American Psychological Association
American Psychology-Law Society

Law and Human Behavior

ISSN: 0147-7307

2021, Vol. 45, No. 4, 271–286
https://doi.org/10.1037/lhb0000450

Guilt Status Influences Plea Outcomes Beyond the Shadow-of-the-Trial in an Interactive Simulation of Legal Procedures

Miko M. Wilford[1], Kelly T. Sutherland[1], Joseph E. Gonzales[1], and Misha Rabinovich[2]
[1] Department of Psychology, University of Massachusetts Lowell
[2] Art and Design Department, University of Massachusetts Lowell

Objective: More than 95% of criminal convictions in the United States are secured by guilty pleas. Our current understanding of the "deals" that lead so many to plead guilty is often tied to the shadow-of-the-trial (SoT) model, which posits that plea outcomes rely solely on the penalty discrepancy they offer (represented as: [trial conviction probability × trial sentence] – plea sentence). This study compared the power of the SoT model to predict plea outcomes with two expanded models. ***Hypotheses:*** We hypothesized that the SoT model's power to predict whether defendants will accept a plea offer could be improved by expanding the model. The first expanded model added a main effect of guilt status, presuming that regardless of penalty discrepancy, the guilty will be more likely to accept plea offers than the innocent. The second expanded model added an interactive effect of guilt status with penalty discrepancy: although greater discrepancies would increase both true and false guilty pleas, the magnitude of the effect would be greater among the innocent. ***Method:*** We recruited student (N = 596, 45.8% female, M = 19.9 years old) and nonstudent (N = 525, 45.1% female, M = 30.9 years old) adult samples to participate in a 2 (guilt status: innocent or guilty) × 3 (conviction probability: 20%, 50%, or 80%) × 3 (plea discount: 6, 12, or 18 months) mixed design. All participants experienced two crime scenarios in a counterbalanced order. We randomized the manipulated variables for each scenario such that each participant was administered two of eighteen potential conditions. ***Results:*** As hypothesized, the predictive power of the SoT model was significantly improved by expanding it to include guilt status—guilty participant-defendants were consistently more likely to accept the plea offer than innocent participant-defendants. However, an interactive effect of guilt status with penalty discrepancy did not significantly improve the power of the SoT model to predict plea outcomes. ***Conclusions:*** The power of the SoT model could be significantly improved by incorporating guilt status as a predictor. Although there are many times at which guilt status is unknown, the acknowledgment of a separate effect of guilt status has important policy implications for the plea process.

Public Significance Statement
The dominant model of plea decision-making has posited that guilty pleas occur "in the shadow-of-the-trial," rendering actual guilt status largely irrelevant. The current research demonstrates that guilt status does meaningfully impact plea decisions. Future research can now focus on ways of capitalizing upon the differences between the innocent and guilty to reform the *system of pleas* in a way that preserves a high true guilty plea rate while better insulating the innocent from the pressure to plead guilty falsely.

Keywords: adjudication, false guilty pleas, legal processes, plea bargaining, virtual simulation

Supplemental materials: https://doi.org/10.1037/lhb0000450.supp

Lora Levett served as Action Editor.

Miko M. Wilford https://orcid.org/0000-0002-8653-8893
Kelly T. Sutherland https://orcid.org/0000-0002-5896-3438
Joseph E. Gonzales https://orcid.org/0000-0003-1051-6185
Misha Rabinovich https://orcid.org/0000-0001-7983-8743

This material is based on work supported by a CAREER Grant from the National Science Foundation (Award: 1844585). We thank Jordi Love, Thomas Nelson, Cas McAuliffe, Jacky McGrath, Marvin Fung, Matt LeBlanc, Michael Pascale, Marcello Barbieri, and Spencer Fournier for their work on the plea simulation. We also thank Rachele DiFava and Morgan Engdahl for their assistance with data coding.

The data are available at https://osf.io/k9amw/files/

The experiment materials are available at https://osf.io/k9amw/files/

Correspondence concerning this article should be addressed to Miko M. Wilford, Department of Psychology, University of Massachusetts Lowell, 850 Broadway Street, Lowell, MA 01854, United States. Email: Miko_Wilford@uml.edu

> [C]riminal justice today is for the most part *a system of pleas*, not a system of trials. (*Lafler v. Cooper*, 2012)

Every U.S. citizen is afforded a number of unique rights. Many of these rights: the right to a jury of our peers, the right to confront our accusers, the right to present our own witnesses, ultimately, the right to a speedy and public trial, are never invoked by the luckiest of us—those who avoid the scrutiny of the justice system. In fact, even those who fall under the suspicion of legal actors rarely exercise these rights. Instead, today's accused typically forgo their right to a trial, and all rights therein, by entering a guilty plea. Prior to the 1980s, approximately 80% of federal cases were adjudicated by guilty plea (Oppel, 2011). Now, as the majority observed in *Lafler v. Cooper*, the prevalence of guilty pleas has reached an unprecedented proportion of 95% or higher. Clearly, pleas are an established trend, and understanding their impact is of critical importance.

False Guilty Pleas

It is no secret that the U.S. justice system is overburdened. Many jurisdictions are confronted with a dilemma in which they must balance each defendant's right to a public trial with their right to a speedy trial. Although the paucity of defendants afforded all of their constitutional rights is troubling, holding defendants in prison for years while they await their day in court is also problematic. And, pleas unquestionably increase judicial efficiency. Consequently, it is not the increase in pleas, but the increase in *false* guilty pleas—cases in which a now known-to-be-innocent defendant pled guilty—that is worrying. The National Registry of Exonerations (2015) has logged a record-breaking number of false guilty pleas for several years. The grsowing number of documented false guilty pleas is especially troubling in light of the difficulty defendants face when challenging plea convictions (e.g., limited avenues for appeal, restricted mechanisms to overturn conviction or be granted legal assistance, etc.; Fisher, 2000; Redlich, 2010; Wilford & Khairalla, 2019).

The Shadow-of-the-Trial

Why would an innocent person plead guilty? According to the dominant model of plea decision-making—the shadow-of-the-trial (SoT) model (Landes, 1971), innocent, as well as guilty defendants will accept plea offers as a function of the penalty discrepancy (PD):

$$log_e \frac{P(Plea_i = Accept)}{1 - P(Plea_i = Accept)} = \beta_1 \times PD_i + \varepsilon_i \quad (1)$$

Penalty discrepancy (see below) is the difference between the sentence offered in exchange for a guilty plea and the product of the trial conviction probability by the sentence if convicted at trial—in other words, penalty discrepancy represents the difference between the expected cost of trial versus the known cost of a guilty plea (Bushway & Redlich, 2012; Bushway et al., 2014):

$$PD = (Trial\ Conviction\ Probability \times Trial\ Sentence) - Plea\ Sentence \quad (2)$$

For instance, the SoT would posit that if a defendant's conviction probability at trial was 50% with an expected 24-month trial sentence, and the plea offer provided a six-month sentence discount (i.e., an 18-month plea sentence), the defendant should consider the plea more costly than trial (Penalty Discrepancy = [.50 × 24] – 18 = –6); thus, they should reject the plea offer and take the case to trial. If, on the other hand, a prosecutor really wanted to secure a guilty plea (and avoid trial), the deal could be easily modified to that end; the defendant could be offered a discount of 18 months (i.e., a six-month plea sentence), making the penalty discrepancy favor pleading guilty (Penalty Discrepancy = [.50 × 24] – 6 = 6). As long as the plea sentence is less than the expected trial sentence, the SoT predicts that defendants will accept the plea—regardless of guilt status. Although it is safe to assume that innocent defendants will typically face weaker cases (and lower conviction probabilities) than guilty defendants (Easterbrook, 1992; Gazal-Ayal, 2006), it is also important to acknowledge that many innocent defendants have faced strong cases (and been convicted). In these cases, the SoT indicates that there is no way to protect the innocent from pleading guilty—however, we posit that an extended SoT model could highlight plea offers that would continue to attract true guilty pleas *without* necessarily attracting false guilty pleas.

Historically, the SoT model has been used to explain variations in the sentence discounts offered during plea negotiations (Redlich et al., 2017; Wilford et al., 2019). But, scholars have applied it as both an explanatory model (accounting for variations in discounts) as well as a predictive model (determining when defendants are, or are not, likely to accept a plea offer). As a normative model of decision-making, we argue that the SoT model excludes important predictive variables. Specifically, the most robust and reliable finding in plea research is that innocent participant-defendants are significantly less likely to accept a plea offer than guilty participant-defendants, regardless of conviction probability and plea discount (Dervan & Edkins, 2013; Redlich & Shteynberg, 2016; Wilford et al., 2020).

