Journal of Experimental Criminology (2022) 18:321–342
https://doi.org/10.1007/s11292-020-09449-4

Diverging from the shadows: explaining individual deviation from plea bargaining in the “shadow of the trial”

Kevin Petersen[1] · Allison D. Redlich[1] · Robert J. Norris[1]

Accepted: 12 October 2020 / Published online: 7 November 2020

Abstract

Objectives The “shadow of the trial” (SOT) theory posits that plea decisions result from mathematical predictions of probability of conviction (POC) at trial and potential trial sentence (TS). Tests of the SOT model often find support in the aggregate, but not at the individual level. This study examines the factors that account for adherence to, or deviation from, the SOT model, such as mathematical competence, a factor not previously examined in tests of the SOT model.
Methods Participants were randomly assigned to one of nine conditions corresponding to manipulations of probability of conviction (10%, 50%, 90%) and potential trial sentence (5, 15, 25 months). After reading a case description, participants were asked whether they would accept a plea offer and how much time in jail they would be willing to spend; a subset of participants was offered a counter plea offer. Participants then answered questions assessing numeracy and about their legal opinions and personal characteristics.
Results Results showed that probability of conviction, but not trial sentence, influenced shadow model adherence. Participants assigned to 50% and 90% POC conditions were significantly less likely to deviate from the SOT model than participants assigned to 10% conditions. This effect did not interact with TS. Additionally, the odds of fitting the SOT model increased significantly as participants’ numeracy scores increased.
Conclusions Our results raise questions about the validity of the SOT model at low POCs and challenge its assumption that defendants are capable of conducting the mathematical calculations required to fit the model.

Keywords Shadow of the trial · Plea bargaining · Guilty pleas · Legal decision-making · Numeracy · Criminal justice

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11292-020-09449-4) contains supplementary material, which is available to authorized users.

✉ Kevin Petersen
kpeter4@masonlive.gmu.edu

[1] George Mason University, Fairfax, VA, USA

Introduction

The leading theory of plea bargaining, the "shadow of trial" model (hereafter "SOT" or "shadow model"), posits that defendants premise their plea decisions on forecasted trial outcomes, meaning that decisions are driven by the perceived likelihood of conviction at trial and the likely sentence if convicted at trial (Bibas 2004; Bushway et al. 2014; Mnookin and Kornhauser 1979). Although valid on its face, the SOT theory has been critiqued as overly simplistic in that it fails to account for structural and psychological factors that influence plea decision-making (see below; Bibas 2004). Furthermore, empirical tests have demonstrated that while the model may hold in the aggregate, there is wide individual-level variation in model adherence (Bushway and Redlich 2012; Bushway et al. 2014). Despite the evidence of such variation, however, there has been relatively little empirical research examining the factors that may lead individuals to adhere to, or deviate from, the shadow model.

In this study, we seek to further specify the conditions under which individuals do and do not adhere to the shadow model. In addition to its core components, we investigate the influence of numeracy on model adherence, whether those with higher level mathematical skills are more likely to adhere to the shadow model. Given that the overwhelming majority of criminal convictions are the result of guilty pleas (National Center for State Courts, n.d.; United States Sentencing Commission 2018), it is important to continue developing the SOT theory to better understand the decision-making processes driving this primary method of case disposition.

Plea bargaining in the "shadow of the trial"

In its simplest terms, the SOT theory approaches plea bargaining as a choice between a certain outcome—an agreed-upon plea deal—and an uncertain one—a trial (Bushway 2019). It is rooted in rational choice, suggesting that "rational parties forecast the expected trial outcome and strike bargains that leave both sides better off by splitting the saved costs of trial" (Bibas 2004, p. 2464). The SOT theory, at its core, is a mathematical one, in that plea decisions are based on (1) the perceived probability of conviction/acquittal at trial and (2) the sentence received if convicted/acquitted at trial. Because the sentence if acquitted is zero, the expected value of a trial is calculated as the probability of conviction (POC) multiplied by the trial sentence (TS). For example, if a defendant perceives their likelihood of conviction to be 70% (POC = .70), and the charged crime carries a sentence of 10 years (TS = 10), then the expected value of trial is 7 years (.70 × 10). Thus, a rational defendant should be willing to accept a plea sentence equal to or lower than this expected value. These two variables—POC and TS—make up the primary elements of the shadow model.

Although the SOT theory is often discussed among legal commentators as a logical extension of rational choice, critics have suggested that it is "far too simplistic." The model fails to account for several factors, including both "structural impediments" and "psychological biases," that influence plea decisions. Structural impediments include "poor lawyering, agency costs, and lawyers' self-interest," as well as rules governing bail, pretrial incarceration, and sentencing (Bibas 2004, pp. 2466–2467; see also, Redlich and Edkins 2019). The model also assumes that those involved in the plea process are rational actors, although that proposition is questionable (Bushway et al. 2014). Psychological biases, such as overconfidence, denial mechanisms, and others (Bibas 2004, p. 2467), hinder rationality and remain unaccounted for in the shadow model. Furthermore, other individual differences, such as defendant race—which is known to affect trial outcomes, plea recommendations, and plea decisions (see Johnson and Richardson 2019)—are also not included in the SOT model. One particular factor, numeracy, or an individual's "facility with basic probability and numerical concepts" (Schwartz et al. 1997, p. 966), may also be important given the mathematical underpinnings of the model (Clatch 2017).

Deviation from the shadow model

Bibas (2004) contended that many pleas "diverge from the shadows of trials" (p. 2467), and that plea outcomes may not be as systematic or predictable as the model suggests, but rather based on individual factors such as wealth, age, education, or intelligence. Discussions of the SOT theory have generally been limited to legal literature (Bushway and Redlich 2012), and there are relatively few direct empirical tests of the model (Clatch 2017). Nonetheless, there is reason to believe that Bibas's critiques are valid, as there is increasing evidence suggesting that not all individuals make plea decisions according to the shadow model.

Aggregate SOT studies Bushway and Redlich (2012) conducted one of the first tests of the SOT theory. Using a dataset of felony cases to allow observed trial outcomes to serve as counterfactual outcomes for similar plea-bargained cases, the authors estimated the POC and TS, and thus computed the expected SOT plea value for these defendants. Their results suggested that, in the aggregate, expected and actual trial outcomes closely aligned, consistent with the SOT theory; whereas, 76.7% of those who went to trial were convicted, the estimated POC for those who pled guilty was quite similar, at 77.1%. However, they noted considerable individual-level deviation in model adherence, with predicted probability of incarceration for individual plea bargainers ranging from significantly lower than, to significantly higher than, those for defendants who went to trial. In fact, 16% of these actual cases were found to have probability of incarceration ratios greater than 1.0.

Bushway et al. (2014) conducted another SOT test using an online survey experiment in which prosecutors, defense attorneys, and judges responded to a hypothetical robbery case. Participants reported the TS, their perceived POC, and the plea sentence they would be willing to offer or accept. Again, the average results were consistent with the shadow model, but there was notable deviation, both between and within legal actor type. Whereas prosecutors and defense attorneys were generally found to bargain in the shadow of trial, judges did not. Judges' estimations were associated with significantly larger plea discounts than would be predicted by the shadow model and were relatively nonresponsive to changes in POC. Furthermore, defense attorneys displayed more nonlinearity than prosecutors across POC levels, as their maximum predicted plea sentence lengths tended to taper off as POC increased.

