04)	(.08)	(.04)
ID	.29**	.41**	.37**	.38**
	(.11)	(.04)	(.08)	(.04)
Confess	.31**	.24**	.21*	.25**
	(.11)	(.04)	(.08)	(.04)
Ln(sentence at trial)	.08	.08*	.20**	.11**
	(.09)	(.03)	(.07)	(.03)
Constant	−1.34**	−1.17**	−1.43**	−1.27**
	(.19)	(.08)	(.19)	(.08)
First stage F	6.82**	39.28**	8.92**	41.83
R^2	.07	.15	.12	.11
N	372	835	378	1,585

Second Stage DV = ln*Y*	Judges (SE)	Defense (SE)	Prosecutors (SE)	Total (SE)
Ln (γ_1) (constant)	−.22	.11	.59	.16
	(.13)	(.11)	(.37)	(.10)
δ_P, Ln (probability of conviction at trial)	.18*	.79**	1.19**	.75**
	(.07)	(.10)	(.27)	(.11)
δ_X, Ln (sentence at trial)	.91**	.83**	.74**	.83**
	(.05)	(.04)	(.11)	(.03)
χ^2 test for $\delta_P = 1$	141.48**	3.91*	.51	5.01*
χ^2 test for $\delta_X = 1$	3.09†	17.16**	5.73*	28.70**
R^2	.59	.45	.27	.37

†$p < .10$; *$p < .05$; **$p < .01$.

bias would lead to an incorrect inference about the importance of the probability of conviction. The top of table 5 presents the first stage of the instrumental variable model described in the Methods section (equation (7a)). The results are interesting in their own right. For the total sample, the existence of a positive eyewitness ID, with all else constant, increases the probability of conviction by a whopping 38 percent, whereas DNA evidence and a confession increase the probability of a conviction by 24 and 25 percent, respectively. As in the OLS findings, the results begin to diverge when the models are estimated by role. In this case, the main difference is that a positive ID is essentially equivalent to the other two evidence measures for judges, whereas it is even stronger, relative to DNA and confession, for defense attorneys and prosecutors.[16]

The results from the second stage of the instrumental variable analysis are reported in the bottom of table 5. They are different from the results in the OLS model (table 4),

16. As anticipated from table 3, the experimental conditions did an excellent job of creating variation in the probability of conviction evidence. The F test on the first stage is a good indicator of the power of these variables to predict the probability of conviction. The F for the full sample is 41.83, which is both large and significant. The F statistic for the defense lawyers (which has the biggest N) is 39.28, and the F for the judges and prosecutors are 6.82 and 8.92, respectively.

which suggest measurement error in the probability of conviction. Specifically, the constants for the total sample and the judges, γ_1, are no longer significantly different from zero, which means the prediction that the constant from the shadow model is equal to one is not rejected. This result is not just a question of power—in each case, the value of the variables increases.

In addition, the coefficient on the probability of conviction increases in each case, suggesting that the results from table 4 are biased downward. For the total sample, the estimate of δ_P went from .31 to .75, whereas the estimate for δ_X stayed basically the same (.83 versus .88). For the total sample, probability of conviction and the sentence at trial have nearly identical impact on the plea sentence, a key prediction of the shadow model.

This result for the total sample is mirrored in the results for defense attorneys and prosecutors. The elasticity for defense lawyers is now nearly the same for both the probability of conviction ($\delta_P = .79$) and the sentence at trial ($\delta_X = .83$), and the elasticity is greater than one (but not significantly so) for the prosecutors. Taken together with the fact that the exponentiated constant, $\gamma_{1,}$ is not statistically different than one, the results demonstrate that the shadow of the trial model seems to fit the defense lawyers particularly well, with some nonlinearity as the expected value of the trial gets large.