The issues with the SoT model are similar to those of several other normative models (e.g., expected utility theory) that have been scrutinized by the burgeoning field of behavioral economics (Kahneman, 2011; Thaler, 2015; Wilson, 2019). Essentially, it seems that there are a number of variables that can have a systematic effect on human decision outcomes, and yet, are omitted from normative models (e.g., reference points). If all defendants were rational (as the SoT model posits), the expected penalty discrepancy between the plea and the trial (taking the conviction probability at trial into account) would be the only systematically-relevant factor in plea decisions. Yet, research has indicated that innocent defendants are less likely to accept plea offers than guilty defendants, regardless of the penalty discrepancy. Thus, it is possible that case variables, such as the plea discount, could impact defendants differently (Bibas, 2004). This could occur because innocent defendants are generally less willing to plead guilty, and/or innocent participants frame the consequences of pleading guilty (vs. going to trial) differently than guilty defendants (Garnier-Dykstra & Wilson, 2021; Redlich et al., 2017). For instance, on average, guilty people might be generally willing to accept pleas that offer sentencing discounts of at least 20%, whereas innocent people might only be willing to accept pleas that offer sentencing discounts of at least 50%. In this hypothetical, if prosecutors offered discounts of 30%, they would secure a high proportion of pleas from the guilty, and avoid false guilty pleas from the innocent; if, instead, prosecutors offered discounts of 50%, they would secure a very high proportion of pleas from the guilty as well as the innocent (Bordens, 1984). Yet, the current system does little to rein in the discounts that prosecutors can

offer defendants, and acceptance of the SoT model would lead one to assume that such reform would have little impact on protecting the innocent anyway.

The Current Research

Thus, we propose that the SoT model be expanded to reflect effects of guilt status on plea outcomes. This expanded model, if accurate, could highlight the disproportionate impact that certain variables (e.g., the plea discount) can have on plea decision-making among the innocent, relative to the guilty, thereby underscoring the importance of regulating the allowable magnitude of plea discounts. The current research employed a novel plea-simulation program to test the predictive power of the shadow-of-the trial (SoT) model against expanded versions, which included guilt status as a predictive variable (for a demo version of the current experiment, go to: https://demo.pleajustice.org/). Although the SoT model has been tested, several tests have relied on real case data (not experimental data), which means they could not test whether guilt status interacted with conviction probability and/or plea discount to moderate plea outcomes (Abrams, 2011; Bushway & Redlich, 2012). Other research that has examined the impact of SoT-relevant factors on plea outcomes has either not focused on testing the predictive validity of the model (Bordens, 1984; Helm & Reyna, 2017; Tor et al., 2010) or focused on legal actors other than the defendant (Bushway et al., 2014; McAllister & Bregman, 1986). Further, although the impact of guilt status has been significant in several experimental studies, its predictive value within a decision-making model has not been tested. Thus, it remains possible that the magnitude of guilt status' effect is too small to justify expanding the currently parsimonious SoT model.

This expanded SoT model also differs from current plea decision-making theories. The trial penalty model posits that plea negotiations are driven by a singular workgroup (consisting of legal actors) whose primary interest is judicial efficiency (McCoy, 2005; Redlich et al., 2017; Wilford et al., 2019). Because this model focuses primarily on the decisions of the workgroup, it does not distinguish between innocent and guilty defendants—all defendants are "punished" for going to trial. Fuzzy-trace theory has been applied to defendants' decision-making but focuses on individual differences (e.g., age, personality traits) that would lend them to engage in more verbatim versus gist-based processing (Helm & Reyna, 2017; Helm et al., 2018). Consequently, there is no predictive effect of guilt status in this model per se; rather, the way in which guilt status is weighed will differ by the type of processing the defendant engages (Helm & Reyna, 2017). Thus, this proposed expansion of the SoT model would provide a novel theory with which to examine defendants' plea decision-making.

Hypotheses

We hypothesized that the traditional SoT model's power to predict whether defendants will accept or reject a plea offer would be significantly improved by the incorporation of guilt status. Specifically, we expected guilt status to affect plea decisions in two ways: (a) participant-defendant guilt status would significantly impact participants' likelihood of accepting a plea offer (i.e., regardless of plea discrepancy—guilty participants would be more likely to accept a plea offer), and (b) participant-defendant guilt status would moderate the effect of penalty discrepancy such that the discrepancy would elicit a stronger effect on innocent (vs. guilty) participants' plea decisions (see Bordens, 1984; Wilford et al., 2020). Thus, we tested the predictive accuracy of the SoT model with two expanded models.

Method

Participants

Six hundred thirty-five undergraduate students enrolled in introductory psychology courses at a large American Northeastern university participated in this experiment in exchange for two course research credits. Three hundred three of them completed the study in-person, and 332 of them completed the study online. We also recruited 600 nonstudent, community participants from Prolific Academic who received $5.00 for the ∼35-minute study. Prolific Academic offers access to more than 70,000 participants worldwide. They boast a higher level of quality control than competing services (e.g., Mechanical Turk), advertising "sophisticated checks" designed to filter out bots, as well as inattentive human participants. Research has also supported Prolific Academic's claims, finding that Prolific Academic participants produce high quality data, and that these participants might be relatively more naïve than participants on Mechanical Turk (Peer et al., 2017).

All participants had to be eighteen years of age or older. Further, community participants had to identify as U.S. residents. Student participants had to pass one attention check to avoid their participation being terminated early (i.e., prior to the start of the simulation). In-person student participants who failed the attention check received one credit for attending the study. Community participants had to accurately respond to two of three attention checks to be compensated for the study (and for their data to be considered valid). We added additional attention checks for the community participants because Prolific Academic discourages evaluating participants based on one attention check; rather, they ask researchers to include multiple attention checks and allow participants some flexibility in their passage rate (Prolific Academic, 2018). Of the 600 initially recruited community participants, we excluded 18 because they failed too many attention checks; of the 332 initially recruited online student participants, one was excluded for failing the attention check. All participants (both online and in-person) had to pass six of eight manipulation checks for their data to be considered viable. In other words, attention checks allowed us to verify that participants actually attended to the study while manipulation checks allowed us to check whether participants could successfully recall details of the study. We excluded 57 community participants for failing to meet the manipulation check criterion, resulting in a final sample of 525 nonstudent adults. We excluded 37 online student participants resulting in a final sample of 294, and only one in-person student participant was excluded from the analysis resulting in a sample of 302. The resulting total sample size for the study was 1,121 (resulting in 2,242 observations). Although we preestablished the attention and manipulation check requirements, we reran study models using different exclusion criteria (see online supplemental materials; for all recruited participants, Table S.1; for participants who passed all attention and manipulation checks, Table S.2; for participants who passed the *guilt status* manipulation check, Table S.3). None of these variations in the study sample impacted model selection or interpretation.

Community participants reported a mean age of 30.9 years, whereas student participants reported an average age of 19.9. Community participants were 51.8% male, 45.1% female, and 1.9% transgender or gender nonconforming; 1.2% of community participants opted not to report gender. Student participants were 51.2% male, 45.8% female and 1.5% transgender or gender nonconforming; 1.5% of student participants opted not to report gender. Community participants identified

as White (65.5%), Asian (12.4%), Black (7.6%), Hispanic or Latinx (5.9%), bi- or multiracial (5.9%), and American Indian or Alaska Native (.8%). Student participants identified as White (60.6%), Asian (14.6%), Black (9.6%), Hispanic or Latinx (8.2%), bi- or multiracial (4.2%), and American Indian or Alaska Native (.2%). A small percentage of participants (2.9% of community participants and 2.6% of student participants) opted not to report their race.

Power Analysis

This study represents the first phase of a multiphase, grant-funded research project. As part of the grant proposal, a power analysis was conducted to determine a sufficient sample to detect hypothesized effects different from those reported in this study. In accordance with that power analysis, we planned to recruit a minimum of 600 college-aged and 600 community participants. For this study (the first phase of the grant project), an additional power analysis was conducted using the simr package (Green & MacLeod, 2016) for R (R Core Team, 2019) to determine whether the present sample ($N_{Subjects}$ = 1,121, $N_{Observations}$ = 2,242) would have sufficient power to detect a small effect (Chen et al., 2010) of between- and within-subject penalty discrepancy, guilt status, and the interaction of guilt status with between- and within-subject penalty discrepancy. We determined that we had sufficient power (>.90) to detect all effects of interest and to select among the SoT model, the SoT model incorporating guilt status, and the SoT model incorporating guilt status and two-way interactions of penalty discrepancy and guilt status (see Table S.4 in the online supplemental materials).

Design

The current study employed a 2 (guilt status: innocent or guilty) × 3 (conviction probability: 20%, 50%, or 80%) × 3 (plea sentence: 6 months, 12 months, or 18 months) mixed design. Potential sentences were informed by federal grand larceny theft sentencing guidelines (the potential range included all study sentences) and Massachusetts law for which leaving the scene of a motor vehicle accident carries a mandatory minimum sentence of no more than 2 years (we could not find federal guidelines concerning hit-and-runs absent injuries to a person). By steadily varying each of these three variables in a fully factorial design, this study can capture what impact each of these variables has on the decision to plead, and even more importantly, whether these effects differ between the innocent and the guilty. Further, the design produces balanced plea outcomes with regard to what the original SoT model would predict, providing a systematic test of its predictive validity (see Table 1). All participants saw both crime scenarios (a theft and a hit-and-run) in a counterbalanced order. The manipulated variables were randomized for each scenario such that every participant was administered a random combination of guilt status, conviction probability, and plea sentence variables for each of the two crime scenarios. Thus, it was possible for participants to be randomly assigned to the same combination of guilt status (50.4%), conviction probability (33.2%), and plea sentence (32.4%) for both scenarios. Across all three conditions, 5.1% of participants had the exact same combination of all three conditions in both scenarios. A reanalysis of our data was conducted removing those participants with identical conditions across both scenarios to confirm they did not bias the study results; our results and conclusions did not differ with the exclusion of these participants (see Table S.5 in the online supplemental materials). These methods of data collection and randomization created several additional variables that could have a systematic impact on the variables of interest. It was, therefore, important for us to test a number of potential control variables to ensure our results were not an artifact of the way in which data was collected. Specifically, these potential control variables included: crime scenario (hit-and-run or theft), scenario order (first or second), study sample, and in-lab versus online administration of the study (for student participants). Notably, exposing each participant to two crime scenarios also allowed us to analyze both the between- and within-subjects effects of the study variables.

