Supplemental Materials

Alternative Outlier Analyses

As stated in the article, we handled the outlier damage awards by taking the extreme outliers (defined as over 3 *SD* above the mean of their condition) and reducing the awards to an award that was 3 *SD* above the mean. We also took two alternative approaches to handling the outliers: including the natural log values of the outlier awards, and excluding the outlier awards. Overall, the patterns were the same and there was just one minor difference in what was significant, so the results presented in the article include the outliers' data reduced to 3 *SD* above the mean. For the sake of completeness, here we report the results of analyses to the two alternate approaches to the outlier damage awards.

Experiment 1, Including Outliers' Natural Log Values. Experiment 1 had 3 extreme outliers. We included the natural log values of the 3 outliers and conducted the same tests of Hypotheses 3, 4, 6, and 7.

Award validity. We conducted an ANOVA including the natural log values of the 3 outliers and using severity, case, and guidance as the independent variables and the natural log damage award as the dependent variable. 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), $F(1,185) = 54.86$, $p < .001$, $\eta_p^2 = .23$, $d = .04$ (*Hypothesis 3*).

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 (M = 11.67, SE = .25) than verbal-only guidance (M = 10.80, SE = .25, $p <$

.01, d = .02) or no-guidance control (M = 10.33, SE = .25, p < .001, d = .03), F(2, 185) = 7.60, MSe = 3.88, p = .001, η_p^2 = .08, Cohen's f = .29. However, there was not a significant interaction between severity and guidance, F(2, 185) = 1.10, MSe = 3.89, p = .33, η_p^2 = .01, Cohen's f = .10 (see Table S1 for means).

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 3 outliers, with 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).

As expected, the size of dollar awards differed in concert with classifications of injury as low, medium, or high (*Hypothesis 6*), F(2, 159) = 8.14, MSe = 3.22, p < .001, η_p^2 = .09, Cohen's f = .31. Awards classified as "low" were significantly lower amounts (M = 9.79, 95% CI [9.11, 10.46], SE = .34) than awards classified as "medium" (M = 11.13 [10.76, 11.49], SE = .18, p < .001, d = .02) or "high" (M = 11.93 [11.10, 12.77], SE = .42, p < .001, d = .06); however, awards classified as "medium" were only marginally lower than those classified as high (p = .08, d = .02).

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) =.56, MSe = 3.23, p = .56, η_p^2 = .02, Cohen's f = .14.

Experiment 1, Excluding Outliers. We excluded the 3 extreme outliers, and conducted the same tests of Hypotheses 3, 4, 6, and 7.

Award validity. We conducted an ANOVA excluding the 3 outliers, using severity, case, and guidance as the independent variables and the natural log damage award as the dependent

variable. 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 = 11.90 [11.51, 12.29], SE = .20) than the less severe case (M = 9.83 [9.44, 10.22], SE = .20), $F(1,182) = 54.91$, $p < .001$, $\eta_p^2 = .23$, $d = .04$ *(Hypothesis 3*).

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 (M = 11.61 [11.14, 12.08], SE = .24) than verbal-only guidance (M = 10.74 [10.26, 11.21], SE = .24, $p < .01$, d = .65) or no-guidance control (M = 10.25 [9.77, 10.74], SE = .25, $p < .001$, d = 1.00), $F(2, 182) = 8.08$, $MSe = 3.64$, $p < .001$, $\eta_p^2 = .08$, Cohen's $f = .29$. However, there was not a significant interaction between severity and guidance, $F(2, 182) = 0.91$, $MSe = 3.64$, $p = .41$, $\eta_p^2 = .01$, Cohen's $f = .10$ (see Table S2 for means).

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 excluding the 3 outliers, with 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).

As expected, the size of dollar awards differed in concert with classifications of injury as low, medium, or high (*Hypothesis 6*), $F(2, 156) = 8.22$, $MSe = 3.02$, $p < .001$, $\eta_p^2 = .10$, Cohen's $f = .33$. Awards classified as "low" (n = 49) were significantly lower amounts (M = 9.79, 95% CI [9.13, 10.44], SE = .33) than awards classified as "medium" (n = 106, M = 11.04 [10.68, 11.39], SE = .18, $p < .001$, d = .84 or "high" (n = 36, M = 11.93 [11.12, 12.74], SE = .41, $p < .001$, d = 1.27). We note that in this analysis with excluded outliers, awards classified as "medium" were also significantly lower than those classified as "high" ($p < .05$, d = .58). Although the pattern of

means is the same, that is a different result from the other two analyses that included the outliers, which found only nonsignificant (p = .08) differences between medium and high.

