

Charged Up and Anchored Down: A Test of Two Pathways to Judgmental and Decisional Anchoring Biases in Plea Negotiations

Stephanie A. Cardenas

John Jay College of Criminal Justice, City University of New York
Department of Psychology, Williams College

Prosecutors sometimes threaten defendants with disproportionately severe potential trial sentences (PTS) to coerce their plea acceptance. Three experiments tested whether the PTS act as an anchor that biases evaluations of the plea deal at deep (vs. shallow) processing level by encouraging asymmetric consideration of information consistent—but not inconsistent—with the appropriateness of the PTS. I expected that a high (vs. moderate) PTS would lead to greater plea acceptance by increasing maximum acceptable sentencing preferences and influencing evidence appraisals. I tested two anchor debiasing strategies: selective consideration of anchor-inconsistent evidence (consider-the-opposite strategy [COS]) and holistic consideration of all evidence (via process accountability). Participants read detailed information about their criminal case and the PTS (high or moderate) before indicating their maximum sentencing preferences, evaluating case attributes, and rendering a plea decision. Mock defendants (Experiment 1; $N = 479$; $d_{\text{relevant}} = 0.83$), defense attorneys (Experiment 2; $N = 155$; $d_{\text{pre-COS}} = 1.19$), and mock attorney–client pairs (Experiment 3; $N_{\text{pairs}} = 265$; $d_{\text{unaccountable}} = 0.60$) were willing to accept (or recommend) longer sentences when presented with a higher (vs. moderate) PTS—even when the PTS was irrelevant to their case (Experiment 1; $d_{\text{irrelevant}} = 0.64$). Mediation analyses across all studies demonstrated that higher maximum sentencing preferences were, in turn, associated with greater willingness to accept a plea deal, but not with biased evaluations of evidence strength. Selective consideration of evidence inconsistent with the high PTS attenuated anchoring effects on sentencing preferences (Experiment 2; $d_{\text{post-COS}} = 0.75$) and plea recommendations but discussing more evidence did not (Experiment 3; $d_{\text{accountable}} = 0.68$). This work challenges the law’s assumption that defense attorneys can debias defendants’ cognitive errors and preliminarily proposes easy-to-implement debiasing solutions to overcharging.

Keywords: anchoring, motivated information processing, expertise, plea bargaining, judgment, decision-making

Supplemental materials: <https://doi.org/10.1037/law0000390.supp>

The Sixth Amendment guarantees criminal defendants the right to trial. Yet over 94% of federal and state convictions result from guilty pleas (Gramlich, 2019; United States Sentencing Commission, 2022), prompting the U.S. Supreme Court to acknowledge that “criminal justice today is for the most part a system of pleas, not a system of trials” (Lafley v. Cooper, 2012, p. 3). To explain the rise of plea-bargaining, legal scholars and practitioners have implicated the combined threats

of mandatory minimum sentences and *strategic overcharging* (e.g., Petegorsky, 2012; Standen, 1993)—an unethical practice that some prosecutors use to threaten defendants who exercise their constitutional right to trial with criminal charges and potential trial sentences (PTS) that are disproportionately severe relative to the defendants’ alleged misconduct (e.g., Alschuler, 1968; Bibas, 2004; Caldwell, 2011; Reimer & Sabelli, 2019). Overcharging has garnered widespread

This article was published Online First May 11, 2023.

Stephanie A. Cardenas  <https://orcid.org/0000-0002-1148-2573>

Portions of this article are adapted from Stephanie A. Cardenas’ doctoral dissertation under the supervision of Dr. Saul Kassin (Chair), Deryn Strange, Steven Penrod, Vanessa Edkins, and Emily Balcetis. This research was supported by a National Science Foundation GRFP Grant 1646736, the Society for Personality and Social Psychology (Heritage Dissertation Award), the American Psychology-Law Society (Student Grant-in-Aid), and CUNY (Doctoral Student Research Grant and Early Research Initiative Catalyst Grant). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Portions of this work were presented at the 2021 annual meeting of the European Association for Psychology and Law, virtually, and the 2022 annual meetings of the American Psychology and Law Society in Denver, CO. Thank you to the invaluable help of research assistants who acted as experimenters and helped code mock attorney–client conversations: Kristen Akin, Marrayah Borcena,

Jonathan Dixon, Briana Irwin, Janna Jensen, Devon Kaat, Min Jeong Kim, Nikki Prascak, Simran Sohal, Merusha Mukherjee, David Sokol, Olivia Winton, Yi-Chen Wu, and Klaudia Zuraw.

Data (Experiment 1: <https://osf.io/djp2a>; Experiment 2: <https://osf.io/4twmu>; Experiment 3: <https://osf.io/6bd7r>), study analysis code (Experiment 1: <https://osf.io/j4ft2>; Experiment 2: <https://osf.io/4my3p>; Experiment 3: <https://osf.io/qsv76>), preregistrations (Experiment 1: <https://osf.io/6ng2m>; Experiment 2: <https://osf.io/y3p2e>; Experiment 3: <https://osf.io/d5wt8>), and online supplemental materials and analyses (<https://osf.io/w24ex>) are publicly available on the OSF (Cardenas, 2021).

Stephanie A. Cardenas served as lead for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—original draft, and writing—review and editing.

Correspondence concerning this article should be addressed to Stephanie A. Cardenas, Department of Psychology, Williams College, 25 Stetson Ct, Williamstown, MA 01267, United States. Email: sac1@williams.edu

disapproval from practitioners and scholars alike (American Bar Association, 2016, Model Rule 3.8), including the late Justice Scalia, who maintained that “prosecutorial overcharging ... effectively compels an innocent defendant to avoid massive risk by pleading guilty to a lesser offense” (Lafler v. Cooper, 2012, p. 12). Legal scholars and psychologists have proposed that the threat of an excessively severe PTS serves to “anchor” defendants to the PTS by shifting the maximum sentence they would be willing to serve toward the PTS (Bibas, 2004; Redlich et al., 2017). However, *how* anchoring might exert its influence in plea contexts, where defendants and attorneys are motivated to obtain lower jail terms and render decisions in a socially dynamic negotiation environment, remains unknown.

Shadow of the Trial Model

Early plea research historically favored a rational actor model that assumes that defendants bargain in the *shadow of the trial* (SOT), that is, basing their plea decisions on the potential sentence if convicted at trial (PTS) and their probability of conviction (POC) (see Bibas, 2004; Bushway et al., 2014; Mnookin & Kornhauser, 1979). Specifically, the SoT model posits that a rational risk-neutral defendant facing a 10-year PTS and an 80% POC (based on evidence strength) should only settle for a plea offer of 8 years (i.e., the expected value of the trial) or less. At the aggregate level, archival and experimental data demonstrate that the PTS and POC account for much of the variance in plea decisions (e.g., Bushway et al., 2014; Kramer et al., 2007; Redlich et al., 2016).

Simultaneously, critics have argued that in its simplicity the shadow model neglects other variables including “structural distortions” (e.g., pretrial detention; Edkins & Dervan, 2018), “psychological pitfalls” (see Redlich et al., 2017), and individual differences (e.g., defendant race, see Johnson & Richardson, 2019) that may explain case-level variation in plea decisions. Indeed, applying basic psychological theory to plea decision-making—such as discounting (Clatch & Borgida, 2021; Edkins & Dervan, 2018), framing (Garnier-Dykstra & Wilson, 2021), social influence (Henderson & Shteynberg, 2020;), and fairness (Khogali et al., 2018; Tor et al., 2010)—has yielded critical new insights about how defendants make decisions that often diverge from SOT model predictions.

Recent work demonstrating how key variables in the SoT model lead to results not predicted by SoT are of particular relevance. For example, although empirical research generally supports the notion that more severe PTS lead to higher guilty plea rates (e.g., Bordens, 1984; Kramer et al., 2007; McAllister & Bregman, 1986; Zimmerman & Hunter, 2018; c.f. Dervan & Edkins, 2013; Petersen et al., 2022), recent work disentangling the influence of the PTS from the plea discount has found otherwise. Holding the percentage of the plea discount constant, Schneider and Zottoli (2019) found a modestly increased guilty plea rate among guilty and innocent mock defendants’ facing a 5- versus 25-year PTS. Field studies have found similar results suggesting that increasing charge severity (and sentence) was associated with decreased plea acceptance (Albonetti, 1990; Meyer & Gray, 1997). Thus, inconsistencies suggest that further work is required to examine whether subjective evaluations of the severity of the charges influence how decision-makers evaluate their case and their willingness to plead.

Recent work examining the influence of POC on plea deal preferences also expands on the SoT model. Using the framework of Prospect Theory (Kahneman & Tversky, 1979), Bartlett and Zottoli

(2021) demonstrated that mock defendants may become *more* risk-seeking as their POC increases, leading them to desire better plea-deal jail terms (relative to the expected value of the trial) as a condition to pleading guilty (for consistent results, see Petersen et al., 2022). The authors note that though these findings are consistent with the SoT model, they offer a valuable addition by enabling more precise estimations of the probability of plea acceptance. This is achieved by incorporating additional information about a defendant’s *subjective* evaluations of their chances of being acquitted, beyond what is accounted for by the traditional SoT model. Building on Bartlett and Zottoli’s findings, the current work examined how the experience of being overcharged may lead to anchoring effects that influence defendants’ subjective evaluations of their likelihood of acquittal (among other appraisals relevant to the plea decision) within the SoT framework.

Anchoring Mechanisms Biasing Legal Decision-Making

Plea negotiations occur under uncertain conditions. Defendants lack information about many factors relevant to their decision to accept or reject plea offers: How strong is the mitigating evidence (such as a first-time offense)? What is the chance of being found guilty at trial? In their pioneering work, Tversky and Kahneman (1974) demonstrated that when people make judgments about unfamiliar topics, any initially available salient value serves as an *anchor* point that can unduly bias those judgments. Following Tversky and Kahneman, anchoring theorists established the robustness of anchoring effects across varied domains (Klein et al., 2014) and proceeded to study how anchors result in judgmental biases (Furnham & Boo, 2011). For instance, proponents of the early *numeric priming hypothesis* suggest that any relevant or irrelevant number in the immediate environment can serve as a reference point, such that people may be willing to pay more for a meal at a restaurant called “Studio 97” than one called “Studio 17” (Cricher & Gilovich, 2008).

More recent theories of anchoring are proposed to occur either at the *deep* or the *shallow* level of cognition (Bystranowski et al., 2021; Wegener et al., 2010). At the *deep* level of cognition, anchoring may result in *or* be the cause of attitudes and judgments that have become assimilated to the anchor. For example, the dominant explanation of anchoring—the *Selective Accessibility Model* (SAM)—suggests that when people consider an anchor they automatically and thoughtfully engage in a selective search of information *consistent* with the hypothesis that the anchor represents the “correct” value (Chapman & Johnson, 1999; Mussweiler & Strack, 1999a, 1999b, 2000, 2001; Strack & Mussweiler, 1997). To demonstrate, English et al. (2006, Study 4) found that legal experts categorized *inculpatory* information faster when exposed to a high rather than low sentencing demand (9- vs. 3-month probation). As “positive hypothesis testing” increases the availability of information *consistent* with the implications of the anchor, people are more likely to base their numeric judgments on a limited information set that excludes anchor-*inconsistent* information (see Strack et al., 2016).

In plea contexts, if anchoring occurs at the *deep* level of cognition defendants facing an excessive PTS might recall more inculpatory information (suggesting the plausibility of the PTS) while failing to consider exculpatory information, thereby biasing their interpretations of the evidence and, in turn, their POC. As defendants and attorneys base their plea decision primarily on the perceived strength of the evidence (Kramer et al., 2007; Redlich et al., 2016), it would be detrimental for them to incorrectly infer from an excessively high

PTS that the prosecution's case against them must be exceptionally strong and that the PTS must be warranted by the evidence.

Alternatively, anchoring biases may result from processes occurring at the *shallow* level of cognition. For instance, the *scale distortion hypothesis* explains anchoring as a distortion of the response scale used for "mapping" judgments rather than a distortion of the underlying mental representation (e.g., beliefs, attitudes, and judgments) of the target value (Mochon & Frederick, 2013, p. 1). Frederick and Mochon (2012) note that the scale distortion model is more appropriate for explaining anchoring that occurs in contexts where people lack domain-relevant knowledge, whereas the SAM is more likely to explain anchoring that occurs in contexts where people have access to an abundance of relevant information. In the current studies, the factually complex experimental stimuli and recruitment of expert attorney decision-makers enabled a test of whether anchors would result in a biased interpretation of the case facts.

In sum, these two conceptual pathways to judgmental biases suggest that anchoring may comprise *several* phenomena characterized by superficially similar outcomes (e.g., assimilation to a numerical anchor) but different non-mutually-exclusive psychological explanations (Epley & Gilovich, 2005, 2006; Turner & Schley, 2016). Understanding how the PTS may lead to anchoring in plea contexts is critical, as the deep and shallow pathways to anchoring lead to different predictions about the causes and consequences of judgmental and decisional biases. Thus, a primary goal of the current work was to measure and debias the influence of the PTS on maximum sentencing preferences, inferences drawn about the case facts, and plea decisions. As debiasing interventions that correct for the exact mechanisms that cause the bias are most effective (Galinsky & Mussweiler, 2001), each experiment was designed to use different measures, methods, and samples to identify the anchoring mechanism(s) that most closely corresponded to the unique circumstances of plea negotiations.

Anchoring of Joint Preferential Judgments in a Plea Negotiation Context

The plea-bargaining context differs from the environment explored in prior anchoring studies on legal decision-making in four critical ways, making it a conceptually rich and novel context in which to examine these alternative pathways to anchoring. First, a recent meta-analysis by Bystranowski et al. (2021) found that studies on anchoring in legal contexts have focused primarily on civil (vs. criminal) cases in which respondents provided monetary damages rather than sentencing preferences (e.g., Hans et al., 2018). This is significant because anchoring effects tend to be consistently larger, though not significantly so, in civil ($k_{\text{effect_sizes}} = 43, d = 1.00$ 95%CI [0.64, 1.36]) compared to criminal cases ($k_{\text{effect_sizes}} = 13, d = 0.79$ 95%CI [0.51, 1.07]) (Bystranowski et al., 2021).

Second, the anchoring literature has focused almost exclusively on the biasing influence of anchors on *numerical* judgments. Research examining whether anchoring biases extend from biased *judgments* to biased *choices* is limited (rare exceptions exist in marketing studies examining willingness to purchase goods; Fudenberg et al., 2012). Thus, one goal of the current work was to examine whether anchors lead decision-makers to assimilate their choices (plea decisions) toward the anchor (PTS) through biased judgments (maximum acceptable sentence).

Third, unlike participants in previous anchoring studies, the parties in plea negotiations have vested interests in particular *desired*

outcomes, not necessarily *factually accurate* outcomes. Judgments accompanied by self-relevant outcomes—especially negative outcomes—may be influenced by motivational biases that lead individuals to seek, evaluate, and recall information in self-serving ways (Kunda, 1990). Excessive trial sentences are, of course, highly threatening to defendants and their attorneys. Defendants who prefer to reduce their jail time may be more naturally inclined to take the *deeper* processing pathway to anchoring. Yet whether anchoring occurs when rendering preferential judgments is still the subject of debate, with studies in the consumer literature finding attenuated effects (Fudenberg et al., 2012) and no effects (Yoon et al., 2019). Furthermore, a few studies have examined whether *negatively* valenced anchors bias self-relevant judgments. For example, Joel et al. (2017) found that anchors were ineffective when they suggested a high likelihood of a negative event occurring. Research by Pogarsky and Babcock (2001) found that civil award anchors were *less* effective when participants assumed the role of defendant compared to plaintiff. These studies, however, do not answer to what extent threatening anchors were ineffective because they triggered a motivated, effortful, and thorough search for *disconfirmatory* evidence. Defendants may be naturally predisposed to search for evidence inconsistent with the implications of a high PTS. Thus, the current work is the first to examine whether anchoring effects persist for legal decision-makers forming preferential judgments and decisions.