The story is slightly different for prosecutors. The large value for γ_1, together with the smaller elasticity on trial length, suggests that plea discounts from the prosecutors will be smaller than the expectations of the defense lawyers, particularly when the trial sentence is low. This finding is consistent with the idea that prosecutors are trying to avoid high fixed costs of trials by offering plea deals.

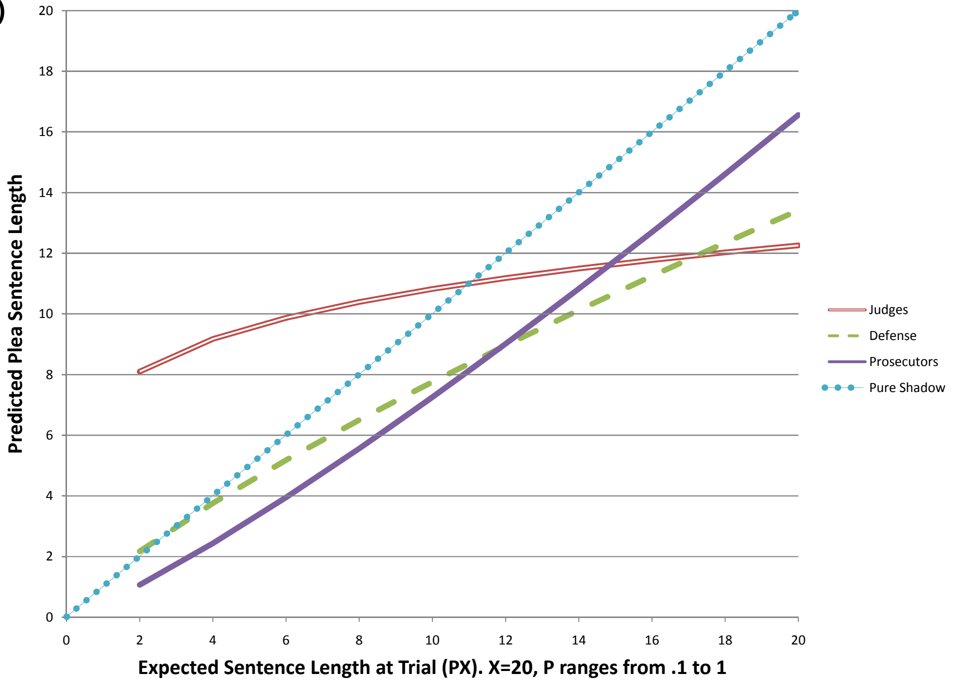
The smallest change from table 4 to table 5 occurs for judges. Although the coefficient on probability of conviction δ_P is now significantly different from zero, it is still small at .18. The results imply that a 10 percent increase in the probability of conviction will only result in a 1.8 percent increase in the value of the plea sentence. Although the instrument bolsters the fit for the simplest shadow model for both prosecutors and defense attorneys, it does little to change the fact that judges seem to be acting in a manner consistent with the fixed discount model.

Figures 3a and 3b provide results from two simulations in an attempt to illustrate more accurately the differences across actors. In the first case, we fix the trial sentence at 20 years and allow the probability of conviction to vary from .1 to 1, and in the second case, we fix the probability of conviction to be .8 and allow the trial sentence to vary from 2.5 to 25 years. In each graph, the horizontal axis is the expected value of the trial (PX), the vertical axis is the plea sentence, and the diagonal line represents the prediction from the simplest shadow of the trial model. Figure 3a demonstrates the relative insensitivity of judges to changes in the probability of conviction. In contrast, the line for the prosecutor is basically parallel to the diagonal, with a slight convex shape. The defense attorneys start out on the diagonal but move below the diagonal as the expected value of the trial gets larger. This happens as a result of the elasticity, δ_P, being less than one.