What remains unclear, however, are the factors that lead to individual deviation. Recent qualitative work provides some insight here. In interviews with defense lawyers in Hong Kong, Cheng and Chui (2015) found that, while POC and TS were commonly considered factors during plea negotiations, trial-related costs and sentence discounts (among other factors) were also important. Similarly, Wright et al.'s (2020) survey of defense attorneys suggests that neither POC nor sentencing considerations are among the most important factors during plea negotiations. Rather, the client's criminal history, knowledge of the relevant case facts, and the client's wants and needs had the highest average importance ratings, respectively.

One issue facing the extant SOT research is that it largely fails to capture the defendant's perspective, instead focusing on legal actors (e.g., Bushway et al. 2014; Pezdek and O'Brien 2014; Wright et al. 2020). While this research is beneficial, the SOT theory ultimately "starts from the perspective of the defendant" (Bushway et al. 2014, p. 726), and should be tested at this individual level (Bibas 2004). To our knowledge, few studies that have focused on shadow model adherence have attended to the defendants who deviate from the model or their possible reasoning. However, several studies have examined defendants' plea decisions more generally.

Willingness to plea studies Although not explicit tests of the SOT theory, experimental studies using participant-defendants have produced evidence that the core aspects of the shadow model—POC and potential TS—are important factors in plea decision-making individually, but that these factors may not interact as predicted by the model. For example, Bordens (1984) manipulated guilt/innocence, POC, TS, and plea sentence in a hypothetical scenario of negligible homicide, and found that all factors significantly affected plea acceptance, with guilty subjects in higher POC conditions and higher TS conditions being more likely to accept plea offers. Importantly, however, Bordens's (1984) participants used the provided pieces of information independently from each other; contrary to his predictions, "the manner in which one piece of information was used was not significantly influenced by the others" (p. 71). Of particular relevance, over 57% of guilty participants in 10% POC conditions accepted plea offers, despite the fact that nearly all such offers were greater than the expected value of the trial under the SOT model.

Using a similar design, Tor et al. (2010) manipulated POC (ranging from 5 to 95%) while holding TS constant. Participants were given plea offers in each POC condition that were exactly equal to the expected value of the trial under the SOT model. Under the baseline SOT model, participants would be predicted to plea at the same rate across conditions; however, Tor et al. found that innocent defendants increasingly accepted plea offers as POC increased, suggesting that these participants did not respond to POC in a linear fashion. In a separate study, Tor et al. (2010) determined that defendants uncertain of their guilt tended to behave as if they are innocent, which the authors noted is "in clear contrast with models of rational defendant behavior" (p. 111).

SOT research with participant-defendants While experimental studies testing participant-defendants' willingness to plead have implications for understanding individual-level deviation from the SOT theory, they are not direct tests of the theory and are not designed to facilitate direct inferences. There is, to our knowledge, only one existing study that provides a direct test of the shadow model using an experimental design with participant-defendants.

Bartlett and Zottoli (in press) manipulated POC (5%, 10%, 15%, 50%, 85%, 90%) and TS (5 years, 10 years) among participants playing the role of a defendant charged with a campus drug offense. Participants were asked to indicate the maximum sentence they would be willing to accept as part of a plea agreement and were then offered a plea deal that was equal to the trial outcome predicted by the shadow model. Ultimately, 91% of participants expressed willingness to accept a plea, though plea values displayed notable variation. The authors created deviation scores for each participant, measuring the distance between the expected SOT trial value and the maximum plea value that participants indicated they would accept. Analyzing the variance in deviation scores across experimental conditions, Bartlett and Zottoli found that, while plea willingness generally increased as POC increased, deviation score was significantly and inversely related to POC. Specifically, the average deviation scores were 390% of the predicted shadow value in the .05 POC conditions, 61% of the predicted value in the .50 POC conditions, and only 42% of the predicted value in the .90 POC conditions. In contrast, potential TS did not exert any independent main effect on deviation scores. Thus, Bartlett and Zottoli's results suggest that shadow model adherence may operate independently of individuals' willingness to plead guilty, and that POC is an important factor in explaining individual-level deviation from the shadow model, independent of TS.

Overall, the existing evidence suggests several important patterns. First, the tenets of the shadow model appear to hold at the aggregate level (Bushway and Redlich 2012; Bushway et al. 2014; see also Bonneau and McCannon 2019) but not at the individual level. Second, the core elements of the SOT theory—POC and TS—are important factors that influence plea decisions among individual mock defendants, though not necessarily in the combination suggested by the shadow model, and not necessarily in isolation of other considerations. Previous individual-level studies, however, have rarely been designed around the formal shadow model and have thus not shown how their findings regarding plea decisions fit within the broader SOT theory. Bartlett and Zottoli (in press) provide much needed insight on the factors that may contribute to individual shadow model deviation. However, it remains important to expand upon this research to include additional factors that may operate beneath the surface.

One potential source of deviation is an individual's ability to make calculations. As the SOT theory is a formal mathematical model, it stands to reason that individuals with increased mathematical ability are more likely to adhere to the shadow model. Recently, Helm et al. (2020) found that mock jurors with higher numeric abilities provided more valid civil award amounts than those with lower abilities. Furthermore, research on medical risk perceptions has found that numeracy can be strongly associated with risk perception (Schwartz et al. 1997; see also Lipkus et al. 2001; Hill and Brase 2012). Given that the plea decision carries an inherent juxtaposition of risk and certainty, and that the SOT model proposes a mathematical calculation of expected trial values, we test whether an individual's mathematical ability affects the relationship between the circumstances of their case and their adherence to, or deviation from, the shadow model.

Lastly, another potential explanation for SOT deviation is the defendant's own perception of POC. In previous studies that have examined the effect of POC on willingness to plead, the probabilities have been manipulated and assigned randomly. For example, both Bordens (1984) and Bartlett and Zottoli (in press) informed participants of their attorneys' estimate of their POC at trial. But, what if the defendant does not agree with this probability? If the defendant's perception of conviction probability differs from their assigned condition, the trial value calculated based on the shadow model will be inaccurate. Thus, the current study expands on this concern by assigning a manipulated POC while also allowing for measurement of participant perceptions.

The present study

In the present study, we test SOT model deviation at the individual-defendant level. We add to Bartlett and Zottoli's (in press) work by including a unique measurement of conviction probability and assessing the relative importance of other factors, including numeracy. We aim to further understand the factors that account for individual-level deviation from the shadow model, including those that, to our knowledge, have not yet been tested. Using a 3 × 3 between-subjects factorial design, we first examine the effects of POC (10%, 50%, 90%) and maximum potential TS (5 months, 15 months, 25 months) on willingness to accept plea offers among participants acting as defendants. We then create three measures of SOT deviation (see below), examining the influence of our manipulated variables and numeracy on adherence to the shadow model. Finally, we examine the method and magnitude of SOT deviation.

Based on prior research (e.g., Bordens 1984; Tor et al. 2010; cf. Schneider and Zottoli 2019), we hypothesized that individuals in higher POC conditions and higher TS conditions would be more likely to accept plea agreements. With regard to model adherence across experimental conditions, we made no a priori assumptions when we designed the study, given the lack of research on the factors underlying individual deviation from the SOT model. However, in line with the new findings of Bartlett and Zottoli (in press), we would expect shadow model deviation to be significantly and inversely associated with POC, while not significantly associated with potential TS. Finally, we hypothesized that individuals with higher numeracy would be more likely to adhere to the shadow model.

