Guiding Jurors' Damage Award Decisions:

Experimental Investigations of Approaches Based on Theory and Practice

V.P. Hans, K. Reed, V. F. Reyna, D. Garavito, and R. K. Helm

Abstract

Theory and practitioner "scaling" advice informed hypotheses that guidance to mock jurors should: (a) increase validity (vertical equity), decrease variability (reliability), and improve coherence in awards; (b) improve subjective experience of jurors' decision-making (rated helpfulness, confidence, and difficulty); and (c) guidance should have greatest impact when it includes both verbal and numerical benchmarks. Three mock-juror experiments (N = 197 students, N = 476 MTurk workers, and N = 391 students) tested novel scaling approaches and predictions from the Hans-Reyna model of damage award decision-making. Jurors reviewed a legal case and provided dollar awards to compensate plaintiffs for pain and suffering following concussions. Experiments varied injury severity (low v. high) and the plaintiff attorney's guidance (no guidance, verbal guidance, numerical guidance, and verbal-plus-numerical guidance) between subjects. Results support predictions that, even without guidance, mock jurors appropriately categorize the gist of injuries as low or high severity, and dollar awards reflect that gist. Participants gave higher awards for more severe injuries, indicating that they extracted the qualitative gist of damages. Also, as expected, guidance, particularly verbal-plus-numerical guidance, had beneficial effects on jurors' subjective experience, with participants reporting that it was a helpful aid in decision making. Numerical guidance, both with and without verbal guidance, reduced award variability in severe injury cases in all three experiments. Scaling

guidance did not improve the already strong gist-verbatim correspondence or award validity. Both grasping the gist of damages and mapping that gist onto numbers are important, but jurors appear to benefit from assistance with numerical mapping.

Keywords: anchoring; concussions; damage awards; fuzzy-trace theory; jury decisions; scales

Guiding Jurors' Damage Award Decisions:

Experimental Investigations of Approaches Based on Theory and Practice

A century ago, a Virginia court noted that "the law wisely leaves the assessment of damages, as a rule, to juries, with the concession that there are no scales in which to weigh human suffering" (*Chesapeake & Ohio Railway Company v. Arrington*, 1919, p. 423). Assessing the dollar value of an injury is a deeply challenging task, yet we ask juries and trial judges to perform this task every day in civil trials. We instruct juries to "price the harm on a case by case basis" (Kalven, 1958, p. 160). From one perspective, this is ideal. Human beings can process the meaning of the facts of each case in context (Reyna, 2020). A jury is also able to incorporate community notions of the worth of an injury into its decision-making about damage awards, thus fulfilling its representational role (Hans, 2014).

However, evaluating a single case in time, without reference to other cases or other injuries, can be difficult for juries. Perhaps it is not surprising that damage awards are highly variable (Greene & Bornstein, 2003; Saks et al., 1997; Wissler et al., 1997) and that civil jurors consider damage award determinations to be one of the most difficult aspects of their decision-making task (Mott et al., 2000). Therefore, it is important to consider how we can best guide jurors in awarding damages for a plaintiff's pain and suffering.

The present project attempts to understand how the courts can help guide jurors in making damage award decisions. The paper first reviews the existing research on guiding damage awards, then describes a theoretical framework, the Hans-Reyna model of damage award decision making, that offers several metrics that can be used to evaluate the effectiveness of jury guidance. We then present three experiments aimed at testing whether a novel form of guidance (attorney scaling advice) assists jurors in reaching better damage award decisions. We conclude

with a discussion of the implications of the findings for law and policy surrounding damage award decision making.

Mechanisms Guiding Damage Awards

Scholars have proposed a variety of approaches to guide damage awards: legal instructions, damage award caps, damages scheduling, comparable cases, and *ad damnum* recommendations. However, each of these approaches has potential limitations. Legal instructions do not provide juries with clear and specific guidance about assigning damages in the case at hand and may be too vague to be useful. Current legal instructions offer only general advice to provide compensation that is "fair and reasonable, neither inadequate nor excessive" (Modern Federal Jury Instructions-Civil, 77.01, 2018; Sherwin & Bray, 2020).

Damage award caps. A number of states have imposed caps or limits on damage awards, primarily for noneconomic damages such as pain and suffering (Avraham, 2016; Black et al., in press). Because they place upper limits on recovery, caps decrease the amounts paid out by civil defendants, yet they have a disproportionate effect on recoveries for the most seriously injured plaintiffs (Black et al., in press). Some studies have found that caps can affect some of the jurors who are aware of the cap, biasing their judgments; and to the degree that the cap is irrelevant to the case, this additional noise will increase variability in awards(see discussion in Bavli, 2017; Saks et al., 1997).

Schedules. Another recommendation is to use award schedules to determine compensation for personal injury and pain and suffering awards, similar to the approach used in workers' compensation (Baldus et al., 1995; Bavli, 2017; Bovbjerg et al., 1989). For example, a judge or a jury in a case in which a plaintiff who lost an arm in a negligently-caused accident would consult the award schedule to determine the amount of compensation the plaintiff would

receive for the injury. The award schedule amount could be binding or nonbinding, and might consist of a specific dollar amount or a range of dollar amounts. However, these remedies may fail to account adequately for the contextual specifics of each case and, even when non-binding, may push decision-makers towards similar awards in cases that are meaningfully different, thereby reducing variability at the expense of validity. Hence, using such schedules may usurp the role of the jury, curtailing the benefits of a contextually-sensitive damage award decision-maker as discussed above.

Additur-remittitur approaches. The courts generally eschew schedules and predetermined award amounts in ordinary tort cases, instead favoring the judicial devices of additur and remittitur to adjust jury awards in response to new trial motions on the grounds that the award is inadequate or excessive (Bavli, 2017; Field et al., 2020). A New York statute (New York State Senate, Civil Practice Law and Rules, 5501(c), no date) instructs judges responding to such motions to assess whether a jury's damage award "deviates materially from what would be reasonable compensation" by considering the awards in comparable cases. It is unclear exactly how judges engage in that comparison process, selecting appropriate comparator cases and identifying relevant dimensions. In addition to the challenges of the comparison process, judicial assessment of the appropriateness of a jury's damage award creates a danger of substituting the judge's intuition for the jury's (Baldus et al., 1995).

Baldus and his colleagues (1995) suggest the development of statewide databases to enable empirically based comparative analysis in additur-remittitur review. Using mock-juror paradigms, Bavli and Mozer (2019) and Saks et al. (1997) conducted useful empirical tests of comparable-case guidance. Information about awards in cases that were identified as similar had significant effects on mock jurors' damage award judgments, in particular reducing the

variability of damage awards. However, the same challenge noted above of identifying the appropriate parameters of similarity across cases applies-- which cases should be considered alike. What is more, the likelihood of introducing this potentially useful reform appears low, given strong political resistance to substantial reform of the civil jury (Schuck, 1993).

Ad damnums. In most jurisdictions, trial lawyers are able to provide guidance to jurors through the mechanism of an *ad damnum* request, the recommendation of specific damage award amounts to civil juries (Campbell et al., 2017). The experimental literature on the impact of *ad damnum* requests generally shows that recommending specific dollar amounts influences damage award amounts (Campbell et al., 2016; Chapman & Bornstein, 1996; Diamond et al., 1998; Hastie, et al., 1999; Hinsz & Indahl, 1995; Malouff & Schutte, 1989; Marti & Wissler, 2000). Even extreme *ad damnum* requests generally affect award judgments, an influence that may be considered problematic. However, extreme cases can merit extreme awards, and thus extremity by itself cannot be taken as problematic, by itself, out of context.

Scaling approaches. In sum, existing approaches to guiding jurors in damage award decisions appear to have limited -- or even arguably negative -- effects. Legal instructions are not specific enough to be helpful. Giving jurors specific dollar amounts, whether in the form of award schedules, *ad damnum* requests, or caps, clearly affects jurors' damage award decisions, although whether they improve damage award decisions is open to question. And although schedules, case comparisons, or judicial actions might reduce variability, they could also operate to undermine the litigants' right to individualized damages assessments.

Juror Subjective Experiences

One generally neglected effect of guidance approaches is how they might improve subjective experience of jurors. Mott et al. (2000) interviewed civil jurors after their trials about

their service. A number of jurors found the award decision to be difficult, reporting surprise and confusion that the legal instructions did not offer them adequate direction: "we were looking for guidance but couldn't get guidance…" (Mott et al., 2000, p. 409). Although subjective experience is not necessarily reflective of accuracy (Reyna et al., 2016, but see Wixted & Wells, 2017), confidence in the process of awarding damages arguably supports trust in the jury system and civic engagement (Bornstein et al., 2020; Gastil et al., 2010; Hans et al., 2014). Thus, we are interested in whether the proposed guidance improves the jurors' subjective experience.

Informed by the findings of this rich literature on guiding or constraining jury damage awards, the current project expands on it in several ways. We draw on the Hans-Reyna model of damage award decision making, a model informed by psychological theory and tested experimentally, to consider from a theoretical perspective what types of guidance might be most effective in generating sound, high quality damage award decision making (Hans & Reyna, 2011). We draw on theory and previous research to identify different dimensions of damage award decision making beyond variability that can be used to assess their quality. Also, we test a novel form of guidance, an attorney's widely taught (but heretofore untested) scaling advice, along these multiple dimensions to determine whether scaling advice promotes sounder damage award decision making.

The Hans – Reyna Model of Damage Award Decision Making

The research program initiated by Hans and Reyna (2011), based on underlying psychological theory, has tested methods that might assist jurors in translating their qualitative sense of the worth of an injury into a corresponding and appropriately matched dollar award. Hans and Reyna (2011) proposed a model of damage award decision making that draws on fuzzy-trace theory (FTT; Reyna & Brainerd, 2011). FTT assumes that people encode both

verbatim and gist mental representations in parallel but retrieve these representations separately, and can rely to differing extents on one or the other depending on the task. Verbatim representations reflect the surface form of the information such as exact numbers—the facts taken literally, whereas gist representations capture the underlying meaning of the information in its context, such as the categorization of numbers as large or small. Research on FTT has shown that adult decision makers typically rely on the simplest gist possible to make decisions (preferring categorical distinctions where this distinguishes between decision options, then ordinal distinctions, then more detailed differences, as the task demands). Therefore, gist-based decision-making is a more natural, intuitive, and meaningful way for most adults to process information that avoids pitfalls of literal or superficial thinking.

Hans and Reyna incorporated these ideas into their multi-stage model of damage award decision making. Although these stages unfold roughly in parallel, they represent distinct processes. According to this model, jurors make a categorical judgment about whether damages are warranted (yes or no) and, if warranted, make ordinal judgments about injury severity and deserved damages. For example, the damage award deserved might be nil (nothing), low, medium, or high (an ordinal gist judgment). Jurors look for a dollar amount that to them is in line with this ordinal gist. Hans and Reyna anticipate that jurors will draw on symbolic numbers from everyday life that already have significance to them as low or high numbers. They will also have available dollar award amounts from the case they are deciding, such as medical costs, lost wages, and attorney *ad damnum* requests (Hastie, 2011). Using these sources of numbers, jurors produce an award amount that fits their ordinal (nil, low, medium, high) gist of the deserved damages.

While jurors are likely to be accustomed to making more gist-based categorical and

ordinal judgments, they are likely to experience difficulty assigning a specific number to these judgments. The number-assignment process may be particularly difficult in jurors with low levels of numeracy, who lack "number sense" – the ability to interpret the meaning of numbers in context (though rote skills are not intuition; Helm et al., 2020; Helm et al., 2017; Peters, 2012; Reyna & Brust-Renck, 2020; Rowell & Bregant, 2014). This could create problems when trying to decide which numerically "low" number to assign to a level of suffering that warrants a "low" level of damages and, analogously, to other levels of suffering. Low numeracy has been associated with a lack of number sense in other contexts (Geary et al., 2011; Liberali et al., 2012; Siegler & Booth, 2004) and with reliance on irrelevant cues in decision making (Peters, 2020; Reyna, Nelson, et al., 2015). Thus, the model anticipates that while jurors may be well equipped to determine whether a "low," "medium," or "high" damage award is warranted, they may struggle to allocate a specific dollar award amount to this ordinal gist. The problem could be most acute for low numeracy jurors, but the matching process could be challenging even for jurors with stronger numerical abilities.

Some may say that it is impossible to identify a "correct" damage award amount in a particular case, since each case is unique. That may be true, but even if one cannot readily identify an absolutely correct award amount, what distinguishes the current approach from this claim is our assumption that one can identify better and worse award amounts. Traditionally, jury research has focused on the variability of awards, with reduced variability being seen as desirable through indicating increased convergence among jurors (*reliability*) and thus increasing levels of horizontal equity (similar awards for similar injuries) (Saks et al, 1997). To a lesser extent, jury research has also examined whether damage awards show vertical equity (*validity*), such that higher awards are given in more severe cases (Saks et al., 1997). The Hans-Reyna

model also suggests a third component, *gist-verbatim coherence.* This component measures the extent to which people's *gist* judgments of the award deserved (i.e., perceptions of damages as low, medium, or high) correspond appropriately to their *verbatim* judgments (dollar award amounts). So as the gist judgment, elicited separately, increases from low to medium to high, the dollar award amount should increase. Fourth, the juror's subjective experience of reaching the award should be considered, notably, the perceived difficulty in determining an award amount and the confidence in the award judgment, for reasons noted above. In sum, the dollar award should match up with the juror's underlying gist of the severity of the injury and the deservingness of damages and do so validly and reliably, and in a way that improves jurors' metacognitive perception of their subjective experiences.

In a series of prior experiments that inform the present work, Hans, Reyna, and their collaborators have tested some predictions from the model, and have examined the efficacy of different approaches to help jurors better match their qualitative judgments about the severity of injury with dollar award amounts (Hans et al., 2018; Helm et al., 2020; Reed et al., 2019; Reyna, Hans, et al., 2015). Using mock juror research methodology, Hans, Reyna, and collaborators found that meaningful anchor numbers, suggestions about the relative magnitude of numbers, and the participants' high level of numeracy all affected features of mock jurors' damage award decision making as expected. These experiments confirmed that qualitative factors, such as meaningfulness and ordinal comparisons, as well numerical ability, helped mock jurors identify damage amounts that matched their underlying sense—the gist—of an individual's pain and suffering. Consequently, as predicted, award variability was reduced and award validity was increased.

