American Psychological Association
American Psychology-Law Society

Law and Human Behavior

© 2021 American Psychological Association
ISSN: 0147-7307

2021, Vol. 45, No. 1, 1–23
https://doi.org/10.1037/lhb0000438

Anchoring Effect in Legal Decision-Making: A Meta-Analysis

Piotr Bystranowski[1], Bartosz Janik[2], Maciej Próchnicki[3], and Paulina Skórska[4]

[1] Interdisciplinary Centre for Ethics, Jagiellonian University
[2] Faculty of Law and Administration, University of Silesia in Katowice
[3] Faculty of Law and Administration, Jagiellonian University
[4] Centre for Evaluation and Analysis of Public Policies, Jagiellonian University

Objective: We conducted a meta-analysis to examine whether numeric decision-making in law is susceptible to the effect of (possibly arbitrary) values present in the decision contexts (anchoring effect) and to investigate which factors might moderate this effect. ***Hypotheses:*** We predicted that the presence of numeric anchors would bias legal decision-makers' judgment in the direction of the anchor value. We hypothesized that the effect size of anchoring would be moderated by several variables, which we grouped into three categories: methodological (type of stimuli; type of sample), psychological (standard vs. basic paradigm; anchor value; type of scale on which the participants assessed the target value), and legal (relevance of the anchor; type of the anchor; area of law to which the presented case belonged; presence of any salient numeric values other than the main anchor). ***Method:*** Twenty-nine studies (93 effect sizes; N = 8,549) met the inclusion criteria. We divided them into two groups, depending on whether they included a control group, and calculated the overall effect size using a random-effects Model with robust variance estimation. We assessed the influence of moderators using random effects metaregression. ***Results:*** The overall effect sizes of anchoring for studies with a control group (z = .27, 95% CI [.21, .33], d = .58, 95% CI [.44, .73]) and without a control group (z = .39, 95% CI [.31, .47], d = .91, 95% CI [.69, 1.12]) were both significant, although we provide some evidence of possible publication bias. We found preliminary evidence of a potential moderating effect of some legally relevant factors, such as legal expertise or the anchor relevance. ***Conclusions:*** Existing research indicates anchoring effects exist in legal contexts. The influence of anchors seems to depend on some situational factors, which paves the way for future research on countering the problematic effect in legal settings.

Public Significance Statement

Our review corroborates the thesis that numeric decisions in law (such as damages or prison terms) are susceptible to the effect of salient numbers present in the decision context. Such anchoring effects might have undesirable consequences, possibly making court rulings biased or erratic. Our results, however, suggest that the effect might be moderated by a number of factors, which might be used by lawmakers to limit the influence of undesirable anchors or by attorneys to calibrate their demands.

Keywords: anchoring effect, meta-analysis, legal decision-making, judges and juries

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Bradley D. McAuliff served as Action Editor.

Piotr Bystranowski https://orcid.org/0000-0001-5539-1342
Bartosz Janik https://orcid.org/0000-0002-7471-6402
Maciej Próchnicki https://orcid.org/0000-0002-2151-9915
Paulina Skórska https://orcid.org/0000-0001-9037-3190

We presented earlier drafts of this article at the 35th Annual Conference of the European Association of Law and Economics, Milan, Italy, and at the conference Law and Mind 2, Krakow, Poland. This research has received funding from the European Research Council (ERC) under the H2020 European Research Council research and innovation program (Grant agreement 805498) as well as from the Priority Research Area Society of the Future under the program "Excellence Initiative – Research University" at the Jagiellonian University in Krakow. Maciej Próchnicki was supported by the National Science Centre, Poland (Grant 2017/25/N/HS5/00944) and Bartosz Janik was supported by the National Science Centre, Poland (Grants 2016/23/N/HS5/00952 and 2020/36/C/HS5/00111 supported literature search and analysis of the results of the study, respectively). We are indebted to Krzysztof Kasparek for indispensable assistance with the meta-analytic techniques employed in this article, to Jeffrey Rachlinski and Michael Saks for helpful comments on earlier drafts, as well as to many scholars who provided us with unreported data from their experimental studies.

Correspondence concerning this article should be addressed to Maciej Próchnicki, Faculty of Law and Administration, Jagiellonian University, Bracka 12, 31-005 Krakow, Poland. Email: maciej.prochnicki@uj.edu.pl

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Theoretical models for the process of generating numerical legal decisions (e.g., the amount of damages or the length of prison term) suggest that this process is typically distorted by salient but not necessarily relevant numerical values present in the decision contexts—also known as *anchors* (Hans & Reyna, 2011; Hastie, 2011; Kahneman et al., 1998). Indeed, the experimental research conducted over the last 30 years seems to clearly indicate that legal decision-makers anchor on salient (but possibly arbitrary) numbers present in the decision environment when rendering numerical decisions in a variety of legal cases. This points to the risk that numerical verdicts are not issued in a predictable and unbiased way, which in turn may lead to a number of socially undesirable consequences (Hastie, 2011; Hastie et al., 1999; Helm et al., 2017). First, if judges and jurors render different numerical verdicts for similar cases, this is obviously unfair and may also undermine the social legitimacy of the legal system. Second, if such inconsistencies increase the variability of rulings (meaning the greater unpredictability of law), it makes legal actors overspend on the management of legal risk and may also decrease the rate of settlement out of court. Finally, if decision-makers' biases are more systematic (simply making the expected sanction too harsh or too lenient), then the result may be the regular overdeterrence of benign activities, or the opposite: the underdeterrence of actions that are actually wrongful.

A court ruling should be treated as a decision with serious consequences at both the individual and societal levels. Therefore, it is crucial that these decisions are made accurately and reliably. In the spirit of Messick's measurement theory (Messick, 1989; Shepard, 1993), which is widely used in psychological and educational research, we consider court verdicts as high-stakes assessment (results lead to major/significant consequences), which should demonstrate consequential validity. It is crucial to investigate to what extent the decision-making process in courts is prone to the influence of factors potentially irrelevant to the case—such as those impacting it through the operation of the anchoring effect.

Experimental research on anchoring in legal decision-making has reached exactly the stage that justifies conducting a systematic research synthesis: On one hand, the number of studies is large enough to allow for some valid meta-analytic inferences, whereas on the other, it is not yet the case that any such inferences are obvious. Even more importantly, the existing literature clearly points to some controversies with regard to which available studies give inconclusive or contradictory answers. From a psychological perspective, for example, it is of major theoretical interest whether asking the legal decision-maker a comparative question (Is the target value more or less than the anchor?) before the absolute question increases the effect of anchoring (so-called standard vs. basic anchoring dispute; Englich, 2006). There are also many crucial questions from a legal perspective: For example, do anchors that are legally irrelevant (or even inadmissible) result in smaller effects than those legally important (Glöckner & Englich, 2015; Wistrich et al., 2005)? Can the influence of the most salient number be effectively mitigated by introducing another anchor in the context (Marti & Wissler, 2000)? Such questions carry great importance for both psychology and law. If legal professionals are less susceptible to the anchoring effect than laypeople (Englich & Soder, 2009; see also Northcraft & Neale, 1987), such a pattern would be interesting both for the general psychology of expert judgment and for the very practical legal question of relative (dis)advantages of a jury system.

These questions are but examples of issues to which the existing experimental research on legal anchoring has given inconsistent answers. To the extent that the present analysis brings us closer to clarifying such issues, it not only makes a substantial contribution to psychology and legal theory but also may be of interest to legal practitioners. For instance, defense attorneys may be interested whether setting another anchor (countering an initial claim) is an effective strategy, and adjudicators may learn about the relative effect of different anchors.

The Psychology of the Anchoring Effect

The *anchoring effect* refers to a situation in which a decision-maker, having been asked to estimate a certain numerical value, tends to ground that value on the first (and/or salient) numerical value they encounter (Orr & Guthrie, 2006). Since the seminal work by Tversky and Kahneman (1974), the anchoring effect has been found to influence numerical estimation tasks in a number of domains, such as negotiations, estimates of product prices, self-efficacy, forecasting, and various general knowledge questions (for a comprehensive review, see Furnham & Boo, 2011).

There are two older psychological models of anchoring. The *insufficient adjustment model* (Epley & Gilovich, 2005; Tversky & Kahneman, 1974) postulates that the decision-maker, being aware that a given anchor might not be an accurate estimate of the target value, tries to adjust the anchor value accordingly but fails to do it to a sufficient degree, thus biasing the final answer. The *numeric priming model* (Jacowitz & Kahneman, 1995; Wilson et al., 1996; Wong & Kwong, 2000) states that the presence of an anchor makes the value suggested by it more mentally accessible, so that it is likely to influence the assessment of the target value. These two models enjoy limited popularity nowadays, as they are seen as narrow in scope and failing to satisfactorily explain many experimental results (Chapman & Bornstein, 1996; Epley & Gilovich, 2005; Mochon & Frederick, 2013; Mussweiler & Strack, 1999a). There are, however, at least two other competing models favored by contemporary scholars.

The *selective accessibility model* (Chapman & Johnson, 1994, 1999; Mussweiler & Strack, 1999b) sees the process of anchoring as consisting of two phases: First, the decision-maker assumes that a target value might be equal to the anchor and accesses as much knowledge supporting this claim as they can. If this makes the decision-maker falsify the assumption that the target value is equal to the anchor, they will move to the second phase where they assess another possible answer. This, however, happens after knowledge consistent with the anchor has been accessed in the first phase (*confirmatory hypothesis testing*; Furnham & Boo, 2011; Strack & Mussweiler, 1997), which influences where the decision-maker looks for alternative answers in the second phase (*semantic priming*; Chapman & Johnson, 1999; Strack & Mussweiler, 1997). This view on anchoring finds some empirical support (Furnham & Boo, 2011). Based on the selective accessibility model, Wegener and colleagues (2010b) proposed the *elaboration-based* account, in which they argue that kinds of elaboration other than selective accessibility may also be useful in explaining anchoring. This account relies heavily on the attitude change research (Petty & Wegener, 1998, 1999) that makes use of persuasion variables (such as the emotions and credibility of the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

person delivering a message) and two different levels of elaboration (thoughtful vs. nonthoughtful) when explaining the decision making process. This relates directly to the shift in the participant's attitude toward the targeted response (Wegener et al., 2010a, 2010b): Message recipients, while not thinking carefully about the message, might develop a favorable attitude toward certain decisions based on their mood or the credibility ascribed to the person delivering the message.

It is worth noting that the two phenomena used in the selective accessibility model occur at the *deep* level of cognition. In contrast, in the *scale distortion theory* (Frederick & Mochon, 2012) anchoring occurs at the *shallow* level of cognition, as a result of a distortion in the mapping of judgments to the provided response scale (Frederick & Mochon, 2012; Mochon & Frederick, 2013). In this model, unlike in the selective accessibility model, numeric anchors do not change the participant's beliefs or representations but simply influence the response scale. The proponents of this model provide a significant number of experiments showing that the process of anchoring seems not to have any effect on the beliefs of the anchored participants and that anchors affect the mapping from stimuli to numbers (Frederick & Mochon, 2012; Mochon & Frederick, 2013).

The selective accessibility model and scale distortion theory might be perceived as competing. The latter offers some evidence suggesting that confirmatory hypothesis testing does not always occur in the anchoring context. On the other hand, the mechanism of hypothesis testing is well established within the general psychological literature (Mussweiler, 2003; Trope & Liberman, 1996), and its ability to explain some cases of anchoring was demonstrated in empirical studies (Strack & Mussweiler, 1997). The available evidence does not favor either of the two models, and one might speculate that each theory offers an explanation of just some aspects of anchoring.

Anchoring and Models of Numerical Verdicts in Law

The anchoring effect is central to all proposed theoretical models that aim at explaining how legal decision-makers determine numerical verdicts. Three such theoretical models have been proposed in the literature to account for the jury damage award (although they can be generalized to the judicial context). The first one, the *shared outrage model*, proposed by Kahneman and colleagues (1998), explicitly aims at accounting for the production of numerical verdicts in a very specific context of punitive damages in tort cases. The model posits that lay jurors, while assessing the defendant's behavior, form a sense of outrage that, combined with the ensuing harm, determines their punitive intent. Finally, they map their punitive intent on an (unbound) dollar scale of punitive damages. Whereas the results of the first two stages (outrageousness and punitive intent) are quite consistent across individuals, the eventual monetary punitive damages are much more capricious and prone to be influenced by various anchors present in the context.

The *Intention + Anchor* model, proposed by Hastie (2011), was conceived as a generalization of the shared outrage model. According to this model, a legal decision-maker first assesses pieces of information present in the proceedings (such as evidence, argumentation, behavior of the parties, etc.) along with the background knowledge and other factors (e.g., instructions for the jury) and then forms a general intention to compensate or punish. There could be various interpretations of this process, including, for example, the construction of a plausible narrative (Pennington & Hastie, 1991); intuitive reactions to some specific elements (Saks & Kidd, 1980); or other types of information processing (Schum & Martin, 1982; Thomas & Hogue, 1976; Weld & Danzig, 1940). Then, an anchor is identified or (very rarely) generated. Finally, the anchor is adjusted in compliance with legally relevant guidance (such as mitigating/aggravating factors or jury instructions). However, the law rarely offers definitive guidance on rendering numerical verdicts. Thus, the process of adjustment is often strongly influenced by two psychological effects. First, there occurs a serial underadjustment (Epley & Gilovich, 2001, 2005), in which sequentially assessed pieces of information that lower or raise the target value are dominated by the presence of an anchor value, which weakens the impact of the information. Second, the anchor biases information sampling: items congruent to the anchor are easier to recall (Mussweiler, 2003).

The last model to be presented here, proposed by Hans and Reyna (2011) draws from research on the fuzzy-trace theory (Reyna, 2012; Reyna & Brainerd, 2011). The central point of this theory is that people encode the information as either verbatim (literal, detailed, veridical) or as gist representations (essential meaning). The main assumption of the proposed model is that jurors engage in gist-based reasoning while deciding on the amount of damages (Reyna et al., 2015). The first stage consists of a categorical gist judgment, in which the decision-maker decides whether damages are warranted. In the second stage, the ordinal gist judgment is made, ranking the damages as high, medium, or low. Based on that, the decision-maker proceeds to the stage of generating numbers that fit the gist magnitude. The exact values are influenced by a multitude of factors, such as anchors present in the context (especially if they are relevant) and symbolic numbers from juror's everyday life that have low or high meaning attached to them (Hans et al., 2018; see Hans & Reyna, 2011; Reyna et al., 2015).