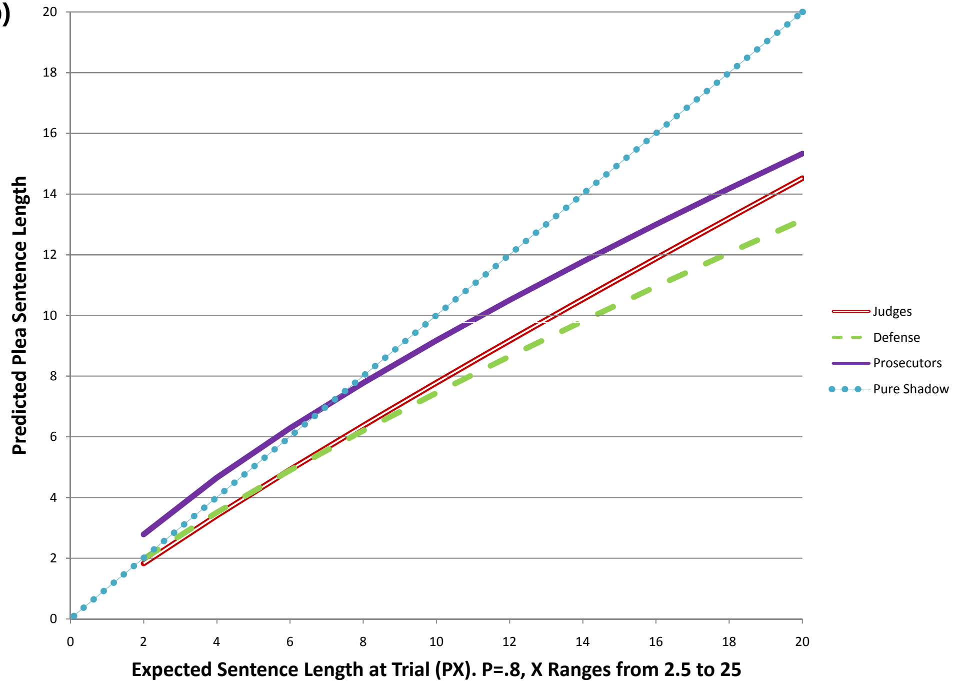
The story for the defense attorneys is nearly identical in figure 3b when compared with figure 3a, which is not surprising given that the elasticities (δ_P and δ_X) were identical for both trial sentence and probability of conviction. The curves for the

Figure 3. Predicted Curves with (a) Changing and (b) Fixed Probability of Conviction and Fixed Trial Length

(a)



(b)



prosecutors and judges resemble the defense attorneys, a reflection of the fact that coefficients on trial sentence were similar across the three actors. The main difference between the actors seems to be the way they respond to variation in the probability of conviction.

The net effect of these regressions is to provide support for a modified version of the shadow model for defense attorneys and prosecutors. The key modification is that the model needs to allow for less than unit elasticity between the components of the expected trial and the sentence length. This could be a reflection of non-neutral risk preferences, or other "nonrational" behavior in the face of these basic choices, suggesting that the basic shadow model needs to be modified somewhat to account for well-known deviations from the basic rational choice model (Albonetti, 1991; Guthrie, Rachlinski, and Wistrich, 2001). Although a detailed exploration of additional models is beyond the scope of this article, the data used here have been given to the National Archive for Criminal Justice Data, and we encourage other researchers to test explicitly other models of behavior that take into account possible deviations from rational behavior.

The case for judges is less subtle; the probability of conviction is clearly less salient for their decision making. Judges are primarily offering a fixed discount of the trial length, a discount that is higher, in percentage terms, when the trial length is long. One possible alternative explanation is that judges are unresponsive to the probability of conviction because they only make decisions on cases from a limited part of the distribution in terms of probability of conviction (close to $P = .5$). Because this information never varies in cases that they see, they were perhaps insensitive to this information when presented (Priest and Klein, 1984).

It is possible that the individual actors are not acting rationally at all and that a different model is at work. As a preliminary test of this idea, we reestimated the instrumental variable model reported in table 5 for each actor, including the age, race, sex, and years of experience of the decision makers as well as some basic facts about their jurisdictions (caseload and county size) and dummy variables for state of practice. These variables often are included in criminological models of plea sentencing. These results are reported in table 6.