Method

Participants

Participants were recruited through the Amazon MTurk service. Eligibility was restricted to users who were at least 18 years of age, located within the USA, and had an approval rating above 95% on all previous MTurk tasks. Prior studies utilizing MTurk samples have often established preemptive criteria for disqualification based on manipulation/attention-check questions and completion time (e.g., see Schneider and Zottoli 2019). Our recruitment and consent documents informed participants of a number of embedded attention/manipulation-check questions that, if failed, would result in disqualification. A total of 387 responses were recorded; however, 33 respondents failed the initial attention-check question (resulting in immediate termination) and 109 respondents failed a subsequent manipulation/attention-check question.

Ultimately, 245 responses were considered valid for purposes of analysis, resulting in cell sizes ranging from 25 to 31 across the nine experimental conditions. The final sample averaged 37.53 years of age (SD = 11.22); 53.1% identified as males and 72.2% identified as white. Additionally, 56.7% of respondents reported having at least a bachelor's degree and 66.5% reported an annual family income between $30,000 and $69,999.

Materials and measures

Vignette and design Participants were asked to play the role of a defendant in a DUI case. We designed the scenario to allow for our analyses while still maintaining an element of realism based on the context of the vignette (e.g., we abided by statutory sentences for similar charges). Participants were told they had been pulled over by the police after having a few drinks and were subsequently arrested after failing a field sobriety test and having a blood alcohol level of 0.09, slightly over the legal limit. They were told that they had prior DUI convictions and thus faced a jail sentence. However, their defense attorney, who had reviewed the arresting officer's body-worn camera footage, suggested that there may not have been sufficient grounds for their arrest, as the field sobriety test was not actually failed, as indicated by the officer. In this manner, participants were not assigned to explicit guilt/innocence conditions but could arrive at a more organic determination. The full text of our vignette is available in the Supplemental Materials.

Independent variables We manipulated two independent variables: POC (10%, 50%, 90%; suggested by their defense attorney) and maximum TS (5, 15, or 25 months).

In addition to our main manipulations, we measured numeracy using questions from the "General Numeracy Scale" (Lipkus et al. 2001), which has primarily been used in studies of medical risk perception (Lipkus et al. 2001; Schwartz et al. 1997; see also Hill and Brase 2012). This 3-item scale consisted of questions involving the calculation/conversion of proportions and percentages. The full questionnaire is available in the online supplementals, but an example question asked, "Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips?" (Schwartz et al. 1997, p. 967). The calculations provided some measure of respondents' ability to engage in numeric reasoning and calculate percentages/proportions, processes similar to the calculations that defendants would make in generating an expected trial value according to the SOT theory. The participant's numeracy score was measured as the aggregate number of questions answered correctly (0–3). Overall, our sample averaged approximately 2 out of 3 correct answers ($M = 1.98$, SD = .996).

Dependent variables We calculated three measures of deviation from the shadow model. The first was whether participants were willing to accept a plea deal and the length of the deal they would be willing to accept. This variable was measured using two separate questions. The first asked the participant about the maximum number of months in jail they would be willing to spend to "avoid the chance of a trial conviction and the possible maximum sentence" (indicated on a slider scale, ranging from 0 to 5/15/25 months, corresponding to condition). The second asked the respondent whether they would be willing to accept a plea agreement at the value they specified (or conversely refuse to accept any plea offer).[1] Based on these responses, participants were initially classified as either adhering to or deviating from the shadow model. If a participant indicated that they would accept a plea deal involving more jail time than the maximum value predicted by the shadow model for their respective condition, or if they indicated they would not accept any plea deal, they were coded as deviating from the model. The maximum predicted plea values under the SOT model for each condition can be seen in Table 1.

Conversely, if the participant indicated that they would accept a plea deal involving less than or equal to the maximum sentence predicted under the shadow model, they were coded as adhering to the model. This is the traditional mathematical view of the shadow model assuming a risk-neutral defendant (see Bushway et al. 2014). The below equation is provided to clarify; PS denotes plea sentence, POC denotes probability of conviction at trial based on experimental condition, and TS denotes maximum trial sentence based on experimental condition. The resulting dichotomous measure became our first dependent measure of SOT deviation.

$$\mathrm{PS} > \mathrm{POC} \times \mathrm{TS} = \text{Deviation from model} \tag{1}$$

$$\mathrm{PS} \leqq \mathrm{POC} \times \mathrm{TS} = \text{Adherence to model} \tag{2}$$

After reading the vignette, participants were also given the chance to indicate whether they disagreed with their defense attorney regarding the POC for their case, and if so, what they believed their true POC to be (indicated via a slider scale from 0 to 100%). This is an important addition, as the plea decision ultimately rests with the defendant, who may or may not agree with their attorney about their likelihood of conviction at trial. If, as previously discussed, a participant's perception differed from that which they were given by their lawyer, the expected trial value predicted by the shadow model would change as well. Thus, for participants who disagreed with their attorney about their POC, we recalculated their model adherence by assessing whether the length of the jail sentence that they indicated they would be willing to accept exceeded the product of their *self-reported POC* and the maximum TS associated with their experimental condition. This served as our second measure of adherence to the shadow model and is represented in eqs. 3 and 4 below. These calculations are identical to those in eqs. 1 and 2 with the caveat that POC at trial based on experimental condition is now replaced by self-reported probability of conviction at trial (SRPOC)[2]:

[1] This question presented the participant with three options: (1) accept a plea offer if it is the value they specified or lower, (2) accept the plea offer even if it higher than the value they specified, or (3) reject the plea offer regardless of the length of the jail sentence. These categories were subsequently collapsed into a dichotomous measure of whether or not the participant was willing to accept a plea offer versus unwilling to accept any plea deal.

[2] Note that participants who indicated that they were unwilling to accept a plea deal of any value were coded as deviating from the shadow model regardless of whether they disagreed with their attorney.

Table 1 Maximum predicted SOT value by condition

	Trial sentence		
POC	5 months	15 months	25 months
10%	Condition 1	Condition 2	Condition 3
	0.5 months	1.5 months	2.5 months
50%	Condition 4	Condition 5	Condition 6
	2.5 months	7.5 months	12.5 months
90%	Condition 7	Condition 8	Condition 9
	4.5 months	13.5 months	22.5 months

$$\mathrm{PS} > \mathrm{SRPOC} \times \mathrm{TS} = \text{Deviation from model} \tag{3}$$

$$\mathrm{PS} \leqq \mathrm{SRPOC} \times \mathrm{TS} = \text{Adherence to model} \tag{4}$$

Finally, in providing a comprehensive test of the shadow model and to mimic plea negotiations more realistically, a subset of eligible participants was given a second opportunity to make a plea decision. Participants who indicated that they would be willing to spend an amount of time in jail less than the maximum allowable sentence under the shadow model were presented with a counteroffer. The counteroffer was simply the maximum allowable sentence predicted under the shadow model for the participant's experimental condition and is similar to the procedures used by Bartlett and Zottoli (in press). This was our third measure of shadow model adherence and was constructed to determine how well the shadow model predicted plea decisions across a larger spectrum of potential values. In other words, we were interested to see whether a participant who indicated willingness to accept a plea deal well-below the maximum predicted value based on the shadow model would still be willing to accept a plea deal at the maximum predicted value.

Covariates Given the lack of knowledge regarding the factors that may influence shadow model deviation, we included several covariates that may be salient during the plea decision-making process. These included demographic items, questions about their experience with and trust in the criminal justice system, trust in the police, and both their personal and vicarious experience with DUI and traffic stops. Trust in the criminal justice system and trust in the police were measured on 7-point Likert scales. A test of the association between these two measures indicated that they were highly correlated, r (245) = .732, $p < .0001$. Thus, they were averaged together, with the resulting overall level of trust in the sample being fairly moderate ($M = 3.53$, SD = 1.32). Level of experience with the criminal justice system was measured on a 5-point Likert scale, and our overall sample was fairly moderate on this measure as well ($M =$ 2.23, SD = .917). Correlations between all covariates can be seen in Table 2.