Overview of the Research

The research on the susceptibility of legal decision-makers to the anchoring effect started with experimental studies on civil proceedings, particularly on the influence of the numbers presented by the parties to a civil dispute on the amount of awarded damages. Considerable experimental evidence suggests that there is a strong dependency of the awarded amount on the initial amount claimed by the plaintiff (*ad damnum* in the Anglo-American legal terminology), which means that the higher the plaintiff's demand, the higher the damages awarded, even if the evidence presented by the plaintiff remains the same (Broeder, 1959; Hastie et al., 1999; Malouff & Schutte, 1989; Zuehl, 1982). It may be argued that *ad damnum* is not a fully irrelevant anchor, since it may convey some information about the severity of harm suffered by the plaintiff, so that legal decision-makers are rational in assimilating their judgments to this value. However, some studies (Chapman & Bornstein, 1996) suggested that increasing *ad damnum* increases the amount of damages without changing the perception of the severity of harm suffered by the plaintiff. On the other hand, some authors (Malouff & Schutte, 1989; Marti & Wissler, 2000) have

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

argued that there exists a *boomerang effect*: If the plaintiff demands an amount that is absurdly high, it decreases their credibility, leading to effectively lower damages. In this context, the anchor produces contrast instead of assimilation. Nonetheless, other studies (Campbell et al., 2016; Chapman & Bornstein, 1996) have shown that participants assimilate their judgments even to an extremely high amount of *ad damnum*, meaning that the assimilation to the anchor in this context is very robust.

Ad damnum is not the only possible anchor that has been studied in the context of civil proceedings. Other values that have been shown to serve as anchors in this context are statutory caps on the amount of noneconomic or punitive damages (Hinsz & Indahl, 1995; Rachlinski et al., 2015; Robbennolt & Studebaker, 1999; Saks et al., 1997); a value appearing in the opinion of an expert witness (Greene et al., 1999; Raitz et al., 1990); a value appearing in an inadmissible piece of evidence (Wistrich et al., 2005); an amount of damages awarded in a similar case on a TV court show (Guthrie et al., 2009; Rachlinski et al., 2015); or even a value appearing in a meritless objection raised by the defendant (Guthrie et al., 2000).

In contrast, relatively little research has been conducted on the role of anchoring in criminal proceedings. One notable exception is a series of studies by Englich and colleagues (Englich, 2005; 2006; Englich & Mussweiler, 2001; Englich & Soder, 2009), which found that participants (including professional judges) decided on the length of prison terms under the influence of anchors, such as the prison term demanded by the public prosecutor, suggested by a journalist or even by a random IT student. What is more, participants tended to assimilate their verdicts to the demand of the prosecutor even when they were fully aware that this value had been generated randomly.

Moreover, experimental studies seem to be corroborated by a respectable number of observational studies in this area (Broeder, 1959; Chang et al., 2016; Diamond et al., 2011; Fariña et al., 2003; Kim & Chae, 2017; Martin & Alonso, 1997) that provide further support for the existence of anchoring in real-life rulings.

Despite the abundance of experimental research on the anchoring effect, no meta-analysis on this topic has been published in the general psychological literature (see Furnham & Boo, 2011 for a literature review). The same can be said about anchoring in legal decision-making: No meta-analysis has been published to date, and the only existing literature review (Englich, 2006) does not include a substantive portion of studies published in recent years.

The existing analysis closest to the present study is the meta-analysis on the anchoring effect in negotiations (Orr & Guthrie, 2006). The authors established that the average size of the effect of anchors on the outcome of simulated negotiations reported in the literature is surprisingly large ($r = .5$). They also assessed the effects of two moderator variables similar to the ones to be analyzed here: experience of participants and presence of numeric values other than the main anchor in the decision context. In both cases, the average effect size was smaller for the professional negotiators and in contexts in which other numeric values were present, although these results did not reach statistical significance.

The Present Research

Usually, the main aim of the meta-analysis is to synthesize the results of the several studies into a single estimate, that is, a summary effect size (Petticrew & Roberts, 2006). However, in the case of anchoring, we had to split this task into two. This stems from the fact that some studies included in this analysis followed the regular paradigm in psychological research of comparing intervention and control groups (i.e., a group exposed to some anchor and a group that did not see any anchor), whereas other studies did not use a control group, merely comparing groups exposed to anchors of different values. Although the employment of a control group is a traditional practice in experimental research (Hinkelmann & Kempthorne, 2008), Englich and colleagues (2006) argued that using two different anchors rather than contrasting the anchored group to the control group is justified by a few reasons: first, to eliminate accidental or self-imposed anchors; second, to eliminate the effect of the presence of mere initial judgment in one of the groups (in the standard paradigm design); and third, to focus on the absolute effect of anchoring rather than determining which anchors produce the stronger effect. To account for this methodological divergence, we conducted all analyses separately for these two groups of studies, to examine whether the effect exists under both types of design and, if so, what its relative strength is under each. Following the authors who have claimed that participants in a control group employ other, possibly self-generated anchors (resulting in a higher variance of target assessments), we predicted that the overall effect size is lower for studies employing a control group.

> *Hypothesis 1:* The introduction of a numeric anchor will bias legal decision-makers' judgment in the direction of the anchor value (anchoring effect). The observed effect size of anchoring will be smaller in studies comparing an intervention group (i.e., a group exposed to an anchor) with a control group than in studies comparing groups exposed to anchors of different values (without a control group).

The second objective of this research synthesis was to examine to what extent the overall effect sizes are prone to the influence of a number of moderators. We proposed the moderator selection guided by three perspectives of research, which are often interconnected. The first one, the methodological perspective, deals with study characteristics that have a general potential to influence the effect size, such as the size of the employed sample or the publication status of a given study. The second—psychological—perspective, focuses on the cognitive aspect of the phenomenon to determine which circumstances may lead to stronger anchoring effects. This includes such study characteristics as the employed paradigm of studying anchoring (basic vs. standard) or the type of scale on which the target value was assessed (bound or unbound). Finally, the third perspective—the legal one—refers to properties especially relevant from the viewpoint of the legal process. This group of variables includes, for example, the legal relevance of the anchor or the branch of law to which the employed case refers.

Our final research question was whether or not overall effect sizes are overestimated as a result of publication bias, that is, a lower chance of publishing studies that report smaller, insignificant results (Joober et al., 2012). We addressed this issue by the trim-and-fill method supplemented by Egger's Test to indicate whether the calculated overall effect sizes should be considered as overestimated (to account for possibly missing studies). We also derived some relevant results from a metaregression analysis (published vs. unpublished condition subgroup analysis).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Moderator Selection

Methodological Characteristics of Primary Research

We introduced four moderators to reflect methodological characteristics of primary research: publication status, sample size, the type of employed stimuli, and the type of employed sample. The reason behind the inclusion of the first two of these variables was to examine the possibility of publication bias. If studies reporting statistically significant results were indeed more likely to be published, then we might expect that the overall effect size for published studies would be larger than for those that were not published. Moreover, a negative dependence between sample sizes and effect sizes might also be an indication of publication bias (assuming that, in the case of studies employing small samples and reporting small effect sizes, it is more likely that such effects would be statistically insignificant; Levine et al., 2009). We introduced the moderator *sample size* measured on a continuous scale.

In all the analyzed studies, the participants were presented with a stimuli in the form of some representation of a legal case to be decided. In most studies, this representation consisted of a written description of a case. Two general approaches are possible while constructing textual stimuli: First, a textual stimulus can be as short as possible in order to help to isolate the operation of the studied effect (Kerr & Bray, 2005), decrease the number of potential confounders, and avoid burdening the participants' information processing (Bieneck, 2009). On the other hand, written stimuli that are too short (possibly oversimplifying the representation of the actual decision context) might lack ecological validity (Bieneck, 2009). In the particular context of anchoring, this might mean that the anchor present in a short vignette is more salient (possibly also through demand characteristics; Nichols & Maner, 2008) than it would be in a real-life context (e.g., of court proceedings), thus inflating its effect size. Such arguments led us to introduce the following levels of the discussed moderator: *short written materials* (i.e., a short summary of main details of the case) and *rich written materials* (i.e., a more comprehensive presentation of case details or a transcript, possibly approaching in length files that a real-world judge would have to examine). Then, however, anchoring studies can also employ nonwritten stimuli: a recording of mock court proceedings. Arguably, such stimuli could increase the ecological validity of a study, more closely resembling the real-life context at least for some types of legal decision-makers, such as lay jurors (Bornstein, 1999). The meta-analysis by Bornstein et al. (2017) indicates that the mode of presentation (written vs. other) influences the outcomes of a number of decision tasks in jury simulation studies (cf. a study by Pezdek et al., 2010 regarding the mode of presenting an expert opinion). For these reasons, we introduced a third level of the discussed moderator: *audio/video recording*. It seems that, in such a mode of presentation, the anchor value may be even less salient, thus reducing the effect size of anchoring.

> *Hypothesis 2:* Exposing participants to short written stimuli will result in a larger effect size of anchoring, compared to other types of stimuli. Exposing participants to audio/video stimuli will result in a smaller effect size of anchoring, compared to other types of stimuli.

Most of the analyzed studies were conducted on mock jurors. The question whether results obtained on lay samples are generalizable to expert populations remains open to debate in cognitive psychology, in general, and in research in anchoring, in particular (see Northcraft & Neale, 1987). In empirical legal research, it is also controversial whether legal expertise has a potential to immunize decision-makers from the influence of undesirable psychological effects (Teichman & Zamir, 2014; Tobia, 2020). There are many possible reasons for professional legal decision-makers to be less susceptible to anchoring than laypeople: professionals' expertise helps them ignore irrelevant information, their training in deliberative reasoning helps them identify actually relevant factors, and their knowledge of existing case law provides them with alternative (and arguably relevant) anchors. Answering the question on the effect of legal expertise bears importance not only for the issue of generalizability of results obtained using a sample from a given population but also for the very practical legal debate on the relative advantages of a lay jury system (Bornstein & Greene, 2017). A direct comparison of samples drawn from lay and professional populations is rare in the research on legal anchoring. To address this issue, we introduced a moderator *type of sample* with two levels: *lay* and *professional*.

> *Hypothesis 3:* The effect size of anchoring will be smaller among legal professionals than among laypeople.

Psychological Variables

We included three moderators in the analysis to address controversies present in the general literature on the anchoring effect: paradigm (standard vs. basic); anchor value; type of scale on which participants assessed the target value (closed vs. open).

There are two general paradigms employed in experimental research on the anchoring effect. Under the standard paradigm, participants are first asked a comparative question (whether the target value is more or less than the provided anchor value) and only subsequently asked to give their estimate of the target value (Englich, 2008). In contrast, under the basic paradigm (Wilson et al., 1996) participants are merely exposed to the anchor value and asked to give their estimate of the target value without having to answer the comparative question first. These different approaches reflect a deeper theoretical disagreement. Proponents of the selective accessibility model assume that a comparative task is a crucial element of the anchoring; thus, the anchoring effect is expected to lead to more robust results under the standard paradigm (Englich, 2008), whereas the basic paradigm anchoring is triggered simply by numeric priming, which is a relatively shallow cognitive mechanism, resulting in a smaller effect. Hence, if studies employing the standard paradigm indeed lead to larger effect size, this would provide evidence useful for choosing among different theoretical accounts of anchoring. This moderator may also be of more direct legal relevance: if the standard paradigm leads to more robust anchoring, it would imply that actual legal decision-makers may anchor more strongly in those legal settings in which they are expected to explicitly evaluate the anchor value.

> *Hypothesis 4:* The effect size of anchoring will be larger in studies employing the standard paradigm than in studies employing the basic paradigm.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

The existing literature is rather unequivocal in its conviction that the anchor value matters, as most of the time people will assimilate their assessment of the target value to the value of the anchor (see, e.g., Hinsz & Indahl, 1995). However, a major point of controversy is whether this dependence is monotonic across the board, that is, whether there will be assimilation to the anchor even if the anchor is absurdly high (Chapman & Johnson, 1994; Strack & Mussweiler, 1997). Some scholars have suggested that the strength of the anchor may be somewhat diminished in such cases, as a value which lies outside some reasonable range loses its credibility as a point of reference (Sugden et al., 2013; Wegener et al., 2001). Earlier research went even further, suggesting a contrast effect; that is, instead of assimilating the target value to the higher anchor, participants may turn in the opposite direction (see Hinsz & Indahl, 1995; cf. Strack & Mussweiler, 1997). This controversy is particularly important in the legal context because of, among others, the question of whether the party starting the court proceedings (plaintiff or prosecutor) can benefit from setting its demand absurdly high. Some of the studies included in this meta-analysis corroborate the thesis that the assessment of the target value increases without a limit correspondingly to the increase in the value of anchor (Campbell et al., 2016; Malouff & Schutte, 1989) even when the anchor is outlandishly high (Chapman & Bornstein, 1996; Marti & Wissler, 2000). To account for the possibility that the dependence between anchor value and the target assessment might be nonlinear, we introduced three levels of the moderator *anchor value: low*, *medium*, and *high* (based on standardized z scores), although we validated our results by repeating the analysis with the continuous moderator.

> *Hypothesis 5:* Medium-valued anchors will result in a larger effect size of anchoring than low- and high-valued anchors.

According to the scale distortion theory (Frederick & Mochon, 2012; Mochon & Frederick, 2013), the main impact of the anchor is to change the response scale. The relation of the type of used scale and the anchoring effect is also central to the shared outrage model (Kahneman et al., 1998), which states that it is the difficulty of translating a qualitative stimulus onto an unbound numeric scale (such as the monetary scale) that allows anchors to play such an important role in legal decision-making. If those claims are correct, we might expect that the anchoring effect is weaker in contexts in which participants are asked to give their responses on a closed scale (such as prison term within the statutorily defined scope or damages not exceeding some specified caps) instead of on an unbound scale (typically present in jury damage award), because it is easier for participants to accurately map their attitudes on such a scale, and the potential shift in scales is limited. Accordingly, we introduced a moderator *limit*, which distinguished studies that provided participants with a closed scale from those that did not.

> *Hypothesis 6:* The effect size of anchoring will be smaller in studies in which the target value is assessed on a bound scale, compared with studies employing an unbound scale.

Legal Moderators

Four moderators were included in the analysis to address controversies typical of the research on anchoring in legal settings: legal relevance of anchor; legal type of anchor (*ad damnum*, damage cap, etc.); area of law to which the presented case belonged (civil vs. criminal); and presence of any salient numeric values other than the main anchor.