The current experiments test the impact of an additional form of potentially valuable

guidance for jurors. We examine the effects of lawyers' scaling guidance on four specific elements: *award validity* (distinguishing between mild and severe cases), *award reliability* (converging on an award amount), and *gist-verbatim coherence* (agreement between the perceived ordinal gist of damages and the award amount). We also measure the extent to which jury guidance increases the positivity of jurors' *subjective experience* (i.e., their metacognitions about reaching an award judgment); in particular, whether they feel confident, experience little difficulty, and find the guidance to be helpful in the task of damage award determination, thus improving their subjective experience. The current experiments use case characteristics (varying whether damages are described as mild or severe), jurors' perceptions of the gist of damages, and jurors' subjective experiences to assess whether scaling guidance helps jurors converge desirable features of damage award decisions, namely, enhancing validity, reliability, gist-verbatim coherence, and experience.

Current Experiments: Testing the Use of Scales

The experiments presented in this article go beyond prior tests of the Hans-Reyna model to examine types of guidance ranging from direct suggestions of award amounts (*ad damnum* recommendations) to those in which jurors are provided with verbal and numerical scales to calibrate the severity of damages, the latter designed to help jurors map perceived severity onto a dollar value. Specifically, the experiments examine whether providing jurors with help identifying the gist of an injury as low, medium, or high (ordinal "gist allocation" judgments in the Hans-Reyna model) and/or providing jurors with help allocating a number to that gist ("number allocation" in the Hans-Reyna model) improves psychometric and coherence properties of the award judgments and helps jurors feel more comfortable making damage award decisions.

The experiments drew inspiration from psychologist Hastie (2011) and trial consultant Ball (2011), both of whom encourage the use of scales in damage award decision making. Hastie (2011) suggested that legal decision making might benefit from a version of scales commonly employed in domains outside of law, such as magnitude scales employed in psychological research for self-reports of pain, although he cautions that scales must reflect the level of needed resolution (nominal, ordinal, interval, or ratio scale). Ball (2011) proposed that attorneys employ a scales approach to help jurors map the severity of the injury onto dollar values. In addition, as discussed below, Ball's guidance dovetails with FTT and the Hans-Reyna model in distinguishing ordinal classifications and magnitudes of damages associated with each classification. The current experiments test various forms of Ball's proposed guidance.

More specifically, Ball suggests that attorneys explain to jurors that the value of each noneconomic harm is based on three elements: (1) how bad the injury is, (2) how long the plaintiff is likely to experience pain or disability; and (3) how much it interferes with the plaintiff engaging in regular activities. He then offers suggested language for the attorney to guide the juror in the details of employing a scale: "Where on the scale of disability does each harm lie? A disability can interfere with functioning anywhere from hardly at all up through total incapacity….let's go up the scale another step to 'high'….In a case like that, I would have to ask you for a verdict of a great deal of money, in the same proportion as the pain" (Ball, 2011, p. 239). He recommends that, in jurisdictions where specifying specific dollar amounts for pain and suffering is allowable, lawyers should integrate these dollar amounts into their arguments. So, for example, rather than saying "I would have to ask you for a verdict of a great deal of money," the lawyer might say "I would have to ask you for a verdict of more like a million dollars**.**"

Ball's approach has the potential to help at two stages of the Hans-Reyna model of damage award decision making. First, it assists jurors with identifying the gist of an injury (how bad it is on an ordinal scale) and the associated ordinal award that is warranted (e.g. as low, medium, or high), the gist allocation stage in the decision-making process. Second, it encourages them to match the injury magnitude to the dollar magnitude. In this connection, in jurisdictions where specifying dollar amounts is acceptable, it advocates recommending numerical amounts for the ordinal gist (e.g. "more like a million dollars"), the number allocation stage in the decision-making process.

Ball's magnitude of outcomes scale is a kind of juror instruction or direction that might be feasible to adopt with actual jurors. In principle, scaling approaches can reduce irrelevant inconsistencies, for example, larger damage awards for similarly-injured plaintiffs who sue richer (compared to poorer) defendants (when actual damages are identical). Drawing attention to key dimensions of the magnitude of an injury, namely, how bad, long, or interfering it is, has the potential to reduce spurious variability in award judgments. However, it is not clear that scaling approaches take appropriate contextual variability into account, as the Hans and Reyna (2011) model suggests they should. Gist representations integrate multiple contextual factors to arrive at a judgment about whether an amount is low or high (Reyna, 2021).

In particular, according to the Hans-Reyna model, verbal guidance aimed at helping jurors classify injury, and the associated award amount, ordinally (without providing numerical values) (*verbal-only guidance*) may help jurors in grasping the gist of a case, particularly in the case of an injury (such as a concussion) where the gist of the injury is often intangible (e.g., social and cognitive impairments). Adding specific dollar award amounts to these instructions has the potential to provide additional help to jurors through also directly assisting with the

mapping of gist meanings onto specific magnitudes, thus calibrating high severity of damages with commensurately high amounts of money (and low damages with low amounts).

However, it is important to note that merely providing arbitrary numbers in instructions—without aiding meaning or mapping--may have a biasing effect by "anchoring" juror decisions inappropriately (Feldman et al., 2016). The psychological phenomenon of meaningless anchoring is a well-known and well-documented heuristic (Kahneman, 2011). Anchoring refers to the process in which a specific number provides a starting point for a judgment. As people determine a final judgment, they make adjustments away from an initial anchor number, but these adjustments are often insufficient (Rachlinski, et al., 2015; Robbennolt & Hans, 2016). Therefore, one is able to observe the impact of providing an anchor number by comparing final judgments when the anchor number was or was not provided. Anchoring research has typically been done with irrelevant numbers, such as outcomes from spins on a roulette wheel or phone numbers. However, as summarized above, previous work has shown that meaningful anchors can have greater impact than meaningless anchors, and that this impact may be beneficial through helping jurors understand the gist of magnitudes and allocate awards more consistently, rather than biasing their judgments. That is, participants may calibrate their awards more appropriately by comparing them to meaningful benchmarks while also reducing variability (Hans et al., 2018; Helm et al., 2020). Hence, lawyers' verbal guidance could create greater meaning for the recommended anchor number and so allow the guidance to function in these beneficial ways, rather than having a biasing influence. These beneficial effects of numerical meaning and mapping should be less likely when numbers are given without verbal guidance, as in *ad damnum* requests.

Here, we test the effect of purely verbal scaling guidance (in all experiments), verbal plus numerical scaling guidance (in all experiments), and purely numerical scaling guidance (in Experiment 3 only, as an *ad damnum*), on damage award amounts and on jurors' subjective perceptions of the experience of damage award decision making. We compare decisions in the guidance conditions to a no-guidance control condition in which jurors are told to use their best judgment. As we discuss below, we selected concussion, and its physical and psychological sequelae, as the focus of our research.

Concussions

The experiment employs a specific type of injury—sports-related concussions—that are challenging yet important to assess. The short and long-term consequences of concussions are of growing concern and the subject of ongoing research (Albicini & McKinlay, 2018; Garavito et al., 2018; McCrory et al., 2017). Concussions can affect a person's life and well-being to varying degrees; some individuals recover quickly yet others have persistent symptoms (Dumke, 2017; Graff et al., 2019; Losoi et al., 2016). At the same time that the medical and psychological research about concussions has been expanding, the legal environment surrounding concussions has also changed, opening up avenues for redress through the courts (Hamilton, 2019; Ward et al., 2017). How one evaluates the pain and suffering of a person who has had a concussion is a major challenge, one that is nonetheless necessary if compensation is called for. In evaluating the severity of the consequences of a concussion, guidance may prove to be especially useful.

Hypotheses

The purpose of this study is to test the effectiveness of varying forms of attorney guidance (verbal only, numerical only, and verbal-plus-numerical) in 1) assisting jurors in getting the gist of the injury, 2) assisting jurors in assigning a dollar amount to the injury, and 3)

improving juror's subjective experiences. Nine hypotheses based on the Hans-Reyna Model fall into these three categories.

Jurors getting the gist (ratings). We predict that jurors will generally get the *gist* of injury severity (due to the natural tendency to encode and rely on gist) as reflected in ordinal ratings, and that attorney guidance will improve further the ability of jurors to get the gist. Specifically, we expect that there will be a main effect of manipulated case severity on severity ratings such that jurors will rate the more severe case as higher in pain and suffering than the less severe case (Hypothesis 1; also a manipulation check). We predict this effect will interact with guidance, such that jurors will differentiate severity ratings between cases more with guidance than without (Hypothesis 2).

Jurors allocating dollar damage awards (dollars). According to the Hans-Reyna Model, we predict that jurors will allocate a numerical damage award consistent with their gist understanding of the case. We hypothesize that guidance, especially verbal-plus-numerical guidance (which helps with both getting the gist and assigning a number to it), will improve a jurors' ability to better match a dollar number to their award, which the modified Ball instructions directly address. Some might argue that plaintiffs, defendants, and their lawyers in civil cases consider a desirable judgment exclusively in terms of the outcome, with the plaintiff's side seeing high damage awards as more desirable and the defendant's side just the opposite. In terms of what is considered desirable in the current experiments, we look specifically at desirable outcomes considered by Helm et al. (2020) and Reyna et al. (2015): *award validity, reliability* and *verbatim-gist coherence*.

That is, we expect jurors will give higher damage awards in the more severe case (i.e., *award validity* or vertical equity; Hypothesis 3). We also expect guidance will increase award

validity, such that participants who receive such guidance will make greater distinctions in their awards between the mild and severe cases (Hypothesis 4). We expect, too, that guidance will improve *reliability*, such that jurors who receive such guidance will converge more on an appropriate award (Hypothesis 5).

We expect jurors will display coherence between their gist classification and their verbatim award. In other words, we expect they will be able to classify their awards in line with the amount allocated (*verbatim-gist coherence*; Hypothesis 6). We expect that guidance will increase verbatim-gist coherence, with jurors being more accurate in their ordinal classification of damage awards (Hypothesis 7). Finally, we expect that numerical requests when unaccompanied by verbal guidance will not have these beneficial effects, but will simply have a biasing anchoring effect (Experiment 3 only; Hypothesis 8).

Subjective experience (ratings). Our third set of hypotheses relates to whether guidance improves the juror's metacognitive experience. We expect that jurors: will find guidance helpful (Hypothesis 9a); will rate the decision as less difficult (Hypothesis 9b); and will express more confidence in their awards (Hypothesis 9c). We expect that these effects will be strongest when verbal-plus-numerical guidance is employed, since these assist with both the judgment of ordinal gist and the assignment of a numerical value to that gist. Verbal guidance may generate less improvement in participants' metacognitive experiences, since it only assists with the more common task of assessing ordinal gist (see Hypothesis 1).

General Method

We conducted three experiments to test these hypotheses. The Institutional Review Board for Human Participants at a northeastern university reviewed and approved the procedure and materials for all three experiments as qualifying for an exemption from full IRB review. The

three experiments had many similarities, so the procedure is described in Study 1, and any differences with subsequent studies are highlighted. For each experiment, we report how we determined our sample size and describe all data exclusions, all manipulations, and all measures that we analyzed in this manuscript. Research materials and data are available from the first author for the purposes of reproducing the results or replicating the procedure. Upon publication, these data will be submitted to a publicly available data archive.

Experiment 1

Method

Participants. In Experiment 1, a total of 203 undergraduate students (77% female; M_{Age} = 19.90, *SD* = 1.94; range: 18-35) took part in the experiment. The sample size was determined through pretesting. Six participants did not find liability and were not included the analyses, resulting in a total of 197 participants. Additionally, three participants provided awards that were extreme outliers (i.e., over 3 *SD* above the mean of their condition). All analyses were conducted in three ways: including these participants as is, excluding the participants completely, and reducing the awards of these participants to an award 3 *SD* above the mean. There were no differences in terms of what was significant, so the data presented include the outliers' data reduced to 3 *SD* above the mean. Participants received course credit in undergraduate courses. Fifty-nine percent of participants identified themselves as White, 25% as Asian, 6% as African American/Black, 6% as mixed/other ethnicity, and 1.5% as Native American/Pacific Islander. Furthermore, 9% of the participants identified themselves as Hispanic/Latinx. Prior research with this population ([redacted]) showed that the majority fulfilled key requirements for jury service in the state (i.e., at least 18 years of age, residence in the area, U.S. citizenship, and facility with English).

Materials

Case summary. This study involved a summary of a hypothetical civil lawsuit involving a sports injury. In the case summary, a high school athlete experienced a concussion during a sports game and left the field. The athlete later returned to the field without being examined and suffered an additional head injury. The athlete sued the coach and school district for his injuries. The details of the case varied based on condition; in addition, one case involved a football injury while the other involved a soccer injury.

The amount of pain and suffering the plaintiff experienced varied based on condition (low or high severity). In the low severity condition, the athlete experienced mild concussion symptoms (headaches, vision problems, difficulty paying attention, mood swings, and discomfort around other people). He was described as unable to attend school for a short period; his grades slipped but then recovered. He experienced some mild problems working and socializing. In the high severity condition, the athlete had severe concussion symptoms. He was unable to attend school for a longer period of time, causing his grades to drop dramatically. He had extreme difficulty working and socializing.

Attorney guidance. Following the case summary, participants read a closing argument by the plaintiff's attorney that varied by condition (no-guidance control, verbal-only guidance, or verbal-plus-numerical guidance; see the Appendix for a complete transcript of guidance in each condition). In the no-guidance control condition, the plaintiff's lawyer told jurors no one could tell them how to decide on an award; instead, they should use their own common sense and best judgment.

In the guidance conditions, the plaintiff's lawyer used language recommended by Ball (2011) to guide the jury in damage award decisions. Specifically, jurors were told to focus on

three factors: 1) how bad the injury was; 2) how long the injury lasted; and 3) how interfering the injury was. For each component (bad, long, and interfering) the attorney advised the jury on how to assess the degree of pain and suffering (low, medium, or high). In the verbal-only guidance condition, the attorney instructed the jury to make an award of a comparable size (very small amount, fairly large amount, or a great deal of money). In the verbal-plus-numeric guidance condition, participants were also given a suggested award number for each size award ($25,000, $250,000, and $1,000,000). These numerical values were chosen based on pretest data.

Case Perceptions. Participants first answered several questions about their award. After providing a damage award, they indicated their confidence in their award on a ten-point scale, and classified the size of their award as nil, low, medium, or high. Participants described how they reached their award and indicated what they believed was the range of acceptable awards in this case.

Participants were then asked questions about their perceptions of the case on seven-point scales. They were asked five questions about their perception of the injury (bad, long, interfering, severity of injury, and severity of pain and suffering), which were averaged together into a composite severity score that had high reliability (Experiment 1 $\alpha = .90$; Experiment 2 $\alpha = .94$; Experiment 3 $\alpha = .92$). Participants also answered questions about their perceptions of the helpfulness of the legal actors (defense attorney, plaintiff's attorney, judge) and the responsibility of the parties.