Table 2 Independent variable correlation matrix

Variable	1	2	3	4	5	6	7	8	9	10	11
1. POC											
2. Trial sent	.02										
3. Numeracy	− .01	.06									
4. Trust	.01	− .09	− .11								
5. CJS exp.	.02	− .03	− .17*	.25**							
6. Pulled over: No reason	− .06	.02	− .02	− .18**	.20**						
7. Pulled over: DUI	.04	.01	− .03	− .01	.14*	.45**					
8. Race	.13*	.02	− .12	− .17**	.01	.10	.06				
9. Sex	.05	− .05	− .05	.02	− .06	− .07	.03	− .08			
10. Age	.02	− .03	.13*	.14*	.01	.04	.06	− .28**	.11		
11. Education	− .02	− .03	− .02	.21**	.15*	− .15*	.00	− .06	− .16*	− .06	
12. Income	.10	− .04	.00	.18**	.16*	− .07	− .08	.03	− .13*	− .03	.34**

Pearson's correlation coefficients were calculated for variables treated as continuous and phi coefficients were calculated for the association between dichotomous measures. All correlations were rounded to two decimal places. A single asterisk (*) indicates $p < .05$. A double asterisk (**) indicates $p < .01$

Results

Given the potential for differential plea willingness to influence shadow model adherence (as participants who were unwilling to plead to any value were considered deviants from the shadow model), we first examined plea rates across our experimental conditions, as shown in Table 3. For this purpose, plea willingness was defined as a binary measure of whether the participant indicated they would be willing to accept a plea deal of any value at the original offer (not the counteroffer). Plea acceptance was more frequent than non-acceptance across all POC and TS conditions, as well as across covariates. Specifically, acceptance rates ranged from 77.8 to 87.3% in POC conditions and 81.5% to 85.4% in maximum TS conditions. In contrast to our expectation, plea willingness did not significantly depend on the level of either variable (plea willingness and POC, $\chi^2(2, N = 245) = 3.143$, $p = .208$, and maximum TS, $\chi^2(2, N = 245) = .470$, $p = .791$).

We also examined plea decisions across all nine experimental conditions (not shown) and acceptance rates ranged from 68 to 96.2%. Due to low expected cell counts, a Fisher's exact test with Freeman-Halton extension was conducted for both POC (with TS as a layered variable), and TS (with POC as a layered variable). Results indicated that plea acceptance was not significantly dependent on any combination of experimental manipulations; POC and TS did not significantly interact for plea willingness. Additionally, participant's race, sex, and prior experiences with being pulled over by police were not significantly related to plea willingness (Table 3). Thus, based on the lack of dependence between experimental conditions and plea willingness, and the random assignment to experimental

Table 3 Plea willingness (categorical independent variables)

Statistic	N	Willing to accept plea ($n = 205$)	Not willing to accept plea ($n = 40$)	X^2
Probability of conviction	245			$X^2(2) = 3.143$, $p = .208$
10%	81	63 (77.8%)	18 (22.2%)	
50%	85	73 (85.9%)	12 (14.1%)	
90%	79	69 (87.3%)	10 (12.7%)	
Sentence at trial	245			$X^2(2) = .470$, $p = .791$
5 months	82	69 (84.1%)	13 (15.9%)	
15 months	81	66 (81.5%)	15 (18.5%)	
25 months	82	70 (85.4%)	12 (14.6%)	
Pulled over: no reason	245			$X^2(1) = .147$, $p = .701$
Yes	83	71 (85.5%)	12 (14.5%)	
No	162	134 (82.7%)	28 (17.3%)	
Pulled over: DUI	245			$X^2(1) = .121$, $p = .728$,
Yes	65	53 (81.5%)	12 (18.5%)	
No	180	152 (84.4%)	28 (15.6%)	
Race	245			$X^2(1) = .065$, $p = .799$,
White	158	131 (82.9%)	27 (17.1%)	
Non-white	87	74 (85.1%)	13 (14.9%)	
Gender	245			$X^2(1) = .195$, $p = .659$,
Male	130	107 (82.3%)	23 (17.7%)	
Female	115	98 (85.2%)	17 (14.8%)	

conditions, we did not include plea willingness as a predictor in our statistical models of shadow model adherence.

Shadow model adherence

Logistic regression analysis was used to assess the factors that predicted who adhered to the shadow model ("shadow followers") and who did not ("shadow deviators"). Separate models were constructed for each of our three dependent variables, all of which are presented in Table 4. Along with our main independent variables (POC, TS, and numeracy), we included all covariates described in Tables 1 and 2.[3] There were no substantive differences to the main findings in our models with and without these covariates.

[3] A post hoc power analysis indicated that our sample size for logistic regression models 1 and 2 was sufficient to detect an odds ratio of 1.68 (power = 0.92), considered equivalent to a Cohen's *d* of 0.2 or higher (see Chen et al. 2010).

Table 4 Logistic regression results for shadow model adherence

	Model 1 Based on experimental manipulations only		Model 2 Accounting for attorney disagreement		Model 3 after counteroffer	
	ß (SE)	Odds ratio	*ß* (SE)	Odds ratio	*ß* (SE)	Odds ratio
Probability (10%)						
50%	2.409*** (0.408)	11.122	1.530*** (0.376)	4.618	− 0.129 (0.559)	0.879
90%	3.105*** (0.467)	22.310	2.192*** (0.434)	8.951	− 0.675 (0.573)	0.509
Sentence (5 months)						
15 months	− 0.125 (0.403)	0.883	− 0.275 (0.388)	0.759	− 0.581 (0.442)	0.559
25 months	0.064 (0.410)	1.066	0.032 (0.397)	1.032	− 0.450 (0.443)	0.638
Numeracy	0.411* (0.173)	1.509	0.463** (0.167)	1.589	− 0.103 (0.191)	0.902
Trust	− 0.217 (0.138)	0.805	−0.199 (0.135)	0.820	0.319* (0.157)	1.375
CJS exp.	0.187 (0.198)	1.206	0.177 (0.193)	1.194	0.214 (0.212)	1.239
Pulled over: no reason	− 0.419 (0.416)	0.658	− 0.123 (0.398)	0.884	− 0.170 (0.467)	0.843
Pulled over: DUI	− 0.448 (0.421)	0.639	− 0.657 (0.401)	0.518	0.065 (0.452)	1.067
Non-white	− 0.085 (0.380)	0.919	0.077 (0.366)	1.080	− 0.071 (0.388)	0.932
Female	0.648^{+} (0.344)	1.912	0.734* (0.332)	2.083	0.419 (0.356)	1.520
Age	0.011 (0.016)	1.011	0.025 (0.016)	1.025	0.019 (0.016)	1.019
Education	− 0.268^{+} (0.143)	0.765	− 0.251^{+} (0.135)	0.778	0.004 (0.148)	1.004
Income	0.059 (0.187)	1.060	0.067 (0.181)	1.069	− 0.149 (0.184)	0.861
Constant	− 1.251 (1.111)	0.286	− 1.136 (1.087)	0.321	− 1.575 (1.240)	0.207
Observations	236		236		158	
Log likelihood	−114.870		−121.406		−98.149	
X^2	90.613***		63.274***		17.006	
Nagelkerke R^2	0.429		0.324		0.138	
Akaike inf. crit.	259.740		272.813		226.297	

$^{+}p < .10$; *$p < .05$; **$p < .01$; ***$p < 0.001$

Model 1 (shown in Table 4) examined the relationship between our predictor variables and shadow model adherence based on experimental manipulations. In total, 58% of participants fit the SOT model in model 1, and 42% of participants deviated from the model. The overall regression significantly predicted shadow model adherence, χ^2 (14, $N = 236$) = 90.613, $p < .001$, with POC and numeracy acting as significant individual predictors. Specifically, increases in POC resulted in increased shadow model adherence, with participants in the 50% POC conditions approximately 11 times as likely ($p < .001$), and participants in the 90% POC conditions over 22 times as likely ($p < .001$), to fit the shadow model than participants in the 10% POC conditions. Numeracy also predicted shadow model adherence, with a single one-unit increase in numeracy score leading to a 51% increase in the odds of fitting the shadow model ($p < .05$). The interaction term for POC and TS was non-significant when added to this model.