Some of the psychological models predict that relevant and irrelevant anchors can have a different impact on the target value assessment. Whereas numeric priming predicts that the relevance of the anchor does not have much impact, elaboration-based models suggest that it is a key factor. When it comes to models of legal decision-making, the meaningfulness of an anchor may correspond more adequately to the gist judgment of decision-makers, as assumed by the Hans–Reyna model (Hans et al., 2018). This implies that relevant anchors may have stronger effects (cf. Glöckner & Englich, 2015). On the other hand, if irrelevant anchors can affect the outcome of legal proceedings nonetheless, it might be an argument in favor of some institutional changes to limit the exposure of legal decision-makers to such anchors. There are two dimensions of relevance of an anchor to the estimation of the target value: the material dimension (to what extent the anchor points to facts that are actually relevant for assessing the target value, e.g., if it points to the magnitude of harm suffered by the plaintiff) and the formal dimension (a given anchor might be explicitly inadmissible, or its inclusion in assessing the target value might be legally required). Because of the limited number of studies, we chose not to analyze these two dimensions separately but rather to combine them into one three-level measure.

The first level, *irrelevant*, includes anchor values that are not directly related to the respective case and not informative with regard to the target value (such as previous awards in unrelated cases) and/or are legally inadmissible. Another aspect of the (ir)relevance of the anchor is its meaningfulness, as described by Hans et al. (2018). From the viewpoint of legal proceedings, some anchors may be meaningful but still irrelevant. Such an example would be the previous award in an unrelated case (Hans et al., 2018; Reyna et al., 2015). Arguably, meaningless anchors may operate differently than irrelevant but still meaningful ones. However, because of the small number of studies in each of these two possible categories, we decided to merge them into one. The second level, *potentially relevant*, refers to anchor values that are given by parties to the dispute. They may be assumed to bear some relevance to the target value, since a strategic party might be expected not to give an outlandish value—but should nonetheless be expected to be biased by the party's interest in the proceedings. The third level, *relevant*, denotes anchor values that are objectively informative while assessing the target value and/or there is a law that explicitly stipulates to take them into account when deciding on the target value (e.g., previous awards in similar cases).

> *Hypothesis 7:* Relevant anchors will result in a larger effect size of anchoring and irrelevant anchors will result in a smaller effect size of anchoring when compared with other levels of anchor relevance.

As with relevance, elaboration-based theories could put an emphasis on the type of anchor unlike numeric priming and other "shallow" models. Existing research on legal anchoring seems to lend support to the first group of theories, as it suggests that the anchor set by the party starting legal proceedings (*ad damnum* demanded by the plaintiff or the penalty demanded by the prosecutor) has a relatively strong effect on the target value assessment

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

(e.g., Hastie et al., 1999; Malouff & Schutte, 1989). In contrast, a counteroffer issued by the defendant is assumed to perform relatively poorly (Campbell et al., 2017; Decker, 2007). Legally binding caps on the damages amount is another type of anchor that seems to result in strong anchoring (e.g., Hinsz & Indahl, 1995; Saks et al., 1997). To corroborate these findings, we introduced a moderator *type of anchor* with five levels: *ad damnum*; prosecutor's demand; defendant's counteroffer; damage caps; and the final, residual category of other anchors.

> *Hypothesis 8:* Anchors in the form of *ad damnum*, prosecutor's demand, or damage caps will result in larger effect sizes of anchoring, when each of them is compared with anchors of all other kinds. Anchors in the form of counteranchors will result in a smaller effect size of anchoring, when compared with anchors of other kinds.

Most of the existing research on judicial anchoring has been conducted in the context of awarding damages in civil cases as opposed to imposing punishment in criminal cases. There are grounds to speculate that anchoring in criminal cases might result in different effects: Apart from different goals in these two types of proceedings, participants in "criminal" studies are asked to put their responses on a closed scale (because the binding range of values in criminal sentences is generally explicitly provided by law) and in temporal (the length of prison term) rather than monetary terms. To explore this possibility, we introduced a moderator *branch of law* with two levels: *civil* and *criminal*.

> *Hypothesis 9:* Studies employing a criminal law context will result in a smaller effect size of anchoring than studies employing a civil law context.

The existing research does not give a conclusive answer to the question as to what extent the anchoring effect of the most salient value present in the decision context can be limited by the presence of other numeric values. For example, although there is some evidence that the defendant issuing a counteroffer might somewhat limit the effect of the plaintiff's *ad damnum*, the actual effectiveness of such a counteranchor is contested, as mentioned above. It also remains unclear whether presenting a number of other values (such as the amount of damages awarded in similar previous cases) can effectively diminish the effect of the main anchor (see Hastie et al., 1999; Saks et al., 1997). To contribute to this discussion, we introduce a moderator *presence of other salient values* with three levels: *mere anchor* (just one salient value present), *counteranchor* (one salient value other than main anchor is present), and *a larger number of salient values* (relevant or irrelevant).

> *Hypothesis 10:* The effect size resulting from the main anchor will be smaller in studies employing a counteranchor or a larger number of other salient values, when compared with other levels of this moderator (see Table 1).

Method

This meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). We applied the PRISMA checklist and a flow diagram demonstrating the four phases of conducting a meta-analysis (identification, screening, eligibility, and included sources).

Literature Search and Inclusion/Exclusion Criteria

Figure 1 provides an overview of the literature search and the inclusion and exclusion criteria. The full list of excluded studies appears in Appendix A.

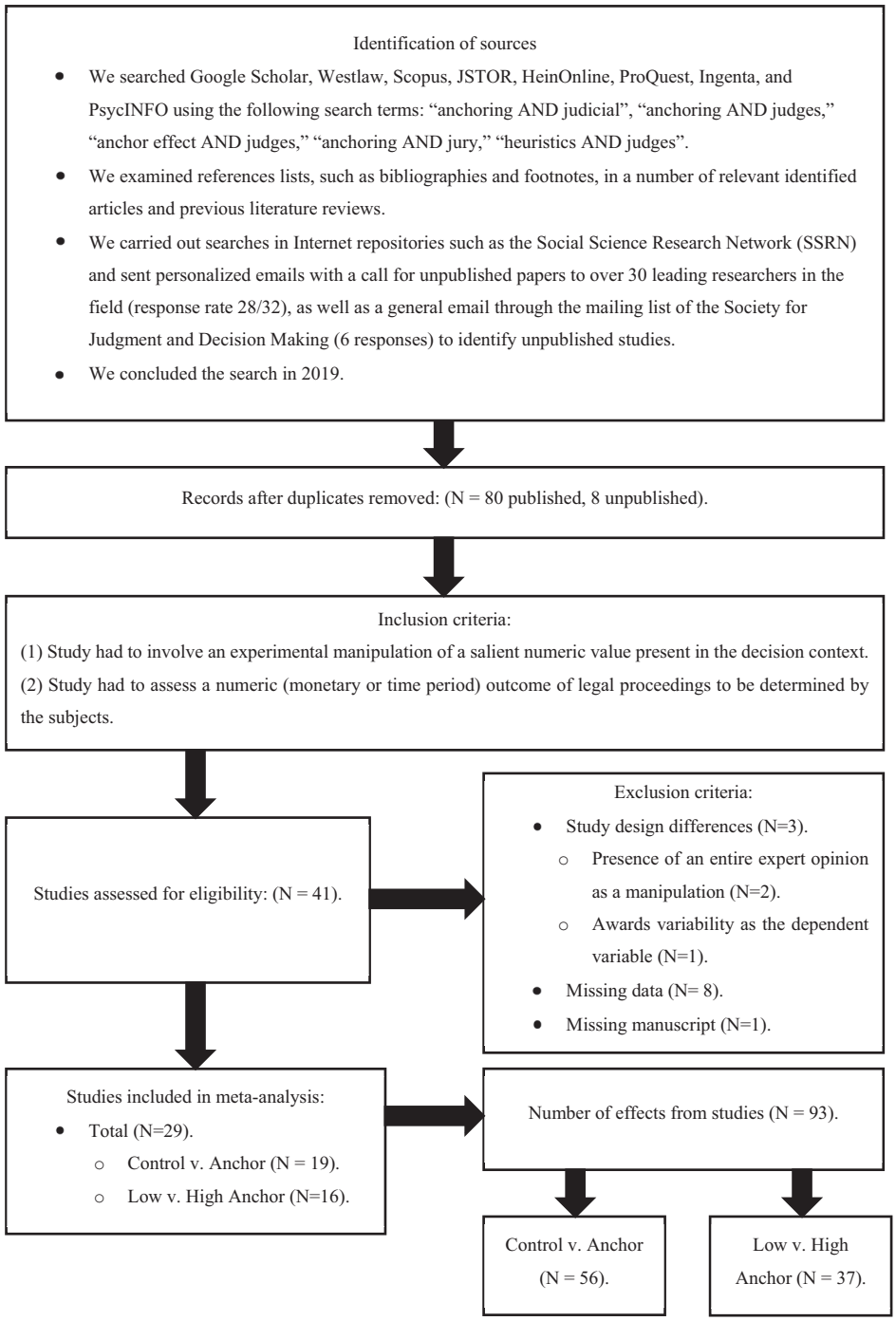
Data Items

To deal with the aforementioned problem of two experimental designs, we divided the calculated effect sizes into two categories: *Control versus Anchor* and *High Anchor versus Low Anchor*, depending on the employed design. From studies having both a control group and more than one anchor group, we extracted effect sizes to be placed under both categories. For example, from a study employing three groups (Control, Low Anchor, High Anchor), we extracted two effect sizes (Low vs. Control and High vs. Control) to be included in the first category and one effect size (High vs. Low) to be included in the second one. Furthermore, if a study compared n groups exposed to anchors of different values, we reported $n - 1$ effect sizes, each based on the comparison of the lowest anchor group with one of the groups with a higher anchor. Because of this procedure, when more than two experimental groups were employed in the same study, we used data obtained from participants belonging to the lowest

Table 1
Moderators and Levels

Moderator	*N* of levels	Level
Publication status	2	Unpublished, published
Sample size	Continuous	
Type of stimuli	3	Short written, rich written, audio/video
Type of sample	2	Lay, professional
Paradigm	2	Basic, standard
Anchor value	3	Low, medium, high
Anchor value	Continuous	
Type of scale	2	Open, closed
Relevance	3	Irrelevant, potentially relevant, relevant
Type of anchor	5	Ad damnum, damage cap, defendant's counteroffer, prosecutor's demand, other
Area of law	2	Civil, criminal
Other salient values	3	Mere anchor, counteranchor, larger number

Figure 1
PRISMA Flow Diagram



Note. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

anchor group or the control group to calculate more than one effect size, thus, in a way, *double-counting* results obtained from those participants. Thus, if a particular study had more than one experimental intervention group with a common control group, we decided to include both comparisons, taking into account the possible correlation between them (dependent effect sizes). Selecting only one comparison (and excluding others) would have resulted in a loss of information and would have been open to results-related choices, which is not generally recommended (Higgins et al., 2019). To address the problem of double-counting

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

results, we employed weights based on sample size and, most importantly, a more formal model taking into account the correlation among dependent effects sizes (the *robust variance estimation* approach) as described in the Summary Measures section.

Data Extraction

Three authors (PB, BJ, MP) simultaneously extracted and coded the data. For most of the employed moderators, we were able to unambiguously code the data on the basis of information explicitly provided in the studies alone. To deal with those moderators that required more complex assessment, we employed the following criteria.

Type of Sample. The *professional* group included samples consisting of law graduates: professional judges and attorneys, but also experienced international arbitrators (Franck et al., 2017) and judicial trainees (*RechtsreferendarInnen*; Englich, 2005). The *lay* group included samples composed of individuals lacking legal education (i.e., students from fields other than law, jury eligible citizens, actual lay jurors, MTurk workers). Samples consisting of law students do not easily qualify as samples of professionals or laypeople and arguably should constitute a third level on its own. This, however, was not feasible because of a small number of studies employing such samples, so we decided to qualify them as professional samples. This follows a more general trend in empirical legal research: Because law students have self-selected themselves to the legal profession and have already received some legal training, they might be assumed to more closely resemble professionals (Tobia, 2020).

Area of Law. We classified the studies which related to the settings of criminal proceedings (including military proceedings) and required the participants to impose a punishment as "criminal law." Other cases, which focused on monetary damages, fell into the umbrella category of "civil law," which included not only torts and contracts but also labor law (Guthrie et al., 2009; Study C.1), bankruptcy law (Rachlinski et al., 2006), corporate law (Feldman et al., 2016; Study 2), and international investment law (Franck et al., 2017).

Case Description. Whenever participants in the study were told to watch/listen to recordings of a simulated court trial instead of/in addition to written stimuli, we classified the stimuli as "audio/video." In the case of written stimuli, we set the cutoff line between "short" and "rich" descriptions at 1,000 words.

Relevance. If a study introduced a value that was objectively informative while assessing the target value, or the law explicitly stipulated taking it into account when deciding on the target value (previous awards in similar cases, median income while deciding on noneconomic damages, reliable expert opinion, and damage caps), the study was coded as "relevant." The second category—"potentially relevant"—included numeric values given by one of the parties to the proceedings (including expert opinions solicited by one of the parties). The last category, "irrelevant," included values that were not directly related to the case and not informative with regard to the target value. This included numbers that were meaningful although irrelevant (previous awards in unrelated cases, damages awarded on a TV court show, prison term suggested by a journalist, a heckler, or a student) or inadmissible according to an explicit legal rule. In addition, in this category, we included anchors that were meaningless (such as the cost of courtroom renovations).

Anchor Value. Taking into account that the studies generally included two distinctive types of anchor/target values (prison terms expressed in units of time, and monetary values), we needed to standardize the relative values within both types. We converted the relative values of anchors—based on the difference between the values of low and high anchors or the absolute value of the anchor in the Control versus anchor category—into z scores. This meant treating "no anchor" as "0 anchor," which might seem to be problematic, because there may be an implicit, default mental "baseline" (or just self-imposed anchor) for a selected question, differing from 0. Moreover, any number accessible to participants could have accidentally worked as an anchor (cf. Englich et al., 2006; also consider the reasons behind the two differing methodologies discussed above). We use the resulting z scores as a continuous measure in subsequent analysis. However, to account for the potential nonlinearity of the relation, we also divide the z score values into three levels: high, medium, and low. We based this division on the distribution of distance scores. The category "low" contained scores up to the 25th percentile of the distribution, "medium" in the 26–74th percentiles, and "high" in the 75th percentile and above.

Final Dataset

We calculated a total of 93 effect sizes from 29 studies, divided into the two groups: Control versus Anchor (56 effect sizes from 19 studies) and Low Anchor versus High Anchor (37 effect sizes from 16 studies). The overall number of participants was 2,177 legal professionals (including 244 law students) and 6,372 laypeople. Publication dates ranged from 1989 to 2018. We concluded the search in 2019.