Two basic findings can be observed from table 6. First, the basic results from table 2 on the key variables of probability of conviction and trial sentence length do not change appreciably after the inclusion of the control variables. For example, δ_X for judges changes from .91 in table 2 to .92 in table 3 and none of the elasticities are significantly different across the two models. Second, the inclusion of the control variables does add explanatory power to the model, with substantial increases in model fit statistics across models. Most coefficients are not significant, but we can safely reject the assumption that they add nothing to the model. Even though a detailed discussion of these variables is beyond the scope of this article, some of the significant results seem to be interesting. For example, we found that judges with more experience are more likely to agree to higher plea amounts relative to less experienced judges, and male defense attorneys have acceptable plea amounts that are 18 percent lower than their female colleagues. The challenge as we view it is to understand how these factors matter within the context of the shadow model.

Table 6. Two-Stage Least-Squares Model Estimates of General Shadow of Trial Model with Control Variables

Variable	Judges (SE)	Defense (SE)	Prosecutors (SE)
Ln (γ_1) (Constant)	1.21	.48	−2.18*
	(1.12)	(.49)	(1.00)
δ_P, Ln (Probability of Conviction at Trial)	.26*	.83**	1.16**
	(.12)	(.10)	(.29)
δ_X, Ln (Sentence at Trial)	.92**	.73**	.65**
	(.05)	(.03)	(.14)
County Size (Omitted Very Small)			
Small to midsize	−.05	.03	.18
	(.10)	(.11)	(.11)
Midsize to large	−.19*	.07	.11
	(.09)	(.14)	(.11)
Very large	.01	−.05	.21†
	(.09)	(.13)	(.13)
Yearly Caseload (Omitted ≤ 100 cases)			
$100 < X \leq 250$ cases	.13	.12*	−.02
	(.11)	(.06)	(.13)
$250 < X \leq 600$ cases	.04	.04	−.13
	(.11)	(.07)	(.16)
600 cases $< X$	.05	.03	−.12
	(.10)	(.14)	(.16)
Male	.17	−.18**	−.04
	(.10)	(.04)	(.08)
Non-White	.05	.02	−.17†
	(.13)	(.08)	(.09)
Age	−.06	−.01	.09†
	(.04)	(.02)	(.05)
Age Squared	.00	.00	.00
	(.00)	(.00)	(.00)
Years Practice (Omitted ≤ 5 Years)			
$5 < X \leq 11.5$ years	.16*	.13	−.09
	(.06)	(.10)	(.14)
$11.5 < X \leq 20$ years	.18**	.20†	−.15
	(.07)	(.11)	(.13)
$X > 20$ years	.22†	.36**	−.22
	(.13)	(.11)	(.20)
First-Stage F Probability of Conviction	24.69**	55.66**	21.84**
R^2	.65	.53	.39
N	372	835	378

NOTE. This regression also contained dummy variables for every state, and dummy variables indicating missing values for caseload, race, gender, and city size. †$p < .10$; *$p < .05$; **$p < .01$.

CONCLUSION

We make at least three unique contributions in this article. First, we provide a formal mathematical explication of several different theoretical frameworks for understanding guilty plea decision making, including the bargaining in the shadow model and the fixed discount model. We present a structural econometric formulation of these models through log transformation, which simultaneously deals with potential non-normality and allows for a general test of the shadow model. Second, we use insights from the field of psychology and law to resurrect an innovative experimental survey approach from the glory days of plea bargaining research. Following the information capture model used by Rossman, McDonald, and Cramer (1980), the case was presented in a realistic fashion that allowed respondents to choose information they needed to make their decision. Experimental variation in evidence and criminal history created variation in factors that the shadow theory uses to predict plea values—probability of conviction and sentence at trial. Unlike previous studies, the approach was delivered to people using the Internet, an approach that simultaneously allowed us to reach nearly five times more people and implement some important process checks. Third, we exploit the experimental variation to correct for possible measurement error in the probability of conviction through an instrumental variable model.