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, $F(4, 156) = .80$, $MSe = 3.02$, $p = .53$, $\eta_p^2 = .02$, Cohen's $f = .14$.

Experiment 2. Including Natural Log Values of Outliers. Experiment 2 had 7 extreme outliers. We included the natural log values of the 7 outliers and conducted the same tests of Hypotheses 3, 4, 6, and 7.

Award validity. As in Experiment 1, we tested award validity through an ANOVA, including the natural log values of the 7 outliers, with 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 ($M_{low} = 10.46$, 95% CI [10.22, 10.71, $SE = .13$; $M_{high} = 13.17$ [12.93, 13.41], $SE = .12$), $F(1, 457) = 240.15$, $p < .001$, $\eta_p^2 = .34$, $d = 2.02$ (*Hypothesis 3*).

Contrary to expectations (*Hypothesis 4*), guidance did not increase award validity, as there was no significant interaction between guidance and severity, $F(2, 457) = 0.63$, $MSe = 3.53$, $p = .53$, $\eta_p^2 < .01$, Cohen's $f = .05$. Unlike Experiment 1, the main effect of guidance on awards was not significant, $F(2, 457) = 0.50$, $MSe = 3.53$, $p = .61$, $\eta_p^2 < .01$, Cohen's $f = .04$ (see Table S3 for means).

Verbatim-gist coherence. In order to test whether ordinal awards were consistent with the classification, we again conducted an ANOVA with the outliers and with 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*), $F(2, 430) = 48.82$, $MSe = 2.61$, $p < .001$, $\eta_p^2 = .19$, Cohen's $f = .48$. Awards classified as high ($n = 93$) were significantly higher ($M = 13.92$ [13.36, 14.47], $SE = .28$) than awards classified as medium ($n = 220$, $M = 12.23$ [12.00, 12.45], $SE = .11$, $p < .001$, $d = 1.07$), which were both significantly higher than awards classified as low ($n = 152$, $M = 10.23$ [9.78, 10.69], $SE = .23$, ps $< .001$, $d_{verbal+numerical} = 1.22$, $d_{verbalonly} = 1.88$).

Contrary to predictions, guidance did not increase verbatim-gist coherence (*Hypothesis 7*) as the interaction between guidance and award classification was not significant, $(2, 430) = 1.27$, $MSe = 2.61$, $p = .28$, $\eta_p^2 = .01$, Cohen's $f = .10$. 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).

Experiment 2, Excluding Outliers. We excluded the 7 outliers and conducted the same tests of Hypotheses 3, 4, 6, and 7.

Award validity. As in Experiment 1, we tested award validity through an ANOVA, excluding the 7 outliers, and 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 ($M_{low} = 10.20$, 95% CI [9.91, 10.49], $SE = .15$; $M_{high} = 12.97$ [12.68, 13.26], $SE = .15$), $F(1, 457) = 174.94$, $p < .001$, $\eta_p^2 = .28$, $d = 1.73$ (*Hypothesis 3*).

Contrary to expectations (*Hypothesis 4*), guidance did not increase award validity, as there was no significant interaction between guidance and severity, $F(2, 457) = 0.29$, $MSe = 5.12$, $p = .75$, $\eta_p^2 < .01$, Cohen's $f = .03$ (see Table S4 for means). Unlike Experiment 1, the main effect of guidance on awards was not significant, $F(2, 457) = 2.24$, $MSe = 5.26$, $p = .11$, $\eta_p^2 = .01$, Cohen's $f = .10$.

Verbatim-gist coherence. In order to test whether ordinal awards were consistent with the classification, we again conducted an ANOVA excluding the outliers and with 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*), $F(2, 423) = 51.25$, $MSe = 2.38$, $p < .001$, $\eta_p^2 = .20$, Cohen's $f = .50$. Awards classified as high ($n = 91$) were significantly higher ($M = 13.88$ [13.35, 14.42], $SE = .27$) than awards classified as medium ($n = 215$, $M = 12.13$ [11.92, 12.35], $SE = .11$, $p < .001$, $d = 15$), which were both significantly higher than awards classified as low ($n = 152$, $M = 10.23$ [9.80, 10.67], $SE = .22$, ps $< .001$, $d_{high} = 1.95$, $d_{medium} = 1.20$).

Contrary to predictions, guidance did not increase verbatim-gist coherence (*Hypothesis 7*) as the interaction between guidance and award classification was not significant, $F(2, 423) = 1.63$, $MSe = 2.38$, $p = .17$, $\eta_p^2 = .02$, Cohen's $f = .14$. 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).

Experiment 3 including Natural Log Values of Outliers. Experiment 3 had 8 extreme outliers. We included the natural log values of the 8 outliers and conducted the same tests of Hypotheses 3, 4, 6, and 7.