Table 1
List of the Possible Experimental Conditions Presented With the Accompanying Plea Decision Prediction (According to the SoT)

Plea sentence	Conviction probability	Trial sentence	Penalty discrepancy	SoT's prediction
6 months	20%	24 months	−1.2	Reject plea
12 months	20%	24 months	−7.2	Reject plea
18 months	20%	24 months	−13.2	Reject plea
6 months	50%	24 months	6	Accept plea
12 months	50%	24 months	0	50:50 accept/reject
18 months	50%	24 months	−6	Reject plea
6 months	80%	24 months	13.2	Accept plea
12 months	80%	24 months	7.2	Accept plea
18 months	80%	24 months	1.2	Accept plea

Note. SoT = shadow-of-the-trial.

Materials

Study materials are available on the Open Science Framework at https://osf.io/k9amw/files/.

Demographics

Prior to beginning the simulation, participants answered a number of demographic questions (e.g., age, gender, race/ethnicity, level of education). They also answered questions assessing their familiarity and experience with the justice system.

Recollection Measures

After completing each simulated scenario, participants had to recall their guilt status (innocent or guilty), the likelihood of conviction provided by the defense attorney (20%, 50%, or 80%), the plea sentence (6, 12, or 18 months), and the potential trial sentence (24 months) for that particular scenario. These measures served primarily as the manipulation checks.

Subjective Measures

We also included subjective measures of each of these variables asking participants: how guilty they thought their avatar had been (on a Likert-type scale from 1 [*not guilty at all*] to 6 [*extremely guilty*]), their perceived probability of conviction (from 0% to 100%), and their perceived severity of the plea and trial penalties (on a Likert-type scale from 1 [*extremely lenient*] to 6 [*extremely severe*]). These scales were only labeled at the endpoints. These subjective measures were exploratory and, as such, we do not report any results concerning them; we had no a priori predictions or preregistered analyses regarding them.

Plea Decision

Our primary dependent measure of interest was plea outcome. Specifically, we were interested in whether participant-defendants chose to plead guilty or reject the plea offer.

Penalty Discrepancy

A penalty discrepancy (see Equation 2) was calculated for each scenario using the conviction probability and plea sentence (in months) to which participants were randomly assigned, as well as a constant trial sentence of 24 months:

$$\begin{aligned} \textit{Penalty Discrepancy} = {} & \textit{Conviction Probability} \times 24 \\ & - \textit{Plea Sentence} \end{aligned} \tag{3}$$

Resulting penalty discrepancy scores ranged from −13.2 to 13.2 (see Table 1), with an average score of .03 (*Mdn* = .00, *SD* = 7.86). Because of the repeated observations of these data, penalty discrepancy scores were partitioned to the represented within-subject penalty discrepancy scores and between-subjects penalty discrepancy scores. This partition allows us to examine the effect of the penalty discrepancy both between participants (examining the average effect of penalty discrepancy across subjects) and within participants (examining the effect of penalty discrepancy within each subject given our repeated-measures design). Between-subjects penalty discrepancy scores (see Equation 4) are the average penalty discrepancy score for each participant, *i*, across their two trials, *t*, and within-subject penalty discrepancy (see Equation 5) scores are each participants' penalty discrepancy scores minus their between-subjects penalty discrepancy score.

$$PD_{BS_i} = \frac{\sum_t PD_{it}}{t_i} \tag{4}$$

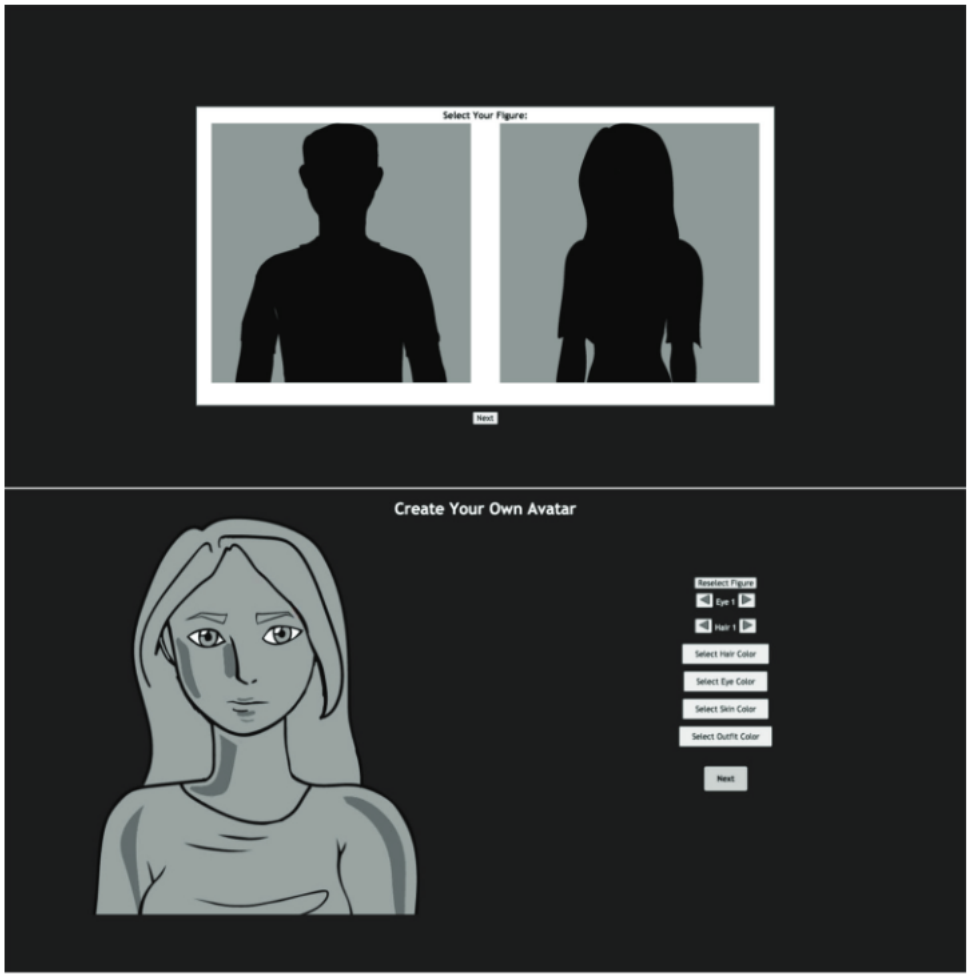
$$PD_{WS_{it}} = PD_{it} - PD_{BS_i} \tag{5}$$

Procedure

The institutional review board at the University of Massachusetts Lowell (IRB no. 18-198-WIL-XPD) approved the methodology for this experiment. After providing informed consent, participants created

Figure 1

The First and Second Pages of the Avatar Customization Process



Note. The graphics were shown in black and white, as pictured, until participants began the customization process.

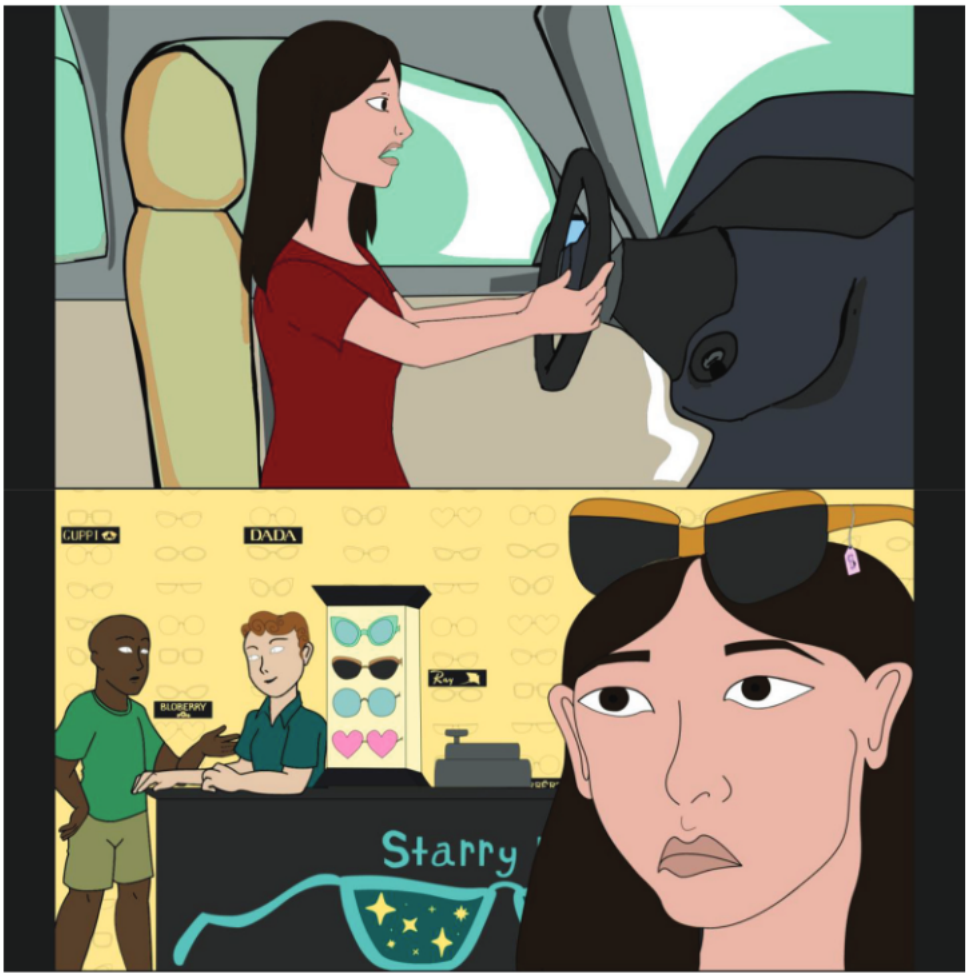
an avatar to represent them in the study (see Figure 1). Upon completing their avatar, they were directed to one of two counterbalanced scenarios: a hit-and-run or a theft. The initial hit-and-run clip showed participant-avatars entering a vehicle and expressing a look of concern (see Figure 2). The initial theft clip showed participant-avatars walking toward a sunglasses store, approaching the cashier, and pointing to a pair of sunglasses behind the counter. The participant-avatars then put the glasses on and walked toward a mirror. While examining themselves, they received a series of text messages from friends. Participant-avatars then look back toward the cashier who is now engaged in conversation with another customer (see Figure 2).