Design

Experiment 1 followed a 2 (injury severity: low v. high) x 2 (case: football v. soccer) x 3 (award guidance: no-guidance control v. verbal-only guidance v. verbal-plus-numerical guidance) mixed groups design. Participants each saw two cases; however, analyses revealed

there were significant order effects. Therefore, for the purposes of these analyses, only between-subjects analyses using the first case the participant decided are presented.

Procedure

Participants were randomly assigned to read the case summary that varied based on injury severity, case, and award guidance. Participants then read judicial instructions which reminded them that both the coach and the school district had been found liable. Participants were also told that the defendants had already paid medical expenses and other economic damages, so their role as jurors was to decide how much to award the plaintiff for pain and suffering.

After indicating whether they agreed with the liability determination, participants decided on a monetary damage award, answered questions about that award and their perceptions of the case, and then answered questions about demographics and individual differences.

Results

Jurors getting the gist (ratings). We expected that jurors would get the gist of the injury on their own, ordinally discriminating mild from severe (*Hypothesis 1*; also a manipulation check). To test this prediction, we conducted an ANOVA using injury severity, case, and guidance as the independent variables and the composite severity score as the dependent variable. As predicted, there was a main effect of severity such that participants rated the severe case as more severe (M = 5.14, 95% CI [4.93, 5.35], SE = .11) than the mild case (M = 3.89 [3.68, 4.10], SE = .11), $F(1, 185) = 68.92$, $MSe = 1.07$, $p < .001$, $\eta_p^2 = .27$).

We expected that attorney guidance would assist jurors in understanding the severity of the injury (i.e., getting the gist) by increasing the difference between severity ratings of the low and high severity cases (*Hypothesis 2*). In order to test this hypothesis, we examined the

interaction between severity and guidance on composite severity score. Contrary to predictions, guidance did not affect the difference between severity ratings for the mild and severe cases, F (2, 185) = 1.84, MSe = 1.07, p = .16, η_p^2 < .02 (*Hypothesis 2*).

There was, however, an unexpected three-way interaction between case severity, case, and guidance on ratings of severity, F(2, 185) = 3.87, MSe = 1.07, p = .02, η_p^2 < .04. Results suggest that participants who received no guidance saw the severe soccer case as less severe (M = 4.66 [4.11, 5.20], SE = .28) than the severe football case (M = 5.57 [5.13, 6.02], SE = .26; p = .01).

Jurors assigning a damage award. Next, we conducted a series of analyses to test whether jurors were able to translate their gist assessment of the case into a numerical damage award. We tested the specific hypotheses that guidance would increase award validity, reliability, and verbatim-gist coherence.

Award validity. We conducted an ANOVA using severity, case, and guidance as the independent variables and the natural log damage award as the dependent variable. We used the natural log transformation because the damage awards were not normally distributed, which violates the assumptions of statistical tests. Consistent with Hypothesis 3, jurors demonstrated high *award validity*. There was a main effect of severity on natural log damage awards, such that participants gave higher awards in the more severe case (M = 11.99 [11.59, 12.39], SE = .20) than the less severe case (M = 9.87 [9.47, 10.27], SE = .20), F(1,185) = 54.86, p < .001, η_p^2 = .23 (*Hypothesis 3;* see Figure 1 for Experiment 1 raw damage awards).

Contrary to predictions (*Hypothesis 4*), guidance did not increase award validity. There was a main effect of guidance on awards, such that awards were higher with verbal-plus-numerical guidance than verbal-only guidance (p < .01) or no-guidance control (p < .001), F(2,

185) = 7.60, *MSe* = 3.89, $p < .001$, $\eta_p^2 = .08$ (see Figure 1). However, there was not a significant interaction between severity and guidance, $F(2, 185) = 1.10$, *MSe* = 3.89, $p = .33$, $\eta_p^2 = .02$.

Award reliability. We expected that guidance would improve reliability by resulting in greater convergence on an award (*Hypothesis 5*). In order to test this hypothesis, we conducted an ANOVA using severity, case, and guidance as the independent variables, and award variability as the dependent variable. To measure variability, we calculated the absolute distance between each award and the mean award in the relevant experimental condition (the AMD). We controlled for award amount by including it as a covariate, since higher awards tend to be more variable. Consistent with this effect of magnitude on variability, the covariate was significant, $F(1,184) = 781.78$, *MSe* = 9.74e^{10}, $p < .001$, $\eta_p^2 = .81$; and, controlling for this effect, awards were significantly more variable for the severe case (*M* = 665,959.09 [600,639, 731,278.44], *SE* = 33,107.62) than for the mild case (*M* = 328,213.98 [263,051.78, 393,376.17], *SE* = 33,027.96), $F(1,184) = 49.26$, *MSe* = 9.74e^{10}, $p < .001$, $\eta_p^2 = .21$.

Partially consistent with expectations (*Hypothesis 5*), there was a main effect of guidance, $F(2,184) = 6.83$, *MSe* = 9.74e^{10}, $p = .001$, $\eta_p^2 = .07$. There was significantly less variability when participants received verbal-plus-numerical guidance (*M* = 381,358.05 [304,464.77, 458,251.33], *SE* = 38,973.95) than with no-guidance control (*M* = 576,100.80 [497,914.11, 654,287.48], *SE* = 39,629.52, $p = .001$) or verbal-only guidance (*M* = 533,800.75 [456,199.175, 611,402.33], *SE* = 393.332.95, $p < .01$); there was no difference between no-guidance and verbal-only guidance ($p > .05$). However, the main effect of guidance was dependent upon severity, such that it only occurred for the severe case but not the mild, $F(2,184) = 5.46$, *MSe* = 5.32e^{10}, $p < .01$, $\eta_p^2 = .06$.

Verbatim-gist coherence. In order to test whether participants' ordinal award classification was consistent with their award (i.e., *verbatim-gist coherence; Hypothesis 6*), we conducted an ANOVA including the manipulations and award classification as the independent variables and the natural log damage award as the dependent variable (i.e., we added award classification to the ANOVA testing Hypothesis 3). Most participants classified their awards as medium (nil: 1.5%; low: 24.9%; medium: 55.3%; high: 18.3%). Because of the small number of participants who classified their awards as "nil," the nil responses were excluded in the gist-verbatim analyses. As expected, the size of dollar awards differed in concert with classifications of injury as low, medium, or high (*Hypothesis 6*), $F(2, 159) = 26.06$, $MSe = 3.16$, $p < .001$, $\eta_p^2 = .09$. Awards classified as "low" were significantly lower amounts ($M = 9.79$, 95% CI [9.12, 10.46], $SE = .34$) than awards classified as "medium" ($M = 11.11$ [10.76, 11.47], $SE = .18$, $p < .001$) or "high" ($M = 11.93$ [11.10, 12.76], $SE = .42$, $p < .001$); however, awards classified as "medium" were only marginally lower than those classified as high ($p = .08$).

Contrary to predictions, guidance did not increase verbatim-gist coherence (*Hypothesis 7*), as the two-way interaction between guidance and award classification was not significant, $F(4, 159) = .77$, $MSe = 3.16$, $p = .55$, $\eta_p^2 = .02$.

Subjective experience. In order to examine whether guidance resulted in a better metacognitive experience, we conducted a series of ANOVAs using severity, guidance, and case as the independent variables and the metacognitive factors (helpfulness, difficulty and confidence) as the dependent variables. As predicted, attorney guidance, particularly verbal-plus-numerical guidance, resulted in a better metacognitive experience (see Table 1 for means, confidence intervals, and standard errors).

Helpfulness. Jurors rated the plaintiff attorney as more helpful with both forms of

guidance than in the no-guidance control (*ps* < .001; *Hypothesis 9a*); there were no differences between types of guidance (p = .21), $F(2, 185) = 28.33$, $MSe = 1.77$, $p < .001$, $\eta_p^2 = .23$. Similarly, jurors who received guidance also found the judges' instructions to be significantly more helpful than jurors who received no guidance (*p*s < .01); there were no differences between types of guidance (p = .47), $F(2, 185) = 6.35$, $Mse = 1.61$, p = .002, $\eta_p^2 = .06$.

For ratings of helpfulness of both the plaintiff's attorney and the judge, there were unexpected interactions with sport. As these results did not bear on our hypotheses, they are presented in Supplemental Materials.

Difficulty. Overall, participants rated the task of damage award decision making as somewhat difficult, an average of 4.96 on a 7-point scale. As expected, guidance did impact ratings of difficulty (*Hypothesis 9b*), $F(2, 185) = 3.97$, $MSe = 1.87$, p = .02, $\eta_p^2 = .04$. Verbal-plus-numerical guidance resulted in the decision being significantly less difficult than with no guidance (p = .007) and marginally less difficult than with verbal-only guidance (p = .053); there were no differences in ratings of difficulty between verbal-only guidance and no guidance (p = .43).

Confidence. Jurors were also more confident in their decisions with verbal-plus-numerical guidance than with verbal-only guidance or no-guidance control (*p*s < .01; *Hypothesis 9c*), which were not different from each other (p = .42), $F(2, 185) = 3.86$, $MSe = 3.47$, $p < .001$, $\eta_p^2 = .12$. Unexpected interactions with the case, injury severity, and guidance conditions on confidence ratings did not bear on our hypotheses and are reported in Supplemental Materials.

Discussion

The results of Experiment 1 showed jurors distinguished the gist of the cases as lower versus higher severity, and that participants' gist ratings of awards (as low, medium, or high)

generally tracked dollar award amounts. However, guidance did not differentially affect these judgments. Verbal-plus-numerical guidance had a positive impact in reducing the variability of damage award amounts in the severe injury case, although it did not affect variability in the mild injury case. Thus, predictions about the influence of guidance on award variability were partially supported, but predictions about the effects of guidance on gist-verbatim coherence and validity were not observed.

Guidance had salutary effects on jurors' subjective experience. Jurors rated the plaintiff's attorney as more helpful when they received either type of guidance. Interestingly, there appears to be a halo effect, with jurors carrying this rating of helpfulness over to the judge as well (even though the judge's instructions did not vary). Although both types of guidance were seen as helpful, only verbal-plus-numerical guidance also made the decision significantly easier and increased confidence. Thus, overall, the verbal-plus-numerical guidance improved jurors' subjective experience more than verbal guidance alone.

Thus, the results of Experiment 1 indicate that, consistent with the Hans-Reyna Model, mock jurors understood the gist of the injury but struggled translating it into a numeric damage award. Attorney guidance was subjectively helpful to jurors, but objectively did not appear to improve their decision making. Moreover, there were unexpected, not theoretically-based differences by sport in decisions. Therefore, we were interested in examining whether these patterns occurred for a community sample without the confound of sport.

Experiment 2

Experiment 2 replicated the methodology of Experiment 1 almost exactly with two primary changes designed to address the limitations of Experiment 1. In Experiment 2, we kept two cases in order to increase generalizability, but both cases involved football.

Second, Experiment 1 used a student sample. Although a recent meta-analysis showed few significant differences between student samples and other samples in jury decision-making, typical student samples differ from the broader population in age, education levels, racial and ethnic group representation, socioeconomic status, intelligence level, and personality characteristics and attitudes (Bornstein et al., 2017). More importantly for our purposes, community members awarded marginally higher damage award amounts compared to student samples in civil cases (Bornstein et al., 2017, p. 22, Table 9). Therefore, in Experiment 2, we recruited a general population sample to test our predictions with a community sample.

Method

Participants. Participants in Experiment 2 were 503 Amazon Mechanical Turk (MTurk) workers from the United States recruited using TurkPrime (Litman, et al., 2017). The sample size was determined through the results of Experiment 1. Participants were paid $5 for their participation. A total of 27 participants were excluded (21 did not find liability; 6 did not answer the attention check questions appropriately), leaving a total of 476 participants. Research on MTurk samples suggests that participants are similar to participants in other online platforms and in-person studies (Irvine, et al., 2018). Participants' ages ranged from 20 to 74 with an average age of 35.76 (*SD* = 10.82) years. Sixty percent were male; 82% reported their race as White; 9% African American/Black; 6% Asian; and 3% mixed/other, with 7.6% reporting Hispanic ethnicity. Most (93.3%) stated they were eligible to serve as jurors; 15% had done so.

Design. Experiment 2 followed a 2 (case: Cooper v. Gray) x 2 (injury severity: low vs. high) x 3 (award guidance: no-guidance control vs. verbal-only guidance vs. verbal-plus-numerical guidance) between-groups design.

Materials and Procedure. The materials and procedure were the same as in Experiment

1 with slight modifications. In Experiment 2, both cases (Cooper and Gray) involved a football injury and two cases were used to ensure results were generalizable. Additionally, Experiment 2 was entirely between-subjects, so participants read only one case.

Results

Jurors getting the gist (ratings). As in Experiment 1, we tested whether jurors got the gist through an ANOVA using the severity manipulation as the independent variables and the composite severity score as the dependent variable. As expected, jurors appeared to get the gist generally. Participants in the high severity case rated the case as more severe (M = 5.51, 95% CI [5.36, 5.65], SE = .07) than participants in the low severity condition (M = 3.56 [3.41, 3.71], SE = .07), $F(1, 464) = 354.98$, $MSe = 1.26$, $p < .001$, $\eta_p^2 = .43$ (*Hypothesis 1;* also manipulation check). However, guidance did not improve differentiating the gist of the injury (its relative severity), as the interaction between guidance and severity was not significant, $F(2, 464) = .57$, $MSe = 1.26$, $p = .56$, $\eta_p^2 < .01$ (*Hypothesis 2*). There were no other significant main effects or interaction effects.

Jurors assigning a damage award (dollars). Next, we again examined how jurors translated their gist assessment into a numerical damage awards by testing whether guidance increased award validity, reliability, and verbatim-gist coherence.

Award validity. As in Experiment 1, we tested award validity through an ANOVA using the manipulations as the independent variable and the natural log damage award as the dependent variable. Consistent with Hypothesis 3, jurors again demonstrated high *award validity*. Participants gave higher awards in the severe case (M_{low} = 10.23, 95% CI [9.93, 10.52, SE = .15; M_{high} = 13.06 [12.76, 13.34], SE = .15), $F(1, 464) = 179.29$, $p < .001$, $\eta_p^2 = .28$ (*Hypothesis 3*; see Figure 2 for Experiment 2 raw damage awards).

Contrary to expectations (*Hypothesis 4*), guidance did not increase award validity, as there was no significant interaction between guidance and severity, $F(2, 464) = 2.80$, $MSe = 5.26$, $p = .59$, $\eta_p^2 < .01$. Unlike Experiment 1, the main effect of guidance on awards was not significant, $F(2, 464) = 1.54$, $MSe = 5.26$, $p = .22$, $\eta_p^2 = .01$.