Fourth, defendants seldom make legal decisions unilaterally, instead relying on the recommendations of their legal counsel to arrive at plea decisions (see Henderson & Levett, 2019). Legal scholars have claimed that defense attorneys act as "debiasers" who guide defendants toward more rational decisions (e.g., Korobkin & Guthrie, 1997; Scott & Stuntz, 1992). The empirical literature unequivocally demonstrates that attorneys remain vulnerable to cognitive errors (Bystranowski et al., 2021). However, it remains unknown whether attorneys are less vulnerable than defendants to the anchoring bias in plea negotiations and, if so, whether they can mitigate its influence on defendant's plea decisions. Using a novel more ecologically valid paradigm, the current work is the first to examine how anchoring biases operate in the context of dynamic verbal information exchange between novice and expert legal actors.

The potential susceptibility of defendants *and* their attorneys to the anchoring bias suggests a need for a debiasing intervention that operates at the dyad level. If predictions derived from the SAM are true—that is, if individual group members exposed to a high PTS will selectively retrieve more *inculpatory* than exculpatory information—then groups may demonstrate a biased exchange of information (de Wilde et al., 2018).

Requirements for Debiasing the Anchoring Effect

Even when defendants and attorneys are motivated to obtain favorable *outcomes*, they may fail to engage in *processes* (e.g., search for anchor-disconfirming evidence) that might elicit such outcomes. Individuals made to be accountable for the *process*, rather than the *outcome* of their decision-making, expect to have to justify their beliefs to others (Tetlock, 1983). Process-accountable individuals are less biased when searching for information (Scholten et al., 2007; Ten Velden et al., 2010), more reluctant to accept unsupported alternatives (Davis et al., 1976), and make more critical self-evaluations, which, in turn, motivates more effortful and deeper processing that impacts the encoding of information (Tetlock, 1983). For example, de Wilde et al.

(2018) found that process-accountable groups (but not control groups) made sentencing decisions that did not significantly differ between those exposed to a high versus low anchor.

Whereas process accountability may debias the anchoring effect by increasing consideration of all the case evidence (c.f., Bibas, 2004, p. 2517), debiasing strategies, such as the *consider-the-opposite strategy* (COS), that selectively increase consideration of only anchor-disconfirming evidence are also likely to work (Adame, 2016; Downen et al., 2019; Stein & Drouin, 2018). For example, in a field study, Mussweiler et al. (2000) asked car sales experts to estimate the value of the latest model of a specific car. Experts who were first asked to list reasons for why a specific car price (high or low anchor) might be inappropriate were less influenced by the price than those not asked to list any reasons. Thus, a successful debiasing strategy motivates the search for information contradicting the PTS.

Overview of the Current Studies

Though legal scholars and psychologists have long suggested that overcharging may elicit powerful anchoring effects that bias plea offer evaluations and coerce defendants to plead guilty, no work has empirically investigated to what extent these claims may be warranted. The current studies were designed to collectively (a) demonstrate to what extent defendants and legal professionals are susceptible to the anchoring bias in plea negotiations, (b) investigate the mechanism underlying these anchoring effects (deep vs. shallow processing pathways), and (c) test the efficacy of two debiasing strategies focused on improving information processing of anchor-disconfirming information. Mock defendants (Experiment 1), practicing criminal defense attorneys (Experiment 2), and mock attorney–client pairs (Experiment 3) read a factually complex, hypothetical plea scenario containing either a *high* or *moderately* sized PTS. I adapted the structure of a case description used in prior plea research (Bartlett & Zottoli, 2021; FindLaw, 2018; Henderson & Shteynberg, 2020; Petersen et al., 2022) by supplementing it with 20 evidentiary circumstances. Materials were similar across all studies, with minor deviations to test for the effect of moderators that might attenuate the anchoring bias: a manipulation of the informational relevance of the anchor (Experiment 1), instructions to *selectively* generate anchor-inconsistent evidence (Experiment 2), and a process accountability induction to promote greater evaluation of *all* evidence (Experiment 3). In each study, legal decision-makers estimated the maximum sentence they would be willing to accept (or recommend, for attorneys) rather than risk a trial conviction, before deciding to accept or reject a plea offer. Notably, because prior work convincingly demonstrates that greater PTS and plea discounts influence plea acceptance (most prominently, Schneider & Zottoli, 2019; but also Bordens, 1984, Study 1; Gregory et al., 1978; McAllister & Bregman, 1986; Tor et al., 2010; Zimmerman & Hunter, 2018), the current work instead sought to examine whether the PTS influences plea decisions specifically by distorting discussion and evaluation of information as predicted by the leading theory of anchoring.

The current research assessed three overarching hypotheses applicable to all studies:

Hypothesis 1 (H1): A high PTS anchor will lead to greater plea acceptances compared to a moderate PTS anchor through

increased maximum acceptable sentencing preferences. Across all experiments, I also predicted an indirect effect of the magnitude of the PTS on plea decisions through maximum sentencing preferences.

Hypothesis 2 (H2): The magnitude of the anchor will influence discussion of the evidence and appraisals of evidence strength. The current work enabled a test of two pathways to judgmental anchoring: deep versus shallow levels of cognitive processing. If mock defendants approach the anchor at the *deep* level of cognition in thoughtful, selective consideration of evidence *consistent* with the PTS, then they ought to discuss and evaluate the strength of the mitigating and aggravating evidence differently in the high compared to moderate anchor magnitude conditions. If instead, still at the deep level of cognition, defendants engage in thoughtful but nonselective consideration of the anchor, then we would expect no significant difference in evidence evaluation across anchor magnitude conditions. Alternatively, if they approach the anchor at the *shallow* level of cognition, then their evidence evaluations should not differ when presented with a relevant compared to an irrelevant PTS. Furthermore, though the SAM proposes decision-makers will engage in a confirmatory search of anchor-*consistent* information (i.e., aggravating evidence), I proposed that the threatening implications of the PTS may cause decision-makers to instead engage in search of anchor-*inconsistent* information (i.e., mitigating evidence). Hence, I made a directionally nonspecific prediction about the influence of the magnitude of the anchor on the quantity (Experiment 3) and evaluation of aggravating and mitigating evidence (Experiments 1 and 3).

Hypothesis 3 (H3): Debiasing strategies that increase consideration of anchor-*inconsistent* evidence will reduce the anchoring bias. If decision-makers presented with a severe PTS selectively retrieve anchor-*consistent* information, then debiasing strategies should undermine the anchoring bias by (a) *decreasing* consideration of anchor-*consistent* information or (b) *increasing* consideration of anchor-*inconsistent* information. In Experiment 1, I predicted that making the PTS *informationally irrelevant* should discourage the selective search and consideration of anchor-*consistent* information, thereby attenuating the anchoring bias. In Experiment 2, I predicted that a selective search for anchor-*inconsistent* information should reverse the anchoring bias. In Experiment 3, I predicted that promoting a holistic search of anchor-*consistent* and *inconsistent* should prevent the anchoring bias from occurring altogether.

The present studies were approved by the institutional review board at John Jay College of Criminal Justice. In keeping with best science practices, all hypotheses, sample sizes, inclusion and exclusion criteria, materials, procedures, data preparation, and code for statistical analyses in R were preregistered and are publicly available on the Open Science Framework (OSF) at osf.io/h6ktg.

Experiment 1

Prior research has established that PTS, POC, and plea discount exert unique influences on plea decisions (e.g., Bartlett & Zottoli, 2021; Schneider & Zottoli, 2019). Experiment 1 complements this research by testing whether the PTS influences decisions via the

experience of *overcharging* and via *anchoring* rather than a cost/benefit analysis of the plea discount.

To ensure that the PTS was perceived as overcharging, the prosecutor threatened an excessively severe sentence that could subjectively be regarded as disproportionate to the alleged misconduct. Thus, overcharging here and in subsequent studies comprised a subjective evaluation of the severity of the PTS relative to the charge. To approximate the range of sentences that participants would subjectively evaluate as “severe,” “appropriate,” and “lenient,” a calibration sample ($N = 87$) read the case materials and indicated their anchor-unbiased maximum acceptable sentencing preferences. Results demonstrated that a 24-month sentence represented a value at the 95th percentile (“severe”) of the distribution of sentences, whereas the 4-month sentence represented a value at the 40th percentile (“appropriate”).

To isolate the effect of anchoring from the plea discount, the current work used two strategies. First, in all studies, mock defendants and attorneys were asked to estimate the longest plea-deal jail term (i.e., their maximum sentencing preference), and they would be willing to accept *before* learning the actual plea sentence offered by the prosecution (for similar procedure, see Bartlett & Zottoli, 2021). Thus, when estimating their maximum sentencing preference, participants had been exposed to the anchor (the PTS) but not the plea discount (as it is not possible to calculate the plea discount without knowing the offered plea sentence). Second, the *irrelevant* anchor condition of Experiment 1 stripped the PTS of its informational relevance by making the source of the PTS unrelated to the participant’s case, such that the PTS gave no meaningful information about the defendant’s expected trial sentence.

Method

Design and Power

Experiment 1 used a 2 (Anchor Magnitude: High [24 months] vs. Moderate [4 months]) $\times$ 2 (Anchor Informational Relevance: Irrelevant vs. Relevant) between-subjects factorial. An a priori power simulation using the *superpower* package (Lakens & Caldwell, 2021) in R (R Core Team, 2021) demonstrated that a total N of 480 participants was sufficient to detect a large-sized effect ($d = 0.69$) for the main effect of the anchor and a medium-sized effect ($d = 0.27$) for the main effect of relevance on with greater than .99 power ($\alpha = .05$). Effect sizes were based on Townson’s (2019) unpublished meta-analysis which found a Cohen’s d of 0.77 for anchoring effects in law contexts and a Cohen’s $d = 0.28$ for informational relevance of the anchor.¹ The simulation using predicted means and SD s based on pilot testing is available at <https://osf.io/p6nmf>. Regrettably, I did not calculate the sample size necessary to power the predicted Anchor by Relevance interaction. A sensitivity analysis conducted in G*Power 3.1.9.3 (Faul et al., 2007) indicates that a sample of $N = 479$ would provide 80% power to detect a small effect of $d_{\text{cohen}} = 0.25$ for a two-way analysis of variance, assuming a two-tailed test and an α of .05.

Participants

There were 561 participants who participated through Prolific.co and were compensated \$3.50. Participants were excluded for reporting inconsistent age information ($n = 3$), failing an instructional attention check ($n = 19$), failing to correctly report the anchor

value ($n = 43$), and providing a maximum acceptable sentence that had a value outside the interval formed by the median plus or minus 4.5 median absolute values ($n = 17$; see Leys et al., 2013). A detailed description of exclusion criteria for all studies is available in the [Supplemental Analyses](#). The final sample consisted of 479 participants (Relevant/Moderate Anchor: $n = 116$; Relevant/High Anchor: $n = 121$; Irrelevant/Moderate Anchor: $n = 124$; Irrelevant/High Anchor: $n = 118$) who were, on average, 32.76 years old ($SD = 12.51$; range 18–73 years); 47.49% identified as male and 47.18% as female, 2.92% as nonbinary; 72.02% identified as White, 13.99% as Asian, 11.27% as Hispanic/Latino, 8.35% as Black/African American, and 2% as Other.

Procedure

Hypothetical Plea Scenario. Participants read a 738-word description of the events leading up to their felony charge for driving under the influence (DUI) and the current discussion with their court-appointed attorney. Participants learned they had become inebriated at a party, crashed their vehicle, and were then arrested followed by the relevant criminal legal code. The vignette detailed 20 circumstances (10 aggravating, 10 mitigating) that could cause the court to impose either a harsher or more lenient penalty. Pretesting ($N = 84$) revealed that all 20 circumstances were sufficiently memorable, relevant to the scenario, and realistic. See [Online Supplemental Materials A and B](#) for a complete list of arguments and the vignette. For details about pilot testing and how materials were constructed, see <https://osf.io/dn75q>.

Participants were then randomly assigned to one of the four between-subjects conditions manipulating the magnitude (either 4 or 24 months) and the relevance of the PTS anchor. Anchor relevance was manipulated in a manner consistent with previous work to reduce potential variability attributable to a new procedure (Glöckner & Englich, 2015; Hans et al., 2018; Reyna et al., 2015). In the *relevant* anchor condition, participants learned the maximum sentence—either 4 or 24 months—they could face if convicted at trial from their attorney who had spoken with the prosecutor, whereas in the *irrelevant* anchor condition, participants learned instead the number of months remaining (4 or 24) for their acquaintance to graduate from law school.

Sentencing Judgments. Next, similar to the standard anchoring paradigm, participants across all conditions provided a *comparative judgment* to the question of “Would you consider a maximum sentence of [4/24] months in jail for your DUI case to be very lenient, appropriate, or very harsh for your case?” Participants then provided an *absolute judgment* to the question “[W]hat is the most severe sentence (in months) that you think you would be willing to accept rather than take your case to trial and face a potentially more severe sentence?” and their confidence in that judgment. The maximum acceptable sentence participants were willing to accept served as the main dependent variable.

Evaluations of Case Attributes. Next, I assessed whether the PTS altered evaluations of related case attributes, including the

¹ A meta-analysis by Bystranowski et al. (2021) published after data collection was complete demonstrated that the average effect size of anchoring in legal context studies ($N = 16$, $k = 37$ effect sizes) employing High versus Low anchors is $d = 0.91$, 95% CI [0.69, 1.12] and for studies employing Irrelevant versus Potentially or Relevant anchors is, $d = 0.52$, [0.42, .74].

estimated likelihood that they would be found guilty if they took their case to trial (from 0% chance, “I would definitely be acquitted,” to 100% chance, “I would definitely be found guilty”), the strength of the aggravating and of the mitigating evidence (1 = *not at all strong* to 7 = *extremely strong*), and the perceived severity of their crime (1 = *not at all serious* to 7 = *extremely serious*).

Plea Decision and Evaluations of Plea Offer. Participants next indicated their *initial* willingness to accept a plea offer at values relative to their maximum acceptable sentence prior to hearing the prosecutors’ plea offer (see Petersen et al., 2022). This measure was collected for exploratory purposes. Results are available in the [Supplemental Analyses](#). Participants then learned the terms of the prosecutors’ proffered plea deal (a 2-month jail sentence), provided a *final* plea decision (accept vs. reject and take case to trial), rated their confidence in their decision, as well as the perceived leniency (1 = *not at all lenient* to 7 = *extremely lenient*) and fairness of the plea offer (1 = *extremely fair* to 7 = *extremely unfair*), and responded to attention checks (described in the [online supplemental material](#)).