While model 1 suggests that POC and numeracy are meaningful factors in explaining shadow model adherence/deviation, it may be the participant's perception of POC that more accurately reflects their adherence. Thus, we recalculated shadow model adherence for all participants who indicated disagreement with their assigned POC condition. Overall, 25% of our sample disagreed with their attorneys' (or assigned) POC.

Model 2 (shown in Table 4) examines our predictor variables against this second dependent variable and can help determine whether the results from model 1 are an artifact of participant disagreement with their randomly assigned POC condition. In total, 65% of participants adhered to the SOT theory in model 2, while 35% of participants did not. The full regression model once again indicated that our covariates significantly predicted shadow model adherence in our sample, χ^2 (14, $N = 236$) = 63.274, $p < .001$. Results for individual predictors were similar to model 1, with POC, numeracy, and participant sex significantly related to shadow model adherence. POC remained positively associated with model adherence, with participants in the 50% POC conditions being just over 4.5 times as likely ($p < .01$), and participants in the 90% POC conditions being nearly 9 times as likely to fit the shadow model ($p < .001$), than participants in the 10% POC conditions. Numeracy remained positively associated with adherence as well, with a one-unit increase in numeracy score leading to nearly a 60% increase in the odds of fitting the shadow model ($p < .001$), and women were roughly two times as likely to fit the shadow model than men ($p < .05$). Thus, even after recalculating the maximum shadow values for participants who disagreed with their attorney regarding conviction probability, two of our three main independent variables (POC and numeracy) remained significant predictors of model adherence, while TS failed to reach statistical significance in both models.[4] Additionally, the interaction effect between POC and TS was not significant in this model.

Finally, we tested the shadow model at its maximum predicted values. That is, whether participants who indicated willingness to plead to values less than the maximum ones predicted by the shadow model would remain willing to plead when offered deals at the maximum values. We conducted a logistic regression model with shadow model adherence after the counteroffer, our third dependent variable (model 3 in Table 4). To reiterate, only those participants who indicated plea values less than the maximum value predicted by the shadow model received the counteroffer ($n = 158$), and in all cases the counteroffer was the maximum sentence predicted by the shadow model, based on the participant's assigned experimental condition. Adherence at this stage was defined as whether the subset of participants accepted or rejected the counteroffer; only 41% of participants adhered to the SOT theory at this stage. Results from model 3 indicated that the overall regression was not a significant improvement over the null model, χ^2 (14, $N = 158$) = 17.006, $p = .256$.

To explore possible explanations underlying the lack of significance in model 3, we plotted plea deviation rates by POC conditions between the initial plea decision and counteroffer decision (Fig. 1). At the initial plea offer phase (i.e., when participants indicated how many months in jail they would be willing to accept), over three quarters

[4] The results of models 1 and 2 remained unchanged when the invalid responses were included.

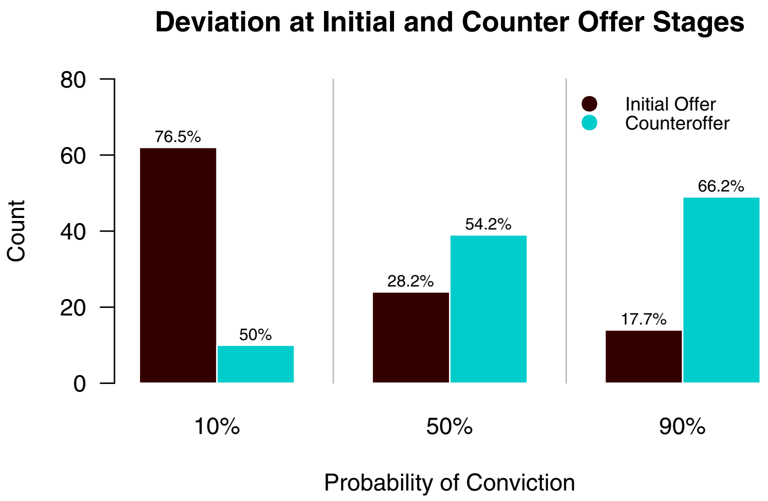

Fig. 1 Shadow model deviation at initial offer ($n = 245$) and counteroffer ($n = 158$)

of participants in the 10% POC conditions deviated from the shadow model (i.e., indicated a sentence higher than the maximum predicted by the shadow model or refused to accept any offer), while the majority of participants in the 50% and 90% conditions adhered to the shadow model.

Conversely, for those participants who received the counteroffer, approximately two-thirds deviated from the shadow model in the 90% POC conditions (i.e., rejected counteroffer) and just over half in the 50% conditions deviated from the model. Because the predominant deviation from the shadow model at the initial plea decision stage occurred in the 10% POC conditions, only 20 of these participants received the counteroffer and were evenly split between deviation and adherence. Thus, trends in model adherence across POC levels generally reversed between initial and counteroffer stages. Whereas none of our primary independent variables significantly predicted shadow model adherence during the counteroffer stage (model 3), these tests may have suffered from low sample size,[5] as only select participants satisfied the conditions to receive the counteroffer, particularly within 10% POC conditions.

Method and magnitude of shadow model deviation

As shown, POC significantly predicted shadow model deviation in our sample, with the majority of participants who deviated doing so in the 10% conditions. While it deserves reiteration that, in total, our test was conducted with a sample in which 83.7% of participants expressed willingness to accept a plea bargain, it remains possible that many of the participants who deviated from the shadow model were simply unwilling to accept a plea deal of any value. Thus, to more fully understand how participants deviated from the shadow model, we determined the specific method of deviation for each participant based on POC condition. We focused on the POC manipulation as it

[5] Estimated statistical power to detect an odds ratio of 1.68 decreased to 0.78 for model 3 due to sample attrition. This is a power level slightly lower than the convention of .80 proposed by Cohen (1992).

was the only core element of the shadow model that significantly predicted adherence in our regression models.

The possible deviation methods and descriptive statistics for each are shown in Table 5. Of the 100 participants who deviated from the original model, 60% indicated willingness to accept a plea value greater than would be predicted by the shadow model, while 40% deviated by refusing to accept any plea offer. In the 10% POC conditions, 44 of 62 participants (71%) deviated by accepting a plea value greater than would be predicted by the shadow model, and 18 of 62 participants (29%) deviated by refusing to accept any plea. Conversely, in the 90% POC conditions, 10 participants (71.4%) deviated by refusing to accept any plea, while only 4 participants (28.6%) deviated by accepting a plea value greater than SOT predictions. In the 50% POC conditions, deviation methods were evenly split.

Because most shadow deviants were still willing to accept a plea, we assessed how close to the maximum shadow values these participants were. Figure 2 plots the plea values for all participants in each assigned condition who indicated that they would accept a plea offer. The horizontal lines in each column represent the maximum predicted plea value under the shadow of the trial model, assuming a risk-neutral defendant. Thus, any points above the horizontal lines in each condition represent deviation from the shadow model.