Award reliability. As in Experiment 1, in order to test award reliability, we conducted an ANOVA using the manipulations as the independent variables, award amount as a covariate, and award variability (AMD) as the dependent variable. As in Experiment 1, the covariate was significant indicating that award size impacts variability, $F(1, 463) = 15129.35$, $MSe = 1.38e^{11}$, $p < .001$, $\eta_p^2 = .97$. Additionally, there was a main effect of severity; however in Experiment 2, awards were more variable in the less severe case (M = 1,068,014.37 [1,018,876.58, 1,117,152.16], SE = 25,005.22) than in the more severe case (M = 895,503.01 [847,012.99, 943,993.02], SE = 24,675.59), $F(1, 463) = 22.93$, $MSe = 1.38e^{11}$, $p < .001$, $\eta_p^2 = .05$.

Partially consistent with Hypothesis 5, there was a main effect of guidance on variability, $F(1, 463) = 6.62$, $MSe = 1.38e^{11}$, $p = .001$, $\eta_p^2 = .03$. As in Experiment 1, there was significantly less variability when participants received verbal-plus-numerical guidance (M = 894,712.26 [837,674.42, 951,750.10], SE = 29,025.40) than with no-guidance control (M = 1,033,700.18 [975,647.08, 1,091,753.28], SE = 29,542.05, $p = .001$) or verbal-only guidance (M = 1,016,863.62 [956,695.56, 1,077,031.68], SE = 30,618.31, $p < .01$); there was no difference between no-guidance and verbal-only guidance ($p > .05$). However, the main effect of guidance depended on severity, such that the main effect of guidance was observed with the severe case, but not the mild case, $F(2, 463) = 3.80$, $MSe = 1.38e^{11}$, $p = .02$, $\eta_p^2 = .02$. There were no other significant main effects or interaction effects.

***Verbatim-gist coherence*.** In order to test whether ordinal awards were consistent with the classification, we again conducted an ANOVA using the manipulations and classification as the independent variables and the natural log damage award as the dependent variable. As expected, jurors' classifications appropriately differentiated the size of their damage awards (*Hypothesis 6*), $F(2, 430) = 50.07$, $MSe = 2.53$, $p < .001$, $\eta_p^2 = .19$. Most participants (51.4%) classified their awards as "medium" (nil = 1.8%; low = 23.8%; high = 23.0%). As in the previous experiment, the small number of nil responses were excluded in the gist-verbatim analyses. Awards classified as high were significantly higher ($M = 13.91$ [13.37, 14.46], $SE = .28$) than awards classified as medium ($M = 12.21$ [11.99, 12.43], $SE = .11$, $p < .001$), which were both significantly higher than awards classified as low ($M = 10.23$ [9.79, 10.68], $SE = .23$, ps $< .001$).

Contrary to predictions, guidance did not increase verbatim-gist coherence (*Hypothesis 7*) as the interaction between guidance and award classification was not significant, $(2, 430) = 1.35$, $MSe = 2.53$, $p > .05$, $\eta_p^2 = .02$. There were no other two-way, three-way or four-way interactions between award classification and the manipulations (although there was still a main effect of severity as described above in the results for Hypothesis 3).

Subjective experience (ratings). Again, participants found it somewhat difficult to make awards, with the average rating of 4.29 out of 7. In order to test metacognitive experience, we conducted the same series of ANOVAs using the manipulations as the independent variables and ratings of helpfulness, difficulty, and confidence as the dependent variables. As predicted, attorney guidance resulted in a better metacognitive experience (see Table 2 for means, standard errors, and confidence intervals).

Helpfulness. As expected, participants rated the plaintiff attorney's guidance as more helpful with verbal-plus-numerical guidance, followed by verbal-only guidance ($p = .001$),

followed by no-guidance control (*p*s < .001), $F(2, 464) = 53.01$, $MSe = 2.74$, $p < .001$, $\eta_p^2 = .19$ *(Hypothesis 9a*). The judges' instructions were also rated as more helpful with either verbal-plus-numerical guidance or verbal-only guidance than with no-guidance control (*p*s < .01), but verbal-plus-numerical and verbal-only guidance did not differ in terms of ratings of judicial helpfulness ($p = .20$), $F(2, 464) = 9.13$, $MSe = 2.70$, $p < .001$, $\eta_p^2 = .04$. There was no difference for ratings of defense attorney helpfulness, $F(2, 464) = .76$, $MSe = 3.53$, $p = .47$, $\eta_p^2 < .01$. There were no other significant main effects or interaction effects on ratings of helpfulness.

Difficulty. As predicted, participants also found it less difficult to compensate the plaintiff with verbal-plus-numerical guidance than verbal-only guidance or no-guidance control (*ps* < .01), $F(2, 441) = 7.85$, $MSe = 3.01$, $p < .001$, $\eta_p^2 = .03$ (*Hypothesis 9b*). There was no difference between verbal-only and no-guidance ($p = .72$). Moreover, manipulated case severity influenced participants' ratings of difficulty making an award, $F(1, 429) = 9.99$, $MSe = 2.99$, $p < .01$, $\eta_p^2 = .02$. Participants found the decision more difficult in the severe case ($M = 4.54$, 95% CI [4.32, 4.77], $SE = .11$) than in the mild case ($M = 4.08$ [3.86, 4.31], $SE = .12$). There were no other significant main or interaction effects on ratings of helpfulness.

Confidence. As expected, participants were more confident in their awards with verbal-plus-numerical guidance than verbal-only guidance or no-guidance control (*p*s < .01), $F(2, 464) = 15.76$, $MSe = 4.06$, $p < .001$, $\eta_p^2 = .06$ (*Hypothesis 9c*). There was no difference between verbal-only guidance and no guidance ($p = .10$). There were also no other significant main effects or interaction effects.

Discussion

As in Experiment 1, Experiment 2 supported predictions about jurors' gist judgments, the impact of guidance on award variability, and metacognitive effects. That is, jurors effectively

differentiated the gist of the cases as lower versus higher severity, and dollar awards corresponded with gist ratings of awards as low, medium, or high. Experiment 2 provided some support for Hypothesis 4 by showing that verbal-plus-numerical guidance significantly increased the reliability of awards in the severe case condition, although effects on other beneficial outcomes were not observed, as in Experiment 1.

Predictions about the positive effects of guidance on jurors' subjective experience were generally supported, replicating the results of Experiment 1. In accordance with Hypothesis 9a, jurors thought the plaintiff's attorney and the judge were more helpful when the plaintiff's attorney offered guidance. Thus, the effect once again transferred to the judge although the judge's instructions did not change, supporting the idea that perceptions of the legal system improve generally when advice is viewed as helpful. This halo effect did not extend to the defense attorney, suggesting that jurors distinguished the plaintiff's attorney and defense attorney. Experiment 2 further emphasized that verbal-plus-numerical guidance improved the subjective experience, resulting in higher ratings of plaintiff's attorney's helpfulness and confidence in award and significantly lower ratings of difficulty compared to verbal-only guidance or no guidance.

Comparing the findings of Experiments 1 and 2, we noticed a striking difference between the average dollar awards in the two experiments' control conditions; see Figures 1 and 2. In Experiment 1, with a student sample, the average dollar awards in the control condition were $32,810 (low severity) and $741,531 (high severity). Experiment 2, with a broader MTurk sample, resulted in substantially higher average dollar damage awards in the control conditions ($526,229 for low severity and $2,143,369 for high severity). This is in line with Bornstein et al.'s (2017) meta-analysis on civil juror decision making showing marginally higher damage

awards in general population samples than in student samples. In addition, the verbal-plus-numeric guidance resulted in lower damage awards but it provided an anchor that was significantly lower than what participants awarded when they were not provided with a numeric anchor, a potential mechanism investigated in Experiment 3.

Experiment 3

Experiment 3 was similar to Experiment 2 but with the following changes. As there were no real differences based on case in Experiment 2, we reduced sample materials to only one case in Experiment 3. We increased the anchors to be more consistent with awards given in Experiment 2; we increased the amounts the plaintiff's lawyer suggested to $250,000, $1 million, and $5 million. And we returned to the student sample.

The major focus of Experiment 3 was identifying why verbal guidance was not having the same effects as verbal-plus numerical guidance. We reviewed our guidance materials and decided to make two substantial changes. First, we tailored the verbal guidance (in both verbal-only and verbal-plus-numerical) to be more specific to the case (i.e., the impact of concussive injuries; see Appendix for wording changes). The goal was to improve the effectiveness of verbal guidance.

Second, we wanted to examine whether verbal-plus-numerical guidance was an improvement on verbal-only guidance as it assisted jurors in the process of assigning a number to the gist, or if the differences stemmed only from providing a numerical anchor. Therefore, we wanted to test whether providing a number without guidance would result in a biasing anchor effect (e.g., only decreasing variability) without the other beneficial attributes (e.g., greater differentiation of severity, increased verbatim-gist coherence; *Hypothesis 8*). To test Hypothesis 8, we included an

additional condition in which participants were given the high award recommendation ($5 million) with no additional guidance (i.e., an *ad damnum* request).

Method

Participants. Participants in Experiment 3 included 411 undergraduate students. The sample size was determined through the results of Experiments 1 and 2. Twenty participants were excluded (14 did not agree with liability, 6 did not answer the key questions), leaving a total of 391 participants. An additional 8 participants had awards that exceeded 3 *SD* above the condition mean, so their awards were trimmed to 3 *SD* above the mean. The student sample was 72% female and 27% male, with an average age of 19.99 (*SD* = 1.48, range: 18–34). As for racial background, 56% reported they were White, 10% said they were African American/Black, 23% were Asian, and 12% indicated mixed or other races. Fourteen percent reported Hispanic ethnicity. Eighty-one percent indicated they would qualify for jury service. Students received course credit for participation. The study was conducted in two successive years; an analysis of the data from the first and second year showed no main or interaction effects in any of the analyses, so the data were combined to test the hypotheses.

Design. Experiment 3 followed a 2 (injury severity: low, high) x 4 (award guidance: no-guidance control, verbal-only guidance, verbal-plus-numerical guidance, numerical-only guidance (*ad damnum*)) between-groups design.

Materials and Procedure. The materials and procedure in Experiment 3 were similar to Experiments 1 and 2. Experiment 3 only involved the Gray football case. The recommended awards in Experiment 3 for the verbal-plus-numerical guidance were increased (low = $250,000; medium = $1,000,000; high = $5,000,000). Additionally, Experiment 3 included a fourth instruction condition (numerical-only guidance) that took the form of an *ad damnum* request in

which the plaintiff's attorney requested an award of 5 million dollars with no additional scaling guidance.

Results

Jurors getting the gist (ratings). As in previous experiments, we tested whether jurors got the gist through an ANOVA using the manipulations (severity and guidance) as the independent variables and the composite severity score as the dependent variable. As predicted, jurors got the general gist of the severity of injury. There was a main effect of case severity such that participants rated the severe case as higher in severity (M = 5.68, 95% CI [5.54, 5.82], SE = .07) than the mild case (M = 4.04 [3.90, 4.17], SE = .07), $F(1, 383) = 279.62$, $p < .001$, MSe = .94, η_p^2 = .42 *(Hypothesis 1*). Guidance did not help participants further differentiate the gist; there was no interaction between severity and guidance, $F(3, 383) = 1.18$, MSe = .94, p = .32, $\eta_p^2 <$ .01 (*Hypothesis 2*).

There was, however, an unexpected main effect of guidance, $F(3, 383) = 3.38$, MSe = .94, p = .01, η_p^2 = .03. Verbal guidance produced lower severity ratings. That is, participants rated the case as significantly more severe when they received no guidance (M = 5.10, 95% CI [4.90, 5.29], SE = .10) than when they received verbal guidance (M = 4.73, 95% CI [4.54, 4.92], SE = .10, $p < .01$) or verbal-plus-numerical guidance (M = 4.69, 95% CI [4.50, 4.88], SE = .10, $p <$ .01); however, there was no difference between no guidance and numerical guidance (M = 4.91, 95% CI [4.72, 5.10], SE = .10, p = .18). The conditions in which guidance was offered did not differ significantly from one another (ps > .05).

Jurors allocating a damage award (dollars). Again, we tested jurors' ability to translate their gist into a numerical award, examining award validity, reliability, and verbatim-gist

coherence. As in Experiments 1 and 2, overall participants provided awards consistent with the gist.

Award validity. As in previous experiments, we tested award validity through an ANOVA using the manipulations as the independent variables and the natural log damage award as the dependent variable. Consistent with expectations, participants gave higher awards in the severe case (M = 13.06 [12.76, 13.34], SE = .16) than in the mild case (M = 10.18 [9.88, 10.48], SE = .15), $F(1, 390) = 174.78$, $p < .001$, $\eta_p^2 = .31$ (*Hypothesis 3;* see Figure 3 for Experiment 3 raw damage awards).

However, guidance did not increase award validity; there was no interaction between severity and guidance, $F(3, 383) = .89$, $MSe = 4.62$, $p = .45$, $\eta_p^2 < .01$; *Hypothesis 4*). There was a significant main effect of guidance on awards, $F(3, 383) = 49.79$, $p < .001$, $\eta_p^2 = .28$. Awards were significantly higher when there was a recommended award amount (i.e., an anchor; $M_{\text{NumericalOnly}}$ = 12.91 [12.48, 13.34], SE = .22; $M_{\text{Verbal+Numerical}}$ = 12.99 [12.57, 13.42], SE = .22) than when there was no recommended award amount ($M_{\text{VerbalOnly}}$ = 10.39 [9.96, 10.81], SE = .22; $M_{\text{NoGuidance}}$ = 10.20 [9.77, 10.63], SE = .22; ps < .001).

Award reliability. As in previous experiments, we tested award reliability using an ANOVA with the manipulations as the independent variables, award amount as a covariate, and award variability (AMD) as the dependent variable. Once again, the covariate was significant, indicating that award size affects variability, $F(1, 382) = 1228.63$, $MSe = 5.21e^{11}$, $p < .001$, $\eta_p^2 = .76$. Additionally, there was a main effect of severity such that awards were more variable in the more severe case (M = 1,496,184.65 [1,388,944.80, 1,603,424.50], SE = 54541.86) than the less severe case (M = 1,013,037.05 [906,846.36, 1,119,227.75], SE = 54,008.26), $F(1, 382) = 36.31$, $MSe = 5.21e^{11}$, $p < .001$, $\eta_p^2 = .09$.