Evaluations of Evidence Strength. Participants rated the extent to which they regarded each of 20 arguments as aggravating or mitigating (−3 = *extremely aggravating*, 0 = *neither aggravating or mitigating*, +3 = *extremely mitigating*). For ease of interpretation, I reverse scored ratings of the 10 aggravating argument items. Based on results of a confirmatory factor analysis, I created a 10-item composite variable for aggravating evidence, $\omega_h = .93$; $\alpha = .82$ [.80, .85], and a five-item composite variable for mitigating evidence, $\omega_h = .79$; $\alpha = .67$ [.62, .71], by taking the average of each set of items. For information about the development of the scales, see the [online supplemental material](#).

Additional Measures. Participants rated how realistic they considered the situation and case materials to be (1 = *strongly disagree* to 7 = *strongly agree*) using a novel five-item scale ($\alpha = .73$; see the [Online Supplemental Material C](#)), completed a manipulation check (described below), and a numeracy skills scale (see Petersen et al., 2022). Additionally, participants reported their age, race/ethnicity, gender, household income, highest level of education attained, whether they or someone they knew had been stopped by the police for driving under the influence, and whether they had ever worked in law enforcement.

Analytic Strategy

To analyze ordinal variables (comparative judgment of anchor harshness) for manipulation checks, I used Mann–Whitney–Wilcoxon tests. To examine whether sentencing count data differed by condition, I specified negative binomial regression models using the `glm.nb()` function in the *MASS* package (Venables & Ripley, 2002) for *R*, supplemented with the *emmeans* package (Lenth, 2021) to compute estimated marginal means and calculated bias-adjusted standardized mean differences (i.e., Cohen’s *d* or Hedges’ *g*; Hedges & Olkin, 1985).² To examine whether dichotomous plea decisions differed by anchor condition, I fit decisions to logistic regressions using the `glm()` function in the built-in *stats* package in *R*. I used the *aod* package in *R* (Lesnoff & Lancelot, 2012) to formally test the overall model fit, and the *DescTools* package (Signorell, 2021) to compute bias-adjusted pseudo R^2 Nagelkerke effect sizes. To examine continuous variables (e.g., case attributes), I used Welch’s *t* test which does not assume equal

variances (Delacre et al., 2017) and linear regressions using the built-in `t.test()` and `lm()` functions, respectively. For each regression analysis with a predicted interaction, I used a likelihood ratio test to compare the model with and without the interaction. I retained and interpreted models that better fit the data. To test the hypothesis that sentencing judgments mediated the relationship between anchor condition and plea decision, I fit a moderated mediation model with the *mediation* package for *R* (Tingley et al., 2014). I built the model by running three independent generalized linear models. I fit the first model (i.e., interaction of independent variables predicting the mediator) using `glm()` function and estimated parameters for a Poisson distribution as the mediate function is not equipped to fit negative binomial models.³ I fit the remaining two generalized linear models (i.e., independent variable and mediator predicting outcome variable) also using `glm()` but estimated parameters for a binomial distribution. For ease of interpretation, I report a narrative summary of the main results and present regression statistics in tables and figures. For descriptive statistics and correlations, see [Table 2 in the online supplemental material](#).

Results

As expected, participants in the *high* anchor condition rated the anchor as harsher than those in the *moderate* anchor condition ($p < .001$); relevance did not influence comparative judgments ($p = .53$; see [Table 1 in the online supplemental material](#)). Participants in the *irrelevant* anchor condition regarded the anchor value as significantly less relevant to their plea decision than those in the *relevant* anchor condition (1 = *not at all relevant to my plea decision* to 7 = *extremely relevant to my plea decision*), $t(431.25) = -24.62$, $p < .001$, $d = 2.22$ [−2.47, −2.01].

Confirmatory Analyses

For all Experiment 1 models, I retained and interpreted the main effects only models as the models with the Anchor by Relevance interaction did not provide a better fit to the data.

Relevant and Irrelevant PTS Influenced Mock Defendants’ Sentencing Preferences (H1, H3). I predicted that participants would say they would be willing to accept longer sentences after receiving high relative to moderate PTS (H1). However, I also

² Results from Pilot Studies 3 and 4 revealed that sentencing count data had conditional variances that vastly exceeded conditional means across conditions and that specifying a negative binomial regression model could better accommodate for such overdispersion than a Poisson regression which assumes equidispersion (Venables & Ripley, 2002). Across all three main studies, the equidispersion assumption was violated and likelihood ratio tests confirmed that the negative binomial regression consistently fit the data better than the Poisson regression ($ps < .001$). I did not log-transform count data and analyze them using OLS regression because this would have failed to model the overdispersion and led to a loss of data (by taking the log of zero values).

³ Though the Anchor by Relevance interaction was previously not significant in the negative binomial model, it was significantly in the Poisson model. I had preregistered the code for a moderated mediation analysis modeling the effect of the anchor along all model paths. However, because the Anchor by Relevance interaction term was only plausibly influential in the model predicting maximum acceptable sentence and not at all in models predicting plea decision, I only tested the moderating effect of relevance along the *a* path. Results from a mediation model without this moderator yielded virtually identical results.

expected an Anchor by Relevance interaction such that the anchoring bias (i.e., the sentencing preferences difference between the high and moderate anchors) would be greater among participants exposed to the relevant compared to the irrelevant anchor (H3). To test these predictions, I fit a negative binomial regression model with anchor magnitude and relevance conditions predicting maximum acceptable sentences. Only the magnitude of the PTS was a significant predictor of sentencing preferences. Supporting H1, mock defendants were willing to accept a significantly longer maximum prison term when presented with a *high* ($M = 7.76$ months, $SD = 6.27$) compared to *moderate* PTS ($M = 3.89$, $SD = 3.44$; $d = 0.74$ 95% CI [0.55, 0.93]). Contrary to H3, the Anchor by Relevance interaction was not significant, indicating that relevance of the PTS did not moderate the effect of the size of the PTS on sentencing preferences. These results preliminarily support the *shallow* pathway to anchoring in a plea context. See Table 1 and Figure 1.

PTS Influence Plea Decisions (H1). Next, I predicted that exposure to the high PTS would be associated with greater plea acceptances relative to the moderate trial sentence (H1). The percentages of participants who accepted the plea deal across conditions were as follows: High/Relevant (85.95%), High/Irrelevant (83%), Moderate/Relevant (75.86%), and Moderate/Irrelevant (75.8%). The main effects logistic regression model predicting plea decisions was significant: As predicted, mock defendants in the *high* anchor condition were more likely to accept a plea offer than defendants in the *moderate* anchor condition, 84.51% versus 75.8%. See Table 2.

Maximum Acceptable Sentencing Preferences Mediates the Relationship Between PTS and Plea Decision (H1). Next, I examined the prediction that the PTS would influence plea decisions through participants' sentencing preferences (H1), moderated by informational relevance of the anchor (H3). Results of this moderated mediation analysis are presented in Figure 2. Consistent with H1, the paths from the anchor condition to the maximum acceptable sentence and from the maximum acceptable sentence to plea decision indicated significant effects. The indirect effects of the anchor condition through the maximum acceptable sentence at each level of the moderator were also significant ($ps < .001$). Contrary to H3, the indirect effects at each level of the moderator did not significantly differ, $b = 0.02$ 95% CI [-0.06, 0.10], $p = .56$, indicating that the effect of the PTS on plea decisions through sentencing preferences did not differ for those who received the *irrelevant* versus the *relevant* anchor. Collapsing across levels of the relevance

moderator, the proportion of the effect of the PTS on plea decision mediated through maximum acceptable sentences was 100% [67%, 100%], $p < .001$. This analysis lends further support to the *shallow* pathway hypothesis.

No Differential Effect of PTS Magnitude or Relevance on Evaluations of Evidence Strength (H2). I predicted that if anchoring influences judgmental decisions via thoughtful, effortful processing of the evidentiary information, participants exposed to different magnitudes of a relevant PTS anchor should differ in their evaluations of the strength of the evidence (H2). However, among participants exposed to an *informationally irrelevant* anchor, the difference in appraisals of evidence strength between high and moderate anchor conditions should be lower than the difference among decision-makers exposed to an *informationally relevant* anchor. To test this expected two-way interaction of anchor and relevance, I report results from composite scales completed *after* the decision-making task reflecting appraisals of the strength of the aggravating and mitigating evidence. Lending further support for the *shallow* pathway but in contrast to H2, results show no significant differences in evaluations of the mitigating evidence based on PTS magnitude ($d = -0.12$ 95% CI [-0.30, 0.06]) or relevance ($d = 0.07$ [-0.11, 0.25]) or in evaluations of the aggravating evidence based on PTS magnitude ($d = 0.13$ [-0.05, 0.31]) or relevance ($d = 0.02$ [-0.16, 0.20]). See Table 3.

Exploratory Analyses

If anchoring exerts its influence through deliberate thoughtful processing, then the Anchor by Relevance interaction should have influenced appraisals of evidence strength. The following exploratory analyses were run to test the possibility that anchors influenced participants' perceptions of other case attributes *prior* to rendering any plea decisions. There were no significant main effects of relevance or of anchor on estimated likelihood of conviction at trial or of evaluations of crime severity, lending further support for the *shallow* processing pathway. In contrast, the models predicting evaluations of plea leniency and plea fairness demonstrated an effect of anchor condition, though not of relevance. Compared to participants in the moderate anchor condition, those in high anchor condition regarded the plea offer as more lenient ($d = 0.26$ 95% CI [0.10, 0.46]) and fairer, ($d = 0.22$ [0.05, 0.40]). These regression analyses are available in the Supplemental Analyses.

Table 1

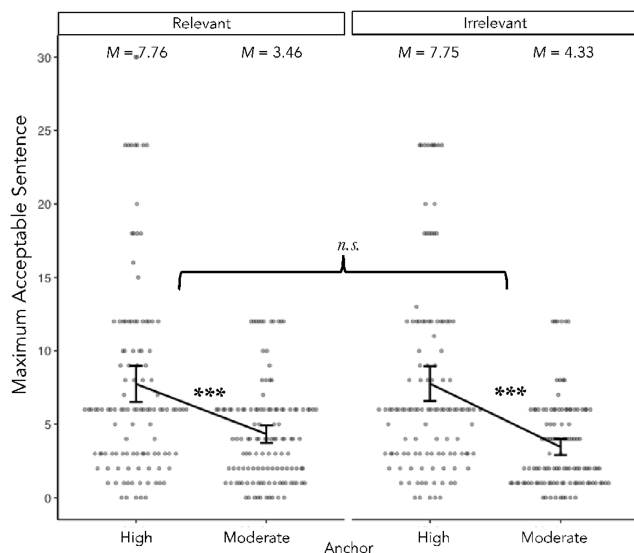
Experiment 1 Negative Binomial Regression Model Results for the Effect of Anchor and Relevance on Maximum Acceptable Sentences

Effect	IRR	95% CI for IRR	<i>b</i>	95% CI for <i>b</i>	<i>SE</i>	<i>z</i>	<i>df</i>	<i>p</i>	<i>R</i> ²
Model 1									
Intercept	8.17	[7.19–9.31]	2.10	[2.21, 2.67]	0.07	32.30	476	<.001	.22
Anchor ^a	0.50	[0.43–0.58]	−0.69	[−0.84, −0.54]	0.08	−8.93	476	<.001	
Relevance ^b	0.90	[0.78–1.05]	−0.10	[−0.26, 0.04]	0.08	−1.32	476	.188	
Model 2									
Intercept	7.75	[6.71–8.98]	2.05	[1.90, 2.19]	0.07	27.58	475	<.001	.22
Anchor	0.56	[0.45–0.69]	−0.58	[−0.79, −0.37]	0.11	−5.40	475	<.001	
Relevance	1.00	[0.82–1.23]	0.002	[−0.20, 0.20]	0.10	0.02	475	0.99	
Anchor × Relevance	0.80	[0.59–1.08]	−0.23	[−0.53, 0.08]	0.15	−1.47	475	.141	

Note. IRR = incidence rate ratio; CI = confidence interval; *R*² = bias-adjusted Nagelkerke effect size.

^a Anchor: *High* = 0 versus *Moderate* = 1. ^b Relevance: *Relevant* = 0 versus *Irrelevant* = 1.

Figure 1
Means and 95% CI for Maximum Acceptable Sentence by Condition



Note. Error bars show 95% confidence intervals. CI = confidence interval.
*** $p < .001$.

Discussion

The current work offers preliminary support for the *shallow* rather than *deep* processing pathway to anchoring in plea contexts. First, mock defendants exposed to a 24- versus a 4-month PTS were willing to serve a longer plea-deal jail term, despite not interpreting the aggravating circumstances as *more* unfavorable or the mitigating circumstances as *less* favorable. A willingness to accept a longer sentence was, in turn, associated with a higher likelihood of accepting the plea offer. Significantly, favoring the *shallow* processing pathway, these results hold when the PTS was, by participants' own admission, irrelevant. What can explain these results? As discussed earlier, defendants rationally weigh the benefits of a plea discount against the costs of pleading guilty when rendering plea decisions. To calculate the plea discount, defendants require two relevant inputs: the PTS and the plea deal sentence. However, participants in the *irrelevant anchor* condition lacked a relevant PTS (i.e., it did not reveal what sentence the prosecutor might threaten to recommend to a judge). If participants with a relevant PTS could calculate a plea discount, whereas those with an irrelevant PTS could not, why then did both arrive at similar plea decisions? One likely explanation

is that all participants—regardless of PTS relevancy—used their anchor-biased maximum sentencing preferences to calculate an unofficial plea discount. However, because the PTS was arbitrary for participants in the irrelevant PTS conditions, any consideration of the gains associated with this unofficial plea discount is irrational.

Experiment 2

Experiment 2 was designed to replicate and generalize the results of Experiment 1's anchor magnitude manipulation to a sample of practicing criminal defense attorneys. Additionally, this experiment tested a debiasing technique (COS) that instructed attorneys to consider anchor-*inconsistent* case facts contradicting the PTS (i.e., mitigating facts in the *high* PTS condition and *aggravating* facts in the *moderate* condition). I predicted that attorneys would be willing to recommend that their client accept longer sentences when exposed to high compared to moderate PTS (H1), but that consideration of anchor-*inconsistent* arguments would lead to *lower* post-COS sentencing recommendations among attorneys in the high anchor condition and *increased* post-COS sentences in the moderate anchor condition (H3). In line with H3, I also predicted that among attorneys exposed to the *high* anchor, generating *more* arguments suggesting that the high PTS was “very harsh” would be associated with *greater* plea rejections, but that among attorneys exposed to the moderate anchor, more arguments suggesting that moderate PTS was “very lenient” would be associated with *lower* plea rejections.

Method

Design and Power

This experiment used a 2 (Anchor: High [24 months] vs. Moderate [6 months]) $\times$ 2 (COS: Pre vs. Post) mixed factorial design in which the latter factor is within-subject. The 24-month sentence represented a value above the 95th percentile of the calibration sample, whereas the 6-month sentence represented a value at the 50th percentile. The moderate anchor was increased to be able to provide a plea offer that was less lenient than the one provided in Experiment 1 (2 vs. 4 months) and increase variability in plea decisions. An a priori power simulation analysis showed that a total N of 148 attorneys ($n = 74$ per cell of the between-subjects factor) was sufficient to detect a medium-sized effect (Cohen's $f = 0.23$) for the Anchor $\times$ COS interaction with greater than .80 power ($\alpha = .05$). See simulation here: <https://osf.io/ptq6y>.