After the initial event, the participant-avatars receive a summons to appear in court two weeks later. In the courtroom, a prosecutor introduces the charges against the defendant (see Figure 3). At this time (and at several points throughout the simulation), participants' actual names appeared within the text to increase their engagement with the scenario. In the hit-and-run scenario, the prosecutor states that damage to the victim's car matched the paint of the participant-avatars' car and shows security camera footage in which the participant-avatars' car appears to contact the victim's car (see Figure 4). In the theft scenario, he states that the store cashier identified the participant-avatar as the person who left the store with the missing glasses and shows security camera footage in which the participant-avatars are walking toward the store exit with the glasses on their heads (see Figure 4). The security footage was identical for innocent and guilty participants for both scenarios. Essentially, the quality of the hit-and-run video made it difficult to discern between a vehicle lightly contacting another vehicle and a vehicle coming within millimeters of another vehicle; similarly, the angle of the theft video could miss someone setting merchandise down immediately before exiting the store. After the prosecutor presents his case, the judge reminds participant-avatars of their rights and remands them to a holding cell to await the assignment of counsel (see Figure 5).

Participant-avatars then experienced a flashback to the day of the incident. In the hit-and-run scenario, innocent participants recall getting very close to the victim's vehicle but narrowly avoiding it (see Figure 6); guilty participants recall having potentially grazed the victim's vehicle, but not thinking they had done that much damage.

Figure 2

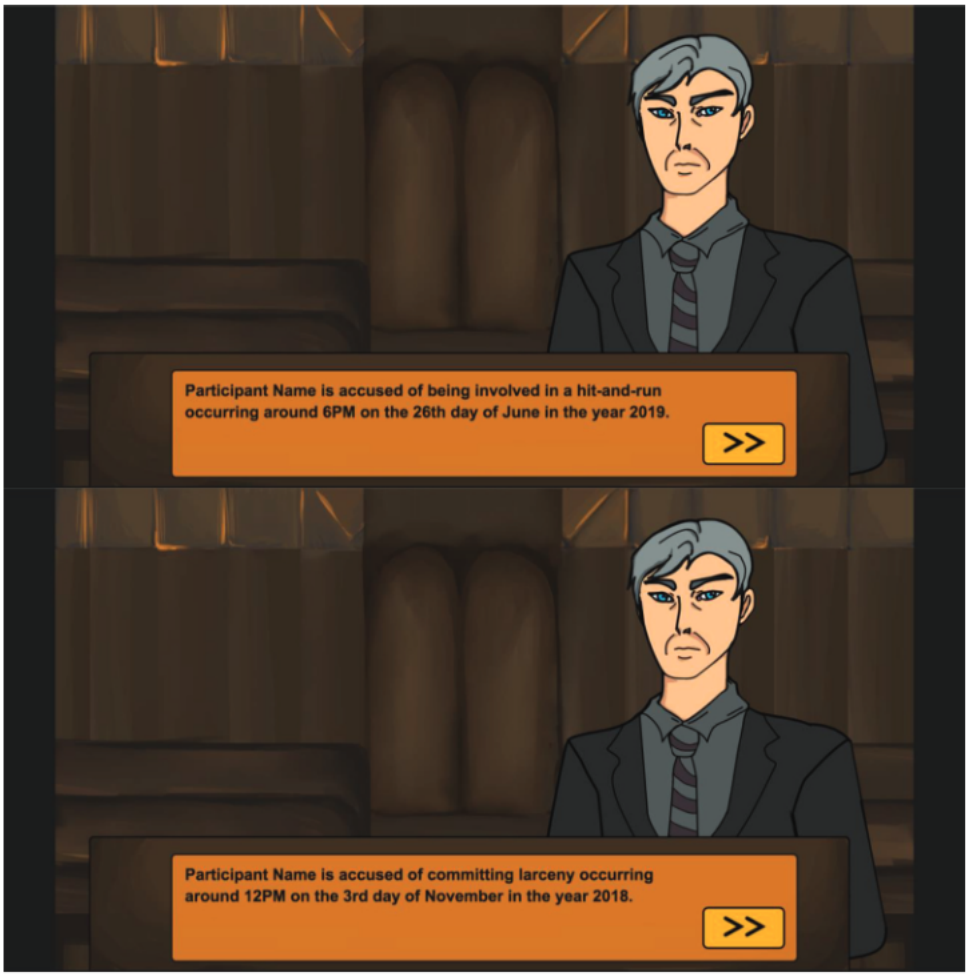
Images From the Opening Sequences for the Hit-and-Run (Top Panel) and Theft (Bottom Panel) Scenarios



Note. See the online article for the color version of this figure.

Figure 3

The Prosecutor Presents the Hit-and-Run (Top Panel) and Theft Charge (Bottom Panel) in Court



Note. See the online article for the color version of this figure.

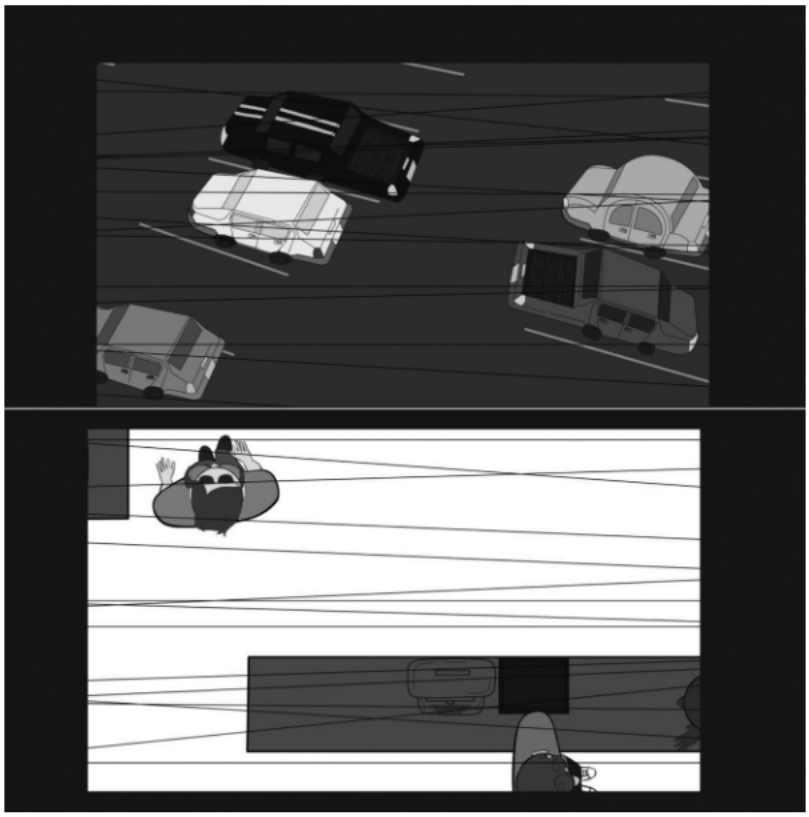
In the theft scenario, innocent participants recall setting the sunglasses on a countertop before leaving the store (see Figure 6); guilty participants recall exiting the store with the glasses atop their heads. In all conditions, the participant-avatars' reflections explicitly revealed their guilt or innocence.