Partially consistent with Hypothesis 5, there was a main effect of guidance on variability, $F(1, 382) = 25.25$, $MSe = 5.21e^{11}$, $p < .001$, $\eta_p^2 = .17$. Awards were significantly less variable when participants received numerical-only guidance *(ad damnum*; M = 783,462.16 [37,422.20, 929,502.13], *SE* = 74,275.48), followed by verbal-plus-numeric guidance (*M* = 1,095,042.66 [948,020.09, 1,242,065.24], *SE* = 74,775.23; *p* = .003) followed by no form of numeric guidance ($M_{\text{Verbal Only}}$ = 1,536,822.35 [1,391,242.59, 1,682,402.11], *SE* = 74,041.42; $M_{\text{No Guidance}}$ = 1,603,116.23 [1,456,198.75, 1,750,033.70], *SE* = 74,721.78; *p*s < .001). However, there was no difference between verbal-only guidance and no guidance (*p* = .52).

The effect of guidance, however, was dependent upon severity, $F(1, 382) = 2.407$, $MSe = 5.21e^{11}$, $p < .001$, $\eta_p^2 = .16$. In examining the pairwise comparisons, the pattern was observed for the severe case. In the severe case, awards were significantly less variable when participants received numerical-only guidance *(M* = 557,653.21 [343,761.79, 771,544.63], *SE* = 108,784.52), followed by verbal-plus-numeric guidance (*M* = 1,227,440.32 [1,010,047.43, 1,444,833.22], *SE* = 110,565.36; *p* < .001) followed by no form of numeric guidance ($M_{\text{Verbal Only}}$ = 2,041,777.32 [1,828,412.04, 2,235,142.60], *SE* = 103,430.96; $M_{\text{No Guidance}}$ = 2,167,867.74 [1,953,709.82, 2,235,142.60], *SE* = 108,920.06; *p*s < .001); there was no difference between verbal-only guidance and no guidance (*p* = .37). In the mild case, there was no difference in variability across any of the guidance conditions (*p*s > .59).

Verbatim-gist coherence. In order to test whether ordinal awards were consistent with the classification, we again added award classification to test Hypothesis 3 using an ANOVA. As expected, jurors accurately assessed the size of their damage awards (*Hypothesis 6*), $F(2, 430) = 50.07$, $MSe = 2.53$, $p < .001$, $\eta_p^2 = .19$. Again, most participants (55.2%) classified their award as "medium" (55.2%) (nil=1.5%; low = 25.1%; high = 18.2%). As before, the small number of

nil responses were excluded in the gist-verbatim analyses. Awards classified as low were significantly lower (M = 9.68, 95% CI [9.20, 10.16], SE = .25) than awards classified as medium (M = 11.97 [11.72, 12.21], SE = .12, $p < .001$), which were significantly lower than awards classified as "high" (M = 12.96 [12.36, 13.55], SE = .30, ps < .01), indicating differentiation.

Contrary to expectations, guidance did not increase verbatim-gist coherence (*Hypothesis 7*) as the interaction between guidance and award classification was not significant, $F(6, 360)$ = .40, MSe = 3.08, p = .88, η_p^2 = .01. There were no other two-way, three-way or four-way interactions between award classification and the manipulations (although there was still a main effect of severity as described in Hypothesis 3).

Subjective experience (ratings). Participants found it moderately difficult to reach an award, averaging 5.05 on a 7 point scale. As predicted, guidance resulted in a better metacognitive experience (see Table 3 for means, confidence intervals, and standard errors).

Helpfulness. Consistent with our prediction (*Hypothesis 9a*), jurors rated the plaintiff attorney's guidance as most helpful in the verbal-plus-numerical guidance condition, followed by verbal-only guidance (p = .02), then numerical-only guidance (ps < .01), and then no guidance (ps < .01), $F(3, 383)$ = 51.50, MSe = 2.20, $p < .001$, η_p^2 = .16. There was also a main effect of severity such that the plaintiff's attorney's guidance was significantly more helpful in the more severe case (M = 4.53 [4.14, 4.74], SE = .11) than in the less severe case (M = 4.05 [3.85, 4.26], SE = .11), $F(1, 383)$ = 9.84, MSe = 2.20, p = .002, η_p^2 = .03.

As in the previous experiments, participants also found the judicial instructions to be more helpful with either type of verbal guidance than without verbal guidance (ps < .03), $F(3, 383)$ = 7.30, MSe = 2.13, $p < .001$, η_p^2 = .05. Verbal-plus-numerical guidance was rated as more

helpful than verbal-only guidance (p = .03); there was no difference between numerical-only guidance and no-guidance (p = .41).

Again, there was no effect of plaintiff's attorney's guidance on ratings of defense attorney helpfulness, $F(3, 383) = .80$, $MSe = 2.26$, $p = .50$, $\eta_p^2 = .01$. There were also no other main effects or interactions on any ratings of helpfulness.

Difficulty. As predicted, participants found it easier to reach an award with guidance, $F(3, 379) = 10.19$, $MSe = 1.91$, $p < .001$, $\eta_p^2 = .08$ (*Hypothesis 9b*). Participants rated it significantly less difficult to select an exact amount to compensate the victim with verbal-plus-numerical guidance than verbal-only guidance, no-guidance control, or numerical-only guidance *(ps* < .01*);* however, there were no differences among verbal-only guidance, no-guidance control, and numerical-only guidance (*p*s > .30).

Confidence. Finally, guidance also resulted in participants being more confident in their awards, $F(3, 383) = 11.99$, $MSe = 4.15$, $p < .01$, $\eta_p^2 = .09$ (*Hypothesis 9c*). Participants who received verbal-plus-numerical guidance were more confident than any other condition (i.e., verbal-only, numerical-only guidance, or no guidance conditions (*ps* = .001). Participants in the verbal-only condition were the least confident compared to the other conditions (*p*s < .02). Numerical-only guidance and no guidance fell in the middle in terms of ratings of confidence, but they were not significantly different from one another (p = .85).

Discussion

Experiment 3 provided further support for predictions about jurors' gist judgments and about the effects of guidance on award variability and jurors' metacognitive experiences. As in the two previous experiments, however, the plaintiff attorney's scaling guidance did not significantly influence gist-verbatim coherence or award validity.

Jurors appropriately rated the gist of the low and high severity injuries, and their awards were in line with their gist judgments, giving higher awards in the more severe case. Guidance did not interact with these judgments. Numerical guidance, whether it was presented by itself as an *ad damnum* or along with verbal guidance, reduced variability of awards in the severe injury case.

Unexpectedly, jurors who received verbal guidance, either with or without numerical information, downgraded the severity of the plaintiff's injury, compared to jurors who did not receive verbal guidance. That result hints that encouraging jurors to consider an injury's length, severity, and interference and offering exemplar cases that are low, medium, or high on these dimensions has the potential to provide context for fact finders. However, verbal guidance did not influence judgments of injury severity in the previous two experiments, so we should be cautious about making too much of this result.

Finally, verbal guidance had positive effects on jurors' subjective experience of damage award decision making. Experiment 3, which allowed a contrast between numerical versus verbal-plus-numerical guidance, showed the superiority of combining verbal and numerical guidance to jurors. Jurors who received both types of information found the plaintiff's guidance more helpful (especially in the serious injury case), found the decision less difficult, and were more confident in their judgments. They also found the (nonvarying) judge's instructions more helpful when the plaintiff offered verbal-plus-numerical guidance.

General Discussion

Civil jurors report that damage awards are among the most difficult and challenging aspects of their complex tasks. The challenge is especially acute for noneconomic damages such as pain and suffering, as opposed to observable physical injuries. The current research explored a

new approach to assisting jurors in the context of concussions, a type of injury with potentially severe and lifelong implications for psychological pain and suffering. We tested the impact of attorneys' recommendations giving verbal, numerical, or verbal-plus-numerical scaling guidance to jurors, drawing on psychological theory and the Hans-Reyna model of damage award judgments. We evaluated whether these forms of attorney guidance helped jurors to decide on appropriate awards. Previous research on guiding damage awards has primarily emphasized the effects of guidance on the variability of damage awards. We consider the variability of damage awards, but also expand the scope of inquiry to take a more expansive perspective on decision quality: One contribution of this paper is to offer a systematic methodology for assessing improvements in award decisions-- testing validity, reliability, gist-verbatim correspondence, and juror confidence and associated metacognitions – to evaluate the effects of different approaches to juror guidance.

Jurors Getting the Gist

We begin by emphasizing the strengths we document in our participants' damage award decision making. The results of these three experiments, with a total of 1,064 participants, support the ability of jurors to reach ordinally distinct damage award decisions that differentiate low from high severity, even for intangible pain and suffering stemming from concussions. In all three experiments, mock jurors rated the case with greater consequences, including more extended pain and suffering, as more severe than mock jurors who had seen a less serious case. These results support the prediction that jurors appreciate the relative magnitude of injuries—the ordinal gist of injuries, as contrasted with the precise amounts of dollar awards.

In addition, consistent with the idea that ordinal gist is apprehended, numerical awards increased as judgments of ordinal gist increased, with average awards increasing as ordinal gist

classifications increased from low to medium to high. Dollar damage awards, although highly variable, corresponded ordinally to participants' assessment of the severity of the plaintiff's injury.

We underscore these regularities in our participants' judgments because they directly contradict some longstanding criticisms of jury damage awards: that jurors' awards are erratic and untethered to the plaintiffs' underlying injuries, especially in noneconomic damages such as pain and suffering. Instead, we observe regularities in both their injury severity assessments and damage award decisions. Although they reported that the award decisions were challenging, our mock juror participants did a credible job distinguishing injuries of different severity and allocating awards in line with these differences. The results support the conclusion that jury judgments are not random with respect to the gist of injuries; they are ordered and systematic with respect to relative magnitude. This result should enhance society's confidence that the jury system is not arbitrary and capricious.

Does Attorney Scaling Guidance Help Jurors Assign a Damage Award?

Even though mock juror awards reflected these regularities, our experiments examined whether we could improve on them by varying attorney scaling guidance and observing whether and how guidance affected the characteristics of awards. Over the three experiments, we observed a salutary effect of verbal-plus-numerical guidance on the reliability of damage awards. We were interested specifically in whether verbal guidance (with or without numbers) would improve awards by reducing variability while increasing reliability and verbatim-gist coherence.

Numeric guidance resulted in reduced award variability in the severe cases, which is where the problem of award variability is typically more significant. Although this would appear to be desirable, there are several concerns. First, rather than assisting jurors in calibrating to an

appropriate number, it appears that this effect could be driven by anchoring. This concern is underlined by examining the difference between verbal and verbal-plus-numerical conditions. Encouraging jurors to think about translating the gist into an award alone without corresponding numbers was not effective. The concern is heightened by the findings in Experiment 3 where we provided an *ad damnum* request devoid of other guidance. In other words, verbal-plus-numeric guidance decreases award variability more similarly to numeric-only guidance than verbal-only guidance.

Second, it is important to consider whether a reduction in variability on its own, without increased validity or verbatim-gist correspondence, is preferable. Guidance did not influence the other two desirable characteristics: validity and gist-verbatim coherence. Offering verbal illustrations about how to distinguish injuries of different severity did not appreciably change the amount that jurors awarded or the ability to distinguish low from high injuries. Jurors appear to already have strong intuitions about the ordinal mapping of injuries, particularly since results suggest that participants were effectively grasping the gist of the plaintiff's pain and suffering as low or high across all guidance conditions, including the no-guidance control. In Experiment 3, verbal guidance affected the rated severity of the plaintiff's injury, suggesting that in some circumstances offering verbal descriptions of exemplars of cases of different magnitude might affect jurors' calibration of plaintiff injuries.

Consider in this light Bavli and Mozer's (2019) carefully-designed empirical project, which showed that providing comparable case information significantly reduced award variability, whether the comparable cases were biased or unbiased representations of community sentiment. Thus, one can reduce variability in damage awards, but it might be at the expense of award validity. The comparability of awards in the verbal-plus-numerical and numerical

conditions suggests that the reduction in variability may be due to an anchoring effect, rather than mock-jurors more effectively converging on a contextually appropriate award. Future research should develop the guidance tested here in order to more effectively influence validity and verbatim-gist correspondence.

Based on the current results, it appears that the recommended scaling guidance does not assist jurors in making better decisions. Apparently, jurors already understand the gist of the case, but they struggle in assigning a number to that gist. The verbal guidance alone did not improve this process. Either type of numeric guidance decreased variability, but did not influence reliability or verbatim-gist coherence, leading to concern that numeric guidance merely serves as a biasing anchor rather than a helpful guide for jurors. However, our manipulation of case severity was strong, and it remains a possibility that verbal guidance might aid ordinal mapping of cases that are less clearly differentiated in severity.

Does Attorney Guidance Affect Jurors' Subjective Experience?

Although verbal guidance did not improve awards, it did improve jurors' subjective experiences. In all three experiments, verbal-plus-numeric guidance resulted in jurors feeling like the attorney and the judge were more helpful, deciding the case was easier, and that they were more confident in their awards. Verbal-only guidance resulted in the attorney and judge appearing to be more helpful, but did not impact the other variables. Numeric-only guidance and no-guidance did not improve any of these subjective factors.

The fact that perceptions of the helpfulness of the judge, whose instructions did not change in any of the experiments, also significantly increased when the plaintiff's attorney offered guidance, suggests that overall perceptions are improved when guidance is given, extending from perceptions of attorneys to perceptions of judges. This result is broadly

consistent with some of the findings from deliberative democracy research, where satisfaction with the jury experience extends to support for the courts (Gastil et al., 2010).

Thus, it appears that jurors appreciate the verbal guidance, particularly when they are given numbers in addition to the verbal guidance. Consistent with the Hans-Reyna Model, jurors appear to be getting the gist but they are looking to the attorneys to help them in their struggle to assign numbers. Hence. the best guidance subjectively for the jurors was the guidance that walked them through the gist of the case and suggested a number. Walking them through the gist without providing a number was less helpful; and just providing a number without associating the gist was not helpful at all subjectively.

Guidance that is targeted at improving mapping between ordinal gist and dollar amounts may have a role to play in helping jurors feel more comfortable and improving their experience of the civil justice system. Improving the subjective experience of jurors also could enhance confidence in the jury system and in civil awards.

Limitations of the Current Experiments

There are some limitations to the mock juror experiments that we conducted to examine the effects of an attorney's scaling guidance. To allow clearer tests of the hypotheses, we informed respondents that liability had been established, that economic losses had already been covered, and that their job was only to decide on pain and suffering awards. Other researchers studying damage awards have similarly focused on a single damages element to achieve clear tests of hypotheses without confounding factors (McAuliff & Bornstein, 2010; Reyna, Hans, et al., 2015; Sunstein et al., 2002). However, jurors in real trials decide on liability, economic and noneconomic damages, and in some cases punitive damages. Robbennolt and Hans (2016; see also Wissler, et al., 2001) summarized research showing that these decisions are related to one

another; someone who strongly believes that the defendant is liable may also endorse a substantial award beyond what might be necessary for the compensation of the plaintiff. Focusing only on pain and suffering damages, as the current experiments have done, avoids confounding, but also might limit ecological validity.