Participants

Practicing defense attorneys were recruited through contacts (e.g., at the [Innocence Project, 2019](#)) and by emailing state and national

Table 2
Experiment 1 Logistic Regression Predicting Likelihood of Final Plea Decision Based on Anchor and Relevance

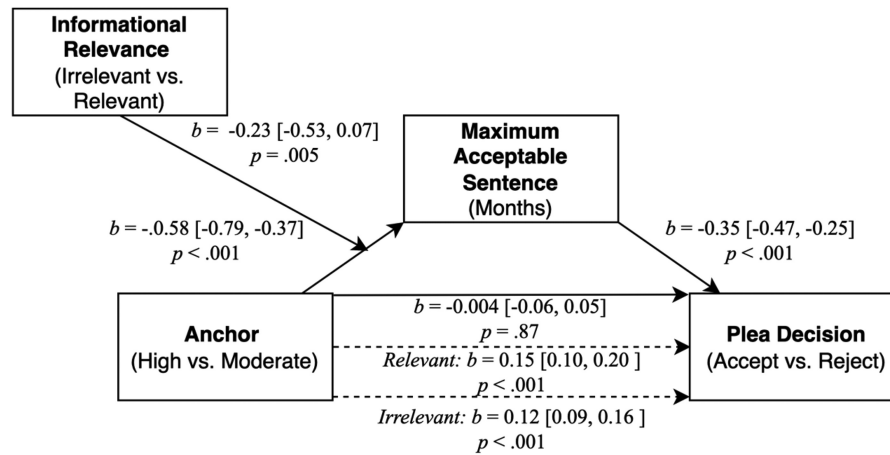
Effect	<i>B</i>	<i>SE</i>	Wald <i>z</i>	<i>df</i>	<i>p</i>	Odds ratio	95% CI for odds ratio		Overall model fit
							Lower	Upper	
Intercept	−1.651	0.212	−7.8	476	<.001	0.19	0.12	0.29	$\chi^2(1) = 5.6, p = .018$
Anchor ^a	0.552	0.234	2.36	476	.018	1.74	1.10	2.76	
Relevance ^b	−0.094	0.231	−0.41	476	.684	0.91	0.58	1.43	

Note. CI = confidence interval.

^a Anchor: *High* = 0 versus *Moderate* = 1. ^b Relevance: *Relevant* = 0 versus *Irrelevant* = 1.

Figure 2

Mediation Model of Anchor Predicting Plea Decision Mediated by Maximum Acceptable Sentence



Note. Unstandardized estimates are displayed, with 95% confidence intervals and *p*-values. Anchor: *High* = 0; *Moderate* = 1; Plea Decision: 0 = *Accept*; 1 = *Reject*. Dashed lines represent the indirect effects of informationally relevant and irrelevant anchors through individually rendered sentencing judgments. Estimates are based on bootstrapping procedure with 10,000 bootstrap samples.

criminal defense attorney associations and public defenders' offices (e.g., Association of Criminal Defense Lawyers in New Jersey, Indigent Defense Services) and paid \$15. Although 274 attorneys consented, only 209 proceeded beyond the consent form, and 173 provided a response to the two main variables (i.e., maximum acceptable sentencing judgments and plea decisions). Data were excluded from attorneys who provided sentencing recommendations constituting outliers per my preregistered criteria ($n = 5$) and for failing to provide anchor-inconsistent facts ($n = 13$). The final sample comprised 155 attorneys from 17 states (56.7% from New York), who were primarily female (54.83%) and White (78%). Ages ranged between 26 and 81 ($M = 43.89$, $SD = 13.45$). Attorneys were primarily public defenders (65%) with an average of 15.63 ($SD = 13.29$) years of experience, and a median number of 1015 cases handled.

Procedure

Procedures for Experiment 2 resembled those of Experiment 1. Attorneys first responded to questions pertaining to their legal

experiences and demographics and then imagined that they were working in a public defenders' office in their own jurisdiction and that they were representing a client with the same circumstances described in Experiment 1. After, half of attorneys were randomly assigned to encounter a *high* ($n = 83$) or *moderate* PTS ($n = 72$). Then, they responded to the comparative judgment question, provided a pre-COS sentencing recommendation, and their confidence in this recommendation.

Consider-the-Opposite Strategy. Next, they were asked to generate anchor-inconsistent arguments—that is, arguments that suggested that the high PTS was “very severe” or that the moderate PTS was “very lenient”:

While seeking advice from a colleague about your client's case, your colleague tells you that they think [6/24] months in jail is [very lenient/very severe] for your client's DUI case. Based on your understanding of your client's case, including all the information you have read, please list as many relevant facts as you can think of that might suggest that [6/24] months is [very lenient/very severe] for this case.

Table 3

Experiment 1 Linear Regression Model Results for the Effect of Anchor and Relevance on Perceptions of the Mitigating and Aggravating Evidence (Composite Measures)

Effect	<i>b</i>	95% CI	<i>SE</i>	<i>t</i>	<i>df</i>	<i>p</i>	Overall model fit
Mitigating evidence							
Intercept	1.44	[1.32, 1.55]	0.06	24.09	476	<.001	$F(2, 476) = 1.241, R^2 = .001, p = .290$
Anchor	−0.09	[−0.23, 0.04]	0.07	−1.37	476	.172	
Relevance	−0.06	[−0.19, 0.08]	0.07	−0.81	476	.418	
Aggravating evidence							
Intercept	1.48	[1.37, 1.60]	0.06	25.03	476	<.001	$F(2, 476) = 1.015, R^2 = .00, p = .363$
Anchor	−0.10	[−0.23, 0.04]	0.07	−1.41	476	.16	
Relevance	0.01	[−0.12, 0.15]	0.07	0.19	476	.85	

Note. Composite measure analyses used the 5-item mitigating composite scale and the 10-item aggravating composite scale completed after rendering a plea decision. CI = confidence interval.

Attorneys were provided with 11 labeled text fields (e.g., “Case Fact 1–10 and Other Relevant Case Facts) where they could list each anchor-*inconsistent* argument. On the following page they were reminded of the maximum sentencing recommendation they had previously provided, presented with the list of all anchor-inconsistent case facts they had generated, and asked whether they did or did not want to revise their original sentence.

Attorneys rated their confidence with this new sentence, estimated their client’s likelihood of conviction, and rated the strength of the aggravating and the mitigating evidence, and the crime severity (in a randomized order). Next, attorneys decided whether to recommend their client accept or reject a nonnegotiable, time-limited plea offer containing a 4-month sentence, rated their decision confidence, the leniency, and fairness of the plea offer, and explained their decision. They also responded to attention and manipulation checks, as well as exploratory measures available in the [online supplemental material](#). Finally, attorneys were given the option to donate their compensation to the Bail Project (<https://bailproject.org/>), a nonprofit organization that combats mass incarceration by helping defendants in pretrial detention who cannot afford bail and were debriefed.

Analytic Strategy

The data analysis strategy for Experiment 2 resembled that of Experiment 1. However, for models involving the repeated measures of maximum acceptable sentences, I tested linear mixed-effects models using the *lme4* and *lmerTest* packages (Bates et al., 2015) in R with condition and covariates as *fixed effects*, and *random intercepts* for participants (i.e., to account for sentences nested within participants). I preregistered inclusion of *random slopes* to estimate the by-participant differences between pre-COS and post-COS sentences. However, because these models failed to converge—likely a result of the near-zero variance explained by this random effect—I report models without random slopes. See Table 3 in the [online supplemental material](#) for descriptive statistics and bivariate correlations.

Results

As expected, attorneys in the *high* anchor condition rated the anchor as harsher than those in the moderate anchor condition, $M_{\text{diff}} = 0.82$, $t_w(100.48) = 5.40$, $p < .001$.

Confirmatory Analyses

In models reported below, I retained and interpreted the model with the Anchor by COS interaction, as the main effects only models did not provide a better fit to the data.

PTS and COS Influenced Attorneys’ Sentencing Recommendations (H1, H3). To test H1, I fit linear mixed-effects models with anchor and COS predicting sentencing recommendations. The expected main effect of the anchor was qualified by the expected Anchor by COS interaction (H3). In partial support of H3, attorneys in the *high* anchor condition provided sentences that were approximately 20 days longer pre-COS compared to post-COS, $M_{\text{pre}} = 2.73$ months vs. $M_{\text{post}} = 2.08$; $p_{\text{tukey}} = .004$. However, contrary to my predictions, attorneys in the moderate PTS anchor condition did not provide significantly different sentences pre-COS compared to post-COS ($M_{\text{pre}} = 0.83$ months vs. $M_{\text{post}} = 0.98$; $p_{\text{tukey}} = .66$).

These results suggest that the COS intervention effectively reduced the anchoring bias among attorneys in the *high* PTS condition, while not causing attorneys in the moderate PTS condition to recommend more severe sentences. See Figure 3 and Table 4.

PTS Predicts Attorneys’ Plea Recommendations (H1). A logistic regression model demonstrated that PTS was significantly associated with attorneys’ plea recommendations, $\chi^2(1) = 13.2$, $p < .001$, $R^2 = .12$. As expected (H1), attorneys in the *high* anchor condition were 3.53 95% CI [1.81, 7.07] times more likely to accept the plea deal than those in the moderate anchor condition (58.5% vs. 28%), $b = 1.26$, $SE = 0.35$, $z = 3.63$, $p < .001$.

Attorneys’ Maximum Sentence Recommendations Mediate the Relationship Between PTS and Plea Decision (H1). To examine the mediating role of maximum acceptable sentences, I report results of two mediation models in Figure 4: one demonstrating the mediating effect of pre-COS sentences and the other the mediating effect of post-COS sentences.⁴ The paths from the anchor condition to maximum acceptable sentences (pre- and post-COS) and from maximum acceptable sentences to plea decision indicated significant effects in both models. The indirect effects of the anchor condition through sentencing judgment were also significant in both models ($ps < .001$). Like Experiment 1, the direct effect from anchor condition was no longer significant in the model when either pre-COS sentence ($p = .23$), or post-COS sentence were included as the mediators ($p = .076$).

Effect of Anchor by Arguments Interaction on Plea Decision (H3). I modeled a two-way interaction of Anchor by Arguments predicting plea decision. The logistic regression model demonstrated a significant two-way interaction. Probing the expected interaction showed partial support for H3: Among attorneys exposed to the *high* anchor, generating more anchor-inconsistent (mitigating) arguments was associated with greater plea rejections ($b = 0.29$ 95% CI [0.08, 0.50]), but among attorneys exposed to the moderate anchor, generating more (aggravating) arguments was not significantly associated with greater plea acceptances ($b = -0.8$ [−0.29, 0.13]). Thus, the COS may be associated with decreased recommendation of plea acceptances among attorneys exposed to a high PTS who generated more mitigating arguments, but not with increased recommendations among attorneys exposed to a moderate PTS. See Table 5.

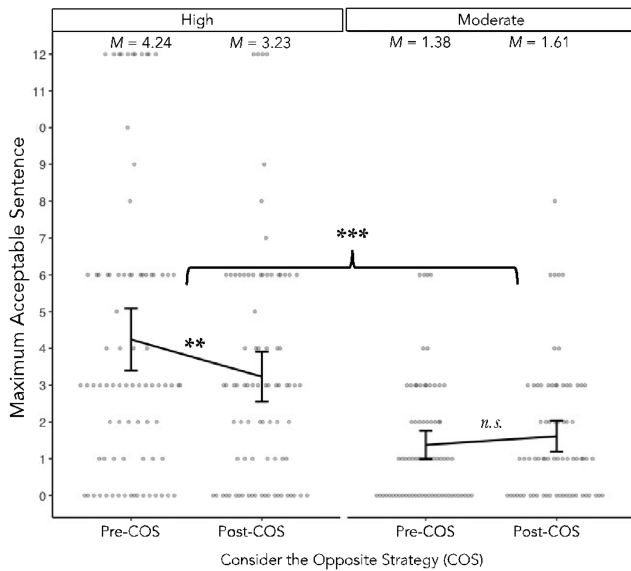
Exploratory Analyses

Attorneys’ evaluations of the case attributes varied systematically based on anchor condition. For example, attorneys in the high anchor condition who were asked to generate evidence suggesting the anchor was severe rated the mitigating evidence as stronger ($d = 0.45$ 95% CI [0.13, 0.78]), the aggravating evidence as weaker ($d = 0.59$ [0.26, 0.92]), and their clients’ alleged crime as less severe ($d = 0.34$ [0.02, 0.67]) than attorneys in the moderate

⁴ I preregistered a moderated mediation model similar to Experiment 1 with the notable difference of fitting two generalized linear mixed models to account for by-participant random intercepts. However, contrary to H3, the generalized linear mixed model fitted with the Anchor by COS interaction and sentencing judgments predicting plea decision did not demonstrate a significant interaction, $b = 0.87$, $SE = 2.73$, $z = 0.319$, $p = .75$. This result indicates that though the COS intervention reduced the anchoring effect it did not directly reduce its effect on plea decisions.

Figure 3

Means and 95% CI for Maximum Acceptable Sentence by Condition



Note. Error bars show 95% confidence intervals. CI = confidence interval. ** $p < .005$. *** $p < .001$.

anchor condition asked to generate aggravating evidence. Interestingly, attorney evaluations of the leniency ($d = 0.21 [-0.11, 0.53]$) and fairness ($d = 0.26 [-0.07, 0.58]$) of the plea offer did not differ by condition. These results are the reverse of those seen in Experiment 1, where mock defendants' perceptions of the case attributes did not significantly vary by anchor condition, but their perceptions of the plea offer did. Attorneys likely arrived at differing case evaluations only *after* completing the COS procedure. For a detailed description of results, see the [Supplemental Analyses](#).

Discussion

Though legal scholars have long claimed that attorneys ought to be capable of assisting clients by presenting them with all relevant information, identifying alternative choices, and “chang[ing] the client’s outlook” to avoid “cognitive errors” (Korobkin & Guthrie, 1997, p. 130), Experiment 2 demonstrated that practicing defense attorneys were susceptible to the anchoring effect. Promisingly, results demonstrate that the COS led attorneys exposed to an exceptionally high PTS to revise their recommendation about what constituted an acceptable maximum sentence for their client to serve. However, despite evaluating the mitigating evidence as *stronger*, the aggravating evidence as *weaker*, and the crime as *less* severe, attorneys facing a higher compared to more moderate PTS did not estimate a lower likelihood of their client being convicted. In fact, attorneys presented with a higher PTS chose to recommend that their client accept the offer, while not even viewing the offer as more lenient or fairer than those exposed to the moderate PTS.

The current work is not without limitations. Though the within-subject nature of the COS intervention had the benefits of reducing the required sample size of a difficult-to-recruit population and enabling the more difficult test of an intervention to debias attorneys’ already anchor-assimilated judgments, it was externally engineered and unlikely to be used spontaneously. Even so, the promising results shown here demonstrate that its implementation among legal experts could serve the same benefits that checklists confer to airlines and hospitals keen on reducing errors of human cognition.