After the flashback, participant-avatars appear in a meeting room with their defense attorney (see Figure 7). The attorney provided participant-avatars with their estimate regarding the chances that they would be convicted at trial (20%, 50%, or 80%) based on the evidence in the case. Although real world defense attorneys might avoid providing their clients with concrete values regarding conviction probability, it would be expected that they try to provide defendants with a general idea regarding their chances at trial when advising them on plea offers. Attorneys also told participants that if the case proceeded to trial, the prosecutor would pursue the maximum penalty of 24 months in jail. The attorney then said that if they pleaded guilty, the prosecutor would be willing to recommend a sentence of 6, 12, or 18 months instead. Participants then chose to accept or reject the plea offer. After each scenario, participants answered a series of questions (e.g., guilt status, probability of conviction, plea offer, trial sentence, etc.), and after they had completed both scenarios, they answered additional questions.

Model Analytic Plan

Because of the repeated observations in the data, general linear mixed modeling was used to nest observations within participants (Gellman & Hill, 2007; Heck & Thomas, 2015; Raudenbush & Bryk, 2002). We conducted all analyses in R ver. 3.6.0 (R Core Team, 2019) using the *lme4* package (Bates et al., 2015). Analyses consisted of two phases. First, we used linear multilevel modeling to confirm that random assignment of participants into conditions did not result in significant dependencies of penalty discrepancy scores as a function of participants, scenario order, or scenario guilt status. Second, we fitted a series of competing logistic multilevel models representing participant plea decision behavior, identified the optimum model to represent the data, and then interpreted the identified model.

Figure 4
Images of the Security Camera Footage From the Alleged Hit-and-Run (Top Panel) and the Alleged Theft (Bottom Panel)



Specifically, to test the original SoT model against competing theoretical models that incorporate guilt status as a predictor of plea decision behavior, we fit a series of hierarchical models: (a) a null, intercept-only model (M0; Equations 6 & 7); (b) M0 with the addition of identified control variables (M1; Equation 8); (c) M1 with the addition of within-, $PD_{WS_{it}}$, and between-subjects penalty discrepancy, PD_{BS_i}—that is, the original SoT model (M2; Equation 9); (d) M2 with the addition of participant

Figure 5
The Judge Reminds Participant-Avatars of Their Rights

Note. See the online article for the color version of this figure.

Figure 6

Participant-Avatars Think Back to the Day of the Alleged Hit-and-Run (Top Panel) or Theft (Bottom Panel)

Note. In these instances, the participant-avatars recall being innocent; in the guilty versions, participants would see damage to the victim's car as their car is shown driving away and would see themselves exiting the store with the sunglasses atop their heads (not pictured). See the online article for the color version of this figure.

guilt status in scenarios (M3; Equation 10); and (e) M3 with the addition of interactions between participant guilt status in scenarios and both within- and between-subjects penalty discrepancy (M4; Equation 11).

$$log_e \frac{P(Plea_{it} = Accept)}{1 - P(Plea_{it} = Accept)} = \beta_{0_i} + \varepsilon_{it}, \text{ where} \tag{6}$$

$$\beta_{0i} = \gamma_{00} + u_{it} \tag{7}$$

$$log_e \frac{P(Plea_{it} = Accept)}{1 - P(Plea_{it} = Accept)} = \beta_{0_i} + \beta_1 \times Order_{it} + \beta_2 \times Student_i + \beta_3 \times Crime_{it} + \varepsilon_{it} \tag{8}$$

$$log_e \frac{P(Plea_{it} = Accept)}{1 - P(Plea_{it} = Accept)} = \beta_{0_i} + \beta_1 \times Order_{it} + \beta_2 \times Student_i + \beta_3 \times Crime_{it} + \beta_4 \times PD_{WS_{it}} + \beta_5 \times PD_{BS_i} + \varepsilon_{it} \tag{9}$$

$$log_e \frac{P(Plea_{it} = Accept)}{1 - P(Plea_{it} = Accept)} = \beta_{0_i} + \beta_1 \times Order_{it} + \beta_2 \times Student_i + \beta_3 \times Crime_{it} + \beta_4 \times PD_{WS_{it}} + \beta_5 \times PD_{BS_i} + \beta_6 \times Guilt_{it} + \varepsilon_{it} \tag{10}$$

Figure 7
The Defense Attorney Introduces Himself to the Participant-Avatars in a New Meeting Room Space

Note. See the online article for the color version of this figure.

$$\log_e \frac{P(Plea_{it} = Accept)}{1 - P(Plea_{it} = Accept)} = \beta_{0_i} + \beta_1 \times Order_{it} + \beta_2 \times Student_i + \beta_3 \times Crime_{it} + \beta_4 \times PD_{WS_{it}} + \beta_5 \times PD_{BS_i} + \beta_6 \times Guilt_{it} + \beta_7 \times PD_{WS_{it}} \times Guilt_{it} + \beta_8 \times PD_{BS_i} \times Guilt_{it} + \varepsilon_{it} \quad (11)$$

We compared models using the likelihood ratio test, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC; Kuha, 2004; Vrieze, 2012), Bayes Factor (Masyn, 2013), and approximate correct model probability (cmP)—the probability that a given model in a group of models is the correct model assuming the true model is among the models considered (Masyn, 2013)—to identify the model that best predicted participants' plea decision outcomes.

Results

Study data are available on the Open Science Framework at https://osf.io/k9amw/files/.

Plea Outcomes

Plea outcomes varied substantially across conditions (see Table 2). The true guilty plea rate ranged from 27.7% to 76.7%, whereas the false guilty plea rate ranged from 4.4% to 30.6%. Thus, whereas guilty participants accepted plea offers at reliably higher rates than innocent participants (as predicted), innocent participants still pleaded guilty with relative frequency. These plea rates were similar to those observed in previous vignette studies that manipulated conviction probability in conjunction with a variety of other variables (Bordens, 1984; Tor et al., 2010; Zimmerman & Hunter, 2018).

Confirmation of Random Assignment

Prior to comparing the original SoT model with our expanded model, we conducted an analysis to confirm that random assignment was effective. The intraclass correlation (ICC; Lorah, 2018) of penalty discrepancy values nested within subjects was 2.7%, indicating an effect below practical significance (Ferguson, 2009), and signifying independence of penalty discrepancy scores and individual participants. Further, we did not find that guilt status in scenarios (B = .26, SE = .33, p = .43), scenario order (B = −.15, SE = .33, p = .66), in-person versus online testing of students (B = −.04, SE = .47, p = .94), student versus nonstudent participant status (B = −.02, SE = .34, p = .95), nor crime scenarios (hit-and-run vs. theft, B = −.06, SE = .33, p = .85) were related to scenario penalty discrepancy scores. Overall, these results suggest that the random assignment procedures did not result in any dependencies of penalty discrepancy scores within-participants or between other experimental conditions.

We also tested for inclusion of the control variables—with respect to the likelihood of accepting a plea. Results of these analyses suggest

Table 2
Proportion of Plea Acceptance Across All Experimental Conditions

Plea sentence	Conviction probability	PD	Guilt status	
			Innocent	Guilty
6 months	20%	−1.2	11.4% (12)	47.5% (56)
12 months	20%	−7.2	10.4% (14)	28.2% (33)
18 months	20%	−13.2	4.4% (6)	29.9% (40)
6 months	50%	6	18.1% (23)	56.5% (70)
12 months	50%	0	8.4% (10)	47.7% (63)
18 months	50%	−6	6.7% (8)	27.7% (33)
6 months	80%	13.2	30.6% (38)	76.7% (112)
12 months	80%	7.2	18.8% (25)	63.8% (74)
18 months	80%	1.2	8.8% (10)	57.3% (71)

Note. The frequency appears in parentheses (n). PD denotes the penalty discrepancy for each set of conditions.

Table 3
Fit Comparisons Between Nested Multilevel Logistic Regression Models of Plea Decision Data

	Likelihood ratio test							
Model	$\Delta\chi^2$	*df*	*p* value	AIC	BIC	BF	cmP	% Accuracy
M0	—	—	—	2,784.8	2,796.23	—	0%	68.90%
M1	17.77	3	<.01	2,773.03	2,801.6	.06	0%	68.90%
M2	150.3	2	<.01	2,626.73	2,666.73	>10	0%	70.40%
M3	378.81	1	<.01	2,249.92	2,295.64	>10	99.94%	78.90%
M4	0.53	2	.77	2,253.39	2,310.54	<.01	0.06%	79%

Note. The change in chi square (χ^2), *df*, and the corresponding likelihood ratio test *p* values in a row indicate a nested model comparison between the model indicated in the row with the model in the row above (e.g., Model 2 is compared with Model 1, and Model 3 is compared with Model 2). Model 3 was selected by the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the likelihood ratio test as the best fitting model of those considered. The % Accuracy is the proportion of plea decisions correctly predicted using the model.