As Fig. 2 shows, the deviation values across conditions are quite variable; some conditions appear tightly clustered around the maximum predicted plea value while others are more widely dispersed (e.g., 10%/15-month and 10%/25-month conditions). Specifically, in the 10% POC conditions, deviations ranged from 0.5 to 22.5 months greater than the maximum predicted plea value, with an average deviation of 4.64 months (SD = 5.96). In the 50% POC conditions, deviations ranged from 0.5 to 7.5 months above the maximum predicted value, with an average deviation of 3.2 months (SD = 2.26). Lastly, in the 90% POC conditions deviations ranged from 0.5 to 2.5 months above the maximum value, with an average deviation of 1.25 months (SD = 0.96). Thus, not only did the majority of participants deviate from the shadow model by accepting plea sentences that exceeded the maximum predicted shadow values but the degree of deviation (i.e., distance between plea value and maximum predicted shadow value) varied inversely with POC, whereby both deviation proportions and deviation magnitude were greatest in the 10% POC conditions. It should be noted, however, that the distribution of deviation amounts was highly skewed, and a Kruskal-Wallis rank sum test indicated that deviation ranks did not differ significantly by POC condition, χ^2 (2) = 2.47, p = .29.

Discussion

Although the shadow of the trial model has been described as the "predominant theory of plea bargaining at the individual level" (Bushway and Redlich 2012, p. 442), it has remained relatively untested in this capacity. Studies finding aggregate support for the model, coupled with significant individual-level deviation, have raised questions regarding the appropriateness of the shadow model as an individual-level explanation of plea decision-making (Bushway and Redlich 2012). In the current study, we sought to

Table 5 Methods for shadow model deviation by probability of conviction

	Probability level			
	10%	50%	90%	Total
Accepted plea above shadow prediction	44 (71%)	12 (50%)	4 (28.6%)	60
Refused to accept any plea	18 (29%)	12 (50%)	10 (71.4%)	40
Total	62	24	14	100

explore factors accounting for deviation from the shadow model among mock defendants; it is, to our knowledge, one of the first studies to do so. Our findings have important implications for the validity of the shadow model, future plea research, and criminal justice policy.

Regarding the main tenets of the shadow model, probability of conviction (POC) significantly predicted SOT model adherence, independent of trial sentence (TS), and controlling for an array of covariates. However, this relationship was inverse, such that deviation was highest in the low POC conditions (10%) and lowest in the high POC conditions (90%). While a critique of such findings may point to the random assignment of participants to POC conditions, these results held even after considering the subjective assessments of POC reported by participants who disagreed with their assigned condition. Furthermore, these results are consistent with those of Bartlett and Zottoli (in press), providing additional evidence of an inverse relationship between conviction probability and shadow model adherence. Our findings also indicated that participants who deviated from the model did so predominately by accepting plea deals that were higher than would be predicted by the SOT model, rather than simply refusing to accept any deal. The plea values indicated by these shadow model deviants varied widely from the expected SOT value in several conditions (see Fig. 2), suggesting that slight changes to the maximum predicted SOT value would not account for this variation.

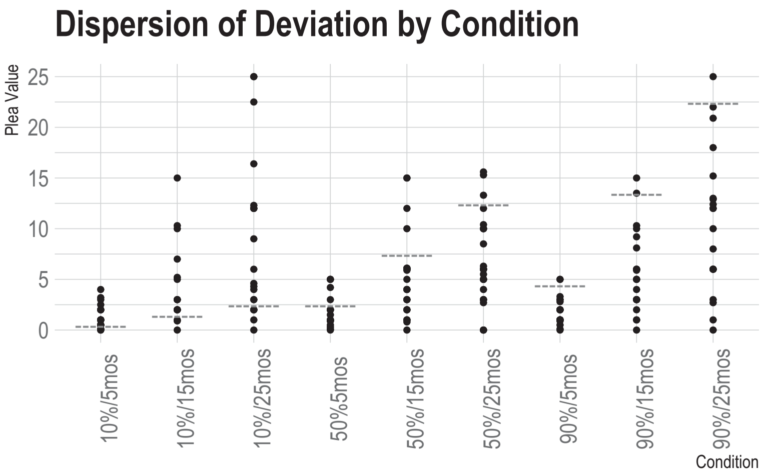

Fig. 2 Plea deviation and maximum shadow value by condition

The relationship between deviation and POC may seem counter-intuitive given findings from prior research suggesting that plea willingness is likely to increase in accordance with increasing POC (Bordens 1984; Gregory et al. 1978). However, it is important to emphasize that our findings are not the product of a lack of willingness to plead. Namely, plea willingness did not significantly increase as POC increased in our sample; plea willingness was high across all conditions, similar to plea rates in real cases. Thus, perhaps the most important finding is that participants in low POC conditions were willing to accept plea values higher than the shadow model would predict. As Bartlett and Zottoli (in press) suggest, the concept of nonlinear probability weighting may hold clues as to the mechanism behind this effect. In the formulation of prospect theory, Kahneman and Tversky (1979) found that individuals attach "psychological weight" (see Wu and Gonzalez 1996, p. 1676) to outcomes, and that such weights may be discrepant with stated probabilities. Specifically, probability weighting may function as an inversed S-shape such that low probabilities are overweighted and high probabilities are underweighted. Indeed, evidence has suggested that overweighting of low probabilities may occur up to probabilities as large as .4, well larger than the probability assigned to participants in our low POC conditions (see Gonzalez and Wu 1999; Wu and Gonzalez 1996).

Thus, it is entirely possible that participants in our study both overweighted POC in the 10% conditions and underweighted POC in the 90% conditions. If this is the case, then perceptions of risk may have moved closer to the middle of the distribution in both instances, perhaps explaining why the majority of shadow deviants in 10% POC conditions were willing to accept plea values higher than would be predicted by the shadow model. This potential is particularly problematic for the validity of the SOT theory in cases where POC approaches 0 or 1, given the theory's inherent assumption of probability linearity. However, it may also explain why, on average, the SOT model appears to hold, as cases with extremely low conviction likelihoods may occur infrequently in the real world.

It should also be noted that prospect theory predicts risk aversion when outcomes are framed as gains and risk-seeking when outcomes are framed as losses (Kahneman and Tversky 1979). When combined with nonlinear probability weighting, there is a predicted "risk aversion for most gains and low probability losses and risk-seeking for most losses and low probability gains" (Wu and Gonzalez 1996, p. 1676). Prior research has suggested that whether a plea is framed as a gain or loss has significant effects on plea acceptance (Garnier-Dykstra and Wilson 2019). While we made no attempt to frame our plea offer explicitly as a gain or a loss, it is possible that individual interpretations varied.

For example, in the 10% POC conditions, participants who interpreted the scenario in terms of losses may have viewed the decision as a choice between a 10% chance of losing their freedom for the full sentence (by risking trial), or the certainty of losing their freedom for a shorter period of time (via plea). With this interpretation, we might predict that participants would be risk averse, overweight the low probability of the loss (the 10% chance of conviction), and thus opt for the certainty of the plea.

Conversely, participants in the same 10% POC conditions who interpreted the situation in the domain of gains may have viewed the same decision as a choice between a 90% chance of gaining their freedom and the certainty of the plea deal. Here, nonlinear probability weighting may predict that these individuals will underweight the large probability of acquittal and thus still display risk aversion in opting for the plea. Either way, the key point is that individuals may not interpret probability in the way predicted by the shadow model, and as such their expected trial outcomes (if they are calculating such outcomes) will likely differ from those of the shadow model, particularly when conviction probabilities are extreme in either direction.