Experiment 3

Experiment 3 was designed to determine whether the effect of PTS on maximum sentencing preferences and the *lack* of an effect of PTS on case evaluations observed in studies reported here would replicate in a more ecologically valid environment. Additionally, two significant alterations enabled a more robust test of the SAM. First, Experiment 3 introduced a novel paradigm requiring mock attorney–client pairs to engage in discussions about the

Table 4

Experiment 2 Negative Binomial Mixed-Effects Regression Model Results for the Effect of Anchor and COS on Maximum Acceptable Sentences

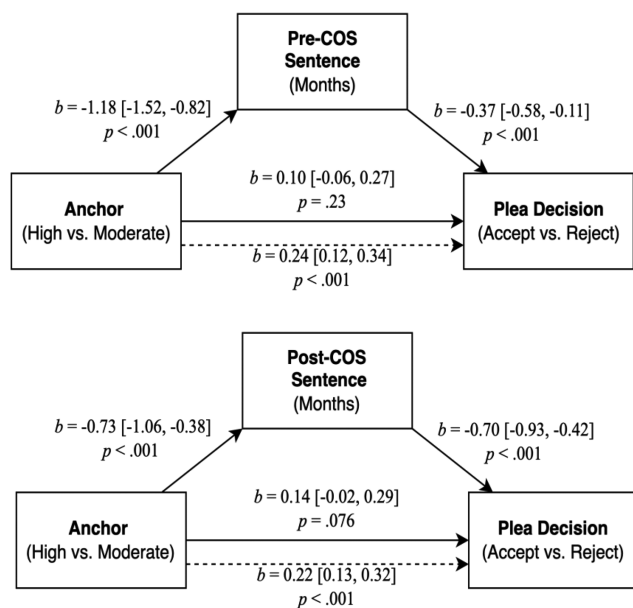
Effect	IRR	95% CI for IRR	<i>b</i>	<i>SE</i>	<i>z</i>	<i>p</i>
Model 1						
Fixed effects						
Intercept	2.60	[2.00–3.38]	0.96	0.13	7.14	<.001
Anchor ^a	0.38	[0.26–0.55]	−0.98	0.19	−5.04	<.001
COS ^b	0.85	[0.74–0.98]	−0.16	0.07	−2.32	.020
Random effects						
Within-participant variance	1.11					
Model 2						
Fixed effects						
Intercept	2.74	[2.11–3.55]	1.01	0.13	7.55	<.001
Anchor	0.30	[0.20–0.46]	−1.19	0.21	−5.53	<.001
COS	0.76	[0.65–0.89]	−0.27	0.08	−3.34	<.001
Anchor × COS	1.54	[1.12–2.12]	0.43	0.16	2.65	.008
Random effects						
Within-participant variance	1.11					

Note. 155 attorneys, 310 observations. COS = consider-the-opposite strategy; IRR = incidence rate ratio; CI = confidence interval.

^a Anchor: High = 0 versus Moderate = 1. ^b COS: Pre = 0 versus Post = 1.

Figure 4

Mediation Models of Anchor Predicting Plea Decision Mediated by Maximum Acceptable Sentence



Note. Unstandardized estimates are displayed, with 95% confidence intervals and *p*-values. Anchor: High = 0; Moderate = 1; Plea Decision: 0 = Accept; 1 = Reject. Dashed lines represent the indirect effects of anchor magnitude through pre- and post-COS sentencing judgments. Estimates are based on bootstrapping procedure with 10,000 bootstrap samples. COS = consider-the-opposite strategy.

plea decision circumstances. Second, half of mock attorney–client pairs were motivated (via process accountability) to elaborate on all their ideas about the case and systematically discuss all the evidence. As actual attorneys and defendants should consider all the evidence, not merely that which contradicts the PTS, this manipulation enabled a more ecologically valid test of the prediction that increasing consideration of anchor-inconsistent evidence should reduce the anchoring bias. If anchoring results from selective retrieval of anchor-consistent evidence, then promoting holistic consideration of all evidence should prevent the anchoring bias from

occurring altogether. More formally, I predicted that among dyads in the (H2) *unaccountable* condition, there would be a difference between anchor conditions in the number of mitigating and/or aggravating arguments discussed, but that (H3) in the *accountable* condition there would be little or no difference in arguments discussed between anchor conditions. Given competing hypotheses, I did not have a hypothesis about the direction of the influence of the anchor on information exchanged. I also hypothesized (H1) that among *unaccountable* dyads, joint maximum sentencing preferences would be higher for high relative to moderate anchors, but among *accountable* dyads, joint sentencing preferences would show either no significant difference between anchor conditions or an attenuated difference (H3). Finally, individual group members also provided a sentencing preference *prior* to exposure to the anchor, testing the prediction that anchor-unbiased sentencing preferences would compete with the externally provided PTS, rendering the latter less influential (Sniezek, 1992).

Method

Design and Power

In a 2 (Anchor: Implausibly High [48 months] vs. Moderate [8 months]) $\times$ 2 (Process Accountability: Present vs. Absent) between-subjects factorial design, 277 attorney–client pairs (554 participants) were randomly assigned to conditions. I chose the 48-month anchor, which corresponded with the sentencing guidelines for an aggravated felony DUI in Arizona (due to the suspended license and injury to a person) because it represented an implausibly high value above the 100th percentile of the calibration sample. An a priori power simulation analysis showed that a total *N* of 256 pairs (*n* = 66 per cell) was sufficient to detect a small to medium effect size (Cohen's *f* = 0.177) with 80% power for the interaction of Anchor $\times$ Process Accountability. See simulation here: <https://osf.io/yo567>. The final sample after exclusions (*N*_{pairs} = 265; 131 accountable, 134 unaccountable).

Participants

Mock defendant participants were Prolific workers (*n* = 262) and undergraduates from an urban commuter college (*n* = 16); mock attorney participants were law students (*n* = 276 from 99 unique law schools) recruited between January and March 2021 for a three-

Table 5

Experiment 2 Logistic Regression Predicting Likelihood of Plea Decision Based on Anchor and Arguments Generated

Effect	<i>B</i>	<i>SE</i>	Wald <i>z</i>	<i>df</i>	<i>p</i>	Odds ratio	95% CI for odds ratio		Overall model fit
							Lower	Upper	
Model 1									
Intercept	−1.13	0.54	−2.11	149	.035	0.32	0.11	0.90	$\chi^2(1) = 15.1, p < .001$
Anchor ^a	1.54	0.40	3.89	149	<.001	4.65	2.19	10.39	
Arguments ^b	0.12	0.08	1.63	149	.103	1.13	0.98	1.32	
Model 2									
Intercept	−2.23	0.73	−3.04	148	.002	0.11	0.02	0.43	$\chi^2(1) = 14.5, p < .001$
Anchor	3.48	0.91	3.81	148	<.001	32.60	5.72	209.86	
Arguments	0.29	0.11	2.73	148	.006	1.34	1.10	1.67	
Anchor × Arguments	−0.37	0.15	−2.45	148	.014	0.69	0.51	0.93	

Note. CI = confidence interval.

^a Anchor: High (24 months) = 0 versus Moderate (6 months) = 1. ^b Arguments: Anchor-inconsistent arguments generated post-COS intervention.

part (70-min) study. Law students were recruited by contacting professors teaching criminal procedure or related courses at every U.S. law school. Prolific participants were paid \$15.00, undergraduates received course credit, and law students received \$25 with the possibility of earning \$5 for each law student (max 3) they referred to the researcher. Sixty-one (22%) law students were recruited via this snowball sampling technique by 54 law student participants. I excluded data from 12 pairs from analyses because the law student failed the accountability manipulation ($n = 2$), could not follow instructions ($n = 2$), experienced a technical issue ($n = 6$), was suspicious ($n = 1$), and had previously participated ($n = 1$).

The final sample ($N = 530$) comprised 265 (54.3% female) mock defendants and 265 mock attorneys (62.3% female): Mock defendants ($M = 34.8$, $SD = 13.09$) were on average a few years older than mock attorneys ($M = 26.54$, $SD = 4.95$). Among mock defendants, 72% identified as White, 13.4% as Asian, as 10.19% Hispanic/Latino, 9.8% as Black/African American, 2.3% as Other. Among mock attorneys, 67.2% identified as White, 22.42% as Asian, 13.96% as Black/African American, 11.7% as Hispanic/Latino, and 2.6% as Other. See [Supplemental Analyses](#) for additional sample characteristics.

Procedure

Role Preparation. In Part 1, participants consented to the study asynchronously via a survey where they used a Calendly meeting scheduler to sign up for Part 2, a synchronous plea role-playing study over Zoom. Lay persons learned they would be asked to role-play a defendant in a criminal case; law students learned they would role-play being a public defender for a mock defendant and read sections of the American Bar Associations' "Criminal Justice Standards for the Defense Function" (see the [Online Supplemental Material D](#)).

Plea Negotiation Role-Playing. One of the 11 experimenters reminded participants of their respective roles and replaced names with role-specific identifiers (e.g., Attorney-02816). Participants individually read the case description, and prior to learning the PTS indicated the maximum sentence they were willing to accept/recommend. Preanchor sentencing judgments were intended to act as competing self-generated anchors that participants could refer to in their attorney–client discussions. This concluded the first individual portion of their participation.

Next, the experimenter began recording the session audio. Participants were instructed that they should follow the guided instructions displayed only on the attorney's survey for the following part of the role-playing task and to temporarily set aside the defendants' survey. Attorneys were responsible for submitting group-level responses. Participants were also told they would read "a very important memo from the researchers before spending a few minutes discussing the case together prior to reaching a joint sentencing decision that considered both of their viewpoints, even if those might initially differ." To simulate a real attorney–client discussion and keep the experimenter blind to condition, the experimenter informed participants that they would temporarily disable their computer's audio system. Attorneys then introduced themselves (as prompted by a script) as the defendant's "court-appointed public defender" and were given access to the rest of the Qualtrics survey.

Manipulation of Process Accountability. I adapted an established procedure for manipulating process accountability (e.g., [De](#)

[Dreu et al., 2000](#); [De Wilde et al., 2018](#)). Mock attorneys were randomly assigned to share a memo with their "client" corresponding to their accountability condition. See the [Online Supplemental Material E](#). The *accountable* condition memo informed participants that they were to sign up for a follow-up poststudy accountability interview about the decision-making task to gain insight into the process by which attorney–client decisions are made. A legal decision-making expert would ask participants to explain in detail *how* their group reached their decision, rather than what that decision was ([Lerner & Tetlock, 1999](#); [Patil et al., 2014](#)). To increase realism of this interview, participants agreed to the storage and review of their audio recording for evaluative purposes, then indicated their mutual availability for the interview. Process accountable memos emphasized four types of process accountability described by [Lerner and Tetlock \(1999\)](#): (a) the presence of another individual (participant expects to have their performance observed), (b) identifiability (participant believes their performance will be linked back to them), (c) evaluation (participant expects to have their performance evaluated on the basis of some preestablished rules and for there to be consequences to this evaluation), and (d) reason-giving (participants expect to explain the rationale behind their responses); unaccountable memos emphasized the opposite of each of these components.

Sentencing Judgments and Perceptions of the Case and Plea Offer. Reading off a script, attorneys communicated the anchor manipulation (the prosecutor's PTS recommendation to a judge) to their "client" via the comparative judgment question. Attorney–client pairs discussed the case for an average of 6.14 ($SD = 4.7$) minutes prior to jointly deciding on a joint maximum acceptable sentence. Participants then spent an average of 3.02 ($SD = 2.05$) minutes making a plea decision, which concluded the collaborative role-playing part of the study. Participants completed the same questionnaire as in prior studies (with some additional questions described below) independently.

Additional Measures. Participants answered nine manipulation checks regarding (a) their assigned accountability condition (*stimulus attention check*: "The memo we read indicated that..."), (b) the extent to which they felt motivated to process information systematically (*construct validity*; e.g., "...we weighed all the pros and cons thoroughly and in a balanced way before making a decision"), and (c) the extent to which they believed they would be outcome accountable (*discriminant validity*; e.g., "...we will have to explain the actual decisions we made, and not the way we arrived at those") ($\alpha = .68$ [.63, .72]). See the [Online Supplemental Material F](#). Finally, mock defendants rated their perceptions of their attorneys using a nine-item scale ($\alpha = .86$ [.83, .88]) and attorneys rated their perceptions of their clients' perceptions of them ($\alpha = .87$ [.84, .89]) before being debriefed. See the [Online Supplemental Material G](#).

Information Exchanged. Coders verified the accuracy of the auto-generated transcript of joint discussions, removed information identifying the accountability condition, and coded whether participants had engaged in a discussion of the case prior to exposure to either of the manipulations. Each verified transcript was coded by one of five pairs of trained coders and myself. All coders were blind to hypotheses and accountability conditions, but not anchor conditions. Using the transcript, Rater 1 and Rater 2 identified all circumstances discussed by participants as influencing their decision-making (see <https://airtable.com/shrWXPsx54WjnFDi>).

Coder pairs independently rated each circumstance as having been discussed as either “mitigating,” “aggravating,” “both mitigating and aggravating” or “neither mitigating or aggravating.” Finally, Rater 1 and Rater 2 met to resolve any disagreements and submit final ratings ($\kappa_s > .72$). I tabulated the proportion of aggravating and mitigating arguments discussed that were contained within the case description.⁵ Additional measures, including collateral consequences spontaneously discussed, can be found in the [Supplemental Analyses](#).

Analytic Strategy

The main dependent variables were collected and analyzed at the group level (sentencing judgment, plea decision, arguments discussed). Dependent variables collected at the individual level (e.g., manipulation checks, evidence strength) were analyzed using linear mixed-effects models with anchor and process accountability as the *fixed* effects, individual difference measures (race, age, gender) as *covariates* effects, and *random intercepts* to account for the variance attributable to group membership. [Tables 4 and 5 in the online supplemental material](#) show descriptive statistics and bivariate correlations.

Results

As expected, there were no significant differences in *preanchor maximum or aspirational sentences* by condition ($ps = .13-.88$) or accountability ($p = .13$). There was an unexpected, marginal difference by anchor condition on participants' *estimates for their likelihood of conviction* prior to exposure to any manipulation ($p = .031$, $d = 0.19$ [0.01, 0.37]). Participants in the high anchor condition rated the anchor as harsher than those in the moderate anchor condition ($p < .001$), whereas accountability did not differentially influence comparative judgments ($p = .20$). Dyads in the process accountable condition indicated having weighed the information more thoroughly and avoiding rushing the discussion or suppressing disagreements compared to members of unaccountable dyads, $p = .032$, $d = 0.19$ [0.03, 0.37]. Discussions among *accountable* dyads were lengthier than discussions among *unaccountable* dyads (13.41 vs. 9.04 min), $p < .001$, $d = 0.69$ [0.51, 0.86]. Additionally, process-accountable groups were less focused on *outcome* accountability ($p < .001$, $d = 1.09$ [0.90, 1.27]).

Confirmatory Analyses

Per my preregistration, variables with significant bivariate correlations (i.e., attorneys' and defendants' individual preanchor preferences and estimated likelihoods of conviction at trial) were included as covariates in all Experiment 3 models. In all models reported below, I retained and interpreted the main effects only model over the model with the Anchor by Accountability interaction, as the latter model did not provide a better fit to the data.

PTS Influenced Mock Attorney–Client Pairs Maximum Sentencing Preferences, but Accountability Did Not (H1 and H3). A main effects only negative binomial regression model with covariates demonstrated that consistent with Studies 1 and 2 and my predictions, attorney–client pairs exposed to the *implausible* PTS preferred maximum sentences ($M = 6.07$, $SD = 4.72$) that were on average 2.8 months greater than those set by participants exposed to the moderate maximum sentence ($M = 3.35$, $SD = 2.88$),

$d = 0.69$ 95% CI [0.45, 0.94]. See [Figure 5](#). Contrary to H3, the motivation-based debiasing intervention was not effective at reducing the anchoring effect on sentencing preferences. However, in partial support of my prediction, defendants', but not attorneys', individual preanchor sentencing estimates were highly positively associated with maximum acceptable sentences. See [Table 6](#).