Summary Measures

We used both Fisher's z and Cohen's d to calculate effect sizes (see Tables 2 and 3). A few studies reported Mann–Whitney test results, which we then recounted (we report all formulas used in the Appendix B). The calculation of effect sizes based on nonparametric tests seemed preferable because of the typical right-skewness of results of experimental studies on anchoring. However, for studies for which we were able to calculate effect sizes based on both the Mann–Whitney test and the difference in means, the difference in obtained sizes was not substantial.

To account for the skewed distribution of Pearson's r reported in the studies, we used Fisher's z transformation, and we calculated the standard error of the distribution of the z statistic. What is more, to obtain the standard error for the distribution of the r statistic, we used the formula provided by Everitt and Skrondal (2002). To calculate Cohen's d, we used the standard formula provided by Cohen (1988). To calculate the confidence interval (CI) for this effect size, we used the general formula from Hedges and Olkin (1984) and from it we calculated the 95% confidence interval for Cohen's d.

To address the issue of studies employing multiple intervention groups, we used a method which takes into account the correlation between effect sizes as recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins et al., 2019), that is, the robust variance estimation (RVE; Hedges et al., 2010). This approach allowed for handling statistically

Table 2
Effect Sizes and Weights: Control Versus Anchor

Study	Id_exp	*N*	Effect Sizes			Weight
			z	95% CI	Cohen's *d*	
Campbell et al. (2017)	1	188	.52	[.38, .67]	1.24	2.40
Campbell et al. (2017)	2	192	.13	[−.01, .27]	.26	2.41
Decker (2007)[a]	1a	142	.04	[−.12, .21]	.09	2.18
Decker (2007)[b]	1b	138	.05	[−.12, .22]	.11	2.16
Decker (2007)[c]	1c	143	.20	[.03, .36]	.41	2.19
Diamond et al. (2000)	1a	43	.46	[.15, .77]	1.07	1.16
Diamond et al. (2000)	1b	43	.77	[.46, 1.08]	2.21	1.16
Diamond et al. (2000)	1c	47	.77	[.48, 1.07]	2.25	1.23
Diamond et al. (2000)[d]	1d	43	.57	[.26, .88]	1.42	1.16
Duke et al. (2015)	2	252	.14	[.01, .27]	.29	2.59
Franck et al. (2017)	2a	104	.09	[−.11, .29]	.18	1.92
Franck et al. (2017)	2b	108	.09	[−.10, .28]	.18	1.96
Franck et al. (2017)	2c	104	.08	[−.11, .28]	.16	1.92
Glöckner and Englich (2015)[e]	1a	30	.32	[−.05, .70]	.69	.88
Glöckner and Englich (2015)[f]	1b	30	.31	[−.07, .67]	.66	.88
Glöckner and Englich (2015)[g]	1c	30	.26	[−.12, .64]	.54	.88
Glöckner and Englich (2015)[h]	2c	44	.23	[−.08, .54]	.48	1.18
Glöckner and Englich (2015)[i]	2d	44	.15	[−.16, .45]	.30	1.18
Glöckner and Englich (2015)[j]	2e	44	.19	[−.11, .50]	.40	1.18
Glöckner and Englich (2015)[k]	2f	44	.43	[.12, .74]	.98	1.18
Guthrie et al. (2000)	1	116	.21	[.02, .39]	.42	2.02
Guthrie et al. (2009)	1	82	.32	[.10, .54]	.69	1.72
Hans et al. (2018)[l]	1a	63	.19	[−.07, .44]	.38	1.48
Hans et al. (2018)[m]	1b	62	.18	[−.08, .43]	.36	1.47
Hans et al. (2018)[n]	1c	62	.40	[.14, .66]	.89	1.47
Hans et al. (2018)[o]	1d	63	.35	[.10, .60]	.76	1.48
Hinsz and Indahl (1995)	1a	74	.45	[.21, .68]	1.02	1.63
Hinsz and Indahl (1995)	1b	75	.55	[.32, .78]	1.33	1.64
Hinsz and Indahl (1995)	3a	154	.35	[.20, .51]	.77	2.25
Hinsz and Indahl (1995)	3b	100	.28	[.08, .48]	.60	1.89
Hinsz and Matz (1997)	1	108	.30	[.10, .49]	.63	1.96
McAuliff and Bornstein (2010)	1a	72	.38	[.14, .61]	.83	1.60
McAuliff and Bornstein (2010)	1b	72	.04	[−.20, .28]	.08	1.60
McAuliff and Bornstein (2010)	1c	72	.28	[.04, .51]	.58	1.60
McAuliff and Bornstein (2010)	1d	72	.31	[.08, .55]	.67	1.60
Rachlinski et al. (2006)	1	103	.18	[−.01, .38]	.37	1.92
Rachlinski et al. (2015)[p]	1	112	.22	[.04, .41]	.46	1.99
Rachlinski et al. (2015)[q]	2	58	.42	[.16, .67]	.95	1.41
Rachlinski et al. (2015)[r]	4a	150	.16	[.00, .32]	.33	2.23
Rachlinski et al. (2015)[s]	4b	149	.07	[−.09, .23]	.14	2.22
Reyna et al. (2015)[t]	1a	137	.38	[.21, .55]	.83	2.15
Reyna et al. (2015)[u]	1b	137	.18	[.01, .35]	.36	2.15
Reyna et al. (2015)[v]	1c	137	.30	[.13, .47]	.64	2.15
Reyna et al. (2015)[w]	1d	137	.06	[−.11, .23]	.12	2.15
Robbennolt and Studebaker (1999)	1	124	.54	[.36, .71]	1.29	2.07
Robbennolt and Studebaker (1999)	2	349	.26	[.15, .36]	.53	2.78
Schweizer (2005)	1a	110	.30	[.11, .49]	.65	1.97
Schweizer (2005)	1b	107	.28	[.09, .47]	.59	1.95
Stein and Drouin (2018)[x]	1a	118	.16	[−.02, .34]	.33	2.03
Stein and Drouin (2018)[y]	1b	122	.08	[−.10, .26]	.16	2.06
Stein and Drouin (2018)[z]	1c	117	.24	[.05, .42]	.50	2.02
Stein and Drouin (2018)[aa]	1d	110	.08	[−.11, .27]	.17	1.97
Stein and Drouin (2018)[ab]	1e	118	.27	[.09, .46]	.57	2.03
Stein and Drouin (2018)[ac]	1f	118	.16	[−.02, .34]	.32	2.03
Wistrich et al. (2005)	1a	100	.28	[.08, .48]	.58	1.89
Wistrich et al. (2005)	1b	75	.39	[.16, .62]	.86	1.64

Note. Id_exp refers to the number of a given study in the article in which it was reported. If there is no explicit numbering of studies in a given article, the first study in the article to deal with anchoring is indicated as 1, and so on. Within the studies employing more than one anchor, we list effect sizes for anchors starting from low to high. If the order of the anchors is different in the original article or we used one of the groups or conditions from the original article, this is indicated by the footnote.
[a] Control vs. 0 anchor. [b] Control vs. 80k anchor. [c] Control vs. 200k anchor. [d] Control vs. 500k cap. [e] Anchors from *Irrelevant* condition. [f] Anchors from *Relevant-prosecutor* condition. [g] Anchors from *Relevant-judge* condition. [h] *Irrelevant* condition, high anchor. [i] *Relevant* condition, high anchor. [j] *Irrelevant* condition, low anchor. [k] *Relevant* condition, low anchor. [l] Control vs. 1,500k anchor. [m] Control vs. 50k anchor. [n] Control vs. 1,500k anchor. [o] Control vs. 50k anchor. [p] Canadian group from Study 2. [q] U.S. group from Study 2. [r] Control vs. Low anchor from Study 4. [s] Control vs. high anchor from Study 4. [t] *Meaningful condition*, low anchor. [u] *Meaningful* condition, high anchor. [v] *Meaningless* condition, low anchor. [w] *Meaningless* condition, high anchor. [x] *Ignore* condition. [y] *Ignore* condition. [z] *Identify* condition. [aa] *Identify* condition. [ab] *Counter* condition. [ac] *Counter* condition.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Table 3
Effect Sizes and Weights: High Versus Low Anchor

Study	Id_exp	*N*	Effect Sizes			Weight
			z	95% CI	Cohen's *d*	
Campbell et al. (2016)	1	108	.65	[.46, .84]	1.71	3.33
Campbell et al. (2016)	2	100	.50	[.30, .70]	1.19	3.27
Chapman and Bornstein (1996)	1a	11	.52	[−.18, 1.21]	1.23	.88
Chapman and Bornstein (1996)	1b	14	.71	[.12, 1.30]	1.94	1.12
Chapman and Bornstein (1996)	1c	14	.82	[.23, 1.41]	2.47	1.12
Diamond et al. (2000)	1a	46	.34	[.04, .64]	.74	2.50
Diamond et al. (2000)	1b	50	.55	[.27, .84]	1.34	2.60
Englich (2005)	1a	90	.41	[.20, .62]	.92	3.18
Englich (2005)	1b	84	.08	[−.14, .30]	.16	3.12
Englich and Mussweiler (2001)	1	19	.62	[.13, 1.11]	1.58	1.46
Englich and Mussweiler (2001)	2	44	.28	[−.03, .59]	.58	2.45
Englich and Mussweiler (2001)	3	16	.41	[−.13, .96]	.92	1.27
Englich et al. (2005)	1	42	.52	[.21, .84]	1.25	2.40
Englich et al. (2005)	2	42	.46	[.15, .77]	1.06	2.40
Englich et al. (2006)	1	42	.38	[.06, .69]	.83	2.40
Englich et al. (2006)	2	39	.33	[.01, .66]	.71	2.31
Englich et al. (2006)	3	52	.37	[.08, .65]	.80	2.64
Feldman et al. (2016)	2	57	.40	[.13, .67]	.89	2.74
Franck et al. (2017)	1	98	.66	[.46, .86]	1.75	3.26
Franck et al. (2017)	2a	114	.02	[−.17, .20]	.03	3.38
Franck et al. (2017)	2b	110	.17	[−.02, .36]	.36	3.35
Glöckner and Englich (2015)[a]	2c	44	.34	[.04, .65]	.74	2.45
Glöckner and Englich (2015)[b]	2d	44	.60	[.29, .90]	1.49	2.45
Hastie et al. (1999)	1	173	.33	[.18, .48]	.71	3.66
Hinsz and Indahl (1995)	2	120	.19	[.01, .37]	.39	3.41
Malouff and Schutte (1989)	1aa	78	.62	[.39, .85]	1.58	3.05
Malouff and Schutte (1989)	1ab	77	.56	[.34, .79]	1.38	3.04
Malouff and Schutte (1989)	1ac	78	.83	[.60, 1.06]	2.53	3.05
Malouff and Schutte (1989)	1ba	80	.73	[.50, .95]	2.01	3.08
Malouff and Schutte (1989)	1bb	78	.63	[.41, .86]	1.63	3.05
Malouff and Schutte (1989)	1bc	80	.46	[.24, .69]	1.06	3.08
Patel (2013)	1	100	.29	[.09, .49]	.61	3.27
Reyna et al. (2015)[c]	1a	44	.42	[.11, .72]	.93	2.45
Reyna et al. (2015)[d]	1c	44	.26	[−.05, .56]	.54	2.45
Stein and Drouin (2018)[e]	1a	122	.12	[−.06, .30]	.25	3.43
Stein and Drouin (2018)[f]	1c	126	.10	[−.07, .28]	.21	3.45
Stein and Drouin (2018)[g]	1e	118	.35	[.17, .53]	.76	3.40

Note. Id_exp refers to the number of a given study in the article in which it was reported. If there is no explicit numbering of studies in a given article, the first study in the article to deal with anchoring is indicated as 1, and so on. Within the studies employing more than one anchor, we list effect sizes for anchors starting from low to high. If the order of the anchors is different in the original article or we used one of the groups or conditions from the original article, this is indicated by the footnote.
[a] *Irrelevant* condition. [b] *Relevant* condition. [c] *Meaningful* condition, low vs. high anchor. [d] *Meaningless* condition, low vs. high anchor. [e] *Ignore* condition. [f] *Identify* condition. [g] *Counter* condition.

dependent effect sizes so that none of the effect sizes had to be omitted when calculating the overall effect sizes. Moreover, this method did not require information about the covariance structure of the effect size estimates (Tanner-Smith & Tipton, 2014). In our model we chose random effects weights and assumed within-study correlation between effect sizes, which is an average correlation between all pairs of observed effect sizes within studies. Then we ran metaregression models of interest, complemented by small sample adjustments (Tipton, 2015), starting with estimation of an unconditional mean effect size and testing whether it is significant. Subsequently, we used a sensitivity analysis approach, focused on reestimation using different values of correlation between 0 and 1 (every .1), obtaining consistent results regardless of changing correlation values. To validate obtained results, we used a classic meta-analytic procedure with additional weights, taking into account the problem of double-counting the participants within each study. Within each of included experiments, we calculated the weight for that comparison based on the overall sample size used in all the comparisons.

Additionally, we used several methods to assess the risk of bias across all the studies. We assessed the influence of moderators through random effects metaregression and comparisons of subgroups assuming robust variance estimators (Tanner-Smith & Tipton, 2014). The study also investigated the publication bias using trim-and-fill (Duval & Tweedie, 2000), which is recommended with observed larger degrees of heterogeneity (Renkewitz & Keiner, 2019).

Results

Reported statistics include the number of studies (k), the number of participants (N), weights, the effect sizes (z, d) and 95% CIs, the significance test, and the heterogeneity measure (i^2). Because we assumed that the studies included

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

might have come from different populations, we calculated the overall effect using a random effects model (two-stage inverse-variance model; Stata package *admetan and robumeta*, where at the first stage we used the experiment outcome, and at the second stage we used the *id* of a given experiment), employing the method of Mantel and Haenszel (1959). We assessed the influence of moderators using a random effects metaregression assuming dependence among estimators (Stata package *robumeta*). The frequency of moderators and the results of a mixed-effects regression analysis are displayed in Tables 4 and 5. For moderator variables, we reported p values of the comparison among effect sizes under the levels of the moderator (in the case of moderators with more than two levels, we compared the effect under a given level and the average effect of the remaining levels). For each level of a given moderator, we also reported the effect size and its associated p value. There were two moderators that we treated as continuous variables: sample size and anchor value (we analyzed the latter both as a continuous variable and as a three-level categorical variable). In these cases, we report mean, range, and regression coefficient and its associated p-value. Furthermore, we provide forest plots for both the Control versus Anchor and Low versus High Anchor groups (Figures 2 and 3).