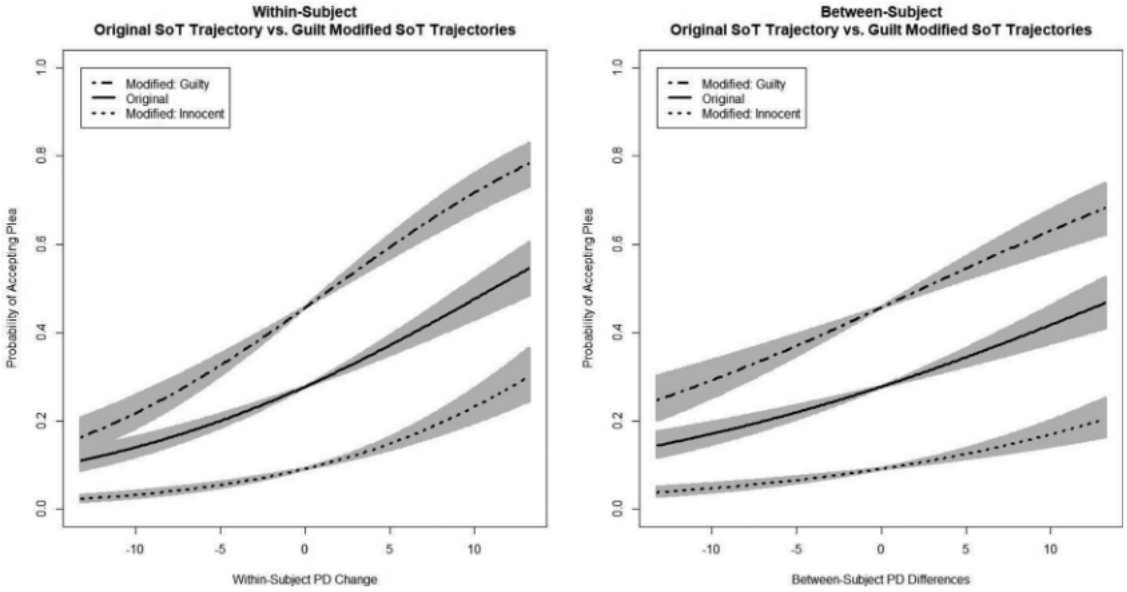
that there were no associations between plea acceptance and hit-and-run versus theft scenarios ($B = -.12$, $SE = .09$, $p = .20$, $OR = .89$) or in-person versus online testing of students ($B = -.07$, $SE = .12$, $p = .56$, $OR = .93$). However, we found that plea acceptance significantly decreases when participants complete their second (vs. first) plea scenario ($B = -.18$, $SE = .09$, $p = .04$, $OR = .83$). Given this significant impact of order, we reran our models looking only at the first scenario to which participants were exposed (refer to Table S.6 in the online supplemental materials). This alternative analysis had no impact on model selection or interpretation. We also found that plea acceptance was significantly higher for students compared with nonstudents ($B = .32$, $SE = .09$, $p < .01$, $OR = 1.38$). Consequently, both the order of scenario completion and student status were incorporated into subsequent models as control variables.

Selection of Competing Plea Decision Models

All model fit comparisons identified an expanded SoT model with the addition of guilt status (M3) as the best representation of plea decisions (see Table 3). Specifically, both likelihood ratio tests and Bayes Factor comparisons of incremental models indicated improved model fit through to M3 but not for M4 (which included an interactive effect of guilt status with penalty discrepancy). Similarly, AIC and BIC both selected M3 as the best fitting model, and cmP indicated that, of the models considered, M3 had a 99.9% probability of being the correct model. This suggests that an expanded SoT model—one that includes guilt status—is the most tenable model of participants' plea decision outcomes.

In comparison with the identified model, issues with the original SoT model (M2) are unmistakable when one examines its predicted trajectory of plea acceptance across penalty discrepancy values (see Figure 8).

Figure 8
Original SoT Trajectories Versus Guilt Modified SoT Trajectories



Note. The black dotted line is the plea acceptance trajectory for innocent participants as a function of PD. The black dot-dash line is the plea acceptance trajectory for guilty participants as a function of PD. The solid black line is the plea acceptance trajectory of participants when guilt status is not incorporated into the model (i.e., the shadow-of-the-trial [SoT] model) and the gray area around the black trajectory lines represents each trajectory's 95% CI. It is evident that the SoT trajectory fails to capture the differential trajectory observed when guilt status is modeled.

These issues can be observed in both the within-subject penalty discrepancy change (left panel), as well as the between-subjects differences in penalty discrepancy (right panel). First, when the classic SoT's plea acceptance trajectory is compared with those predicted when incorporating guilt status (M3), it is evident that the aggregate trajectory of the classic SoT model is masking important variations in plea trajectories as a function of guilt status. Clearly, there are consistent, dramatic differences in the plea rates between the innocent and the guilty across penalty discrepancy values, and these trajectories significantly differ from that predicted by the classic SoT model—failing to represent the plea probability of all participants as a function of penalty discrepancy status and other study conditions.

Second, the SoT theory predicts the probability of plea acceptance is 50% when penalty discrepancy is zero (i.e., $\mathbb{P}$[Accept Plea|PD = 0] = 50%). However, in the original SoT model, the predicted plea acceptance probability when penalty discrepancy is zero is significantly less than 50% ($\mathbb{P}$[Accept Plea|PD = 0] = 28.9%, 95% CI [25%, 33%]). In contrast, the expanded SoT model we identified estimates plea probability rates consistent with the original SoT theory when penalty discrepancy is zero for guilty participants ($\mathbb{P}$[Accept Plea|PD = 0 ∩ Guilty] = 47.3%, 95% CI [39.6%, 55.1%]). Together, this suggests that the original SoT model underestimates plea rates significantly compared with theoretical expectations, with theory-consistent plea rates observed only when guilt status is incorporated into the model.

Interpretation of the Identified Plea Decision Model

Interpretation of the identified model (see Table 4) indicates that plea decision likelihood increases for participants as a function of both within-subject changes in penalty discrepancy between-scenarios (B = .11, SE = .01, $p < .01$, OR = 1.12) and as penalty discrepancy increases between subjects (B = .07, SE = .01, $p < .01$, OR = 1.07). As predicted, guilt status also had a large effect on plea decision likelihood (Chen et al., 2010; Ferguson, 2009), such that guilty participants were significantly more likely to plea than innocent participants (B = 2.13, SE = .14, $p < .01$, OR = 8.41).

There was a significant effect of order (B = −.23, SE = .11, p = .03, OR = .79), such that participants were less likely to plea in their second session than in their first. There was also a significant effect of student status (B = .45, SE = .11, $p < .01$, OR = 1.56), such that student participants were more likely to accept a plea offer than adult community participants. It is worth noting, however, that the effects of order and participant type (student vs. adult) were less than a small effect size (Chen et al., 2010), or below what is considered a practically significant effect (Ferguson, 2009), whereas the effect of guilt status is considered large (Chen et al., 2010) or strong (Ferguson, 2009).

Discussion

The current research indicates, as predicted, that guilt status provides a predictive effect of plea outcomes beyond conviction

Table 4
Model Log-Odds Parameter Estimates, Standard Errors, and Odds Ratios

	Model 1				Model 2			
Predictor	*B*	*SE*	*p* value	*OR*	*B*	*SE*	*p* value	*OR*
Intercept	−0.82	0.09	<.01	0.44	−0.90	0.1	<.01	0.41
Scenario order	−0.19	0.09	.04	0.83	−0.19	0.1	.05	0.83
Student	0.32	0.09	<.01	1.38	0.34	0.1	<.01	1.41
Crime scenario	−0.12	0.09	.20	0.89	−0.12	0.1	.21	0.89
PD W	—	—	—	—	0.09	0.01	<.01	1.09
PD B	—	—	—	—	0.06	0.01	<.01	1.06
Guilt	—	—	—	—	—	—	—	—
PD W × Guilt	—	—	—	—	—	—	—	—
PD B × Guilt	—	—	—	—	—	—	—	—
	Model 3				Model 4			
Predictor	*B*	*SE*	*p* value	*OR*	*B*	*SE*	*p* value	*OR*
Intercept	−2.24	0.16	<.01	0.11	−2.23	0.16	<.01	0.11
Scenario order	−0.23	0.11	.03	0.79	−0.24	0.11	.03	0.79
Student	0.45	0.11	<.01	1.56	0.45	0.11	<.01	1.56
Crime scenario	−0.13	0.11	.23	0.88	−0.13	0.11	.24	0.88
PD W	0.11	0.01	<.01	1.12	0.10	0.02	<.01	1.11
PD B	0.07	0.01	<.01	1.07	0.07	0.02	<.01	1.07
Guilt	2.13	0.14	<.01	8.41	2.11	0.14	<.01	8.27
PD W × Guilt	—	—	—	—	0.02	0.02	.47	1.02
PD B × Guilt	—	—	—	—	0.0,001	0.02	1.00	1.00

Note. Models 1–4 refer to a model with control variables only, the original shadow-of-the-trial (SoT) model, the SoT model incorporating guilt status, and the SoT model incorporating guilt status and guilt status' moderating effect on penalty discrepancy (PD), respectively. *B* indicates the log-odds parameter estimates, *SE* is the standard error for the log-odds parameter estimates, *p* value is the probability of the observed coefficient compared with a null assumption of a coefficient of zero, and *OR* is the corresponding odds ratio of the log-odds parameter estimates. The Scenario Order parameter indicates the change in likelihood of accepting a plea in the second (vs. first) participant trial. The Student parameter indicates the change in the likelihood of accepting a plea for students (vs. nonstudents). The Crime Scenario parameter indicates the change in the likelihood of accepting a plea during the hit-and-run scenario.

probability and plea discount (measured as penalty discrepancy). However, interestingly, we found no evidence of a moderating effect of guilt status. Although penalty discrepancy alone was inferior in predicting plea behavior relative to the combination of penalty discrepancy and guilt status, the effect of penalty discrepancy was consistent for the innocent and the guilty. That is, although guilt status is critical to the prediction of plea behavior, it did not moderate penalty discrepancy's relation with plea decision outcomes.