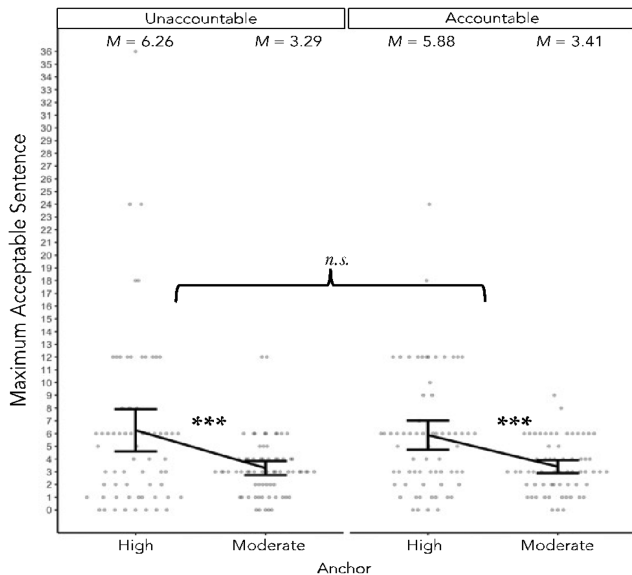
PTS Influenced Plea Decisions, but Accountability Did Not (H1 and H3). The percentages of attorney–client pairs who accepted the plea deal across conditions were as follows: High/Accountable (59.70%), High/Unaccountable (52.30%), Moderate/Accountable (28.13%), and Moderate/Unaccountable (31.88%). As expected, attorney–client pairs in the *implausible* anchor condition were more likely to accept a plea offer than pairs in the *moderate* anchor condition, 56.06% versus 30.01%. Accountability did not moderate the influence of the Anchor. See [Table 7](#).

Maximum Acceptable Sentencing Preferences Mediate the Relationship Between PTS and Attorney–Client Pairs' Plea Decisions (H1 and H3). Next, I tested a mediation model to assess whether jointly decided maximum acceptable sentences would mediate the relationship between anchor condition and final plea decision. The paths from the anchor condition to joint sentencing judgments and from sentences to plea decision indicated significant effects, and the indirect effect of the anchor condition through joint sentencing judgments was also significant. Unlike Studies 1 and 2, but still supporting H1, the direct effect of anchor on plea decision remained significant and the proportion of the effect mediated was partial, 38.58% 95% CI [7%, 73%], $p = .018$, rather than complete. See [Figure 6](#).

Accountability Predicted Number of Arguments Discussed as Mitigating and Aggravating, but Anchor Magnitude Did Not (H2 and H3). Recall, I predicted that among pairs in the *unaccountable* condition, there would be a difference between anchor conditions in the number of mitigating and/or aggravating arguments discussed (H2), but that pairs in the *accountable* condition there would be little or no difference between anchor conditions (H3). To test these predictions, I fit negative binomial regression models with anchor magnitude and accountability conditions predicting number of aggravating and of mitigating arguments discussed. See [Table 8](#). In contrast to H2, but consistent with Experiment 1, there was no main effect of anchor condition in either model. This result provides the clearest evidence that the PTS does not lead decision-makers to selectively retrieve information either consistent or inconsistent with the PTS. The models also demonstrated a main effect of accountability, such that participants under process *accountability* discussed *more* arguments as mitigating ($M_{\text{diff}} = 0.75$, $d = 0.28$ 95% CI [0.036, 0.52]) and as aggravating ($M_{\text{diff}} = 0.57$, $d = 0.34$ [0.09, 0.58]) than those under process *unaccountability*. However, in contrast to H3, the Anchor by Accountability interaction was not significant in either model predicting evidence discussed. Thus, even under conditions of (un)accountability, attorney–client pairs did not discuss the evidence differently based on the magnitude of the PTS.

⁵ The sum of these two information types do not comprise 100% of information exchanged. Unexpectedly, 72% ($n = 193$) of attorney–client pairs engaged in discussion of at least one additional circumstance not explicitly provided to them that could influence their sentencing judgements or plea decision-making (e.g., willingness to publicly express remorse, legal actors' dispositions, dangers of being incarcerated during a global pandemic).

Figure 5
Means and 95% CI for Maximum Acceptable Sentence by Condition



Note. Error bars show 95% confidence intervals. CI = confidence interval.
*** $p < .001$.

Exploratory Analyses

Exploratory analyses found no effects of anchor or accountability on mock attorneys' or defendants' evaluations of the proximal case attributes. However, like Experiment 1, mock defendants and attorneys tended to view the plea deal as fairer ($d = 0.28$ 95% CI [0.11, 0.46]) and more lenient ($d = 0.41$ [0.24, 0.58]) when exposed to the high compared to the moderate trial sentence.

Table 6

Experiment 3 Negative Binomial Regression Model Results for the Effect of Anchor and Process Accountability on Joint Maximum Acceptable Sentences

Effect	<i>b</i>	95% CI	<i>SE</i>	<i>z</i>	<i>df</i>	<i>p</i>
Model 1						
Intercept	1.79	[1.63, 1.96]	0.08	21.35	262	<.001
Anchor ^a	−0.60	[−0.79, −0.40]	0.10	−5.90	262	<.001
Accountability ^b	0.02	[−0.18, 0.22]	0.10	0.18	262	.85
Model 2						
Intercept	1.77	[1.59, 1.96]	0.01	18.54	261	<.001
Anchor	−0.55	[−0.83, −0.26]	0.14	−3.80	261	<.001
Accountability	0.06	[−0.20, 0.32]	0.14	0.46	261	.664
Anchor × Accountability	−0.1	[−0.49, 0.30]	0.20	−0.48	261	.63
Model 3						
Intercept	0.64	[0.14, 1.13]	0.25	2.6	240	.01
Anchor	−0.71	[−0.90, −0.52]	0.10	−7.20	240	<.001
Accountability	0.14	[−0.048, 0.33]	0.10	1.46	240	.15
Baseline sentence ^c	0.01	[−0.003, 0.02]	0.004	1.52	240	.13
Baseline sentence ^d	0.07	[0.043, 0.10]	0.01	5.93	240	<.001
Likelihood of conviction ^c	0.01	[0.004, 0.01]	0.002	3.58	240	<.001
Likelihood of Conviction ^d	0.003	[−0.002, 0.01]	0.002	1.33	240	.18

^a Anchor: High (48 months) = 0 versus Moderate (8 months) = 1. ^b Accountability: Process accountability: accountable = 0 versus Unaccountable = 1.
^c Mock attorney (law student) preanchor response. ^d Mock defendant preanchor response.

Results of these and other analyses can be found in the [Supplemental Analyses](#).

Discussion

Results of Experiment 3 were consistent with those of Experiments 1 and 2, further supporting the *shallow* processing pathway to anchoring in the plea domain. Contrary to H3, a debiasing strategy focuses on increasing consideration of *all* evidence, rather than merely anchor-inconsistent mitigating evidence was ineffective. However, because process-accountable groups discussed, on average, just one argument more than unaccountable groups, further work may be needed to confidently rule out the efficacy of this strategy. Finally, consistent with the competing anchors hypothesis (Sniezek, 1992) and with prior work (e.g., de Wilde et al., 2018; Experiment 1, Epley & Gilovich, 2006; Whyte & Sebenius, 1997), anchor-unbiased sentencing preferences may have served as additional reference points when jointly determining an appropriate maximum acceptable sentence, thereby attenuating the anchoring effect. This finding suggests an easy-to-implement solution to reduce the anchoring influence of the PTS.

General Discussion

PTS Influence Plea Acceptance via Increased Maximum Acceptable Sentences

The first goal of the current work was to examine the extent to which mock defendants and legal professionals are susceptible to the anchoring bias. Results of three studies with varied methodologies and samples are consistent with the SoT model and prior empirical research: Legal decision-makers were willing to accept jail sentences that were significantly longer when exposed to a high compared to moderate PTS anchor (e.g., Bordens, 1984; Kramer et al., 2007; McAllister & Bregman, 1986; Zimmerman & Hunter, 2018; cf., Schneider & Zottoli, 2019). Although preliminary, the

Table 7*Experiment 3 Logistic Regression Predicting Likelihood of Final Plea Decision Based on Anchor, Accountability, Anchor × Accountability*

Effect	<i>B</i>	<i>SE</i>	Wald <i>z</i>	<i>df</i>	<i>p</i>	Odds ratio	95% CI for odds ratio		Overall model fit
							Lower	Upper	
Model 1									
Intercept	−0.28	0.22	−1.30	262	.192	0.75	0.49	1.15	$\chi^2(1) = 17.7, p < .001, R^2 = .09$
Anchor ^a	1.09	0.26	4.21	262	<.001	2.96	1.80	4.94	
Accountability ^b	0.08	0.26	0.31	262	.759	1.08	0.65	1.79	
Model 2									
Intercept	−0.39	0.25	−1.58	261	.115	0.68	0.41	1.09	$\chi^2(1) = 12.7, p < .001$
Anchor	1.33	0.37	3.57	261	<.001	3.79	1.85	8.01	
Accountability	0.30	0.35	0.86	261	.393	1.35	0.68	2.70	
Anchor × Accountability	−0.48	0.52	−0.93	261	.354	0.62	0.22	1.70	

Note. CI = confidence interval.

^a Anchor: *High* (48 months) = 0 versus *Moderate* (8 months) = 1. ^b Accountability: *Process accountability: Accountable* = 0 versus *Unaccountable* = 1.

contribution of Experiment 1 further demonstrates that legally uninformative anchors similarly altered sentencing preferences. Anchor-biased sentencing preferences, in turn, explained participants' choice to accept prosecutors' discounted plea offers at higher rates in the high compared to moderate anchor conditions. As Experiment 1 defendants were not presented with a relevant PTS they could not calculate a case-relevant plea discount further demonstrating that absent any alternative numeric sentencing values even case-irrelevant discounts may influence plea decisions.

Magnitude of the PTS Does Not Differentially Influence Appraisals of Evidence Strength

A second goal of the current work was to clarify the mechanism underlying the anchoring effect in plea negotiations. In contrast to predictions derived from the SAM, exposure to PTS of varying magnitudes—whether presented as a prosecutors' demand or as an arbitrarily chosen value—did not differentially bias judgments of related case attributes including perceived crime severity, evidence strength, or estimated likelihood of conviction. These results further suggest

that anchoring effects in plea contexts may be the result of less thoughtful processes rather than a deeper change to the underlying mental representation of attributes related to the PTS. This finding is important as studies using self-report, experimental, and archival methods demonstrate that perceived evidence strength is a critical determinant of defendants' (e.g., Champion, 1989; Peterson-Badali & Abramovitch, 1993; Viljoen et al., 2005) and defense attorneys', prosecutors', and judges' plea decisions (e.g., Bushway et al., 2014; Kutateladze et al., 2015; Redlich et al., 2016). Instead, the PTS seems to inflate participant defendant's perceptions of the leniency of the plea deal even when no relevant PTS is offered to judge the relative fairness of the deal. Only attorneys who generated evidence contradicting the appropriateness of the PTS were inoculated from judging the plea deal as fairer.

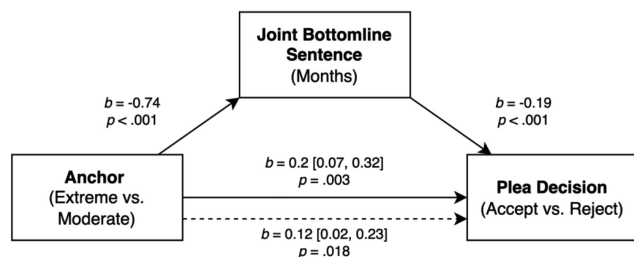
Increased Consideration of Anchor-Inconsistent Evidence Reduces the Anchoring Bias, but Consideration of All Evidence Does Not

In line with my predictions and prior work, instructions to consider anchor-inconsistent (i.e., mitigating) evidence led to a reduction of the anchoring bias, such that a quarter of defense attorneys exposed to the *high* PTS anchor reduced their sentencing recommendations by an average of 4 months. Reduced maximum sentences in the high (but not moderate) anchor condition were, in turn, associated with a lower rate of plea acceptances. Experiment 2 shows interventions designed to increase consideration of *mitigating* evidence can reduce the arbitrary influence of a high PTS.

In contrast, Experiment 3 demonstrated that more thorough information processing did not change the type or quantity of information discussed by attorney-client pairs across anchor conditions, let alone their sentencing judgments and plea decisions. Though the inefficacy of the accountability manipulation in the present work is consistent with prior work in nonlegal settings (Brewer et al., 2007; Whyte & Sebenius, 1997), it is inconsistent with de Wilde et al.'s (2018) findings. One explanation may be that while increased accountability works to debias sentencing judgments made by mock jurors about a third-party defendant, it may be less effective for debiasing anchoring in more preferential decision-making. Results from these interventions are in line with the handful of studies demonstrating that explicit instructions to consider the

Figure 6

Mediation Model of Anchor Predicting Plea Decision Mediated by Maximum Acceptable Sentence



Note. Unstandardized estimates are displayed, with 95% confidence intervals and *p*-values. Anchor: *High* = 0; *Moderate* = 1; Plea Decision: 0 = *Accept*; 1 = *Reject*. Dashed line represents the indirect effect of anchor magnitude through sentencing judgments controlling for covariates (accountability condition, attorney and defendants' individual baseline sentencing estimates, and estimated likelihood of conviction) across all paths of the model. Estimates are based on bootstrapping procedure with 10,000 bootstrap samples.

Table 8

Experiment 3 Negative Binomial Regression Model Results for the Effect of Anchor and Process Accountability on Number of Mitigating and Aggravating Arguments Discussed

Effect	<i>b</i>	95% CI	<i>SE</i>	<i>z</i>	<i>df</i>	<i>p</i>
Mitigating arguments						
Model 1						
Intercept	1.55	[1.42, 1.68]	0.07	23.31	262	<.001
Anchor ^a	−0.05	[−0.20, 0.11]	0.08	−0.62	262	.533
Accountability ^b	−0.18	[−0.33, −0.03]	0.08	−2.29	262	.022
Model 2						
Intercept	1.51	[1.36, 1.66]	0.08	19.75	261	<.001
Anchor	0.03	[−0.18, 0.25]	0.11	0.30	261	.765
Accountability	−0.10	[−0.31, 0.12]	0.11	−0.87	261	.384
Anchor × Accountability	−0.17	[−0.48, 0.14]	0.16	−1.08	261	.279
Aggravating arguments						
Model 1						
Intercept	0.95	[0.81, 1.10]	0.07	12.81	262	<.001
Anchor	0.01	[−0.16, 0.19]	0.09	0.13	262	.89
Accountability	−0.24	[−0.42, −0.07]	0.09	−2.73	262	.006
Model 2						
Intercept	0.95	[0.79, 1.12]	0.08	11.29	261	<.001
Anchor	0.01	[−0.23, 0.25]	0.12	0.09	261	.929
Accountability	−0.25	[−0.50, 0.002]	0.13	−1.94	261	.053
Anchor × Accountability	0.003	[−0.35, 0.35]	0.18	−0.01	261	.988

^a Anchor: *High* (48 months) = 0 versus *Moderate* (8 months) = 1. ^b Accountability: *Process accountability: Accountable* = 0 versus *Unaccountable* = 1.

information *contradicting* the anchor are more effective for debiasing the anchoring effect than other less direct accuracy warnings (Adame, 2016; Mussweiler et al., 2000; Stein & Drouin, 2018).