One of our main findings is that judges do not seem to act in a manner that is consistent with the shadow model. Rather, their behavior is most consistent with a model that offers fixed discounts to people who plead relative to what they could expect to get at trial. An alternative interpretation is that this experiment is particularly artificial and irrelevant for judges, given that they may only make decisions in cases with probability of conviction near .5. As a result, their behavior could still be consistent with the shadow model, despite the findings of the study.

In contrast, we found that defense lawyers behaved in a manner that was largely consistent with the simplest specification of the bargaining in the shadow model, although there was some unexpected nonlinearity as the expected value of the trial increases, leading to larger plea discounts at higher values of the expected trial. The prosecutors had similar coefficients as the defense lawyers, but the instrumental variable model was much less powerful for the prosecutors than for the defense lawyers, which might indicate that prosecutors responded to the experiment differently than the defense attorneys. In a limited exploration of this issue, we found some evidence that additional controls for the characteristics of the individual actors and their jurisdictions added explanatory value to the model, although the control variables did not affect the key coefficients from the structural model.

The main advantage of this survey approach relative to research using administrative data is that it allows for an explicit or structural test of theories of plea decision making, using measures of the key variables (expected outcome at trial, probability of conviction) in these theories. Estimates of these variables are not available in case files. However, we recognize that this benefit comes at the cost of limited external validity relative to prior studies that used actual case data (Abrams 2011; Bushway and Redlich, 2012; Smith, 1986). We acknowledge that the behavior of the actors may differ substantially from their responses in a survey when confronted with the real constraints of the criminal justice system.

One way to make our approach more realistic would be to simulate a negotiation between the actors, essentially developing the research analogy of the mock trial (see Devine et al., 2001) for the plea bargaining context. Alternatively, researchers could attempt to capture valid estimates of the expected value of the trial during actual plea negotiations. Regression-based estimates using actual trial outcomes for those who go to trial require heroic assumptions about the comparability of those who go to trial and those who plead guilty (Abrams, 2011; Smith, 1986). At the same time, actors may face legal barriers to providing truthful answers to questions about ongoing cases. As the collection of case data becomes more systematic and more common, we are hopeful that future research can generate more direct insight into how cases are resolved. A mixed-methods approach that combines this experimental approach in a jurisdiction where case-level data on actual outcomes is available might be a promising first step.

Future research using hypotheticals or actual case data should explore alternative ways to solicit information from actors about the expected outcome at trial. We selected a "sliding bar" approach for the probability of conviction to minimize clumping, but our responses were still clumpy. The instrumental variable approach is a helpful post hoc solution, but it would be trumped by an approach that avoided the problem in the first place.

Subsequent research on plea bargaining also might consider testing more sophisticated versions of the theories specified in this article. Our general econometric model allowed for nonunit elasticities, but our theoretical development did not fully explore the meaning of the apparent nonlinearity that shows up as the expected value of the trial increases. We also did not fully integrate control variables into a theoretical or empirical model in part because our sample did not represent courtroom workgroups in a systematic or representative way. Would the results change if the sample allowed for a more systematic inclusion of the dynamics of the courtroom workgroup? Finally, we look forward to formal theoretical work that either specifies how structural features of the jurisdictions (Bibas, 2004; Ulmer, Eisenstein, and Johnson, 2010) can be integrated into the shadow model or replaces the shadow model with an alternative explanation of plea decision making.

REFERENCES

Abrams, David. 2011. Is pleading really a bargain? *Journal of Empirical Legal Studies* 8:200–21.

Albonetti, Celesta A. 1986. Criminality, prosecutorial screening, and uncertainty: Toward a theory of discretionary decision making in felony case processings. *Criminology* 24:623–44.