Overall Effect (H1)

The overall effect size of anchoring on numeric legal decisions for studies employing the Control vs. Anchor design according to the RVE procedure was significant: $z = .27$, 95% CI [.21, .33], $d = .58$, 95% CI [.44, .73], $p < .001$. For the High vs. Low Anchor, the same test yielded the following statistically significant results: $z = .39$, 95% CI [.31, .47], $d = .91$, 95% CI [.69, 1.12], $p < .001$. Validation conducted by the classical meta-analytic procedure with additional weights (accounting for the double-counting problem) supported these results ($z = .26$, $d = .57$, $p < .001$ in the Control vs. Anchor Condition and $z = .41$, $d = 1.00$, $p < .001$ in High vs. Low condition). The homogeneity test statistic was significant for both conditions at $p < .001$. Based on the analyzed studies, we could conclude that the anchoring effect was present in both groups of studies (Control vs. Anchor and High vs. Low Anchor).

The i^2 heterogeneity measure, that is, the diversity between studies, yielded the following results: for Anchor vs. Control, $i^2 = 56\%$, and for High vs. Low Anchor, $i^2 = 68\%$. According to Higgins and colleagues (2019) this might indicate moderate or substantial heterogeneity. This level of heterogeneity should not however be considered surprising—although all of the analyzed studies examined the same psychological effect in the same area (legal decision-making), the diversity of employed scenarios, types of anchors, target values, studied populations, and other peculiarities of the experimental design was substantial. To investigate in-depth systematic differences between the primary research included in the meta-analysis, we also conducted metaregression and subgroup analyses, as described below (see Tables 4 and 5).

Moderator Analysis

Because we analyzed almost all moderators using effect sizes from both categories, the overall assessment of the moderator analysis

Table 4
Moderator Analysis for the Control Versus Anchor Design

					Effect sizes					
Moderator	P	k	Level	z'	p [z']	CI [z']	d	p [d]	CI [d]	I^2 [b]
Publication status	.54	10	unpublished	.33	.06[a]	[−.03, .68]	.73	.09[a]	[−.19, 1.64]	.57
		46	published	.25	<.001	[.20, .30]	.54	<.001	[.42, .66]	
Sample size	.13	56	mean = 104, range = 30−349	coef = −.13	.13	[−.30, .05]	coef = −.30	.16	[−.75, .16]	.53
Type of stimuli	.73	17	short written	.26	<.001	[.19, .33]	.55	<.001	[.38, .72]	.57
	.05	29	rich written	.20	<.001	[.14, .27]	.43	<.001	[.28, .58]	.52
	.11	10	audio/video	.46	.04[a]	[.05, .86]	1.10	.06[a]	[−.12, 2.30]	.43
Type of sample	.30	42	lay	.29	.29	[.20, .38]	.64	<.001	[.41, .86]	.56
		14	professional	.23	<.001	[.15, .31]	.49	.001	[.30, .67]	
Paradigm	.99	42	basic	.27	<.001	[.19, .35]	.58	<.001	[.39, .77]	.56
		14	standard	.27	.002[a]	[.19, .35]	.59	.003[a]	[.37, .80]	
Anchor value	.36	4	low	.21	.05[a]	[.02, .40]	.43	.05[a]	[.01, .85]	.57
	.09	42	medium	.28	<.001	[.20, .35]	.59	<.001	[.43, .77]	.54
	.09	5	high	.01	.07[a]	[−.04, .24]	.19	.08[a]	[−.09, .49]	.53
Anchor value	.12[a]	51	mean = 2.45, range = −.90−5.99	coef = −.04	.12[a]	[−.11, .04]	coef = −.08	.13[a]	[−.26, .10]	.54
Type of scale	.21	34	open	.25	<.001	[.18, .33]	.54	<.001	[.36, .72]	.56
		22	closed	.36	.002	[.21, .50]	.79	.003	[.41, 1.17]	
Relevance	.40	12	irrelevant	.24	.003	[.20, .36]	.52	.003	[.42, .74]	.58
	.72	29	potentially relevant	.27	<.001	[.16, .37]	.57	.001	[.32, .82]	.59
	.58	12	relevant	.28	.002[a]	[.18, .38]	.59	.003[a]	[.34, .84]	.57
Type of anchor	.01	18	ad damnum	.37	<.001	[.24, .50]	.82	.001	[.48, 1.16]	.46
	.053	5	damage cap	.43	.03[a]	[.11, .74]	.99	.04[a]	[.11, 1.87]	.55
	.10	4	defendant's counteroffer	.12	.11[a]	[−.15, .39]	.24	.11[a]	[−.30, .79]	.52
	.51	9	prosecutor's demand	.24	.19[a]	[−.69, 1.17]	.51	.21[a]	[−1.64, 2.65]	.56
	.09	20	other	.21	<.001	[.14, .27]	.43	<.001	[.28, .58]	.54
Area of law	.48[a]	43	civil	.28	<.001	[.21, .34]	.60	<.001	[.43, .76]	.56
		13	criminal	.22	.15[a]	[−.45, .88]	.46	.16[a]	[−1.05, 1.96]	
Other salient values	.05	47	mere anchor	.29	<.001	[.22, .36]	.63	<.001	[.46, .81]	.55
		8	counteranchor	.20	.006[a]	[.09, .30]	.40	.007[a]	[.18, .62]	

[a] Degrees of freedom in small number could result in untrustworthy p values (Tanner-Smith & Tipton, 2014). [b] I^2 statistic as estimated by meta-analysis with additional weights. With RVE estimation stata package '*robumeta*' were used, which does not enable direct I^2 estimation.

Table 5
Moderator Analysis for the High Versus Low Anchor Design

Moderator	*p*	*k*	Level	*z'*	Effect Sizes *p* [*z'*]	CI [*z'*]	*d*	*p* [*d*]	CI [*d*]	I^2 [b]
Publication status	.77	3	unpublished	.37	.14[a]	[−.64, 1.37]	.81	.16[a]	[−1.91, 3.53]	.69
		34	published	.40	<.001	[.31, .49]	.92	<.001	[.68, 1.16]	
Sample size	.07	37	mean = 7.2. range = 11−173	coef = −.002	.07	[−.01, .01]	coef = −.005	.07	[−.01, .01]	.64
Type of stimuli	.93[a]	10	short written	.40	.40[a]	[−.10, .90]	.95	.10[a]	[−.42, 2.32]	.67
	1.00	21	rich written	.39	<.001	[.29, .50]	.91	<.001	[.62, 1.20]	.66
	.92	6	audio/video	.39	.02[a]	[.12, .65]	.87	.03[a]	[.16, 1.57]	.69
Type of sample	.54	23	lay	.41	<.001	[.28, .51]	.94	<.001	[.59, 1.22]	.67
		14	professional	.37	<.001	[.27, .49]	.84	<.001	[.60, 1.16]	
Paradigm	.11	24	basic	.35	<.001	[.24, .45]	.78	<.001	[.50, 1.06]	.67
		13	standard	.46	<.001	[.38, .54]	1.06	<.001	[.84, 1.29]	
Anchor value	.73	5	low	.37	.02[a]	[.14, .60]	.82	.03[a]	[.18, 1.45]	.69
	.11	25	medium	.43	<.001	[.34, .51]	1.00	<.001	[.77, 1.23]	.60
	.15	7	high	.33	.09[a]	[−.10, .77]	.77	.13[a]	[−.42, 1.95]	.59
Anchor value	.90	37	mean = −.24, range = −.62 – .81	coef. = −.02	.91[a]	[−2.32, 2.27]	coef = −.07	.89[a]	[−5.50, 5.35]	.63
Type of scale	.12	20	open	.45	<.001	[.32, .58]	1.07	<.001	[.70, 1.45]	.61
		17	closed	.34	<.001	[.24, .44]	.75	<.001	[.50, 1.00]	
Relevance	.19	12	irrelevant	.33	.002	[.18, .47]	.72	.002	[.37, 1.06]	.64
	.08	24	potentially relevant	.46	<.001	[.35, .57]	1.08	<.001	[.76, 1.40]	.62
	.40[a]	2	relevant	.30	.23[a]	[−1.12, 1.72]	.65	.25	[−2.76, 4.06]	.68
Type of anchor	.12	16	ad damnum	.50	<.001	[.32, .67]	1.20	.002	[.65, 1.76]	.51
	.39	10	prosecutor's demand	.35	.003[a]	[.20, .50]	.79	.005[a]	[.40, 1.17]	.65
	.21	12	other	.32	.001	[.21, .44]	.71	.001	[.44, .98]	.64
Area of law	.31	21	civil	.43	<.001	[.30, .47]	1.00	<.001	[.64, 1.36]	.64
		16	criminal	.35	<.001	[.24, .46]	.79	<.001	[.51, 1.07]	
Other salient values	.77	25	mere anchor	.37	<.001	[.26, .47]	.84	<.001	[.57, 1.12]	.65
	.38[a]	10	counteranchor	.50	.02[a]	[.17, .82]	1.18	.03[a]	[.18, 2.18]	.63
	.24[a]	2	larger number	.31	.05[a]	[.03, .59]	.66	.051[a]	[−.01, 1.33]	.68

[a] Degrees of freedom in small number could result in untrustworthy *p* values (Tanner-Smith & Tipton, 2014). [b] I^2 statistic as estimated by meta-analysis with additional weights. With RVE estimation stata package '*robumeta*' were used, which does not enable direct I^2 estimation.

should take into account whether a similar pattern of results occurred for both groups of effect sizes. Because of the relatively small number of effect sizes under any category, we assumed that the replication of the same pattern would provide some evidence in favor of a given hypothesis, even if statistical significance had been reached only under one category (or even under none).

Methodological Moderators (H2–H3)

As for the type of stimuli, the analysis showed that only in the Control vs. Anchor design were there differences between effect sizes in the levels of the moderator (under the High vs. Low design, the values were almost the same across the levels). Although, contrary to our hypothesis, the audio/video level yielded the largest effect size in the Control vs. Anchor group, the effect size in the short written level was larger than in the rich written level of the moderator, a result which partially supported our hypothesis.

As for the hypothesis on the effect of sample type, the differences under both groups went in the predicted direction, with legal professionals anchoring less, although the differences did not reach statistical significance in either group. Thus, the evidence favoring H3 was limited.

Psychological Moderators (H4–H7)

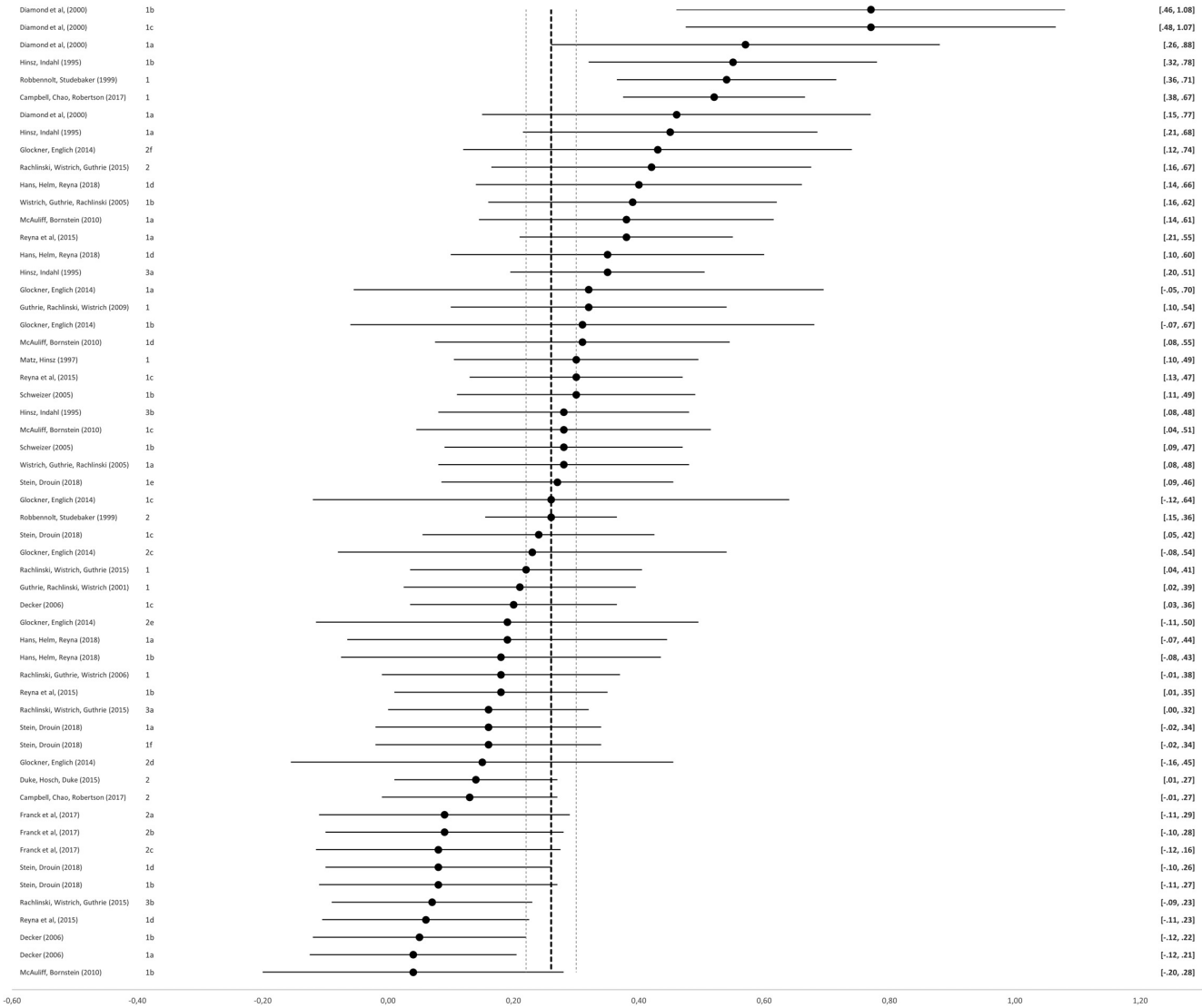
As for the paradigm moderator, there were no differences between the effect sizes in the Control vs. Anchor design, and the effect size under the standard paradigm in the High vs. Low Anchor was noticeably larger than under the basic paradigm, although the difference did not reach statistical significance. Hence, the hypothesis predicting stronger anchoring under the standard paradigm did not find any empirical support.

We treated the moderator anchor value as both a categorical and continuous variable to test H5. The analysis of this moderator using the categorical design yielded the following results: In both groups, anchors of medium value led to the largest effects (although in neither group was this level's effect size significantly larger than average), thus providing some evidence in favor of the hypothesis of the nonlinearity of the relation between the anchor value and the effect size. When treating the moderator as a continuous variable, the slope coefficient was not significant in either group or, if anything, took a negative value in both groups. This observation provides further support to the hypothesis that the relation between anchor value and the effect size is nonlinear, reaching its maximum for anchors of moderate values.