Theoretical Contributions

Deep Versus Shallow Cognitive Processing Pathways to Anchoring

The current work makes theoretical contributions by extending our understanding of the context and mechanism through which numerical anchors influence numerical sentencing judgments and choice decisions. Overall, results support the *shallow* processing pathway to anchoring in a preferential decision-making plea context. Lacking significant bargaining power, mock defendants and legal experts may pragmatically treat the PTS—regardless of its relevance—as a “hint” to a plausibly acceptable sentence (Blankenship et al., 2008; Wegener et al., 2001). Though prior work has demonstrated that inadmissible or meaningless anchors bias mock jurors’ and legal experts’ punitive award damages (Hans et al., 2018; Reyna et al., 2015) and sentencing judgments (e.g., Englich, 2006; Glöckner & Englich, 2015), the current work is the first to demonstrate the impact of irrelevant anchors on a nonnumeric choice outcome (plea decision) for participants rendering preferential decisions.

The current work also fails to support the *deep* processing pathway to anchoring described by the SAM by finding that anchor magnitude did not bias interpretations of the case facts. Prior work examining the impact of anchor-biased monetary damages on perceptions of related case attributes has yielded mixed results in the civil arena (e.g., Marti & Wissler, 2000). Chapman and Bornstein (1996) found that a high (vs. low) demand for damages changed how defendants saw the plaintiff (i.e., more selfish and less generous), rather than how they viewed the case. Thus, it may be fruitful

to test whether the PTS influences the attributions defendants make about the prosecutor’s character, intentions, evaluations of the case, and future offers. Defendants may make an *external* attribution about the reason for threatening a severe PTS rather than the *internal* attribution about the case facts predicted here (Jones & Davis, 1965). Indeed, some participants speculated that the prosecutors’ intent in charging so severely was related more to their personal motives or sexism than to the defendants’ case: “Defendant-292: Must be an election year . . . does the prosecutor want me to make an anonymous donation to his reelection campaign?... I think the prosecutor’s bluffing. I think he’s trying to take advantage of you [attorney], because you’re a young woman and he’s trying to bully you.”

The absence of a relationship between the numeric anchor and qualitative judgments about the case is theoretically important as it suggests that a dominant explanation for anchoring effects—the SAM—fails to explain anchoring that occurs in plea contexts among decision-makers rendering preferential judgments. To wit, the current work does not discount the possibility that anchoring in plea contexts is the result or cause of any kind of thoughtful processes, but it does provide strong evidence against selective retrieval. If thoughtful processes are responsible for anchoring, we would expect differences in case evaluations not based on the *magnitude* of the anchor but rather based on the *availability* of an anchor. For example, participants may approach the task of determining a maximum acceptable sentence thoughtfully—regardless of anchor’s magnitude or presence—when the context calls for imagining oneself as the principal actor in a plea negotiation. However, perhaps when participants under self-interested circumstances are *not* presented with an anchor, then that same thoughtful estimation process may lead to more variable case evaluations that differ from those of participants exposed to an anchor. Future work could test this prediction by including a control condition that omits an anchor in self-interested contexts.

Implausibly High Anchors

The current work also furthers our understanding of the impact of implausibly high anchors on numerical judgments and choice decisions. Prior work has demonstrated that implausible anchors sometimes lead to larger anchoring effects (Chapman & Johnson, 1994; Mussweiler & Strack, 2000, 2001; Strack & Mussweiler, 1997), attenuated effects (Bystranowski et al., 2021; Sugden et al., 2013), or even backfire effects (Wegener et al., 2001). Though legally plausible, the high PTS in Experiment 3 of 48 months far exceeded the distribution of unanchored sentencing preferences reported in pilot testing. The finding that decision-makers exposed to this harsh PTS considered the ways in which the anchor was *inappropriate*—by discussing more anchor-inconsistent (i.e., mitigating) evidence than the SAM would have predicted, but not significantly more than those in the 8-month anchor condition—suggests the kindling for a backfire effect was present. However, the 48-month anchor, which one defendant referred to as “insulting,” still caused a powerful anchoring effect—a finding that argues *against* the backfire effect in plea negotiations and suggests that even absurdly high anchors may unduly influence mock defendants in plea contexts (see Chapman & Johnson, 1994; Mussweiler & Strack, 2000, 2001; Strack & Mussweiler, 1997).

However, what can explain the lack of a backfire effect of this “insulting” trial sentence on nonbinding sentencing preferences? One explanation may be that though participants were not told that their stated sentencing preferences would directly influence the prosecutor’s subsequent plea offer, they may nonetheless have anticipated a causal effect of their stated preference. Attorney–client pairs exposed to the high anchor may not have provided lower sentencing preferences because they did not want to risk prosecutorial retaliation in the form of other disagreeable plea offers or the termination of plea negotiations altogether.⁶ Such fears might well be justified.

Legal Implications: Attorneys as “Debiasers”

The *Brady* Court stipulated that offers of leniency must not be so enticing as to “overbear[] the will of the defendant” (p. 750) and that defendants should, when assisted by legal counsel, be capable of approaching plea negotiations “rationally” (p. 750). The current work replicates earlier studies demonstrating that legal experts, primarily in the civil area of law (e.g., Bystranowski et al., 2021; Englich et al., 2006; Englich & Soder, 2009; Xifen & Yong, 2018) and negotiation experts (Orr & Guthrie, 2006) show similar susceptibility to anchoring biases. By examining legal experts working closely with their clients, this work is the first to test these earlier claims in a novel context involving groups of novice and expert plea decision-makers. However, why were real and mock attorneys incapable of protecting defendants against cognitive errors as legal scholars have argued (e.g., Bibas, 2004; Gigerenzer, 1991; Scott & Stuntz, 1992)? Ironically, one reason may be their faithfulness to the criminal defense practice of deferring to their client’s wishes, a finding reflected by the lower variance in joint sentencing preferences explained by attorney’s (individually provided, preanchor) sentencing preferences compared to that of their clients’ (see Kramer et al., 2007). Thus, this work suggests that despite their best intentions to assist clients, defense attorneys may be vulnerable to the same cognitive errors that likely afflict their clients.

Strengths, Limitations, and Future Directions

Though mock plea decision-making studies cannot completely model true plea negotiations, the current studies were designed to generalize to the legal domain (Sales & Krauss, 2015). Several methodological and sampling choices enhanced the internal and ecological validity of the current work, including the factually complex nature of the plea scenario stimuli necessary to test predictions of selective retrieval of information, as well as the recruitment of expert legal decision-makers. Additionally, the third experiment is the first to examine the plea negotiation decision-making process in a novel paradigm with role-playing attorney–client pairs, which is important because attorneys are known to influence defendant decision-making (see Henderson & Levett, 2018).

Other choices that increased internal validity, such as use of the standard anchoring paradigm (Strack & Mussweiler, 1997; Tversky & Kahneman, 1974), ensured that participants would process and consider the provided anchor value, may have reduced the practical applicability of this study. Future research should investigate if defendants and attorneys naturally form judgments about the harshness of the prosecutor’s threatened charge(s) and PTS, as well as the maximum sentence they would be willing to serve. Finally, the finding that mock defendants in Experiment 1 regarded the aggravating evidence as overall stronger than the mitigating evidence whereas the opposite was true of attorney–client pairs in Experiment 3 suggests that the modality of asynchronous vignette studies may lead to qualitatively different results than more interactive paradigms. This discrepancy suggests a need for more immersive paradigms that may more accurately reflect the decision-making of truly self-interested decision-makers (see also Wilford et al., 2021).

Moreover, it is also important to examine the effect of anchors using alternative nonnumeric modalities related to strategic overcharging, such as charges and punishment outcomes (e.g., community service, probation, legal fines). Some work demonstrates a biasing effect of cross-modal anchors—images priming magnitude (Oppenheimer et al., 2008), objects denoting monetary value (Mochon & Frederick, 2013), and damages (Chapman & Bornstein, 1996). However, other work has failed to find anchoring effects when the anchor did not match the provided response scale (Chapman & Johnson, 1994; Frederick & Mochon, 2012). It may be that, as Chapman and Bornstein (1996) have argued, anchoring is more likely to occur when the anchor and the response scale are compatible. As a higher criminal charge is likely to be appraised as compatible with a higher sentencing judgment (especially if a sentencing range is provided), overcharging should elicit a cross-modal anchoring effect.

Conclusion

Recent years have seen a rise in efforts among legal practitioners and justice-involved citizens to bring public awareness to the issue of

⁶ One mock attorney eloquently articulated this very fear to his client (“The risk here is that if you go to trial, the DA is going to be butthurt that you didn’t take the deal and recommend a higher sentence”). In another instance, a defendant related the same concern to her attorney: “My brain wants to say one [month], but I don’t know if that would make somebody on the prosecuting side really angry and want to slap me with more time as retaliation for saying one month.”

overcharging (Reimer & Sabelli, 2019). Simultaneously, the psychological field has begun to consider how its vast accumulation of theoretical insights can enhance our understanding of psychological and structural issues that engender coercion in plea decision-making (Redlich et al., 2017). The current studies challenge the assumption that attorneys are the solution to preventing coercion in the face of constitutionally permissible threats and leniency and provide preliminary support for two low-cost interventions that could easily be incorporated into internal office policies. Attorneys and defendants could first generate an evidence-informed preference for appropriate punishment *prior* to hearing the prosecutors' PTS and then systematically consider the case circumstances that *contradict* a PTS resulting from overcharging. Together, these strategies may decrease the evaluation of appropriate punishment and reduce plea acceptances. In sum, this work joins critics of the coercive impact of overcharging in their critique of the immense discretionary power of prosecutors by providing an evidence-based argument for issuing charges proportional to the alleged misconduct (e.g., Brown, 2012; Devers, 2011).

References

- Adame, B. J. (2016). Training in the mitigation of anchoring bias: A test of the consider-the-opposite strategy. *Learning and Motivation*, 53, 36–48. <https://doi.org/10.1016/j.lmot.2015.11.002>
- Albonetti, C. A. (1990). Race and the probability of pleading guilty. *Journal of Quantitative Criminology*, 6(3), 315–334. <https://www.jstor.org/stable/23365599>
- Alschuler, A. W. (1968). The prosecutor's role in plea bargaining. *The University of Chicago Law Review*, 36(1), 50–112. <https://doi.org/10.2307/1598832>
- American Bar Association. (2016). *Model rules of professional conduct*. https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/model_rules_of_professional_conduct_table_of_contents.html
- 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., & Walker, S. (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 trial. *Harvard Law Review*, 117(8), 2463–2547. <https://doi.org/10.2307/4093404>
- Blankenship, K. L., Wegener, D. T., Petty, R. E., Detweiler-Bedell, B., & Macy, C. L. (2008). Elaboration and consequences of anchored estimates: An attitudinal perspective on numerical anchoring. *Journal of Experimental Social Psychology*, 44(6), 1465–1476. <https://doi.org/10.1016/j.jesp.2008.07.005>
- 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/s15324834basop0501_4
- Brewer, N. T., Chapman, G. B., Schwartz, J. A., & Bergus, G. R. (2007). The influence of irrelevant anchors on the judgments and choices of doctors and patients. *Medical Decision Making*, 27(2), 203–211. <https://doi.org/10.1177/0272989X06298595>
- Brown, D. K. (2012). American prosecutors' powers and obligations in the era of plea bargaining. In E. Luna & M. Wade (Eds.), *The prosecutor in transnational perspective* (Oxford Scholarship Online ed., pp. 200–213). <https://doi.org/10.1093/acprof:osobl/9780199844807.003.0015>
- 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>
- Bystranowski, P., Janik, B., Próchnicki, M., & Skórska, P. (2021). Anchoring effect in legal decision-making: A meta-analysis. *Law and Human Behavior*, 45(1), 1–23. <https://doi.org/10.1037/lhb0000438>
- Caldwell, H. M. (2011). Coercive plea bargaining: The unrecognized scourge of the justice system. *Catholic University Law Review*, 61(1), 63–96. <https://scholarship.law.edu/lawreview/vol61/iss1/2>
- Cardenas, S. A. (2021). *Charged up and anchored down: Structural and psychological processes influencing defendant and defense attorney plea decision-making* [Data sets for Studies 1–3]. Open Science Framework. <https://osf.io/h6ktg>
- Champion, D. J. (1989). Private counsels and public defenders: A look at weak cases, prior records, and leniency in plea bargaining. *Journal of Criminal Justice*, 17(4), 253–263. [https://doi.org/10.1016/0047-2352\(89\)90031-7](https://doi.org/10.1016/0047-2352(89)90031-7)
- Chapman, G. B., & Bornstein, B. H. (1996). The more you ask for, the more you get: Anchoring in personal injury verdicts. *Applied Cognitive Psychology*, 10(6), 519–540. [https://doi.org/10.1002/\(SICI\)1099-0720\(199612\)10:6<519::AID-ACP417>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1099-0720(199612)10:6<519::AID-ACP417>3.0.CO;2-5)
- Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, 7(4), 223–242. <https://doi.org/10.1002/bdm.3960070402>
- Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes*, 79(2), 115–153. <https://doi.org/10.1006/obhd.1999.2841>
- Clatch, L., & Borgida, E. (2021). Plea bargaining: A test of dual discounting preferences for non-monetary losses. *Personality and Social Psychology Bulletin*, 47(7), 1039–1056. <https://doi.org/10.1177/0146167220952228>
- Critcher, C. R., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making*, 21(3), 241–251. <https://doi.org/10.1002/bdm.586>
- Davis, J. H., Stasser, G., Spitzer, C. E., & Holt, R. W. (1976). Changes in group members' decision preferences during discussion: An illustration with mock juries. *Journal of Personality and Social Psychology*, 34(6), 1177–1187. <https://doi.org/10.1037/0022-3514.34.6.1177>
- de Dreu, C. K. W., Koole, S. L., & Steinel, W. (2000). Unfixing the fixed pie: A motivated information-processing approach to integrative negotiation. *Journal of Personality and Social Psychology*, 79(6), 975–987. <https://doi.org/10.1037/0022-3514.79.6.975>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's *t*-test instead of Student's *t*-test. *International Review of Social Psychology*, 30(1), 92–101. <https://doi.org/10.5334/irsp.82>
- Devers, L. (2011, January 24). *Research summary: Plea and charge bargaining*. Bureau of Justice Assistance. <https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/PleaBargainingResearchSummary.pdf>
- de Wilde, T. R. W., Ten Velden, F. S., & De Dreu, C. K. W. (2018). The anchoring-bias in groups. *Journal of Experimental Social Psychology*, 76, 116–126. <https://doi.org/10.1016/j.jesp.2018.02.001>
- Downen, T., Furner, Z., & Cataldi, B. (2019). The effects on anchoring of increasing quantities of disconfirming evidence. *International Journal of Management and Decision Making*, 18(3), 309–331. <https://doi.org/10.1504/IJMDM.2019.100536>
- Edkins, V. A., & Dervan, L. E. (2018). Freedom now or a future later: Pitting the lasting implications of collateral consequences against pretrial detention in decisions to plead guilty. *Psychology, Public Policy, and Law*, 24(2), 204–215. <https://doi.org/10.1037/law0000159>
- Englich, B. (2006). Blind or biased? Justitia's susceptibility to anchoring effects in the courtroom based on given numerical representations. *Law and Policy*, 28(4), 497–514. <https://doi.org/10.1111/j.1467-9930.2006.00236.x>
- Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, 32(2), 188–200. <https://doi.org/10.1177/0146167205282152>