Further, administering two scenarios to each participant (with varying study conditions) also allowed us to examine the effects of the study variables both between- and within-subjects. Notably, the observed effects were fairly consistent at both levels indicating that penalty discrepancy and guilt status produce change in plea acceptance both across and within defendants. Although previous research has likely assumed that between-participants effects would manifest within-participants, this is one of the first studies to confirm this assumption (methodologically and analytically), thereby further illustrating the robust nature of these effects. We also found that participants were generally less likely to plead guilty in the second scenario versus the first (regardless of the order of the crime scenarios). Although we cannot offer a theoretical explanation for this finding, some research has suggested that defendants with a criminal history are less likely to plead guilty than those newer to the system (Testa & Johnson, 2020). Future research should continue to employ repeated measures to determine the robustness of this effect.

Participant Differences

Notably, by recruiting two different study samples using two different modes of data collection, we were able to make a number of additional observations. First, we did not observe any meaningful differences between the in-person and online student sample. Granted, this null effect was observed after excluding a larger proportion of online student participants who failed the manipulation checks. Regardless, this finding supports the notion that data quality can be preserved even when the simulation is deployed online (as long as sufficient attention and manipulation checks are incorporated into the study measures).

Second, we observed a difference in plea rates between the student and community sample; namely, student participants were about 1.56 times more likely to accept the plea offer than nonstudent participants. The community participants differed from the student participants in at least two theoretically important ways: age and level of education. We found that the average age of student participants (M = 19.86, Mdn = 19, SD = 3.50) was significantly lower than the average age of adult participants recruited from Prolific Academic (M = 30.94, Mdn = 28, SD = 11.27; B = −11.08, SE = .49, $p < .01$, R^2 = .32). Further, student participants had lower levels of educational attainment—34.3% were high school graduates and 59.8% had completed some college—compared with adult participants recruited from Prolific Academic—22.8% reported completing some college and 40.2% reported receiving a four-year degree. Also, 17% of community participants reported having completed either professional or graduate education, but no student participants had such educational attainment. Either of these differences could underlie the higher tendency among students to accept the plea offer. Previous research has indicated that youth can increase the likelihood of pleading guilty (see Redlich et al., 2019 for a review), and that increased education can reduce the likelihood of pleading guilty (Wilford & Khairalla, 2019). This result highlights the importance of further research that involves both student and nonstudent samples. It is possible that samples with demographics that resemble incarcerated populations more closely could respond to study variables differently.

Guilt Status

Although the effect of penalty discrepancy on plea decisions was not moderated by guilt status, as we had predicted, we did observe considerable differences in the estimated probability of plea acceptance across the range of penalty discrepancy values as a function of guilt status' large effect. For example, at the lowest penalty discrepancy score the difference in predicted plea acceptance probability between innocent and guilty participants was around 10% within-participants and 20% between. However, at the highest level of penalty discrepancy the difference between innocent and guilty participants' predicted plea acceptance probability is much larger—closer to 50% both within- and between-participants. Thus, although the effect of penalty discrepancy is not moderated by guilt status, the impact in terms of plea acceptance probability is disproportionate between the innocent and the guilty. These variations are made more interesting by recent research demonstrating that the impact of conviction probability on plea outcomes (or the perceived value of plea offers) among guilty defendants is nonlinear in nature (Bartlett & Zottoli, 2021). Consequently, it remains possible that the combination of conviction probability and plea discount influence innocent versus guilty defendants differently at certain points.

Future Directions and Limitations

We limited the range of penalty discrepancy values in the current study to maintain a balanced design. When penalty discrepancy becomes more exaggerated, it is possible that the predicted interaction could emerge. Thanks to the existence of mandatory minimum sentences, there have been a growing number of cases wherein defendants can be offered massive sentencing discounts in exchange for waiving their right to a trial (e.g., Horwitz, 2015). These types of sentencing discounts would produce penalty discrepancies greatly exceeding the largest employed in the current experiment. Future research should explore the impact of more exaggerated penalty discrepancy values on plea outcomes among the innocent and the guilty. If an interaction between guilt status and penalty discrepancy is found when a larger range is employed, that would provide support for reforms reining in the sentencing discounts prosecutors can offer during plea negotiations. A recent national survey of statutes, regulations, and court rules concerning guilty pleas found very few restrictions on sentence differentials (Zottoli et al., 2019). Of the 52 (50 states, D.C., and the Federal government) jurisdictions examined, only two included any language that addressed the magnitude of the sentence or charge differential prosecutors could offer to defendants.

Another limitation of the current research is that we do not know how participant or case characteristics might influence guilt status in real-life scenarios. That is, guilt status could correlate with other characteristics that might impact plea decision-making in a systematic way. For instance, people who commit crimes might be less sensitive to factors influencing conviction probability, which might make them more likely to plead guilty even in relatively weak cases. It is also probable that (as many legal scholars posit, Easterbrook, 1992) innocent defendants typically face weaker cases than guilty defendants. Similarly, this research did not account for other relations that may exist among crime scenario, sentencing guidelines, conviction probability, and resulting penalty discrepancy ranges.

The way in which we manipulated conviction probability also has limitations. We opted to have conviction probability communicated by the defense attorney as an exact percentage. We believed the defense attorney was the most appropriate source for this information because they would likely be considered both knowledgeable and trustworthy (Henderson, 2021; Henderson & Shteynberg, 2019). That said, both in our study and in the real world, defense attorneys cannot know a defendant's exact probability of conviction. Consequently, it is possible that defense attorneys avoid conveying conviction probabilities. Instead, they might rely on providing their clients with a complete picture of the evidence and allowing them to draw their own conclusions regarding their chances of conviction. However, defendants could weigh evidence very differently. An eyewitness identification could be perceived as strong evidence to one defendant but weak to another. Further, research has shown that other variables can have a meaningful impact on how probabilities are evaluated (e.g., framing, Helm & Reyna, 2017). In fact, even guilt status can impact defendants' perceived probability of conviction (Wilford et al., 2020). Thus, being that conviction probability is a critical component to the SoT model, we opted to manipulate it directly in this study.

Future research should examine the impact of manipulating conviction probability more organically (e.g., via evidence strength). Researchers could also examine the impact of introducing conflicting information regarding conviction probability. For instance, having a prosecutor appear confident in their ability to convict the defendant, whereas a defense attorney appears confident in their ability to prevail at trial. The current study omitted advice from the defense attorney. Although he conveyed information regarding his perceived conviction probability, he did not advise participants to accept or reject the plea. It would be interesting to examine the impact of attorney advice both in the presence and absence of information regarding conviction probability. We also believe that examining the impact of advice from nonlegal actors could be important given the impact that these individuals can have on legal choices, particularly among juveniles (e.g., peers, parents; Daftary-Kapur & Zottoli, 2014).

The current research employed a novel plea simulation that can now be used to investigate a myriad of plea-related questions (Wilford et al., 2021). Although the simulation cannot be modified as easily as a vignette, it does provide some flexibility. The simulation script (i.e., anything said by the judge, prosecutor, or defense attorney) is entirely changeable. Thus, researchers could further manipulate any of the variables included in the current research or build a completely new study that manipulates entirely new variables. The current simulation can employ one or both of the existing crime scenarios and includes three characters (i.e., judge, defense attorney, and prosecutor) whose features (i.e., gender, ethnicity) are changeable. Further, any researcher with access to animation students could create their own assets (e.g., characters, environments, scenarios, etc.) to add to their own versions of the simulation (see internal documentation for guidance on incorporating new assets: https://pleajustice.org/internal). They could then share these assets with others using the simulation. These new assets could then be deployed within the existing simulation framework. The simulation is also designed to interact with Qualtrics such that variables assigned within the simulation (e.g., guilt status, conviction probability) can be passed to a Qualtrics data file. Interested researchers are encouraged to go to https://researcher.pleajustice.org/ to create an account and start designing their own simulation studies.

Conclusion

The current research clearly demonstrates that guilt status matters beyond the shadow-of-the-trial. Thus, future research can now focus on ways of improving the system's ability to separate the innocent from the guilty by finding variables that interact with guilt status to moderate plea outcomes. Our expanded SoT model can be seen as a springboard for this new line of inquiry because we anticipate that other variables could be added to this expanded model to meaningfully improve its predictive power. If we can discover variables that have a differential impact on plea outcomes among the innocent versus the guilty, then we could open the door to plea offers that are better designed to encourage the guilty to plead guilty without necessarily compelling the innocent to do the same. Consequently, we would encourage researchers to prioritize testing system variables (e.g., sentencing discounts, bargaining strategies) that are under the direct control of the justice system (Wilford & Wells, 2013). We offer our plea simulation as a new method of studying these variables. If we uncover system variables that can differentiate the innocent from the guilty, we could inform changes to the plea system that would better ensure that the guilty plead guilty while the innocent go to trial.