The effect sizes with respect to the type of scale moderator were as follows. Under the Control vs. Anchor design, the closed scale level of the moderator presented a larger effect size than the open scale level, contrary to our hypothesis. This situation was reversed in the High vs. Low Anchor design, where the difference went in the predicted direction, although it did not reach statistical significance. Such conflicting evidence implies that H6 lacks empirical support.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Figure 2
Effect Sizes for Control Versus Anchor Design



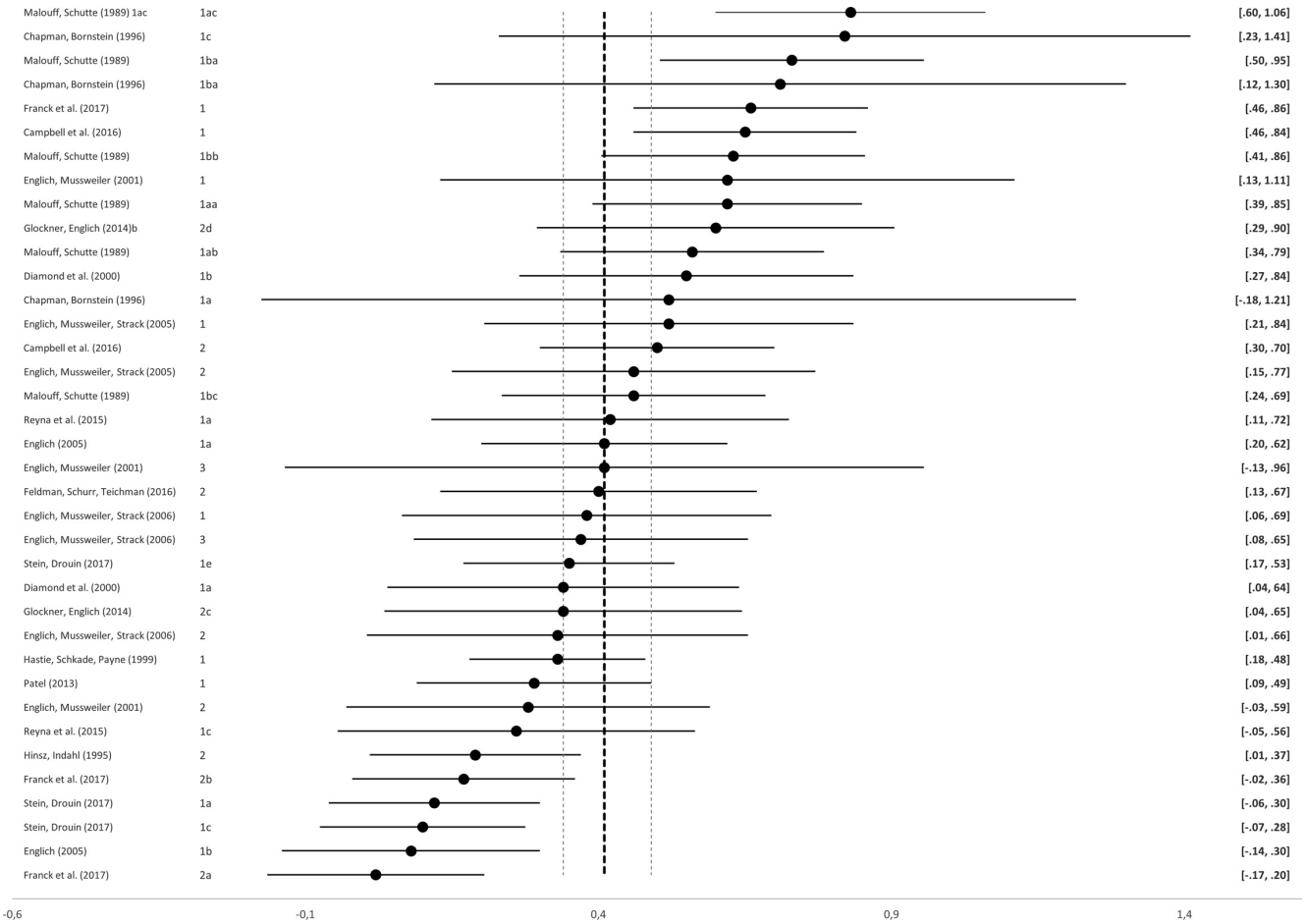
Legal Moderators (H7–H10)

For the relevance moderator, in the Control vs. Anchor design, the trend went exactly in the predicted direction, with irrelevant anchors leading to the smallest effect, and the relevant anchors resulting in the largest effects. The differences between levels, however, were very small and did not reach significance. This pattern was partially replicated under the High vs. Low design, where potentially relevant anchors resulted in larger effects than irrelevant anchors. Surprisingly, in this group, the relevant anchors led to the smallest effects, although this result was not reliable because of the very small number of effects included in this level. Thus, we obtained only limited evidence in favor of the predicted role of relevance.

The number of levels of the moderator type of anchor varied across designs, from five in the Control vs. Anchor to three in the High vs. Low Anchor design. In the latter group the differences between the levels and mean effect in others were not significant, but the effect size in the *ad damnum* level was higher than in other levels (even if the comparison was not statistically significant). In the Control vs. Anchor design, the situation was more complicated. The effect sizes in the *ad damnum* and damage cap levels were higher than in the other levels. The difference between *ad damnum* compared with the rest of the level was significant, and the difference in the damage cap condition is noticeable although not significant. The differences for the other levels were not significant and only the effect size in the level *other* was statistically significant. One should note the small number of studies in the counteroffer and prosecutor's demand levels of this moderator. To sum up, the results only partially corroborate H8, as the effect sizes resulting from *ad damnum* and damage caps were

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Figure 3
Effect Sizes for High Versus Low Anchor Design



consistently larger. We did not corroborate the part of the hypothesis concerning the prosecutor's demand, but this could have been the result of the limited number of studies in which this level of moderator was present.

As for the area of law to which the studies were referring (H9), the effect in civil cases was larger than in the criminal level for both designs (consistent with our hypothesis) even if the differences were not significant. Thus, we obtained some limited evidence in favor of the hypothesis in question.

The last hypothesis dealt with the number of salient values present in the decision context (H10). In the Control vs. Anchor group, we located only studies qualifying for two levels of the moderator: Mere anchor and Counteranchor with the difference between the two levels significant, and the trend in the size of the effects going in the predicted direction. In the High vs. Low Anchor design, the effect size was the largest in the case when the counteranchor was present, and smallest when multiple other salient values were present. However, differences between the levels of the moderator were insignificant. We should emphasize that the number of studies in both categories in which at least one anchor was present was very low, thus affecting the reliability of the analysis. In the Control vs. Anchor design, the statistically significant difference between Mere anchor and Counteranchor supports H10, but the situation in the High vs. Low Anchor design, in which the effect size in one counteranchor level was the largest, does not support this hypothesis.

Publication Bias

The trim-and-fill method (Duval & Tweedie, 2000) indicated the presence of significant publication bias (we validated the results via Egger's Test for small-study effects), both in the Control vs. Anchor and High vs. Low Anchor designs. In the Control vs. Anchor design, the trim-and-fill method suggested 14 additional effect sizes missing due to bias. The overall mean effect size was thus overestimated and had to be corrected to $z = .19$ ($p < .001$; 95% CI [.15, .24]) compared with $z = .27$, whereas in the classical meta-analysis the correction for Cohen's d equaled .40 ($p < .001$; 95% CI [.29, .52]). In the High vs. Low Anchor design, the diagnosis revealed as well the 14 filled effect sizes and the correction of z from its original value of .39 to .29 ($p < .001$; 95% CI [.21, .37]) and Cohen's d to .62 ($p < .001$; 95% CI [.40, .84]).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Filled funnel plots for both designs are shown in Figure 4. Additionally, the effect of the moderator sample size went in the direction indicating the presence of publication bias (with larger effect sizes obtained with smaller samples), although this result turned out to be insignificant for both groups.

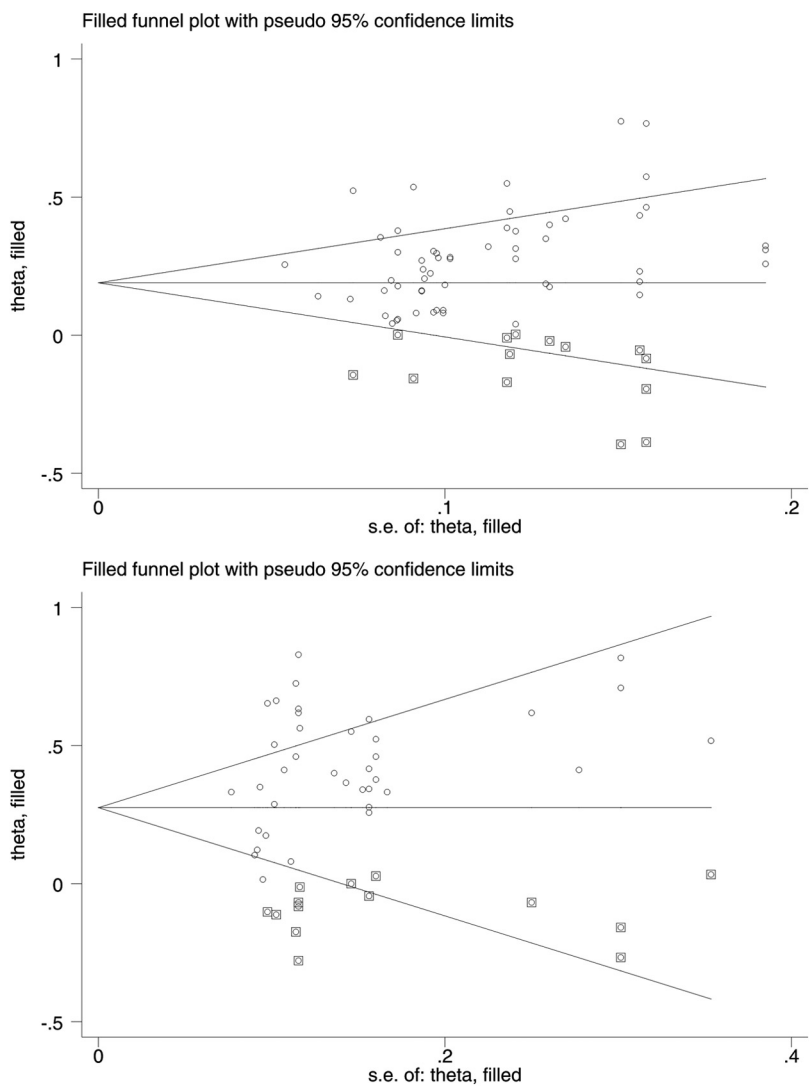
Discussion

Overall Effect

The overarching goal of this analysis was to examine whether the experimental research corroborated the thesis that salient numeric values present in the decision context affect numeric outcomes of legal proceedings. Our results indeed supported this thesis with solid evidence. Despite the necessity to split analyzed effect sizes into two groups—mostly because of fundamental experimental design differences—we found significant overall effect sizes of anchor presence in both groups.

Because the estimates of the average anchoring effect presented in the article may be biased as a result of the presence of a publication bias, we are inclined to interpret our results with caution (although one should also remember that estimation of publication bias itself may be problematic, because none of the detection methods test "directly" for publication bias [Renkewitz & Keiner, 2019] and they perform poorly when true effect sizes were heterogeneous, which is the norm rather than the exception [McShane et al., 2019]). Thus, we want our study to be treated not as concluding the investigations on legal anchoring but rather as inspiring further research in the field. Subsequent studies should focus on replication of existing primary studies both to

Figure 4
Funnel Plots for Control Versus Anchor (Up) and High Versus Low Anchor (Down) Design



Note. s. e. = standard error.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

verify the extent to which currently available data are skewed by the publication bias and to examine the sources of noticeable heterogeneity.

It is hard to assess how the overall effect sizes obtained in this analysis compare with the effect of anchoring in other contexts, because this is one of the first meta-analyses of research on anchoring in general. We cannot reliably compare our results with those from the meta-analysis on anchoring in negotiations (Orr & Guthrie, 2006) because of methodological divergences. In that meta-analysis, the researchers managed to treat anchor value as a continuous variable, which most likely made their overall effect size larger than it would have been under our approach. In our case, the Pearson *r* correlation coefficient for the High versus Low group (which can be compared with the studies on negotiations, where no control group was employed) was .38, making it lower than the result reported by Orr and Guthrie (2006; $r = .5$). Nevertheless, according to the general rules of thumb for interpreting the magnitude of effect size (Cohen, 1988; Sawilowsky, 2009), we can consider the obtained effect sizes as medium (Control vs. Anchor) and large (High vs. Low).

The necessity to split effect sizes into two groups was, on the one hand, a limitation of the current study because the limited number of studies in both groups restricted our ability to make more solid inferences. On the other hand, it allowed us to compare overall effect sizes under the two study designs. This was important because these two different types of design were not the result of research convenience but rather represent a more fundamental theoretical dispute (Englich et al., 2006). For all the measures that we employed, the overall effect size for the High vs. Low group was noticeably larger than the value under the Control vs. Anchor design (although there is a small overlap between the respective confidence intervals). This is certainly a result that can contribute to resolving that dispute. Despite the fact that the interpretation of this outcome depends on one's theoretical preconceptions, we might assume that it stemmed from the fact that the participants in the control group were still under the influence of some (potentially self-generated) anchors that increased the variability of target value assessments in that group, thus decreasing the observed overall effect size (Englich et al., 2006).

Type of Stimulus

Whereas in the High vs. Low design the estimates for all three levels of the moderator were almost equal, the differences for the Control vs. Anchor Group seem large enough to deserve a mention. In the latter case, as predicted, we found that short vignettes were somewhat better at isolating the effect of anchoring than longer written materials, but they might have done that at the cost of decreased ecological validity. What is interesting, contrary to our hypothesis, we observed a noticeably, although not significant, larger effect with audio/video materials. One could argue that perception of images causes increased attention and thus better acquisition of information, thus making the anchor more salient and easier to recall.

Type of Sample

Contrary to H3, we did not observe significant differences between lay and professional samples. This implies that concerns about making generalizations from one population to another might not be justified. We noted, however, a uniform trend under both analyzed groups of studies, with professional samples anchoring a bit less than lay ones. This result tracks an analogical trend observed in the meta-analysis on anchoring in negotiations (Orr & Guthrie, 2006). It may be the case that expertise insulates from some factors that reinforce the effect size of anchoring in the general sample (For instance, experts could be less influenced by their mood, see Englich & Soder, 2009. This conjecture would be consistent with the predictions of the elaboration-based model, see Wegener, 2010b.). We thus conclude that anchoring is a universal phenomenon, prevalent in different populations, yet further research on the moderating effect of legal expertise is warranted.