In Experiment 3, we compared verbal-plus-numerical guidance with numerical-only guidance. We crafted the two conditions to reflect real-world application, so they differed in both the number and the amounts of dollar recommendations. The verbal-plus-numerical guidance included three different dollar recommendations (for example, in Experiment 3, the dollar amounts were $250,000, $1 million, and $5 million) whereas the numerical-only guidance condition included the single $5 million recommendation. This maintained verisimilitude in the way that scaling guidance and an *ad damnum* request would be made in the courtroom. But whether the provision of multiple dollar amounts (within the same guidance condition) contributed to differences between these two conditions is worthy of additional study. In particular, in the verbal-plus-numerical condition, the provision of three dollar amounts might have pulled participants in different directions, interfering with the ability of the guidance to positively affect gist-verbatim correspondence and award validity.

One way in which the scaling approach differs from some other ways of guiding jurors is that the verbal and numerical information is provided by a partisan advocate rather than a more neutral source. We did not study whether a judge's advice might have produced more effective guidance than a lawyer's. However, the pattern of significant effects on damage award amounts in these experiments, combined with the significant findings in previous *ad damnum* studies, suggest that this is not a concern.

Other limitations include the reliance on a relatively short written trial stimulus and the

absence of jury deliberation. We studied individual judgments rather than group decisions, even though jury awards are a group product. A number of studies have confirmed that individual juror preferences strongly predict jury outcomes (for a review, see Devine, 2012), supporting this approach. Even so, exactly how individual preferences about damage awards are combined in deliberation remains something of an open question. Some experiments have found that jury awards do not appear to be a simple averaging of individual jurors' award preferences. Diamond et al. (1998) found in a mock jury study that jury awards were higher than the mean and median award preferences of the individual jurors within these juries. In addition to these combinatorial processes that we did not study, deliberation could be a locus of influence for scaling guidance. Ball observed that the instructions would also "arm[ ] your favorable jurors with a concrete way to fight for the amount they want to allow" (Ball, 2011, p. 238). We hope in future research to address these and other limitations by testing theoretical predictions using realistic mock jury deliberations in civil cases.

Policy Implications

The variability in award judgments that we document means there is significant potential for an unusually high – or low – award. Courts have different rules for what type of guidance is permitted – some allow attorneys to make *ad damnum* requests, others do not (Campbell et al., 2017). The present studies can inform court policies about whether to permit attorneys to request a specific amount of money. We examined whether instructions might improve damage award decision making in desirable ways, specifically, in discriminating severity of injuries, decreasing variability in perceptions of injures and of awards, and increasing the coherence between qualitative perceptions of the gist of injuries and dollar awards. The results of these studies indicate that monetary requests can bias decision-making but are also appreciated by jurors.

The results indicate that allowing attorneys to include a monetary request serves as an anchor for jurors. It has the beneficial effect of reducing award variability; however, numeric guidance does not have the other important metrics of increased reliability and verbatim-gist coherence. Consequently, in these studies, the numeric request has served as a biasing anchor. In Experiments 1 and 3, the anchor was high compared to the Control condition average, bringing awards up; in Experiment 2, the anchor was low compared to the Control condition average, bringing awards down.

On the other hand, verbal-plus numerical guidance improved jurors' subjective experiences in the case. Jurors say identifying a damage award is among the most challenging parts of civil jury service. Our guidance that helped them translate the gist into a numeric award (i.e., verbal-plus-numeric guidance) to be more helpful, making the decision easier and jurors more confident in their decisions. Satisfaction with the jury experience is an important factor; it leads to greater confidence in the jury system and the courts (Gastil et al., 2010). Greater confidence and other metacognitive phenomena are not always reflective of greater accuracy (see, for example, Reyna et al., 2016; cf. Wixted & Wells, 2017). Nonetheless, trust in juries and trust in verdicts are likely linked to confidence in the process of awarding damages, which in turn supports civic behavior, such as reported intentions to comply with jury summonses (Bornstein et al., 2020). In addition, jury service has been linked to greater participation in other forms of civic engagement, including voting (Gastil et al., 2010; Hans et al., 2014). It is therefore important for policy makers to consider whether the increase in juror satisfaction is worth the trade-off of a biasing anchoring effect.

More research is necessary to identify a method that attorneys can use that would simultaneously result in more desirable awards (i.e., decreased variability, increased reliability

and verbatim-gist coherence) and an improvement in jurors' subjective experiences. One possible option is contextualizing the numbers given in the guidance more effectively, by referencing external benchmarks. The Reyna, Hans, et al. (2015) project found that providing generally meaningful anchor numbers (for example, median income) aided their study participants' damage award decision making (including improving both validity and verbatim-gist correspondence, see Helm et al., 2020). In addition, a relative perception manipulation, in which they described a specific dollar amount as, for example, "way below" or "way above" the median income, was also effective in guiding damage award decisions. We conclude that one productive avenue for future research is testing a variety of approaches to framing the magnitude of dollar amounts in damage award recommendations.

It is useful to reflect on these results and recommendations in light of previous efforts to guide jury damage award decision making, including on the one hand, the general and often vague directions found in legal instructions and on the other hand, the more precise *ad damnums*, comparable cases, and damages scheduling. Typical legal instructions exhort jurors to make fair and reasonable decisions, but do not help jurors with what they say is most challenging, appropriately calibrating the damage award amounts. In contrast, precise recommendations such as damages scheduling and comparable cases, even if they are advisory, have the risk of potentially pushing jurors toward similar award judgments even if the cases are meaningfully different. Leaving aside the challenge of finding appropriate comparison cases, the influence of comparable cases could be so powerful as to reduce the jury's contextual sensitivity as a fact finder. We need to find a middle ground, guidance that is helpful but does not overly bind or control the jury's decision making. Scaling guidance based on that examined in this study, but with additional benchmarks to help the jury by putting dollar amounts in context holds some

promise.

The combined results offer some practical implications for attorneys. The starting point is to effectively and powerfully convey the characteristics of an injury and the extent to which the injury interferes with the plaintiff's activity so that jurors appreciate the gist, which has been shown to be helpful in other domains of decision making (e.g., Blalock & Reyna, 2016). The significant connections among injury severity, ordinal gist judgments, and damage awards show that helping jurors get the gist of an injury is a critically important step in moving them to adopt a desired damage award.

Concussions

As a final note, in these experiments, we incorporated sports-related concussions, an increasingly important topic in public health. We chose to use concussions not only because of the impact on public health but also because of the varying ways and degrees in which concussions can affect a person's life. Concussion injuries vary in severity and intangibility of impairment, which present challenges for the legal system (Giza et al., 2013; McCrory et al., 2017; Rasmussen et al., 2018). Our results indicate that participants were able to distinguish between mild and severe cases, both in terms of injury severity and damage award amounts. Nevertheless, participants also indicated that they had greater difficulty coming up with dollar awards for severe injuries, though numerical guidance reduced this perceived difficulty.

As one of the first empirical studies examining the ability of potential jurors to examine concussions in a legal context, these results are informative. Major athletic organizations, school districts, coaches, and athletic trainers have been subject to increased liability in recent years. While public school districts and employees may fall under sovereign or qualified immunity in

many cases, there has been success in finding liability when an actor, particularly a coach or athletic trainer, acts recklessly or otherwise increases the risks inherent in the sport (Hamilton, 2019). Public policy efforts to provide greater protections against sport-related concussions are still being developed and whether these laws will be able to bring about desired reductions in the burden of concussive injuries is still subject to debate (Shen, 2018). Thus, additional work examining concussions in a legal context, including jurors' abilities in injury evaluation and award allocation, is worthwhile.

References

Albicini, M., & McKinlay, A. (2018). A review of sideline assessment measures for identifying sports-related concussion. *Journal of Concussion, 2*, 1–11. doi: 10.1177/2059700218784826.

Avraham. R. (2006). Putting a price on pain and suffering damages: A critique of the current approaches and a preliminary proposal for change. *Northwestern University Law Review, 100*, 87-120.

Baldus, D. & MacQueen, J. C., & Woodworth, G. (1995). Improving judicial oversight of jury damages assessments: A proposal for the comparative additur/remittitur review of awards for nonpecuniary harms and punitive damages. *Iowa Law Review, 80,* 1109-1267.

Ball, D. (2011). *David Ball on damages* (3rd ed.). National Institute for Trial Advocacy.

Bavli, H. J. (2017). The logic of comparable-case guidance in the determination of awards for pain and suffering and punitive damages. *University of Cincinnati Law Review, 85*, 1-31.

Bavli, H. J., & Mozer, R. (2019). The effects of comparable-case guidance on awards for pain and suffering and punitive damages: Evidence from a randomized controlled trial. *Yale Law & Policy Review, 37*, 405-458.

Black, B. S., Hyman, D. A., Paik, M., Sage, W. M., & Silver, C. (in press). *Medical malpractice litigation: How it works; why tort reform hasn't helped.* Cato Institute.

Bovbjerg, R. R., Sloan, F. A., Blumstein, J. F. (1988-1989). Valuing life and limb in tort: Scheduling pain and suffering. *Northwestern University Law Review, 83*, 908-976.

Bornstein, B. H. (2017). Jury simulation research: Pros, cons, trends, and alternatives. In M. B. Kovera (Ed.), *The psychology of juries* (pp. 207-226). American Psychological Association. http://dx.doi.org/10.1037/0000026-010

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

Bornstein, B. H., Hamm, J. A., Dellapaolera, K. S., Kleynhans, A., & Miller, M. K. (2020). JUST: A measure of jury system trustworthiness. *Psychology, Crime & Law, 26*(8), 797-822. https://doi.org/10.1080/1068316X.2020.1740222

Campbell, J., Chao, B., & Robertson, C. (2017). Time is money: An empirical assessment of non-economic damages arguments. *Washington University Law Review, 95*, 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*, 543-571.

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*, 519-540. https://doi.org/10.1002/(SICI)1099-0720(199612)10:6<519::AID-ACP417>3.0.CO;2-5

Chesapeake & Ohio Railway Company v. Arrington, 101 S.E. 415 (Va. 1919).

Devine, D. J. (2012). *Jury decision making: The state of the science*. NYU Press.

Diamond, S. S., Saks, M. J., & Landsman, S. (1998). Juror judgments about liability and damages: Sources of variability and ways to increase consistency. *DePaul Law Review, 48*, 301-325.

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

Dumke, H. A. (2017). Posttraumatic headache and its impact on return to work after mild traumatic brain injury. *Journal of Head Trauma Rehabilitation, 32*(2), E55–E65. doi:

10.1097/HTR.0000000000000244

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

Field, R. H., Kaplan, B., & Clermont, K. M. (2020). *Materials for a basic course in civil procedure* (13th ed.). Foundation Press.

Garavito, D. M. N., Reyna, V. F., & DeTello, J. E. (2018). A concussion by any other name: Differences in willingness to risk brain injury by label and level of participation in high-school and college sports. *Applied Cognitive Psychology, 33*(4), 646-654. doi: 10.1002/acp.3481.

Gastil, J., Dees, E. P., Weiser, P. J., & Simmons, C. (2010). *The jury and democracy: How jury deliberation promotes civic engagement.* Oxford University Press.

Geary, D. C., Berch, D. B., Boykin, A. W., Embretson, S., Reyna, V., & Siegler, R. S. (2011). Learning mathematics: Findings from the national (United States) mathematics advisory panel. In N. Canto (Ed.), *Issues and proposals in mathematics education* (pp. 175-221). Gulbenkian.

Giza, C. C., Kutcher, J. S., Ashwal, S., Barth, J., Getchius, T. S. D., Gioia, G. A., … Zafonte, R. (2013). Summary of evidence-based guideline update: Evaluation and management of concussion in sports: Report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology, 80*(24), 2250–2257. doi: 10.1212/WNL.0b013e31828d57dd

Graff, H. J., Siersma, V., Møller, A., Kragstrup, J., Andersen, L. L., Egerod, I., & Malá Rytter, H. (2019). Labour market attachment after mild traumatic brain injury: nationwide cohort study with 5-year register follow-up in Denmark. *BMJ Open, 9*(4), e026104. doi:

10.1136/bmjopen-2018-026104

Greene, E. & Bornstein, B. H. (2003). *Determining damages: The psychology of jury awards*. American Psychological Association. http://dx.doi.org/10.1037/10464-000

Hamilton, V. E. (2019). Play now, pay later?: Youth and adolescent collision sports. *Hastings Law Journal, 71*, 151-196.

Hans, V. P. (2000). *Business on trial: The civil jury and corporate responsibility*. Yale University Press.

Hans, V. P. (2014). What's it worth? Jury damage awards as community judgments. *William & Mary Law Review, 55,* 935-969. https://scholarship.law.wm.edu/wmlr/vol55/iss3/8

Hans, V. P., Gastil, J., & Feller, T. (2014). Deliberative democracy and the American civil jury. Journal of Empirical Legal Studies, 11(4), 697-717. https://doi.org/10.1111/jels.12053

Hans, V. P., Helm, R. K, & Reyna, V. F. (2018). From meaning to money: Translating injury into dollars. *Law and Human Behavior, 42,* 95-109. doi: 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* (Special Issue 1), 120-147. https://doi.org/10.1111/j.1740-1461.2011.01233.x

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

Hastie, R., Schkade, D., & Payne, P. W. (1999). Juror judgments in civil cases: Effects of plaintiff's requests and plaintiff's identity on punitive damage awards. *Law & Human Behavior, 23*, 445-470.