- Englich, B., & Soder, K. (2009). Moody experts—How mood and expertise influence judgmental anchoring. *Judgment and Decision Making*, 4(1), 41–50. <https://doi.org/10.1017/S1930297500000693>
- Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making*, 18(3), 199–212. <https://doi.org/10.1002/bdm.495>
- Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, 17(4), 311–318. <https://doi.org/10.1111/j.1467-9280.2006.01704.x>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- FindLaw. (2018). *Examples of two drunk driving cases*. <https://www.findlaw.com/dui/charges/the-tale-of-two-drunk-driving-cases.html>
- Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General*, 141(1), 124–133. <https://doi.org/10.1037/a0024006>
- Fudenberg, D., Levine, D. K., & Maniadi, Z. (2012). On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics*, 4(2), 131–145. <https://doi.org/10.1257/mic.4.2.131>
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42. <https://doi.org/10.1016/j.soec.2010.10.008>
- Galinsky, A. D., & Mussweiler, T. (2001). First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology*, 81(4), 657–669. <https://doi.org/10.1037/0022-3514.81.4.657>
- 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>
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases”. *European Review of Social Psychology*, 2(1), 83–115. <https://doi.org/10.1080/14792779143000033>
- Glöckner, A., & Englich, B. (2015). When relevance matters. *Social Psychology*, 46(1), 4–12. <https://doi.org/10.1027/1864-9335/a000214>
- Gramlich, J. (2019, June 11). *Only 2% of federal criminal defendants go to trial, and most who do are found guilty*. Pew Research Center. <https://www.pewresearch.org/fact-tank/2019/06/11/only-2-of-federal-criminal-defendants-go-to-trial-and-most-who-do-are-found-guilty/>
- 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>
- Hans, V. P., Helm, R. K., & Reyna, V. F. (2018). From meaning to money: Translating injury into dollars. *Law and Human Behavior*, 42(2), 95–109. <https://doi.org/10.1037/lhb0000282>
- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Henderson, K. S., & Levett, L. M. (2018). Investigating predictors of true and false guilty pleas. *Law and Human Behavior*, 42(5), 427–441. <https://doi.org/10.1037/lhb0000297>
- Henderson, K. S., & Levett, L. M. (2019). Plea bargaining: The Influence of counsel. In B. H. Bornstein & M. K. Miller (Eds.), *Advances in psychology and law: Volume 4* (pp. 73–100). Springer. https://doi.org/10.1007/978-3-030-11042-0_3
- Henderson, K. S., & Shteynberg, R. V. (2020). Plea decision-making: The influence of attorney expertise, trustworthiness, and recommendation. *Psychology, Crime and Law*, 26(6), 527–551. <https://doi.org/10.1080/1068316X.2019.1696801>
- Innocence Project. (2019). *All cases*. <https://www.innocenceproject.org/all-cases/>
- Joel, S., Spielmann, S. S., & MacDonald, G. (2017). Motivated use of numerical anchors for judgments relevant to the self. *Personality and Social Psychology Bulletin*, 43(7), 972–985. <https://doi.org/10.1177/0146167217702613>
- 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.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). Academic.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–292. <https://doi.org/10.2307/1914185>
- Khogali, M., Jones, K., & Penrod, S. (2018). Fairness for all? Public perceptions of plea bargaining. *Applied Psychology in Criminal Justice*, 14(2), 136–153.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š, Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cernalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., ... Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology*, 45(3), 142–152. <https://doi.org/10.1027/1864-9335/a000178>
- Korobkin, R., & Guthrie, C. (1997). Psychology, economics, and settlement: A new look at the role of the lawyer. *Texas Law Review*, 76(1), 77–142. <https://scholarship.law.vanderbilt.edu/faculty-publications/703>
- Kramer, G. M., Wolbransky, M., & Heilbrun, K. (2007). Plea bargaining recommendations by criminal defense attorneys: Evidence strength, potential sentence, and defendant preference. *Behavioral Sciences and the Law*, 25(4), 573–585. <https://doi.org/10.1002/bsl.759>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Kutateladze, B. L., Lawson, V. Z., & Andiloro, N. R. (2015). Does evidence really matter? An exploratory analysis of the role of evidence in plea bargaining in felony drug cases. *Law and Human Behavior*, 39(5), 431–442. <https://doi.org/10.1037/lhb0000142>
- Lafler v. Cooper, 566 U.S. 156 (2012). <https://www.oyez.org/cases/2011/10-209>
- Lakens, D., & Caldwell, A. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, 4(1). <https://doi.org/10.1177/2515245920951503>
- Lenth, R. V. (2021). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.6.0) [Computer software]. <https://CRAN.R-project.org/package=emmeans>
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>
- Lesnoff, M., & Lancelot, R. (2012). *aod: Analysis of overdispersed data* (R package Version 1.3.1) [Computer software]. <http://cran.r-project.org/package=aod>
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- Marti, M. W., & Wissler, R. L. (2000). Be careful what you ask for: The effect of anchors on personal-injury damages awards. *Journal of Experimental Psychology: Applied*, 6(2), 91–103. <https://doi.org/10.1037/1076-898X.6.2.91>
- McAllister, H. A., & Bregman, N. J. (1986). Plea bargaining by prosecutors and defense attorneys: A decision theory approach. *Journal of Applied Psychology*, 71(4), 686–690. <https://doi.org/10.1037/0021-9010.71.4.686>
- Meyer, J. A., & Gray, T. (1997). Drunk drivers in the courts: Legal and extra-legal factors affecting pleas and sentences. *Journal of Criminal Justice*, 25(2), 155–163. [https://doi.org/10.1016/S0047-2352\(96\)00058-X](https://doi.org/10.1016/S0047-2352(96)00058-X)

- 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>
- Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes*, 122(1), 69–79. <https://doi.org/10.1016/j.obhdp.2013.04.002>
- Mussweiler, T., & Strack, F. (1999a). Comparing is believing: A selective accessibility model of judgmental anchoring. *European Review of Social Psychology*, 10(1), 135–167. <https://doi.org/10.1080/14792779943000044>
- Mussweiler, T., & Strack, F. (1999b). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, 35(2), 136–164. <https://doi.org/10.1006/jesp.1998.1364>
- Mussweiler, T., & Strack, F. (2000). The “relative self”: Informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology*, 79(1), 23–38. <https://doi.org/10.1037/0022-3514.79.1.23>
- Mussweiler, T., & Strack, F. (2001). The semantics of anchoring. *Organizational Behavior and Human Decision Processes*, 86(2), 234–255. <https://doi.org/10.1006/obhd.2001.2954>
- Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, 26(9), 1142–1150. <https://doi.org/10.1177/01461672002611010>
- Oppenheimer, D. M., LeBoeuf, R. A., & Brewer, N. T. (2008). Anchors aweigh: A demonstration of cross-modality anchoring and magnitude priming. *Cognition*, 106(1), 13–26. <https://doi.org/10.1016/j.cognition.2006.12.008>
- Orr, D., & Guthrie, C. (2006). Anchoring, information, expertise, and negotiation: New insights from meta-analysis. *Ohio State Journal on Dispute Resolution*, 21(3), 597–628.
- Patil, S. V., Vieider, F., & Tetlock, P. E. (2014). Process versus outcome accountability. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford handbook of public accountability* (pp. 69–89). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199641253.013.0002>
- Petegorsky, M. N. (2012). Plea bargaining in the dark: The duty to disclose exculpatory Brady evidence during plea bargaining note. *Fordham Law Review*, 81(6), 3599–3650. <https://ir.lawnet.fordham.edu/flr/vol81/iss6/13>
- Petersen, K., Redlich, A. D., & Norris, R. J. (2022). Diverging from the shadows: Explaining individual deviation from plea bargaining in the “shadow of the trial”. *Journal of Experimental Criminology*, 18(2), 321–342. <https://doi.org/10.1007/s11292-020-09449-4>
- Peterson-Badali, M., & Abramovitch, R. (1993). Grade related changes in young people’s reasoning about plea decisions. *Law and Human Behavior*, 17(5), 537–552. <https://doi.org/10.1007/BF01045072>
- Pogarsky, G., & Babcock, L. (2001). Damage caps, motivated anchoring, and bargaining impasse. *The Journal of Legal Studies*, 30(1), 143–159. <https://doi.org/10.1086/468114>
- R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Redlich, A. D., Bibas, S., Edkins, V. A., & Madon, S. (2017). The psychology of defendant plea decision making. *American Psychologist*, 72(4), 339–352. <https://doi.org/10.1037/a0040436>
- Redlich, A. D., Bushway, S. D., & Norris, R. J. (2016). Plea decision-making by attorneys and judges. *Journal of Experimental Criminology*, 12(4), 537–561. <https://doi.org/10.1007/s11292-016-9264-0>
- Reimer, N. L., & Sabelli, M. A. (2019). The tyranny of the trial penalty: The consensus that coercive plea practices must end. *Federal Sentencing Reporter*, 31(4–5), 215–221. <https://doi.org/10.1525/fsr.2019.31.4-5.215>
- Reyna, V. F., Hans, V. P., Corbin, J. C., Yeh, R., Lin, K., & Royer, C. (2015). The gist of juries: Testing a model of damage award decision making. *Psychology, Public Policy, and Law*, 21(3), 280–294. <https://doi.org/10.1037/law0000048>
- Sales, B. D., & Krauss, D. A. (2015). Using valid and reliable scientific methods. In B. D. Sales & D. A. Krauss (Eds.), *The psychology of law: Human behavior, legal institutions, and law* (pp. 95–116). American Psychological Association. <https://doi.org/10.1037/14593-006>
- Schneider, R. A., & Zottoli, T. M. (2019). Disentangling the effects of plea discount and potential trial sentence on decisions to plead guilty. *Legal Criminological Psychology*, 24(2), 288–304. <https://doi.org/10.1111/lcrp.12157>
- Scholten, L., van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. W. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, 43(4), 539–552. <https://doi.org/10.1016/j.jesp.2006.05.010>
- Scott, R. E., & Stuntz, W. J. (1992). Plea bargaining as contract. *The Yale Law Journal*, 101(8), 1909–1968. <https://doi.org/10.2307/796952>
- Signorell, A. (2021). *DescTools: Tools for descriptive statistics* (Version 0.99.41) [Computer software]. <https://cran.r-project.org/package=DescTools>
- Sniezek, J. A. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes*, 52(1), 124–155. [https://doi.org/10.1016/0749-5978\(92\)90048-C](https://doi.org/10.1016/0749-5978(92)90048-C)
- Standen, J. (1993). Plea bargaining in the shadow of the guidelines. *California Law Review*, 81(6), 1471–1540. <https://doi.org/10.2307/3480956>
- Stein, C. T., & Drouin, M. (2018). Cognitive bias in the courtroom: Combating the anchoring effect through tactical debiasing. *University of San Francisco Law Review*, 52(3), 393–427. <https://heinonline.org/HOL/P?h=hein.journals/usflr52&i=411&a=ampheS5jdW55LmVkdQ>
- Strack, F., Bahník, Š., & Mussweiler, T. (2016). Anchoring: Accessibility as a cause of judgmental assimilation. *Current Opinion in Psychology*, 12, 67–70. <https://doi.org/10.1016/j.copsyc.2016.06.005>
- Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, 73(3), 437–446. <https://doi.org/10.1037/0022-3514.73.3.437>
- Sugden, R., Zheng, J., & Zizzo, D. J. (2013). Not all anchors are created equal. *Journal of Economic Psychology*, 39, 21–31. <https://doi.org/10.1016/j.joep.2013.06.008>
- Ten Velden, F. S., Beersma, B., & De Dreu, C. K. W. (2010). It takes one to tango: The effects of dyads’ epistemic motivation composition in negotiation. *Personality and Social Psychology Bulletin*, 36(11), 1454–1466. <https://doi.org/10.1177/0146167210383698>
- Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, 46(4), 285–292. <https://doi.org/10.2307/3033716>
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59(5), 1–38. <https://doi.org/10.18637/jss.v059.i05>
- 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/jels.2010.7.issue-1>
- Townson, C. (2019). *The anchoring effect: A meta-analysis* (Publication No. 2311653546) [Doctoral dissertation, Michigan State University]. ProQuest Dissertations & Theses Global. <https://doi.org/10.25335/k9wb-sg48>
- Turner, B. M., & Schley, D. R. (2016). The anchor integration model: A descriptive model of anchoring effects. *Cognitive Psychology*, 90, 1–47. <https://doi.org/10.1016/j.cogpsych.2016.07.003>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>

- United States Sentencing Commission. (2022). *2022 Annual report and sourcebook of federal sentencing statistics*. <https://www.ussc.gov/sites/default/files/pdf/research-and-publications/annual-reports-and-sourcebooks/2022/2022-Annual-Report-and-Sourcebook.pdf>
- Venables, W. N. and Ripley, B. D. (2002) *Modern applied statistics with S* (4th ed.). Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>
- Viljoen, J. L., Klaver, J., & Roesch, R. (2005). Legal decisions of preadolescent and adolescent defendants: Predictors of confessions, pleas, communication with attorneys, and appeals. *Law and Human Behavior*, 29(3), 253–277. <https://doi.org/10.1007/s10979-005-3613-2>
- Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010). Elaboration and numerical anchoring: Implications of attitude theories for consumer judgment and decision making. *Journal of Consumer Psychology*, 20(1), 5–16. <https://doi.org/10.1016/j.jcps.2009.12.003>
- Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W. B. G. (2001). Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology*, 37(1), 62–69. <https://doi.org/10.1006/jesp.2000.1431>
- Whyte, G., & Sebenius, J. K. (1997). The effect of multiple anchors on anchoring in individual and group judgment. *Organizational Behavior and Human Decision Processes*, 69(1), 75–85. <https://doi.org/10.1006/obhd.1996.2674>
- Wilford, M. M., Sutherland, K. T., Gonzales, J. E., & Rabinovich, M. (2021). Guilt status influences plea outcomes beyond the shadow-of-the-trial in an interactive simulation of legal procedures. *Law and Human Behavior*, 45(4), 271–286. <https://doi.org/10.1037/lhb0000450>
- Xifen, L., & Yong, M. (2018). Sentencing recommendations, anchoring effect and fairness in criminal justice: An empirical study based on a sample of 520 sentences in K city. *Social Sciences in China*, 39(3), 149–170. <https://doi.org/10.1080/02529203.2018.1483109>
- Yoon, S., Fong, N. M., & Dimoka, A. (2019). The robustness of anchoring effects on preferential judgments. *Judgment and Decision Making*, 14(4), 470–487. <https://doi.org/10.1017/S1930297500006148>
- 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.2018.23.issue-1>

Received July 29, 2022

Revision received March 17, 2023

Accepted March 19, 2023 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <https://my.apa.org/portal/alerts/> and you will be notified by e-mail when issues of interest to you become available!