An alternative explanation for the inverse relationship between deviation and POC may simply involve differing amounts of acceptable error. That is, as POC increases, the likelihood of the shadow model being violated naturally decreases, because the POC dictates the range of values consistent with SOT predictions. For instance, in our 10%/25-month condition, the range of predicted SOT values only encompasses values of less than or equal to 2.5 months, whereas the 90%/25-month condition encompasses values of less than or equal to 22.5 months. Thus, higher POC conditions simply have more room for error, and if other factors influence the amount of time an individual is willing to accept in a plea deal, even by small increments, it may be enough to violate the assumptions of the SOT model in low POC conditions.

Related to this explanation, the majority of participants who received the counter-offer in our study (i.e., the maximum predicted trial value based on the SOT model) refused to accept it, despite being initially willing to accept a plea offer of a different value. In other words, the majority of our subsample did not fit the shadow model when forced to do so at the maximum predicted value, and this effect was primarily concentrated in the 90% POC conditions. Similar results were noted by Bartlett and Zottoli (in press), who found that average plea sentences became increasingly smaller percentages of the predicted SOT value as POC increased. This raises questions about the validity of the maximum plea values predicted by the SOT model. While the shadow model appears to hold in higher POC conditions, it may simply be the result of benefitting from a larger range of theoretically consistent plea values. Future research should continue to examine the validity of the upper bound of SOT model predictions.

In attempting to explain SOT model adherence/deviation, we examined participants' numeric abilities. We found that participants' numeracy was significantly and positively related to their model adherence. Participants may have attempted to determine expected trial outcomes as predicted by the shadow model, but that those predictions were contingent on mathematical ability. In other words, the shadow model seemingly assumes that individuals engaging in plea bargaining know their conviction probability and potential trial sentence (or are provided with educated predictions from their attorneys), and can then use those values to accurately calculate their trial value. Our results raise questions about this assumption and the possibility that individual deviation from the shadow model may, in part, be due to a lack of ability to perform the mathematical calculations required by the model. Scholars have discussed the role of education during plea negotiations in the past (Bibas 2004; Redlich et al. 2017). However, perhaps the key is not education in general, but mathematical abilities in particular. Indeed, numeracy does not seem to be a proxy for education, but rather a unique construct given their near-zero correlation in our sample ($r = -.02$). This is particularly salient to the SOT theory given its mathematical basis, but may also be salient to measures of fairness and comprehension in plea bargaining more generally. The varying terms and types of sentences contained in plea agreements likely require corresponding levels of understanding with respect to numbers, the conversion of units of time (e.g., months to years), perception of risk, etc. Future research on both the SOT model and plea bargaining comprehension should seek to incorporate measures of numeracy.

Limitations and future research

The results of this study must be considered alongside its limitations. First is our sample, which we recognize may be considered small by some. However, the concern with small sample sizes is generally related to issues of statistical power, or the probability of finding a true effect if present (Kraemer and Thiemann 1987). In theory, the smaller the sample, the larger the error between the sampling distribution and population, and thus the harder it is to find a statistically significant effect (Weisburd and Britt 2014). It is therefore meaningful that we found significant relationships despite the smaller sample size. Furthermore, a post hoc power analysis indicated that we had sufficient probability of finding a small effect (power = 0.92 for finding an odds ratio of 1.68) for our main regression models (see Chen et al. 2010; Cohen 1992).

A related limitation is the use of an MTurk sample. While the use of such a sample may provide a more diverse collection of participants than can be obtained administering surveys to, for example, university students, our participants were still predominately white and educated. However, we designed our vignette to describe a fairly common and relatable situation. Self-report studies have found that whites with higher education and/or higher household income may be the most likely to engage in drinking and driving (Fan et al. 2019; Oh et al. 2020), and DUI arrests appear to be concentrated among white males (Federal Bureau of Investigation 2018). In this regard, our sample may have overrepresented females (47%). Yet, in recent decades, the proportion of female DUI arrestees has more than doubled (FBI 2018; Gregory 2013), and we did not find significant differences between males and females on willingness to plea in our sample. Moreover, given our results concerning numeracy, and the possibility that many defendants in the real world may score lower on this measure than our sample, our findings may be conservative.

Our sample may have differed from real-world criminal defendants in other ways, however, including characteristics such as mental illness, substance abuse, and cognitive deficits. The presence of such characteristics, which we did not measure, may lead defendants to make decisions that seemingly violate rational models of plea decision-making, and as such could be contributing factors to our results. Future studies, particularly ones that test the SOT model with different crime types and sentences, may benefit from considering these characteristics and their influence on model adherence.

We were also interested in testing the baseline shadow model assuming a risk-neutral defendant (see Bushway et al. 2014), and as such, we did not measure risk-seeking. Research has suggested that the relationship between probability of conviction and maximum plea sentence is mediated by risk preferences (see Bjerk 2008). Thus, it is certainly possible that variation in model adherence resulted from varying risk preferences among participants, such that risk-averse participants may have been willing to accept plea offers higher than the baseline SOT model would predict. Differing risk preferences may also lead to plea decisions that violate the baseline

SOT model but are not necessarily irrational. Given that an individual's risk preference may change the expected value of the trial under the SOT model, for some defendants, accepting plea values higher than the baseline SOT model would predict may still be rational decisions (see Bushway et al. 2014).

It is also possible that risk preferences are influenced by numeracy, and indeed prior research has found more accurate emotional responses to risk among those with higher numeracy (Petrova et al. 2014). While we did provide controls for these and other measures that have been linked to risk-seeking behavior, such as age (Mather et al. 2012) and gender (Eckel and Grossman 2008), and randomized participants to conditions (which should have reduced the chance of any particular group differing significantly on risk-seeking tendencies), future studies would benefit from including comprehensive measures of risk-seeking.

Additionally, the current study does not include a manipulation of guilt/innocence. Our hypothetical was intentionally designed to allow participants to make their own determinations of guilt in an attempt to improve experimental realism. It is possible that explicitly innocent and guilty defendants follow different patterns of shadow model adherence. Given research suggesting that this distinction creates unique effects for plea decision-making more generally (Dervan and Edkins 2013; Garnier-Dykstra and Wilson 2019; Tor et al. 2010), the incorporation of such a manipulation would be an important addition to future studies. We also acknowledge that, despite our attempts to construct a realistic scenario, our participants did not have to live with the repercussions of their decisions, and thus it may have been easier for them to accept or reject plea offers knowing they would not have to deal with the real-world consequences of these decisions.

Overall, much remains unknown about how criminal defendants weigh decisions to plead guilty or go to trial. While the shadow of the trial model has been shown to be a valid explanation in the aggregate (Bonneau and McCannon 2019; Bushway and Redlich 2012), its utility for individual-level explanation remains under-studied. There is now growing evidence suggesting that probability of conviction alone predicts individual adherence to the shadow model, but that probability is not interpreted in a strictly rational fashion (Bartlett and Zottoli in press). Additionally, our findings open the door to the possible effects of an individual's mathematical ability on their plea decision-making. Additional research is needed, but these findings provide preliminary evidence that we must move toward a more nuanced understanding of individual-level plea decisions.

Acknowledgments We would like to thank Tina Zottoli and Jennifer Bartlett for sharing their work with us and for their continued communication and insight throughout the course of this study.

References

Bartlett, J., & Zottoli, T. (in press). The paradox of conviction probability: mock defendants want better deals as risk of conviction goes up. *Law and Human Behavior*.