Helm, R. K., Hans, V. P., & Reyna, V. R. (2017). Trial by numbers. *Cornell Journal of Law and Public Policy*, *27,* 107-143. https://scholarship.law.cornell.edu/cjlpp/vol27/iss1/3

Helm, R. K., Hans, V. P., Reyna, V. R., & Reed, K. (2020). Numeracy in the jury box: Numerical ability, meaning, and damage award decision making. *Applied Cognitive Psychology, 34,* 434-448. https://doi.org/10.1002/acp.3629

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

Irvine, K., Hoffman, D.A., & Wilkinson-Ryan, T. (2018). Law and psychology grows up, goes online, and replicates. *Journal of Empirical Legal Studies, 15*, 320-355. https://doi.org/10.1111/jels.12180

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

Kalven, H., Jr. (1958). The jury, the law, and the personal injury damage award. *Ohio State Law Journal*, *19*, 158-178.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making, 25*, 361-381. https://doi.org/10.1002/bdm.752

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowd sourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*, 433-442. https://doi.org/10.3758/s13428-016-0727-z

Losoi, H., Silverberg, N. D., Wäljas, M., Turunen, S., Rosti-Otajärvi, E., Helminen, M., … Iverson, G. L. (2016). Recovery from mild traumatic brain injury in previously healthy adults. *Journal of Neurotrauma, 33*(8), 766–776. doi: 10.1089/neu.2015.4070

Malouff, J., & Schutte, N. S. (1989). Shaping juror attitudes: Effects of requesting different

damage amounts in personal injury trials. *Journal of Social Psychology, 129*(4), 491-497. https://doi.org/10.1080/00224545.1989.9712067

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

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,* 164-174. https://doi.org/10.1007/s10979-009-9178-8

McCrory, P., Meeuwisse, W., Dvorak, J., Aubry, M., Bailes, J., Broglio, S., … Vos, P. E. (2017). Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in Berlin, October 2016. *British Journal of Sports Medicine,* bjsports-2017-097699. doi: 10.1136/bjsports-2017-097699

Modern Federal Jury Instructions-Civil (2018).

Mott, N. L., Hans, V. P., & Simpson, L. (2000). What's half a lung worth? Civil jurors' accounts of their award decision making. *Law and Human Behavior, 24*, 401-419. https://doi.org/10.1023/A:1005540229224

New York State Senate, Civil Practice Law and Rules (no date), https://www.nysenate.gov/legislation/laws/CVP.

Peters, E. (2020). *Innumeracy in the wild: Misunderstanding and misusing numbers.* Oxford University Press.

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, *21*(1), 31-35. https:/doi.org/10.1177/0963721411429960

Peters, E., Dieckmann, N. F., Dixon, A., Slovic, P., Mertz, C. K., & Hibbard, J. H. (2007). Less is more in presenting quality information to consumers. *Medical Care Research and Review*, *64*(2), 169-190. doi:10.1177/1077558707064002030l

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

Rasmussen, C., Diekmann, S., Egan, C., Johnson, T., & Shen, F. X. (2018). How dangerous are youth sports for the brain? A review of the evidence. *Berkeley Journal of Entertainment and Sports Law, 7*(67), 81-191. doi: 10.15779/Z38MS3K20S

Reed, K., Hans, V. P., & Reyna, V. F. (2019). Accounting for awards: An examination of juror reasoning behind pain and suffering awards. *Denver Law Review, 96*, 841-867.

Reyna, V. F. (2020). Of viruses, vaccines, and variability: Qualitative meaning matters. *Trends in Cognitive Sciences*. doi: https://doi.org/10.1016/j.tics.2020.05.015

Reyna, V. F. (2021). A scientific theory of gist communication and misinformation resistance, with implications for health, education, and policy. *Proceedings of the National Academy of Sciences, 118,* (15) e1912441117; first published April 20, 2020. https://doi.org/10.1073/pnas.1912441117

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

Reyna, V. F., & Brust-Renck, P. G. (2020). How representations of numbers and numeracy predict decision paradoxes: A fuzzy-trace theory approach. *Journal of Behavioral Decision Making.* Early View; first published April 1, 2020, 1-23. https://doi.org/10.1002/bdm.2179

Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition, 5*, 1-9. https://doi.org/10.1016/j.jarmac.2015.12.003

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,* 280-294. doi:10.1037/law0000048

Reyna, V. F., Nelson, W. L., Han, P. K., & Pignone, M. P. (2015). Decision making and cancer. *American Psychologist, 70*, 105-118. http://dx.doi.org/10.1037/a0036834

Robbennolt, J. K., & Hans, V. P. (2016). *The psychology of tort law*. NYU Press.

Rowell, A., & Bregant, J. (2014). Numeracy and legal decision making. *Arizona State Law Journal, 46*, 13-29.

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

Schuck, P. H. (1993). Mapping the debate on jury reform. In R. E. Litan (Ed.), *Verdict: Assessing the civil jury system* (pp. 306-340). The Brookings Institution.

Shen, F. X. (2018). Are youth sports concussion statutes working? *Duquesne Law Review, 56*, 7–33.

Sherwin, E., & Bray, S. L. (2020). *Ames, Chafee, and Re on remedies, cases, and materials* (3rd ed.). West Academic.

Siegler, R. S. & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428-444. https://doi.org/10.1111/j.1467-8624.2004.00684.x

Sunstein, C. R, Hastie, R., Payne, J. W., Schkade, D. A., & Viscusi, W. K. (2002). *Punitive*

damages: How juries decide. University of Chicago Press.

Ward, J., William, J., & Manchester, S. (2017, July 25). 110 N.F.L. brains. *New York Times*, https://www.nytimes.com/interactive/2017/07/25/sports/football/nfl-cte.html.

Wissler, R. L., Rector, K. A., & Saks, M. J. (2001). The impact of jury instructions on the fusion of liability and compensatory damages. *Law and Human Behavior, 25,* 125-139.

Wissler, R. L., Saks, M. J., Hart, A. J., & Evans, D. C. (1997). Reducing variability in civil jury awards. *Law and Human Behavior, 21,* 243-256.

Wixsted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18*(1), 10-65. Doi: 10.1177/1529100616686966

Tables

Table 1

Experiment 1: Juror Metacognitive Experience as a Function of Attorney Guidance

	Means [95% CI LL, UL], *SE*		
Measure	Verbal+Numerical	Verbal-Only	No-Guidance Control
Helpfulness, Plaintiff	4.93 [4.60, 5.26], .17	4.63 [4.30, 4.96], .17	3.25 [2.92, 3.59], .17
Helpfulness, Judge	3.86 [3.55, 4.17], .16	3.70 [3.38, 4.01], .16	3.10 [2.78, 3.41], .16
Difficulty	4.58 [4.24, 4.91], .17	5.05 [4.71, 5.39], .17	5.24 [4.90, 5.89], .17
Confidence	6.22 [5.76, 6.68], .23	4.96 [4.50, 5.42], .23	4.69 [4.22, 5.15], .24

Note. Helpfulness and difficulty rated on 7-point scales (1 = not at all; 7 = extremely); confidence rated on 10-point scale (1= not at all; 10 = extremely).

Table 2

Experiment 2: Juror Metacognitive Experience as a Function of Attorney Guidance

	Estimated Marginal Means [95% CI LL, UL], *SE*		
Measure	Verbal+Numerical	Verbal-Only	No-Guidance Control
Helpfulness, Plaintiff	5.60 [5.35, 5.85], .13	5.00 [4.73, 5.27], .14	3.74 [3.48, 4.00], .13
Helpfulness, Judge	4.96 [4.71, 5.21], .13	4.72 [4.45, 4.98], .14	4.20 [3.94, 4.45], .13
Helpfulness, Defense	3.67 [3.39, 3.96], .15	3.59 [3.29, 3.89], .15	3.38 [3.09, 3.67], .15
Difficulty	3.87 [3.60, 4.13], .14	4.50 [4.21, 4.78], .15	4.57 [4.29, 4.86], .14
Confidence	8.05 [7.75, 8.36], .16	6.82 [6.50, 7.15], .17	7.20 [6.89, 7.52], .16

Note. Helpfulness and difficulty were rated on 7-point scales (1 = not at all; 7 = extremely); confidence was rated on a 10-point scale (1= not at all; 10 = extremely).

Table 3

Experiment 3: Juror Metacognitive Experience as a Function of Attorney Guidance

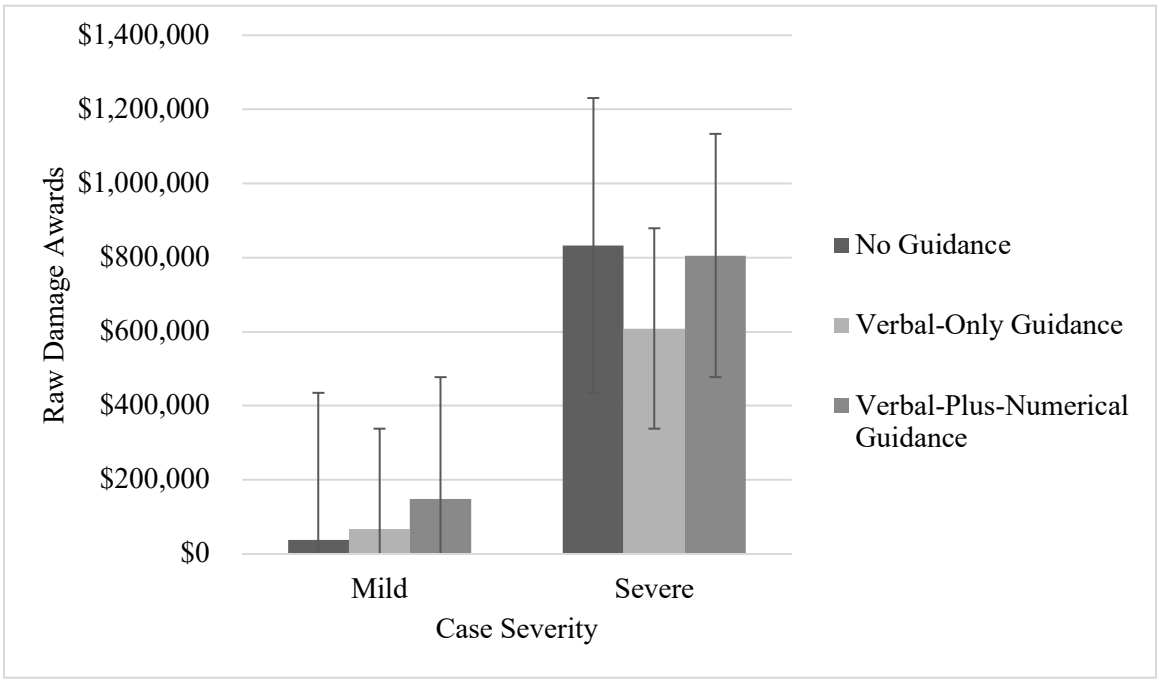
	Means [95% Confidence Interval], *SE*			
Measure	Verbal+Numerical	Verbal-Only	Numerical-Only	No-Guidance Control
Helpfulness, Plaintiff	5.12 [4.82, 5.41], .15	4.61 [4.32, 4.90], .15	3.99 [3.69, 4.29], .15	3.45 [3.15, 3.74], .15
Helpfulness, Judge	4.57 [4.28, 4.86], .15	4.12 [3.83, 4.41] .15	3.83 [3.54, 4.13], .15	3.66 [3.36, 3.95], .15
Helpfulness, Defense	3.69 [3.25, 4.14], .23	3.77 [3.3, 4.22], .23	3.48 [3.02, 3.94], .23	3.40 [2.93, 3.87], .23
Difficulty	4.40 [4.13, 4.68], .14	5.36 [5.09, 5.64], .15	5.28 [5.00, 5.55], .15	5.22 [4.94, 5.50], .14
Confidence	6.25 [5.85, 6.66], .21	4.53 [4.13, 4.93], .21	5.26 [4.86, 5.67], .21	4.53 [4.13, 4.93], .21

Note. Helpfulness and difficulty were rated on 7-point scales (1 = not at all; 7 = extremely); confidence was rated on a 10-point scale (1= not at all; 10 = extremely).

Figures

Figure 1

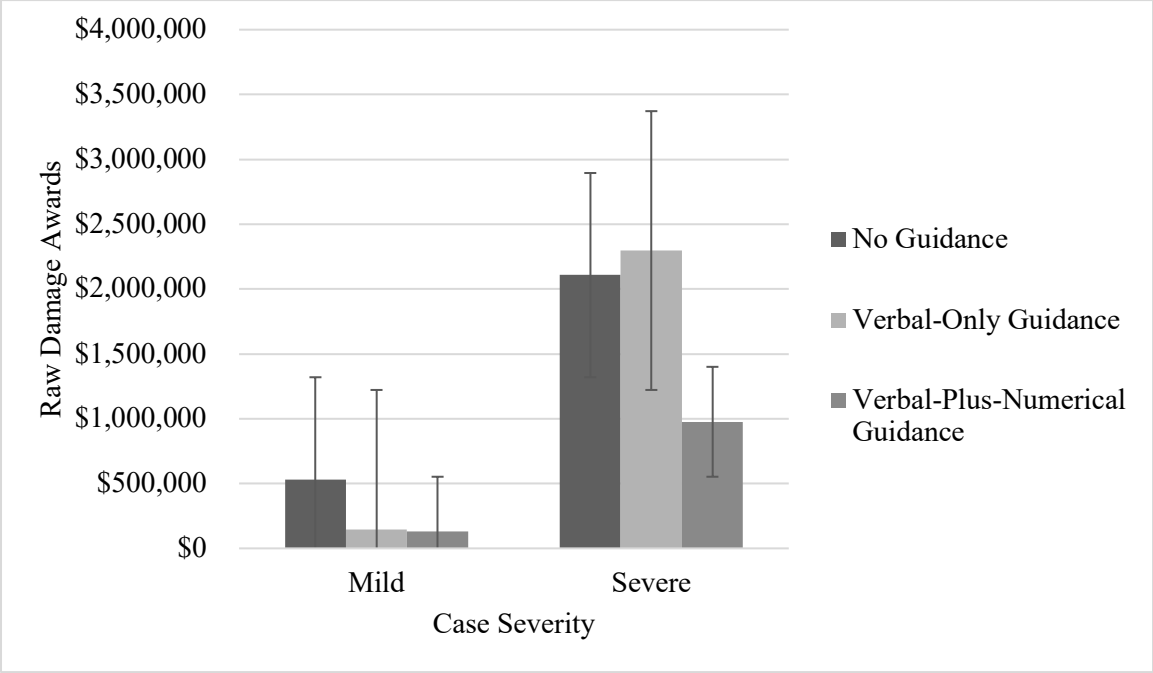
Experiment 1: Damage Awards by Case Severity and Guidance



Note. Figure 1 displays the raw damage awards. Statistical tests reported in the text are based on the natural log data.