Award validity. As in previous experiments, we tested award validity through an ANOVA including the natural log values of the 8 outliers and 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 ($M = 13.08$ [12.78, 13.39], $SE = .15$) than in the mild case ($M = 10.23$ [9.94, 10.53], $SE = .15$), $F(1, 382) = 174.56$, $MSe = 4.52$, $p < .001$, $\eta_p^2 = .31$, $d = 1.94$ (*Hypothesis 3*).

However, guidance did not increase award validity; there was no interaction between severity and guidance, $F(3, 382) = .59$, $MSe = 4.52$, $p = .62$, $\eta_p^2 < .01$, Cohen's $f = .07$ (*Hypothesis 4,* see Table S5). There was a significant main effect of guidance on awards, $F(3, 382) = 50.23$, $p < .001$, $\eta_p^2 = .28$, Cohen's $f = .62$. Awards were significantly higher when there was a recommended award amount (i.e., an anchor; $M_{\text{numericalonly}} = 12.97$ [12.54, 13.40], $SE = .22$; $M_{\text{verbal+numerical}} = 12.99$ [12.57, 13.42], $SE = .21$) than when there was no recommended award amount ($M_{\text{verbalonly}} = 10.39$ [9.97, 10.81], $SE = .21$; $M_{\text{noguidance}} = 10.28$ [9.85, 10.71], $SE = .22$; ps $< .001$).

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 with the 8 outliers included. As expected, jurors accurately assessed the size of their damage awards (*Hypothesis 6*), $F(2, 360) = 44.83$, $MSe = 3.13$, $p < .001$, $\eta_p^2 = .20$, Cohen's $f = .50$. Awards classified as low ($n = 93$) were significantly lower ($M = 9.68$, 95% CI [9.19, 10.17], $SE = .25$) than awards classified as medium ($n = 201$, $M = 11.97$ [11.72, 12.21], $SE = .13$, $p < .001$, $d = 1.51$), which were significantly lower than awards classified as "high" ($n = 90$, $M = 13.00$ [12.40, 13.60], $SE = .30$, ps $< .01$, $d_{\text{low}} = 1.78$ $d_{\text{medium}} = .61$), 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, $F(6, 360) = .50$, $MSe = 3.13$, $p = .81$, $\eta_p^2 < .01$, Cohen's $f = .10$. 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).

Experiment 3, Excluding Outliers. We excluded the 8 outliers and conducted the same tests of Hypotheses 3, 4, 6, and 7.

Award validity. As in previous experiments, we tested award validity through an ANOVA excluding the 8 outliers and 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 (n = 192, M = 13.06 [12.76, 13.36], SE = .15) than in the mild case (n = 190, M = 10.06 [9.75, 10.436], SE = .15), $F(1, 375) = 190.69$, $p < .001$, $\eta_p^2 = .34$, $d = 2.05$ (*Hypothesis 3*).

However, guidance did not increase award validity; there was no interaction between severity and guidance, $F(3, 375) = .42$, $MSe = 4.62$, $p = .74$, $\eta_p^2 < .01$, Cohen's $f = .05$ (*Hypothesis 4*; see Table S6). There was a significant main effect of guidance on awards, $F(3, 375) = 45.35$, $p < .001$, $\eta_p^2 = .27$, Cohen's $f = .61$. Awards were significantly higher when there was a recommended award amount (i.e., an anchor; $M_{\text{numericalonly}} = 12.72$ [12.28, 13.16], $SE = .22$, $n = 90$; $M_{\text{verbal+numerical}} = 12.92$ [12.50, 13.35], $SE = .22$, $n = 97$) than when there was no recommended award amount ($M_{\text{verbalonly}} = 10.39$ [9.97, 10.81], $SE = .21$, $n = 99$; $M_{\text{noguidance}} = 10.20$ [9.77, 10.63], $SE = .22$, $n = 96$; ps < .001) ($d_{\text{numericalonlyvverbalonly}} = 1.58$; $d_{\text{numericalonlyvnoguidance}} = 1.68$; $d_{\text{verbal+numericalvverbalonly}} = 1.68$; $d_{\text{verbal+numericalvnoguidance}} = 1.78$).

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 excluding the 8 outliers. As expected, jurors accurately assessed the size of their damage awards (*Hypothesis 6*), $F(2, 352) = 42.56$, $MSe = 3.07$, $p < .001$, $\eta_p^2 = .20$, Cohen's $f = .50$. Awards classified as "low" (n = 93) were significantly lower (M = 9.68, 95% CI [9.20, 10.16], SE = .25) than awards classified as "medium" (n = 198, M = 11.92 [11.67, 12.17], SE = .13, $p < .001$, d = 1.48), which were significantly lower than awards classified as "high" (n = 85, M = 12.89

[12.28, 13.50], *SE* = .31, low v. high: $p < .001$, $d = 1.72$*;* medium v. high: $p = .004$, $d = .57$), 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, $F(6, 352) = .35$, *MSe* = 1.08, $p = .91$, $\eta_p^2 < .01$, Cohen's $f = .08$. 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).