Albonetti, Celesta A. 1987. Prosecutorial discretion: The effects of uncertainty. *Law & Society Review* 21:291–314.

Albonetti, Celesta A. 1990. Race and the probability of pleading guilty. *Journal of Quantitative Criminology* 6:315–34.

Albonetti, Celesta A. 1991. An integration of theories to explain judicial discretion. *Social Problems* 38:247–66.

Albonetti, Celesta A. 1997. Sentencing under the federal sentencing guidelines: An analysis of the effects of defendant characteristics, guilty pleas, and departures, 1991–1992. *Law & Society Review* 31:601–34.

Alschuler, Albert W. 1979. Plea bargaining and its history. *Columbia Law Review* 79:1–43.

Alschuler, Albert W. 2002. Straining at gnats and swallowing camels: The selective morality of Professor Bibas. *Cornell Law Review* 88:1412–24.

Baumer, Eric. 2013. Reassessing and redirecting research on race and sentencing. *Justice Quarterly* 30:231–61.

Bibas, Stephanos. 2004. Plea bargaining outside the shadow of trial. *Harvard Law Review* 117:2463–547.

Bradley-Engen, Mindy, Rodney L. Engen, Chris Shields, Kelly R. Damphousse, and Brent L. Smith. 2012. The time penalty: Examining the relationship between time to conviction and trial vs. plea disparities in sentencing. *Justice Quarterly* 29:829–57.

Breen, Patricia. 2011. The trial penalty and jury sentencing: A study of Air Force courts-martial. *Journal of Empirical Legal Studies* 8:206–35.
Brereton, David, and Jonathan D. Casper. 1982. Does it pay to plead guilty? Differential sentencing and the functioning of criminal courts. *Law & Society Review* 16:45–70.
Bushway, Shawn D., and Robert Apel. 2010. Instrumental variables in criminology and criminal justice. In *Handbook of Quantitative Criminology*, eds. Alex R. Piquero and David L. Weisburd. New York: Springer.
Bushway, Shawn D., and Allison D. Redlich. 2012. Is plea bargaining in the "shadow of the trial" a mirage? *Journal of Quantitative Criminology* 28:437–54.
Cohen, Thomas H., and Tracey Kyckelhahn. 2010. *Felony Defendants in Large Urban Counties, 2006*. Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.
Devine, Dennis J., Laura D. Clayton, Benjamin B. Dunford, Rasmy Seying, and Jennifer Pryce. 2001. Jury decision-making: 45 years of empirical research. *Psychology, Public Policy and Law* 7:622–727.
Dixon, Jo. 1995. The organizational context of criminal sentencing. *American Journal of Sociology* 100:1157–98.
Eisenberg, Theodore, and Valerie P. Hans. 2009. Taking a stand on taking the stand: The effect of prior criminal record on the decision to testify and on trial outcomes. *Cornell Law Review* 94:1353–90.
Eisenstein, James, and Herbert Jacob. 1977. *Felony Justice: An Organizational Analysis of Criminal Courts*. Boston: Little, Brown.
Engen, Rodney L., and Sara Steen. 2000. The power to punish: Discretion and sentencing reform in the war on drugs. *American Journal of Sociology* 105:1357–95.
Guthrie, Chris, Jeffrey Rachlinski, and Andrew Wistrich. 2001. Inside the judicial mind. *Cornell Law Review* 86:777–830.
Halberstam, Malvina. 1982. Towards neutral principles in the administration of criminal justice: A critique of supreme court decisions sanctioning the plea bargaining process. *The Journal of Criminal Law and Criminology* 73:1–49.
Kramer, Greg M., Melinda M. Wolbransky, and Kirk Heilbrun. 2007. Plea bargaining recommendations by criminal defense attorneys: Evidence strength, potential sentence, and defendant preference. *Behavioral Sciences and the Law* 25:573–85.
Kutateladze, Besiki, Nancy R. Andiloro, and Brian D. Johnson. 2014. Opening Pandora's box: How does defendant race influence plea bargaining? *Justice Quarterly*. E-pub ahead of print.
Kutateladze, Besiki, Nancy R. Andiloro, Brian D. Johnson, and Cassia Spohn. 2014. Cumulative disadvantage: Examining racial and ethnic disparity in prosecution and sentencing. *Criminology* 52:514–51.
Landes, Williams M. 1971. An economic analysis of the courts. *Journal of Labor Economics* 14:61–107.
McAllister, Hunter A., and Norman J. Bregman. 1986. Plea bargaining by prosecutors and defense attorneys: A decision theory approach. *Journal of Applied Psychology* 71:686–90.
Miller, Herbert S., William McDonald, and James A. Cramer. 1978. *Plea Bargaining in the United States*. Washington, DC: U.S. Department of Justice, National Institute of Justice.