Figure 2

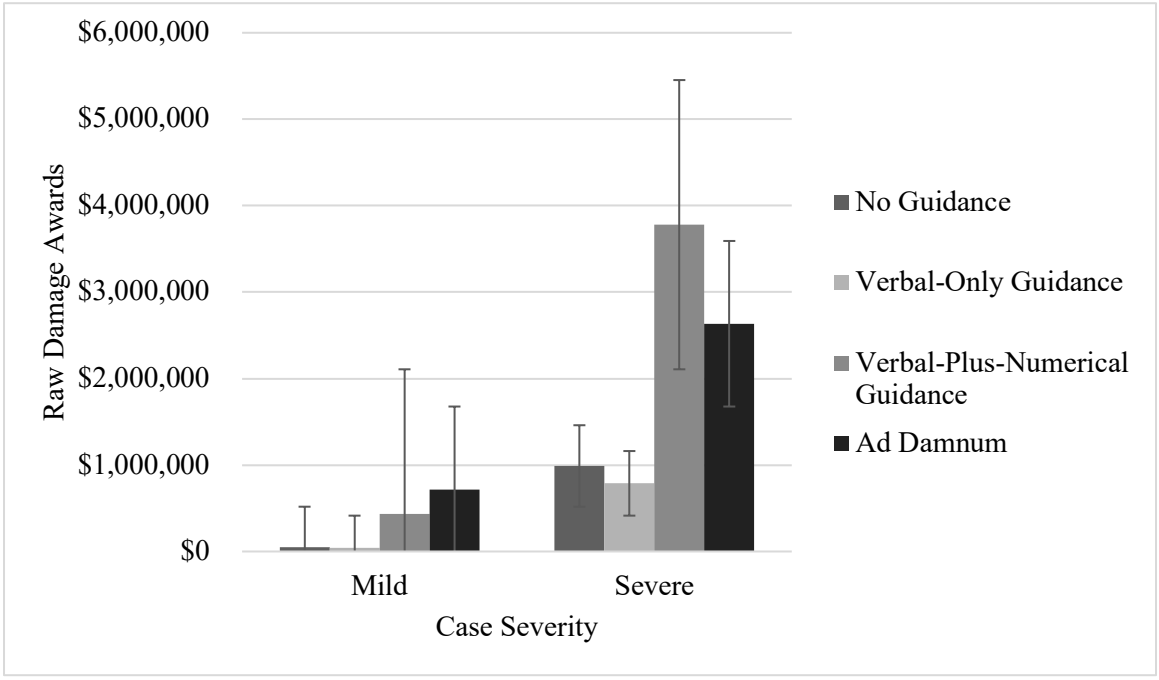
Experiment 2: Damage Awards by Case Severity and Guidance



Note. Figure 2 displays the raw damage awards. Statistical tests reported in the text are based on the natural log data.

Figure 3

Experiment 3: Damage Awards by Case Severity and Guidance



Note. Figure 3 displays the raw damage awards. Statistical tests reported in the text are based on the natural log data.

Appendix

Plaintiff Attorney's Guidance

No-Guidance Control – Experiments 1, 2, & 3

Members of the jury, let me advise you about the calculation of an amount for pain and suffering. No one can tell you how to arrive at an amount of money for pain and suffering. I cannot tell you. The judge cannot. No one can. You just have to use your own good common sense and your best judgment.

Verbal-Only Guidance & Verbal-plus-Numerical Guidance – Experiments 1 & 2

Members of the jury, let me advise you about the calculation of an amount for pain and suffering. The value of pain and suffering is based on three factors: **How bad? How long? And how interfering?**

First, <u>how bad</u> is the pain and suffering? How much does it hurt? Where on the scale from mild to horrendous does the intensity of the harm lie? Pain can be minor, medium, or extremely bad. If negligence hurts someone's back and the pain is low, not too bad, there should be money to make up for that pain, but not much. "Not too bad" means low on the scale. [Verbal: That'd be the kind of case in which I'd have to ask you for a verdict of a very small amount.] [Numerical: That'd be the kind of case in which I'd have to ask you for a verdict of an amount like **$25,000**.]

Now let's move up the scale. If the pain and suffering are worse, if it's the kind of hurting a person can't put out of their mind, maybe bad enough to even keep him from doing certain things, keeps him awake nights even when he takes all the medications they give him, then it takes more money to make up for the pain. [Verbal: That would make it the kind of case in which I'd have to ask you for a verdict of a fairly large amount.] [Numerical: That would make it the kind of case in which I'd have to ask you for a verdict of more like **$250,000.]**

Now let's go up the scale another step to "high." What happens when the pain and suffering are so bad it takes over everything? Nothing makes it go away, the person can't do anything—the kind of searing, constant pain they call "the window into Hell." [Verbal: That high up the scale, a small amount or even a fairly large amount cannot make up for it. In that kind of case, I would have to ask you for a verdict of a great deal of money in the same proportion as the pain.] [Numerical: That high up the scale, $25,000 or $250,000 cannot make up for it. In that kind of case, I would have to ask you for a verdict of more like **a million dollars.**]

Second, <u>how long?</u> What length of time does the pain and suffering last? Pain and suffering that lasts for a short time, a few weeks, maybe a month or two, is low on the scale. Takes just a little to make up for it. [Verbal: So that'd be the kind of case in which I'd have to ask for maybe a very small amount. But what if it lasts for two or three years? Then that small amount is not enough, because the pain has intruded on a big part of the person's life. That would make it the kind of case in which I'd be forced to ask for a fairly large amount depending on how long.] [Numerical: So that'd be the kind of case in which I'd have to ask for maybe a verdict of **$25,000**. But what if it lasts for a medium amount of time, say for two or three years? Then that **$25,000** is not enough, because the pain has intruded on a big part of the person's life. That would make it the kind of case in which I'd be forced to ask for **$250,000** or more**,** depending on how long.]

But at the top of the scale? Not just for a few weeks or a substantial part of a year, but the rest of the person's life? Thirty more years or more? Living every day, every hour with this pain? Getting worse with time, as arthritis and other problems set in to make it hurt more? Then we go higher on the scale. [Verbal: We're beyond making up for it with just a fairly large amount. Now, the person can't even lie there at night and take comfort in hope – the one thing we cling to

when we have nothing else - the hope that "someday this pain will be gone." Here, too, I would have to ask you for a verdict of a great deal of money for the pain and suffering.] [Numerical: We're beyond making up for it with just $25,000 or $250,000. Now the person can't even lie there at night and take comfort in hope - the one thing we cling to when we have nothing else - the hope that "someday this pain and suffering will be gone." Again, **a million dollars** or more would be warranted.]

Third and finally, <u>how interfering?</u> That is, how much does the pain and suffering interfere with things a person normally does? Just a little? If the person can do all but a few things he used to do, then it does not take much to make up for it. [Verbal: Say, just a twinge of pain but he can still walk and run; that's maybe a very small amount.] [Numerical: Say, just a twinge of pain but he can still walk and run; that's maybe **$25,000.]** But if there are lots of things the pain keeps him from doing, things that had been important parts of his life, that made his life worth living and now they're taken away, it takes more money. If it's too painful to work around the house or in the yard, if he can't walk without assistance or ride in a car for more than a few minutes because of the pain, if he can't pick up his little girl, that's a major intrusion. [Verbal: That's the middle of the scale, which puts it in the range of a fairly large amount.] **[**Numerical: That's the middle of the scale, which puts it in the range of **$250,000**.**]** Up at the top of the scale is the kind of case in which the pain is so bad that the person can't do anything. Everything his body used to be able to do, now someone else has to do for him. [Verbal: In this case, we're at the top of the high scale, where it takes a great deal of money to make up for what he can't do anymore.] [Numerical: In this case, we're at the top of the high scale, where it takes **a million dollars** to make up for what he can't do anymore.]

Members of the jury, as you decide on a damage award for my client's pain and suffering,

I recommend that you keep in mind these three dimensions – How bad? How long? How interfering? It's up to you how you consider them and how you combine them to reach a fair and just award for my client.

Verbal-Only & Verbal-plus-Numerical Guidance – Experiment 3

*Note. Changes from the language in Experiments 1 and 2 are underlined.

Members of the jury let me advise you about the calculation of an amount for pain and suffering. The value of pain and suffering is based on three factors: **How bad? How long? And how interfering?**

First, how bad is the pain and suffering? How much does it hurt? Where on the scale from mild to horrendous does the intensity of the harm lie? Pain can be minor, medium, or extremely bad. <u>If negligence gives someone a concussion</u> and their pain is LOW, not too bad, there should be money to make up for that pain, but not much. "Not too bad" means low on the scale. [verbal: That'd be the kind of case in which I'd have to ask you for a verdict of a very small amount.] [numeric: That'd be the kind of case in which I'd have to ask you for a verdict of an amount like $25,000.]

Now let's move up the scale to a MEDIUM level of pain. If the pain and suffering are worse, if it's the kind of hurting a person can't put out of his mind, maybe bad enough to even keep him from doing certain things, keeps him awake nights even when he takes all the medications they give him, then it takes more money to make up for the pain. [verbal: That would make it the kind of case in which I'd have to ask you for a verdict of [verbal: a fairly large amount.] [numeric: an amount like <u>1 million dollars</u>.]

Now let's go up the scale another step to a HIGH level of pain. What happens when the pain and suffering are so bad it takes over everything? Nothing makes it go away, the person

can't do anything—the kind of searing, constant pain they call "the window into Hell." That high up the scale, [verbal: a small amount or even a fairly large amount] [numeric: $25,000 or even 1 million dollars] cannot make up for it. In that kind of case, I would have to ask you for a verdict of [verbal: a great deal of money in the same proportion as the pain] [numeric: more like 5 million dollars].

Second, how long is the pain and suffering? What length of time does the pain and suffering last? Pain and suffering from a concussion that lasts for a SHORT time, a few weeks, maybe a month or two, is low on the scale. Takes just a little to make up for it. So that'd be the kind of case in which I'd have to ask for maybe [verbal: a very small amount] [numeric: a verdict of $25,000]. But what if it lasts a MEDIUM amount of time, say for half a year or more? Then that small amount is not enough, because the pain has intruded on a big part of the person's life. That would make it the kind of case in which I'd be forced to ask for [verbal: a fairly large amount] [numeric: 1 million dollars] depending on how long. But at the TOP of the scale? Not just for a few weeks or a few years, but the rest of the person's life? Thirty more years or more? Living every day, every hour with this pain? Getting worse with time, as a person's brain ages and the concussion interacts with other problems to make it worse? Then we go higher on the scale. We're beyond making up for it with just [verbal: a fairly large amount] [numeric: 1 million dollars]. Now the person can't even lie there at night and take comfort in hope - the one thing we cling to when we have nothing else - the hope that "someday this pain will be gone." Here, too, I would have to ask you for a verdict of [verbal: a great deal of money] [numeric: 5 million dollars or more] for the pain and suffering.

Third and finally, how interfering is the pain and suffering? That is, how much does the pain and suffering interfere with things a person normally does? Just a LITTLE? If the person

can do all but a few things he used to do, then it does not take much to make up for it. Say, very mild dizziness but he can still walk and run, still go to school**.** That's maybe [verbal: a very small amount] [numeric: $25,000]. But if there are lots of things the pain keeps him from doing, things that had been important parts of his life, that made his life worth living and now they're taken away, it takes more money. If it's too painful to walk around school, if he can't walk without assistance, or ride in a car for more than a few minutes because of the pain, if he can't be physically active or do the other things he loves, that's a major intrusion. That's the MIDDLE of the scale, which puts it in the range of [verbal: a fairly large amount] [numeric: 1 million dollars]. Up at the TOP of the scale is the kind of case in which the pain is so bad that the person can't do anything. Everything his body used to be able to do, now someone else has to do for him. In this case, we're at the top of the high scale, where it takes [verbal: a great deal of money] [numeric: 5 million dollars] to make up for what he can't do anymore.

Members of the jury, as you decide on a damage award for my client's pain and suffering, I recommend that you keep in mind these three dimensions – How bad? How long? How interfering? It's up to you how you consider them and how you combine them to reach a fair and just award for my client.

Numerical-Only Guidance (*Ad Damnum* condition) – Experiment 3

On behalf of my client, I would like to ask you for a damage award in the amount of 5 million dollars to compensate him for the pain and suffering he has experienced as a result of the defendant's negligence.

Supplemental Materials

Helpfulness and confidence analyses, Experiment 1. Data analysis revealed unexpected interactions with the case. There was an unexpected three-way interaction between severity, guidance and case that was not relevant to our hypotheses, $F(2,185) = 3.57$, $MSe = 1.77$, $p = .03$, $\eta_p^2 = .04$. Guidance was more helpful than no guidance in all conditions (ps < .03). None of the pairwise case-severity comparisons showed a statistically significant difference between verbal-plus-numerical and verbal-only guidance (ps > .05); however, in the mild soccer case, verbal-plus-numerical guidance ($M = 5.14$, [4.44 – 5.85], $SE = .36$) was marginally more helpful than verbal-only guidance ($M = 4.25$, [3.66 – 4.84], $SE = .30$, $p = .06$).

In evaluations of the helpfulness of the judge's guidance, participant ratings also showed an unexpected interaction effect, this time between the injury severity and the case, $F(2, 185) = 12.11$, $Mse = 1.61$, $p = .001$, $\eta_p^2 = .06$. In examining the pairwise comparisons, it appears that, in the mild condition, the judge's guidance was rated as more helpful for football ($M = 4.02$ [3.65, 4.40], $SE = .19$) than for soccer ($M = 3.42$ [3.07, 3.77], $SE = .18$; $p = .02$). However the opposite pattern occurred for the severe case; in the severe injury condition, the judge's guidance was rated as more helpful for soccer ($M = 3.72$ [3.36, 4.08], $SE = .18$) than for football ($M = 3.04$ [2.68, 3.41], $SE = .19$; $p = .01$).

Finally, in the confidence ratings, there was an unexpected two-way interaction between the case and injury severity, $F(1, 185) = 3.98$, $MSe = 3.47$, $p < .05$, $\eta_p^2 = .02$. The pairwise comparisons indicated a marginal effect of the case, such that confidence was higher for the football case than for the soccer case, but only in the mild injury conditions ($M_{\text{soccer}} = 5.01$ [4.50, 5.53], $SE = .26$; $M_{\text{football}} = 5.73$ [5.18, 6.28], $SE = .27$; $p = .06$). There was an unexpected three-way interaction between the case, severity, and guidance, however it did not clarify the two-way

interaction, $F(2, 185) = 3.84$, $MSe = 3.47$, $p = .04$, $\eta_p^2 = .04$. In the three-way interaction, the only significant effect occurred for the no guidance condition with the severe case, for which participants were significantly more confident in their award in the soccer case ($M = 5.64$ [4.66, 6.63], $SE = .50$) than in the football case ($M = 4.05$ [3.25, 4.85], $SE = .41$; $p = .01$); but, there were no significant differences with either type of guidance or in the severe case.

The presence of these unexpected interactions was puzzling. We employed two cases describing two different sports, football and soccer, to examine generalizability of the findings. It is possible that participants' distinctive (but unmeasured) reactions to the two sports helped to produce these interactions with our theoretically relevant variables. Therefore, to remove this potential source of variability, in subsequent experiments, we revised the materials to use only football cases.