Paradigm

According to the assumption associated with the selective accessibility model, studies employing the standard paradigm should lead to larger effects than those obtained in the basic paradigm (Englich, 2008). However, our analysis showed that only in the High vs. Low Anchor group the difference went in the predicted direction, and even there it was not significant. The null result in the Control vs. Anchor group possibly suggests that either the issue at hand might be more complicated (meaning that another factor could be at play here) or that the *shallow* aspect of anchoring could perhaps explain the result, which would open the debate on the role of conscious and subconscious processing of anchors.

Anchor Value

Although moderate anchors led, as predicted, to the largest effect sizes (even if the result was not significant), high anchors produced effects smaller than the medium or low anchors. Moreover, treating the anchor value as a continuous variable failed to establish significant results. This, to some extent, contradicts the predictions of the selective accessibility model (Strack & Mussweiler, 1997). On the other hand, the present findings might find theoretical explanation in the elaboration-based account (Wegener et al., 2010b). One possible mechanism that might facilitate this effect is the role of plausibility of the anchor which is directly employed as a factor mitigating the effect in an elaboration-based account of anchoring. This issue clearly needs further research as one might test different levels of extremity of the anchors and perceived plausibility of the anchors as factors mitigating the magnitude of the effect.

Type of Scale

We hypothesized that, in accordance with scale distortion theory, the effect sizes would be smaller when the target value was assessed on a bound scale. Our highly inconclusive results in this case mean that this study does not provide further evidence in favor of this theory. Moreover, the fact that the overall effect sizes of anchoring for bound scales (in both groups) are significant means that it is not the open character of some scales used in law that is the primary reason behind the omnipresence of the anchoring effect in legal contexts and merely setting statutory boundaries (such as caps) would not insulate the decision-maker from the undesirable influence of an anchor.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Overall, the analysis of the psychological moderators confirmed our earlier conjecture that forming a unified theory of the anchoring effect may be problematic, as different points of analysis might support competing theories. Thus far, in most of the cases, the selective accessibility model supplemented by the elaboration-based model might account for the effect in question, but none of the studies of anchoring in legal settings explicitly employed a design based on scale distortion theory, so that conclusion may be premature.

Relevance

The current analysis offers very limited evidence in favor of the thesis that relevant anchors result in stronger effects. We still find this pattern of results important because it points out that, in line with some models of numeric judgment in law (Hans & Reyna, 2011), anchoring in legal contexts can have some desirable features. On the other hand, the fact that the overall effect sizes for irrelevant anchors remain significant corroborates the thesis that even absolutely immaterial numeric values can influence the outcome of legal proceedings (Englich et al., 2006). However, because differences between levels of this moderator were rather small and mostly failed to reach statistical significance, the category of relevance should be subject for further studies.

Presence of Other Salient Values

The results are inconclusive, mostly because of a small number of studies employing more than one anchor. This points out the important theoretical question that has been thus far rather neglected in the literature. Despite the fact that both theoretical models of anchoring and numerical decision-making in law describe the process of utilizing the anchor to get a target value, they devote little attention to the process of anchor selection. In a typical study, only one or two numbers appear as viable anchors. This may be not the case in real-life decision-making, when a plethora of numbers may be equally salient (see the study by Hastie et al., 1999, which analyzes how study participants can choose between different potential anchors).

Type of Anchor, Area of Law

In line with the arguments raised in the literature, our study highlights the importance of *ad damnum.* The opportunity to start with a numeric demand gives the plaintiff a clear first-move advantage. Furthermore, our results corroborate the concern that damage caps do not insulate from anchoring, as they are itself a potent anchor. Noticeably smaller effect sizes resulting from prosecutor's demand suggest that such a first-mover advantage is less of an issue in criminal law, although the pattern of results for the moderator area of law clearly shows that, in general, anchoring influences decisions in the criminal law context as well.

Limitations

Despite the importance of conclusions drawn from the present research, some limitations should be noted. The main factor that might influence the conclusions drawn is that we were forced to split the studies into two main groups, in accordance with the employed design. Although under both groups we obtained significant overall effect sizes, this procedure also diminished the number of studies in some of the moderator levels and resulted in a lower power of statistical analysis (limiting our ability to draw valid inferences regarding our hypotheses). What is more, in some cases, the pattern of results of the moderator analysis under the two designs went in the opposite directions making it impossible to reach univocal conclusions.

Moreover, because of the necessity of using the two design categories and because of the limited number of studies, some of the potentially important variables could not be treated separately and had to be included in one category. For instance, the moderator *relevance* was less fine-grained than we had hoped, ignoring, for example, the distinctive nature of formally inadmissible anchors (Wistrich et al., 2005) and the differences between meaningful and meaningless anchors (Hans et al., 2018). This potentially limits the depth of the conclusions offered and introduces a certain need for further studies.

Also, we faced a difficulty similar to that of other meta-analyses: We did not manage to get to all unpublished studies. Despite that we used methods for correcting this bias, one should be aware that the adjusted estimates are approximate. We had to deal with a lack of important data in published articles that were essential for overall effect estimation and its evaluation. However, we believe that the presented results will be the starting point for further discussion, and effect estimates will be updated with new research in subsequent meta-analyses, allowing us to dispel the doubts expressed in our article.

Implications for Researchers and Legal Professionals

It is of utmost importance that numerical legal rulings are issued in a biased and unpredictable way. The existing literature has suggested that the prevalence of the anchoring effect in legal contexts might be a major obstacle for achieving those goals (although one should remember that some anchors might have a positive role in legal decision-making; Helm et al., 2020). This synthesis of the experimental research provides further evidence suggesting the robustness of this effect in legal contexts. The effect of anchoring remains significant independently of the employed research design or branch of law used to construct stimuli. Its relative strength derives from its meaning, as some anchors are more potent than others—especially the initial demand of the party, which reveals a troubling effect from the viewpoint of the fairness of the proceedings.

The results of the moderator analysis point to some potential directions of further research. One example is the effect of the anchor's relevance to the case to be decided: It remains a much understudied topic (Glöckner & Englich, 2015; Hans et al., 2018; and Reyna et al., 2015 seem to be the only studies in the field to directly manipulate this variable), whereas the present study clearly indicates it has an impact on the effect size. Assessing the impact of relevance could also help to elucidate anchoring mechanisms in situations in which more salient values are present. For example, in a study by Hastie et al. (1999), in posttrial questionnaires, a number of participants admitted using anchoring and adjustment as a conscious strategy, arbitrarily choosing some of the numbers present in a defendant's financial statement as a basis for their calculations. Our review showed the scarcity of studies dealing with such issues as the effect of the inadmissibility of the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

values that are materially relevant and meaningful. If the saliency of the anchor plays a role here, exclusionary rules may increase it. The categories of inadmissibility and meaningfulness and their combinations may be only a germ for future studies, and the current results possibly justify the hope for an outcome that would be less worrying from the legal viewpoint.

Another example of a moderator worth further research is the susceptibility of professional legal decision-makers to anchoring. The common wisdom in the field is that they do not perform significantly better than laypeople (e.g., Englich & Mussweiler, 2001; Englich & Soder, 2009, comparing professional judges and students subject to the same experiment). This study, however, found slight and insignificant, but nonetheless consistent, results pointing to professionals' smaller susceptibility to anchoring, thus indicating there might be more to this story and that further research directly comparing legal professionals and laypeople would be worthwhile. A wholly new direction of research would be comparing the relative strength of anchoring in different branches of law, and this meta-analysis provides a good reason to believe that the effect might operate differently in different legal contexts. Last but not least, an important methodological issue arises from the present synthesis: the substantial effect of the method by which participants are familiarized with the case. To account for the pattern of results described here and to decide which experimental setting leads to a greater ecological validity is not only crucial to research on anchoring in the courtroom but could also prove important to experimental research on legal decision-making in general, as well as shedding some light on the cognitive basis of anchoring mechanisms.

Strikingly, in the present study we did not analyze the potentially moderating effect of employing some debiasing techniques, such as Larrick's (2004) cognitive (*training in biases*, *training in representations*, etc.) or motivational (accountability, etc.) strategies. Such analysis would be worthwhile, especially because there are at least two methods of debiasing anchoring whose effectiveness in legal contexts has found some corroboration: making individuals aware of existence and potential influence of anchoring (an example of training in biases; Stein & Drouin, 2018); and making them aware of their accountability for assessing the target value accurately (Guthrie et al., 2009; on the effect of accountability, see Lerner & Tetlock, 1999). Another potential debiasing measure (that could be classified as a case of training in representations) employed in the studies was changing the format of the anchor. In legal settings, this might be achieved by comparing anchors in the form of *lump sum* and *per diem* (Campbell et al., 2017; McAuliff & Bornstein, 2010). However, because of the small number of studies employing any of these debiasing measures, we were forced not to engage with this topic in the moderator analysis. Nevertheless, this suggests the urgent need for further research in this area, as effective debiasing techniques could prove invaluable in legal settings.

When it comes to implications for legal practice, the crucial issue is whether the parties should set their demands as high as possible. It seems that the dependence between the anchor value and effect size has its maximum, so the absurdly high anchors may not work as intended. However, even if the strength of the effect is somewhat diminished when an anchor is set very high, the assimilation persists. Therefore, setting the demand amount high (but not absurdly high) provides a solid advantage for the party initiating the proceedings. Despite the fact that, in the criminal context the effects are somewhat weaker than in civil proceedings, the pattern is similar. Can the defendant effectively respond by setting a counteranchor? Counteranchors appear to work to some extent, yet seem unable to substantially diminish the effect of the main anchor. It appears that the right to the first move favors the party initiating the proceedings.

Another consideration concerns whether the lawmakers can balance the scales, proving a level playing field for both parties to legal proceedings. The typical tool used to achieve this is to introduce damage caps. Our analysis corroborated the thesis that caps themselves work as anchors and, therefore, might work in a different way than intended (potentially lifting the damage amount instead of lowering it). On the other hand, debiasing methods seemed to limit the effect, but the data on their usage remain scarce, and it would be premature to base the policy on this assumption. Some more structural solutions to address this problem may be applicable. For instance, a promising proposition would be to provide the decision-maker with some values that would serve as meaningful and relevant anchors—such as the mean amount of damages awarded in a given type of case (Saks et al., 1997). The potential positive role of some kinds of anchors was recently underlined by Helm and colleagues (2020). As shown by our synthesis, multiple salient numbers present in the decision context seem to limit the anchoring.

Despite decades of research on anchoring in legal decision-making, a number of important issues remain unsolved, as we still do not have a single psychological theory encompassing all the complexities of this phenomenon. The most important directions for future research are related to the role which numbers play in the decision-making context: Which numbers serve as anchors in decision-making? and What is the role of the saliency of the number, and how does it drive the effect? After solving more theoretical controversies about the cognitive basis of the effect, methods of debiasing should be developed, as limiting the role of irrelevant or biased anchors is crucial to guarantee the fairness of legal proceedings.

Conclusion

The anchoring effect in legal decision-making seems to be a robust effect, as we estimated large and significant overall effect sizes even after having to split analyzed studies into two separate groups on methodological grounds. However, we also obtained evidence suggesting that research in this area might suffer from a publication bias, which hints that more further studies are needed before we can declare that legal anchoring is indeed a robust phenomenon. Although our moderator analysis largely did not yield significant results, the obtained pattern of results indicates the possible moderating effect of some legally important factors (such as legal expertise, the relevance of the anchor, or the anchor value), which will hopefully also lead to further research in this area. Moreover, the results of our analysis provide grounds for criticizing the two dominating theories of anchoring: scale distortion theory (i.e., because providing participants with a bound scale did not lead to decreased effect size) and selective accessibility model (e.g., studies using the standard paradigm did not lead to larger effect sizes).

References

References marked with an asterisk indicate studies included in the meta-analysis.

Bieneck, S. (2009). How adequate is the vignette technique as a research tool for psycho-legal research. In M. E. Oswald, S. Bieneck, & J. Hupfeld-Heinemann (Eds.), *Social psychology of punishment of crime* (pp. 255–271). Wiley-Blackwell.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Bornstein, B. H. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior*, *23*(1), 75–91. https://doi.org/10.1023/A:1022326807441

Bornstein, B. H., & Greene, E. (2017). *The jury under fire: Myth, controversy, and reform*. Oxford University Press.

Bornstein, B. H., Golding, J. M., Neuschatz, J., Kimbrough, C., Reed, K., Magyarics, C., & Luecht, K. (2017). Mock juror sampling issues in jury simulation research: A meta-analysis. *Law and Human Behavior*, *41*(1), 13–28. https://doi.org/10.1037/lhb0000223

Broeder, D. W. (1959). The University of Chicago jury project. *Nebraska Law Review*, *38*, 744.

*Campbell, J., Chao, B., & Robertson, C. (2017). Time is money: An empirical assessment of non-economic damages arguments. *Washington University Law Review*, *95*(1), 1–52.

*Campbell, J., Chao, B., Robertson, C., & Yokum, D. V. (2016). Countering the plaintiff's anchor: Jury simulations to evaluate damages arguments. *Iowa Law Review*, *101*(2), 543–571.

Chang, Y., Chen, K.-P., & Lin, C.-C. (2016). *Anchoring effect in real litigation: An empirical study* (SSRN Scholarly Paper ID 2726903). Social Science Research Network. https://papers.ssrn.com/abstract=2726903

*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

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

Daftary-Kapur, T., & Berry, M. (2010). The effects of outcome severity, damage amounts and counterfactual thinking on juror punitive damage award decision making. *The American Journal of Forensic Psychology*, *28*(1), 21–45.

Decker, T. L. (2007). *Effects of counter-anchoring damages during closing argument* [Unpublished manuscript]. Department of Communication Studies and the Faculty of the Graduate School, University of Kansas.

*Diamond, S., Ellis, L., Saks, M., & Landsman, S. (2000). Ad damnums and caps: Assistance or merely influence *Meeting of American Psychology-Law Society*.

Diamond, S. S., Rose, M. R., Murphy, B., & Meixner, J. (2011). Damage anchors on real juries. *Journal of Empirical Legal Studies*, *8*(Suppl. 1), 148–178. https://doi.org/10.1111/j.1740-1461.2011.01232.x

*Duke, M. C., Hosch, H. M., & Duke, B. (2015). The effect of liability stipulation on damage awards in a personal injury case. *Psychology, Public Policy, and Law*, *21*(3), 265–279. https://doi.org/10.1037/law0000050

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Ebbesen, E. B., & Konecni, V. J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology*, *32*(5), 805–821. https://doi.org/10.1037/0022-3514.32.5.805

Ellis, L. A. (2002). *Don't find my client liable, but if you do. . .: Defense recommendations, liability verdicts, and general damage awards* [Unpublished manuscript]. Department of Psychology, University of Illinois.