Mnookin, Robert H., and Lewis Kornhauser. 1979. Bargaining in the shadow of the law: The case of divorce. *Yale Law Journal* 88:950–97.

Nagel, Stuart S., and Marian Neef. 1979. *Decision Theory and the Legal Process*. Lexington, MA: Lexington.

Nardulli, Peter, James Eisenstein, and Roy Fleming. 1988. *The Tenor of Justice: Criminal Courts and the Guilty Plea Process*. Urbana: University of Illinois Press.

Plous, Scott. 1993. *The Psychology of Judgment and Decision-Making. McGraw-Hill Series in Social Psychology*. New York: McGraw-Hill.

Priest, George, and Benjamin Klein. 1984. The selection of disputes for litigation. *Journal of Legal Studies* 13:1–55.

Redlich, Allison D., Shawn D. Bushway, and Robert J. Norris. 2013. Bargaining in the shadow of trial? Exploring the reach of evidence. Albany, NY: University at Albany, School of Criminal Justice. Unpublished manuscript.

Rossman, Henry H., William F. McDonald, and James A. Cramer. 1980. Some patterns and determinants of plea-bargaining decisions: A simulation and quasi-experiment. In *Plea Bargaining*, eds. William F. McDonald and James A. Cramer. Lexington, MA: Lexington.

Smith, Douglas A. 1986. The plea bargain controversy. *The Journal of Criminal Law and Criminology* 77:949–68.

Spetzler, Carl, and Carl-Axel Von Holstein. 1975. Probability encoding in decision analysis. *Management Science* 22:340–58.

Steffensmeier, Darrell, Jeffery T. Ulmer, and John Kramer. 1998. The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology* 36:763–98.

Tita, George E., Tricia L. Petras, and Robert T. Greenbaum. 2006. Crime and residential choice: A neighborhood level analysis of the impact of crime in housing prices. *Journal of Quantitative Criminology* 22:299–317.

Tversky, Amos, and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–31.

Ulmer, Jeffery T., and Mindy S. Bradley. 2006. Variation in trial penalties among serious violent offenses. *Criminology* 44:631–70.

Ulmer Jeffery T., James Eisenstein, and Brian D. Johnson. 2010. Trial penalties in federal sentencing: Extra-guidelines factors and district variation. *Justice Quarterly* 27:560–92.

Wallsten, Thomas S., and David V. Budescu. 1983. Encoding subjective probabilities: A psychological and psychometric review. *Management Science* 29:151–73.

Shawn D. Bushway is a professor of public administration and policy in the Rockefeller College of Public Affairs and Policy at the University at Albany (SUNY). His current research interests include the relationship between work and crime, plea bargaining, and the process of desistance and dynamic change in offending behavior.

Allison D. Redlich is an associate professor in the School of Criminal Justice, University at Albany (SUNY). Her research focuses on identifying the dispositional and situational factors that lead to true and false admissions in the interrogation and guilty plea contexts, and on evaluating the effectiveness of mental health courts.