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

Bjerk, D. (2008). Glass ceilings or sticky floors? Statistical discrimination in a dynamic model of hiring and promotion. *The Economic Journal, 118*(530), 961–982. https://doi.org/10.1016/j.irle.2007.12.005.

Bonneau, D., & McCannon, B. C. (2019). Bargaining in the shadow of the trial? Deaths of law enforcement officials and the plea bargaining process. https://doi.org/10.2139/ssrn.3457809.

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. (2019). Defendant decision-making in plea bargains. In V. A. Edkins & A. D. Redlich (Eds.). *A system of pleas: social sciences contributions to the real legal system* (pp. 24–36). Oxford University Press.

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.

Cheng, K. K.-y., & Chui, W. H. (2015). Beyond the shadow-of-trial: decision-making behind plea bargaining in Hong Kong. *International Journal of Law, Crime and Justice, 43*(4), 397–411. https://doi.org/10.1016/j.ijlcj.2014.10.001.

Clatch, L. (2017). Shining a light on the shadow-of-trial model: a bridge between discounting and plea bargaining note. *Minnesota Law Review, 102*(2), 923–968.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155.

Dervan, L. E., & Edkins, V. A. (2013). The innocent defendant's dilemma: an innovative empirical study of plea bargaining's innocence problem criminal law. *Journal of Criminal Law and Criminology, 103*, 1–48.

Eckel, C. C., & Grossman, P. J. (2008). Men, women and risk aversion: experimental evidence. *Handbook of Experimental Economics Results, 1*, 1061–1073. https://doi.org/10.1016/S1574-0722(07)00113-8.

Fan, A. Z., Grant, B. F., Ruan, W. J., Huang, B., & Chou, S. P. (2019). Drinking and driving among adults in the United States: results from the 2012–2013 national epidemiologic survey on alcohol and related conditions-III. *Accident Analysis & Prevention, 125*, 49–55.

Federal Bureau of Investigation (2018). *Crime in the United States, 2018*. Retrieved (September 2020), from (https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/persons-arrested).

Garnier-Dykstra, L. M., & Wilson, T. (2019). Behavioral economics and framing effects in guilty pleas: a defendant decision making experiment. *Justice Quarterly*, 1–25. https://doi.org/10.1080/07418825.2019.1614208.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*(1), 129–166. https://doi.org/10.1006/cogp.1998.0710.

Gregory, T. (2013). DUI demographics point to higher mix of women. Chicago tribune. Retrieved from https://www.chicagotribune.com/news/ct-xpm-2013-09-12-ct-met-dui-demographics-20130912-story.html

Gregory, W. L., Mowen, J. C., & Linder, D. E. (1978). Social psychology and plea bargaining: applications, methodology, and theory. *Journal of Personality and Social Psychology, 36*(12), 1521–1530. https://doi.org/10.1037/0022-3514.36.12.1521.

Helm, R. K., Hans, V. P., Reyna, V. F., & Reed, K. (2020). Numeracy in the jury box: numerical ability, meaningful anchors, and damage award decision making. *Applied Cognitive Psychology, 34*(2), 434–448.

Hill, W. T., & Brase, G. L. (2012). When and for whom do frequencies facilitate performance? On the role of numerical literacy. *The Quarterly Journal of Experimental Psychology, 65*(12), 2343–2368. https://doi.org/10.1080/17470218.2012.687004.

Johnson, B. D. & Richardson, R. (2019). Race and plea bargaining. In V. A. Edkins & A. D. Redlich (Eds.) *A system of pleas: social sciences contributions to the real legal system* (pp. 83–106). Oxford University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47*(2), 263–291. https://doi.org/10.2307/1914185.

Kraemer, H. C., & Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, Calif: Sage.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*(1), 37–44. https://doi.org/10.1177/0272989X0102100105.

Mather, M., Mazar, N., Gorlick, M. A., Lighthall, N. R., Burgeno, J., Schoeke, A., & Ariely, D. (2012). Risk preferences and aging: the "certainty effect" in older adults' decision making. *Psychology and Aging, 27*(4), 801. https://doi.org/10.1037/a0030174.

Mnookin, R. H., & Kornhauser, L. (1979). Bargaining in the shadow of the law: the case of divorce. *The Yale Law Journal, 88*(5), 950–997. https://doi.org/10.2307/795824.

National Center for State Courts (n.d.). *Trends in state courts.* Retrieved from https://www.ncsc.org/trends

Oh, S., Vaughn, M. G., Salas-Wright, C. P., AbiNader, M. A., & Sanchez, M. (2020). Driving under the influence of alcohol: findings from the NSDUH, 2002–2017. *Addictive Behaviors, 108*, 106439. https://doi.org/10.1016/j.addbeh.2020.106439.

Petrova, D. G., Van der Pligt, J., & Garcia-Retamero, R. (2014). Feeling the numbers: on the interplay between risk, affect, and numeracy. *Journal of Behavioral Decision Making, 27*(3), 191–199. https://doi.org/10.1002/bdm.1803.

Pezdek, K., & O'Brien, M. (2014). Plea bargaining and appraisals of eyewitness evidence by prosecutors and defense attorneys. *Psychology, Crime & Law, 20*(3), 222–241. https://doi.org/10.1080/1068316X.2013.770855.

Redlich, A. R. & Edkins V. A. (2019). Moving forward in a system of pleas. In V. A. Edkins & A. D. Redlich (Eds.). *A system of pleas: social sciences contributions to the real legal system* (pp. 187–197). Oxford University Press.

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.

Schneider, R. A., & Zottoli, T. M. (2019). Disentangling the effects of plea discount and potential trial sentence on decisions to plead guilty. *Legal and Criminological Psychology, 24*(2), 288–304. https://doi.org/10.1111/lcrp.12157.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine, 127*(11), 966–972.

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://onlinelibrary.wiley.com/doi/abs/10.1111/j.1740-1461.2009.01171.x.

United States Sentencing Commission (2018). *2018 federal sentencing statistics.* Retrieved from https://www.ussc.gov/research/data-reports/geography/2018-federal-sentencing-statistics

Weisburd, D., & Britt, C. (2014). *Statistics in criminal justice* (4th ed.). New York: Springer.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science, 42*(12), 1676–1690 .https://pubsonline.informs.org/doi/abs/10.1287/mnsc.42.12.1676.

Wright, R. F., Roberts, J., & Wilkinson, B. (2020). The shadow bargainers. Cardozo Law Review*, Forthcoming*. Available at SSRN: https://ssrn.com/abstract=3577322

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Kevin Petersen is a doctoral student in the Department of Criminology, Law and Society at George Mason University. After receiving his M.S. in Criminal Justice, he worked as a victim/witness coordinator for the Roanoke County (VA) Commonwealth's Attorney's Office. His research interests include policing, crime and place, evidence-based crime interventions, guilty pleas, and legal decision-making.

Allison D. Redlich is a professor in the Department of Criminology, Law, and Society at George Mason University. Her research is concerned with miscarriages of justice, both wrongful convictions and whether legal decision-making is knowing, intelligent, and voluntary. More specifically, she studies admissions made during interrogations and guilty pleas. She is currently President of the American Psychology-Law Society, a division of the American Psychological Association.

Robert J. Norris is an assistant professor in the Department of Criminology, Law, and Society at George Mason University. His research explores miscarriages of justice, social change and legal reform, and decision-making in the criminal legal process. He is the author of several articles and books dealing with wrongful convictions, including *Exonerated: A History of the Innocence Movement* (NYU Press, 2017), as well as criminal admissions and public opinion.