References

Abrams, D. S. (2011). Is pleading really a bargain? *Journal of Empirical Legal Studies*, *8*(s1), 200–221. https://doi.org/10.1111/j.1740-1461.2011.01234.x

Bartlett, J. M., & Zottoli, T. M. (2021). The paradox of conviction probability: Mock defendants want better deals as risk of conviction increases. *Law and Human Behavior*, *45*(1), 39–54. https://doi.org/10.1037/lhb0000432

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bibas, S. (2004). Plea bargaining outside the shadow of a trial. *Harvard Law Review*, *117*(8), 2463–2547. https://doi.org/10.2307/4093404

Bordens, K. S. (1984). The effects of likelihood of conviction, threatened punishment, and assumed role on mock plea bargaining decisions. *Basic and Applied Social Psychology*, *5*(1), 59–74. https://doi.org/10.1207/s15324834basp0501_4

Bushway, S. D., & Redlich, A. D. (2012). Is plea bargaining in the 'shadow of the trial' a mirage? *Journal of Quantitative Criminology*, *28*(3), 437–454. https://doi.org/10.1007/s10940-011-9147-5

Bushway, S. D., Redlich, A. D., & Norris, R. J. (2014). An explicit test of plea bargaining in the "shadow of the trial. *Criminology*, *52*(4), 723–754. https://doi.org/10.1111/1745-9125.12054

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics. Simulation and Computation*, *39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Daftary-Kapur, T., & Zottoli, T. M. (2014). A first look at the plea deal experiences of juveniles tried in adult court. *International Journal of Forensic Mental Health*, *13*(4), 323–336. https://doi.org/10.1080/14999013.2014.960983

Dervan, L. E., & Edkins, V. A. (2013). The innocent defendant's dilemma: An innovative empirical study of plea bargaining's innocence problem. *Journal of Criminal Law and Criminology*, *103*(1), 1–48. https://doi.org/10.2139/ssrn.2071397

Easterbrook, F. H. (1992). Plea bargaining as compromise. *The Yale Law Journal*, *101*(8), 1969–1978. https://doi.org/10.2307/796953

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology, Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Fisher, G. (2000). Plea bargaining's triumph. *The Yale Law Journal*, *109*(5), 857–1086. https://doi.org/10.2307/797483

Garnier-Dykstra, L. M., & Wilson, T. (2021). Behavioral economics and framing effects in guilty pleas: A defendant decision making experiment. *Justice Quarterly*, *38*(2), 224–248. https://doi.org/10.1080/07418825.2019.1614208

Gazal-Ayal, O. (2006). Partial ban on plea bargains. *Cardozo Law Review*, *27*(5), 2295–2351.

Gellman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models* (1st ed.). Cambridge University Press.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM Approaches using Mplus* (3rd ed.). Routledge.

Helm, R. K., & Reyna, V. F. (2017). Logical but incompetent plea decisions: A new approach to plea bargaining grounded in cognitive theory. *Psychology, Public Policy, and Law*, *23*(3), 367–380. https://doi.org/10.1037/law0000125

Helm, R. K., Reyna, V. F., Franz, A. A., & Novick, R. Z. (2018). Too young to plead? Risk, rationality, and plea bargaining's innocence problem in adolescents. *Psychology, Public Policy, and Law*, *24*(2), 180–191. https://doi.org/10.1037/law0000156

Henderson, K. S. (2021). Examining the effect of case and trial factors on defense attorneys' plea decision-making. *Psychology, Crime & Law*, *27*(4), 357–382. https://doi.org/10.1080/1068316X.2020.1805744

Henderson, K. S., & Shteynberg, R. V. (2019). Plea decision-making: The influence of attorney expertise, trustworthiness, and recommendation. *Psychology, Crime & Law*, *26*(2), 527–551. https://doi.org/10.1080/1068316X.2019.1696801

Horwitz, S. (2015, August 15). Unlikely allies. *The Washington Post*. http://www.washingtonpost.com/sf/national/2015/08/15/clemency-the-issue-that-obama-and-the-koch-brothers-actually-agree-on/

Kahneman, D. (2011). *Thinking fast and slow*. Farrar, Straus & Giroux.

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research*, *33*(2), 188–229. https://doi.org/10.1177/0049124103262065

Lafler v. Cooper, 132 U.S. 1376 (2012).

Landes, W. (1971). An economic analysis of the courts. *Journal of Labor Economics*, *14*(1), 61–107. https://doi.org/10.1086/466704

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, *6*(1), 8. https://doi.org/10.1186/s40536-018-0061-2

Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. *2*, pp. 551–611). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199934898.013.0025

McAllister, H. A., & Bregman, N. J. (1986). Plea bargaining by defendants: A decision theory approach. *The Journal of Social Psychology*, *126*(1), 105–110. https://doi.org/10.1080/002245451986.9713576

McCoy, C. (2005). Plea bargaining as coercion: The trial penalty and plea bargaining reform. *Criminal Law Quarterly*, *50*(1–2), 1–41.

National Registry of Exonerations. (2015, November 24). *Innocents who plead guilty*. http://www.law.umich.edu/special/exoneration/Documents/NRE.Guilty.Plea.Article1.pdf

Oppel, R. A. Jr. (2011, September 26). Sentencing shift gives new leverage to prosecutors. *The New York Times*. http://www.nytimes.com/2011/09/26/us/tough-sentences-help-prosecutors-push-for-plea-bargains.html?_r=0

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Prolific Academic. (2018, September 12). *Using attention checks as a measure of data quality*. https://researcher-help.prolific.co/hc/en-gb/articles/360009223553-

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Sage, Inc.

R Core Team. (2019). R. A language and environment for statistical computing (Version 3.6.0) [Computer software]. https://www.R-project.org/

Redlich, A. D. (2010). False confessions, false guilty pleas: Similarities and differences. In G. D. Lassiter & C. Meissner (Eds.), *Interrogations and confessions: Current research, practice, and policy* (pp. 49–66). APA Books. https://doi.org/10.1037/12085-003

Redlich, A. D., & Shteynberg, R. V. (2016). To plead or not to plead: A comparison of juvenile and adult true and false plea decisions. *Law and Human Behavior*, *40*(6), 611–625. https://doi.org/10.1037/lhb0000205

Redlich, A. D., Wilford, M. M., & Bushway, S. (2017). Understanding guilty pleas through the lens of social science *Psychology, Public Policy, and Law*, *23*(4), 458–471. https://doi.org/10.1037/law0000142

Redlich, A. D., Zottoli, T., & Daftary-Kapur, T. (2019). Juvenile justice and plea bargaining. In V. A. Edkins & A. D. Redlich (Eds.), *A system of pleas: Social science's contribution to the real justice system* (pp. 107–131). Oxford University Press. https://doi.org/10.1093/oso/9780190689247.003.0007

Testa, A., & Johnson, B. D. (2020). Paying the trial tax: Race, guilty pleas, and disparity in prosecution. *Criminal Justice Policy Review*, *31*(4), 500–531. https://doi.org/10.1177/0887403419838025

Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. Norton Co., Inc.

Tor, A., Gazal-Ayal, O., & Garcia, S. M. (2010). Fairness and the willingness to accept plea bargain offers. *Journal of Empirical Legal Studies*, *7*(1), 97–116. https://doi.org/10.1111/j.1740-1461.2009.01171.x

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Wilford, M. M., Frazier, A., Sutherland, K. T., Gonzales, J. E., & Rabinovich, M. (2021). *A system of pleas: Testing a role-playing computer simulation of plea-bargaining*. Manuscript in preparation.

Wilford, M. M., & Khairalla, A. (2019). Innocence and plea bargaining. In V. A. Edkins & A. D. Redlich (Eds.), *A system of pleas: Social science's contribution to the real justice system* (pp. 132–152). Oxford University Press.

Wilford, M. M., & Wells, G. L. (2013). Eyewitness system variables: Revisiting the system-variable concept and the transfer of system variables to the legal system. In B. L. Cutler (Ed.), *Reform of eyewitness*

identification procedures (pp. 1–29). American Psychological Association. https://doi.org/10.1037/14094-002

Wilford, M. M., Wells, G., & Frazier, A. (2020). Plea-bargaining law: The impact of innocence, trial penalty, and conviction probability on plea outcomes. *American Journal of Criminal Justice*. Advance online publication. https://doi.org/10.1007/s12103-020-09564-y

Wilford, M., Shestak, A., & Wells, G. (2019). Plea bargaining. In N. Brewer & A. Bradfield Douglass (Eds.), *Psychological science and the law* (pp. 266–292). Guilford Press.

Wilson, T. (2019). The promise of behavioral economics for understanding decision-making in the court. *Criminology & Public Policy*, *18*(4), 785–821. https://doi.org/10.1111/1745-9133.12461

Zimmerman, D. M., & Hunter, S. (2018). Factors affecting false guilty pleas in a mock plea bargaining scenario. *Legal and Criminological Psychology*, *23*(1), 53–67. https://doi.org/10.1111/lcrp.12117

Zottoli, T. M., Daftary-Kapur, T., Edkins, V. A., Redlich, A. D., King, C. M., Dervan, L. E., & Tahan, E. (2019). State of the States: A survey of statutory law, regulations and court rules pertaining to guilty pleas across the United States. *Behavioral Sciences & the Law*, *37*(4), 388–434. https://doi.org/10.1002/bsl.2413

Received September 2, 2020
Revision received May 20, 2021
Accepted May 22, 2021 ■