Table S1

Experiment 1: Mean Natural Log Damage Awards by Guidance and Severity (Hypothesis 4)

Including Outliers As-Is

	Means [95% CI LL, UL], *SE*		
Severity	Verbal+Numerical (n = 65)	Verbal-Only (n = 67)	No-Guidance Control (n = 65)
Low Severity (n = 98)	10.78 [10.09, 11.47], .35	9.87 [9.22, 10.53], .33	8.97 [8.24, 9.69], .37
High Severity (n = 99)	12.57 [11.89, 13.24], .34	11.73 [11.00, 12.46], .37	11.69 [11.02, 12.36], .34
d	1.29	1.30	1.90

Table S2

Experiment 1: Mean Natural Log Damage Awards by Guidance and Severity (Hypothesis 4)

Excluding Outliers

	Means [95% CI LL, UL], *SE*		
Severity	Verbal+Numerical (n = 64)	Verbal-Only (n = 66)	No-Guidance Control (n = 64)
Low Severity (n = 97)	10.78 [10.11, 11.45], .34	9.74 [9.11, 10.38], .32	8.97 [8.26, 9.67], .36
High Severity (n = 97)	12.44 [11.77, 13.11], .34	11.73 [11.02, 12.43], .36	11.54 [10.87, 12.20], .34
d	1.20	1.53	1.81

Table S3

Experiment 2: Mean Natural Log Damage Awards by Guidance and Severity (Hypothesis 4)

Including Outliers As-Is

	Means [95% CI LL, UL], *SE*		
Severity	Verbal+Numerical (n = 166)	Verbal-Only (n = 146)	No-Guidance Control (n = 157)
Low Severity (n = 230)	10.72 [10.32, 11.12], .20	10.36 [9.93, 10.80], .22	10.31 [9.87, 10.75], .22
High Severity (n = 239)	13.15 [12.73, 13.57], .21	13.19 [12.76, 13.62], .22	13.16 [12.76, 13.56], .20
d	1.87	2.11	1.48

Table S4

Experiment 2: Mean Natural Log Damage Awards by Guidance and Severity (Hypothesis 4)

Excluding Outliers

	Means [95% CI LL, UL], *SE*		
Severity	Verbal+Numerical (n = 166)	Verbal-Only (n = 142)	No-Guidance Control (n = 154)
Low Severity (n = 229)	10.72 [10.34, 11.10], .19	10.36 [9.94, 10.78], .21	10.20 [9.78, 10.63], .22
High Severity (n = 233)	13.15 [12.75, 13.55], .20	12.99 [12.56, 13.41], .22	13.09 [12.70, 13.41], .20
d	1.93	2.05	2.12

Table S5

Experiment 3: Mean Natural Log Damage Awards by Guidance and Severity (Hypothesis 4)

Including Outliers As-Is

	Means [95% CI LL, UL], *SE*			
Severity	Verbal+Numerical (n = 99)	Verbal-Only (n = 99)	Numerical-Only (n = 96)	No-Guidance Control (n = 96)
Low Severity (n = 197)	11.68 [11.07, 12.30], .31	8.88 [8.29, 9.47], .30	11.71 [11.11, 12.30], .30	8.66 [8.08, 9.24], .30
High Severity (n = 193)	14.31 [13.73, 14.88], .29	11.90 [11.30, 12.49], .30	14.23 [13.62, 14.84], .31	11.90 [11.27, 12.53], .32
d	1.76	2.02	1.70	2.13

Table S6

Experiment 3: Mean Natural Log Damage Awards by Guidance and Severity (Hypothesis 4)

Excluding Outliers

	Means [95% CI LL, UL], *SE*			
Severity	Verbal+Numerical (n = 97)	Verbal-Only (n = 99)	Numerical-Only (n = 90)	No-Guidance Control (n = 96)
Low Severity (n = 190)	11.54 [10.92, 12.15], .31	8.88 [8.30, 9.46], .29	11.30 [10.69, 11.92], .31	8.66 [8.10, 9.23], .29
High Severity (n = 192)	14.31 [13.74, 14.87], .29	11.90 [11.31, 12.48], .30	14.14 [13.54, 14.75], .31	11.90 [11.28, 12.52], .31
d	1.87	2.06	1.93	2.03