*Englich, B. (2005). Geben Sie ihm doch einfach fünf Jahre!": Einflüsse parteiischer Zwischenrufer auf richterliche Urteile [Give him five years!" - Influences of Partisan Hecklers on judges' sentencing decisions] *Zeitschrift Für Sozialpsychologie*, *36*(4), 215–225. https://doi.org/10.1024/0044-3514.36.4.215

Englich, B. (2006). Blind or biased? Justitia's susceptibility to anchoring effects in the courtroom based on given numerical representations. *Law & Policy*, *28*(4), 497–514.

Englich, B. (2008). When knowledge matters—Differential effects of available knowledge in standard and basic anchoring tasks. *European Journal of Social Psychology*, *38*(5), 896–904. https://doi.org/10.1002/ejsp.479

*Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom 1. *Journal of Applied Social Psychology*, *31*(7), 1535–1551.

*Englich, B., Mussweiler, T., & Strack, F. (2005). The last word in court—A hidden disadvantage for the defense. *Law and Human Behavior*, *29*(6), 705–722. https://doi.org/10.1007/s10979-005-8380-7

*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.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12*(5), 391–396. https://doi.org/10.1111/1467-9280.00372

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

Everitt, B., & Skrondal, A. (2002). *The Cambridge dictionary of statistics* (Vol. 106). Cambridge University Press.

Fariña, F., Arce, R., & Novo, M. (2003). Anchoring in judicial decision making. *Psychology in Spain*, *7*(1), 56–65.

*Feldman, Y., Schurr, A., & Teichman, D. (2016). Anchoring legal standards. *Journal of Empirical Legal Studies*, *13*(2), 298–329. https://doi.org/10.1111/jels.12116

*Franck, S. D., van Aaken, A., Freda, J., Guthrie, C., & Rachlinski, J. J. (2017). Inside the arbitrator's mind. *Emory Law Journal*, *66*(5), 1115–1173.

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

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *Journal of Socio-Economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

*Glöckner, A., & Englich, B. (2015). When relevance matters: Anchoring effects can be larger for relevant than for irrelevant anchors. *Social Psychology*, *46*(1), 4–12. https://doi.org/10.1027/1864-9335/a000214

Greene, E., Downey, C., & Goodman-Delahunty, J. (1999). Juror decisions about damages in employment discrimination cases. *Behavioral Sciences & the Law*, *17*(1), 107–121.

*Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2000). Inside the judicial mind. *Cornell Law Review*, *86*(4), 777–830.

*Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2009). The "hidden judiciary": An empirical examination of executive branch justice. *Duke Law Journal*, *58*(7), 1477–1531.

*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

Hans, V. P., & Reyna, V. F. (2011). To dollars from sense: Qualitative to quantitative translation in jury damage awards. *Journal of Empirical Legal Studies*, *8*(S1), 120–147. https://doi.org/10.1111/j.1740-1461.2011.01233.x

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Hastie, R. (2011). The challenge to produce useful "legal numbers." *Journal of Empirical Legal Studies*, *8*(1), 6–20. https://doi.org/10.1111/j.1740-1461.2011.01224.x

*Hastie, R., Schkade, D. A., & Payne, J. W. (1999). Juror judgments in civil cases: Effects of plaintiff's requests and plaintiff's identity on punitive damage awards. *Law and Human Behavior*, *23*(4), 445–470. https://doi.org/10.1023/A:1022312115561

Hedges, L. V., & Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin*, *96*(3), 573–580. https://doi.org/10.1037/0033-2909.96.3.573

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Helm, R. K., Hans, V. P., & Reyna, V. F. (2017). Trial by numbers. *Cornell Journal of Law and Public Policy*, *27*(1), 107–143.

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

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). Wiley.

Hinkelmann, K., & Kempthorne, O. (2008). *Introduction to experimental design*. Wiley-Interscience.

*Hinsz, V. B., & Indahl, K. E. (1995). Assimilation to anchors for damage awards in a mock civil trial. *Journal of Applied Social Psychology*, *25*(11), 991–1026. https://doi.org/10.1111/j.1559-1816.1995.tb02386.x

*Hinsz, V. B., & Matz, D. C. (1997, May). *Group member and individual reactions to anchors for damage awards.* [Paper presentation]. Meeting of the Midwestern Psychological Association, Chicago, IL.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*(11), 1161–1166. https://doi.org/10.1177/01461672952111004

Joober, R., Schmitz, N., Annable, L., & Boksa, P. (2012). Publication bias: What are the challenges and can they be overcome? *Journal of Psychiatry, & Neuroscience*, *37*(3), 149–152.

Kahneman, D., Schkade, D., & Sunstein, C. (1998). Shared outrage and erratic awards: The psychology of punitive damages. *Journal of Risk and Uncertainty*, *16*(1), 49–86. https://doi.org/10.1023/A:1007710408413

Kerr, N. L., & Bray, R. M. (2005). Simulation, realism, and the study of the jury. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 322–364). Guilford Press.

Kim, J., & Chae, S. (2017). Anchoring effect of the prosecutor's demand on sentence: Evidence from Korean sexual crime cases. *KDI Journal of Economic Policy*, *39*(3), 1–18.

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–338). Blackwell.

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

Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against nonsignificant findings. *Communication Monographs*, *76*(3), 286–302. https://doi.org/10.1080/03637750903074685

*Malouff, J., & Schutte, N. S. (1989). Shaping juror attitudes: Effects of requesting different damage amounts in personal injury trials. *The Journal of Social Psychology*, *129*(4), 491–497. https://doi.org/10.1080/00224545.1989.9712067

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*(4), 719–748.

Marti, M. W. (2000). *Anchoring biases and corrective processes in personal injury damages awards* [Unpublished manuscript]. Graduate College, University of Iowa.

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

Martin, E. G., & Alonso, C. H. (1997). Influence of the prosecutor's plea on the judge's sentencing in sexual crimes: Hypothesis of the theory of anchoring by Tversky and Kahneman. In S. Redondo, V. Garrido, & J. Pérez (Eds.), *Advances in psychology and law: International contributions* (pp. 215–226). Walter de Gruyter.

*McAuliff, B. D., & Bornstein, B. H. (2010). All anchors are not created equal: The effects of per diem versus lump sum requests on pain and suffering awards. *Law and Human Behavior*, *34*(2), 164–174. https://doi.org/10.1007/s10979-009-9178-8

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, *73*(Suppl. 1), 235–245. https://doi.org/10.1080/00031305.2018.1527253

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan.

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

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097–6. https://doi.org/10.1371/journal.pmed.1000097

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, *110*(3), 472–489. https://doi.org/10.1037/0033-295x.110.3.472

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

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, *135*(2), 151–166. https://doi.org/10.3200/GENP.135.2.151-166

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, *39*(1), 84–97. https://doi.org/10.1016/0749-5978(87)90046-X

Orr, D., & Guthrie, C. (2006). Anchoring, information, expertise, and negotiation: New insights from meta-analysis. *Ohio State Journal on Dispute Resolution*, *21*, 597–628.

Patel, S. (2013). *Off the scale: Scale distortion theory and the lack of scale-specificity of anchoring effects* [Unpublished manuscript]. Department of Experimental Psychology, University College London.

Pennington, N., & Hastie, R. (1991). A cognitive theory of juror decision making: The story model. *Cardozo Law Review*, *13*(2/3), 519–558.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing.

Petty, R. E., & Wegener, D. T. (1998). Matching versus mismatching attitude functions: Implications for scrutiny of persuasive messages. *Personality and Social Psychology Bulletin*, *24*(3), 227–240. https://doi.org/10.1177/0146167298243001

Petty, R. E., & Wegener, D. T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 37–72). Guilford Press Press.

Pezdek, K., Avila-Mora, E., & Sperry, K. (2010). Does trial presentation medium matter in jury simulation research? Evaluating the effectiveness of eyewitness expert testimony. *Applied Cognitive Psychology*, *24*(5), 673–690. https://doi.org/10.1002/acp.1578

*Rachlinski, J. J., Guthrie, C., & Wistrich, A. J. (2006). Inside the bankruptcy judge's mind. *Boston University Law Review*, *86*(5), 1227–1265.

*Rachlinski, J. J., Wistrich, A. J., & Guthrie, C. (2015). Can judges make reliable numeric judgments? Distorted damages and skewed sentences. *Indiana Law Journal*, *90*(2), 695–739.

Raitz, A., Greene, E., Goodman, J., & Loftus, E. F. (1990). Determining damages. *Law and Human Behavior*, *14*(4), 385–395. https://doi.org/10.1007/BF01068163

Renkewitz, F., & Keiner, M. (2019). How to detect publication bias in psychological research. *Zeitschrift Fur Psychologie Mit Zeitschrift Fur Angewandte Psychologie*, *227*(4), 261–279. https://doi.org/10.1027/2151-2604/a000386

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in Fuzzy-trace theory. *Judgment and Decision Making*, *7*(3), 332–359.

Reyna, V. F., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review*, *31*(2-3), 180–206. https://doi.org/10.1016/j.dr.2011.07.004

*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

*Robbennolt, J. K., & Studebaker, C. A. (1999). Anchoring in the courtroom: The effects of caps on punitive damages. *Law and Human Behavior*, *23*(3), 353–373. https://doi.org/10.1023/a:1022312716354

Saks, M. J., Hollinger, L. A., Wissler, R. L., Evans, D. L., & Hart, A. J. (1997). Reducing variability in civil jury awards. *Law and Human Behavior*, *21*(3), 243–256. https://doi.org/10.1023/A:1024834614312

Saks, M. J., & Kidd, R. F. (1980). Human information processing and adjudication: Trial by heuristics. *Law & Society Review*, *15*(1), 123–160.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–474. https://doi.org/10.22237/jmasm/1257035100

Schmitz, J. L. (2007). *The effect of random anchors on damage awards* [Unpublished manuscript]. LaGrange College.

Schum, D. A., & Martin, A. W. (1982). Formal and empirical research on cascaded inference in jurisprudence. *Law & Society Review*, *17*, 105–151.

Schweizer, M. (2005). *Kognitive Täuschungen vor Gericht: eine empirische Studie* [Cognitive illusions in court: An empirical study] [Unpublished manuscript]. Faculty of Law, University of Zurich.

Shepard, L. A. (1993). Evaluating test validity. In L. Darling-Hammon (Ed.), *Review of research in education* (Vol. 19, pp. 405–450). AERA.

*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), 393–428.

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

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*(1), 13–30. https://doi.org/10.1002/jrsm.1091

Teichman, D., & Zamir, E. (2014). Judicial decision-making. A behavioral perspective. In E. Zamir & D. Teichman (Eds.), *The Oxford handbook of behavioral economics and the law* (pp. 664–702). Oxford Handbooks.

Thomas, E. A., & Hogue, A. (1976). Apparent weight of evidence, decision criteria, and confidence ratings in juror decision making. *Psychological Review*, *83*(6), 442–465. https://doi.org/10.1037/0033-295X.83.6.442

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Tobia, K. P. (2020). *Legal concepts and legal expertise* (SSRN Scholarly Paper ID 3536564). Social Science Research Network.

Trope, Y., & Liberman, A. (1996). Social hypothesis testing: Cognitive and motivational mechanisms. In *Social psychology: Handbook of basic principles* (pp. 239–270). Guilford Press.

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

Viscusi, W. K. (2001). The challenge of punitive damages mathematics. *The Journal of Legal Studies*, *30*(2), 313–350. https://doi.org/10.1086/322059

Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010a). Elaboration and numerical anchoring: Breadth, depth, and the role of (non-)thoughtful processes in anchoring theories. *Journal of Consumer Psychology*, *20*(1), 28–32. https://doi.org/10.1016/j.jcps.2009.12.007

Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010b). 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

Weld, H. P., & Danzig, E. R. (1940). A study of the way in which a verdict is reached by a jury. *The American Journal of Psychology*, *53*(4), 518–536. https://doi.org/10.2307/1417631

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*(4), 387–402. https://doi.org/10.1037/0096-3445.125.4.387

*Wistrich, A. J., Guthrie, C., & Rachlinski, J. J. (2005). Can judges ignore inadmissible information? The difficulty of deliberately disregarding. *University of Pennsylvania Law Review*, *153*(4), 1251–1345.

Wong, K. F. E., & Kwong, J. Y. Y. (2000). Is 7300 m equal to 7.3 km? Same semantics but different anchoring effects. *Organizational Behavior and Human Decision Processes*, *82*(2), 314–333. https://doi.org/10.1006/obhd.2000.2900

Zuehl, J. J. (1982). *The ad damnum, jury instructions, and personal injury damage awards* [Unpublished manuscript]. Department of Sociology, University of Chicago.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Appendix A

Excluded Studies

Citation	Rationale
Daftary-Kapur and Berry (2010)	missing data (unable to contact the authors)
Ebbesen and Konecni (1975)	missing data (author contacted but unable to provide missing data)
Ellis (2002)	missing data (author contacted but unable to provide missing data)
Englich and Soder (2009)	missing data (author contacted but unable to provide missing data)
Greene et al. (1999)	presence of an entire expert opinion as a manipulation
Marti (2000)	missing data (author contacted but unable to provide missing data)
Marti and Wissler (2000)	missing data (author contacted but unable to provide missing data)
Raitz et al. (1990)	presence of an entire expert opinion as a manipulation
Saks et al. (1997)	awards variability as the dependent variable
Schmitz (2007)	missing data (unable to contact the author)
Viscusi (2001)	missing data (author contacted but unable to provide missing data)
Zuehl (1982)	missing manuscript

Appendix B

Statistical Formulas

Formula used to recount Mann Whitney test statistic: $r = \frac{Z}{\sqrt{N}}$

Fisher's z transformation: $z = .5\left[ln(1+r) - ln(1-r)\right]$

Standard error of the distribution of z statistic: $se'_z = \frac{1}{\sqrt{N-3}}$

Standard error of the distribution of r statistic: $se_r = \sqrt{\frac{(1-r^2)}{(n-2)}}$

Cohen's d: $d = \frac{(M_1 - M_2)}{SD_{POOLED}}$

Pooled standard deviation: $SD_{POOLED} = \sqrt{\frac{(SD_1^2 + SD_2^2)}{2}}$

CI for Cohen's d: $\sigma(d) \ = \sqrt{\frac{N_1+N_2}{N_1 \times N_2} + \frac{d^2}{2(N_1+N_2)}}$

95% CI for Cohen's d: $[d - 1.96 \times \sigma(d), d + 1.96 \times \sigma(d)]$

Received January 29, 2020
Revision received December 3, 2020
Accepted December 17, 2020 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.