

## Supplemental Materials

### Experiment 1

#### Additional Demographic Information

In Experiment 1, eight percent of participants had served as a juror in a criminal trial. When asked about their highest level of education, 0.9% of participants reported that they had completed some high school, 19.9% had completed high school, 8.9% had completed a certificate, 18.2% had completed a diploma/advanced diploma, 34% had completed a bachelor degree, 3.4% had completed a graduate certificate/diploma, and 14.5% had completed a postgraduate degree.

#### Verdict

We preregistered the intention to analyse the dichotomous guilt variable using logistic regression. However, there is compelling evidence to suggest that linear regression is just as suitable for dichotomous outcome variables (Gomila, 2019) and more easily interpretable, thus we have chosen to conduct a linear regression on the dichotomous guilt variable. Thus, a multiple linear regression was conducted to predict guilt based on the two independent variables: Evidence Expression (Likelihood Ratio vs. Diagnostic Information), where Likelihood Ratio was coded as -1 and Diagnostic Information was coded as 1, and Evidence Strength (5 levels), where the increasing evidence strengths were coded as -2, -1, 0, 1, and 2. The overall regression equation was significant,  $F(3,732) = 7.23$ ,  $p < .001$ , with an  $R^2$  of .029. Evidence expression was a significant predictor of guilt,  $\beta = -.085$ ,  $p = .021$ , indicating that, overall, participants were less likely to conclude that the defendant was guilty when the evidence was expressed diagnostically rather than as a likelihood ratio. Strength of expression was also a significant predictor of guilt,  $\beta = .111$ ,  $p = .002$ , indicating that, overall, participants were more likely to conclude that the defendant was guilty as the strength of the evidence increased. Finally, the Evidence Expression by

Evidence Strength interaction was significant,  $\beta = .097$ ,  $p = .008$ . When the evidence was expressed diagnostically, participants were more likely to conclude that the defendant was guilty as the strength of the expression increased,  $t(1,734) = 39$ ,  $p < .001$ , however strength of expression did not affect guilt ratings when the evidence was expressed as a likelihood ratio,  $t(1,734) < .01$ ,  $p > .999$ .

## Experiment 2

### Additional Demographic Information

In Experiment 2, 7% of participants had served as a juror in a criminal trial. When asked about their highest level of education, 1% of participants reported that they had completed some high school, 25.5% had completed high school, 8.6% had completed a certificate, 11.8% had completed a diploma/advanced diploma, 38.3% had completed a bachelor degree, 3.6% had completed a graduate certificate/diploma, and 11.2% had completed a postgraduate degree.

### Error

There was no main effect of Likelihood Ratio,  $F(1, 394) = .032$ ,  $p = .857$ ,  $\eta^2_p = <.001$ , therefore participants were equally likely to conclude that an error occurred during analysis when a Likelihood Ratio was Present ( $M = 31.44$ ,  $SD = 27.75$ ) or Absent ( $M = 32.15$ ,  $SD = 26.71$ ). However, there was a main effect of Diagnostic Information,  $F(1, 394) = 154.64$ ,  $p = <.001$ ,  $\eta^2_p = .282$ . Participants were less likely to conclude that an error occurred during analysis when Diagnostic Information was High ( $M = 17.30$ ,  $SD = 26.06$ ) compared to Low ( $M = 46.15$ ,  $SD = 19.70$ ). Contrary to our prediction, there was no significant Likelihood Ratio  $\times$  Diagnostic Information interaction,  $F(1, 394) = .019$ ,  $p = .890$ ,  $\eta^2_p = <.001$ .

### Comparing Across Likelihood Ratio Conditions

A one-way ANOVA comparing all five experimental conditions revealed a significant difference between conditions on the measure of error likelihood,  $F(4,496) = 38.29$ ,  $p <$

.001. As per H4, we hypothesised that the Likelihood Ratio Only condition would only differ from conditions where a Likelihood Ratio was absent (i.e., diagnostic information only) compared to present. However, contrary to our hypothesis, post hoc multiple comparisons using Tukey's range test revealed that participants in the Likelihood Ratio Only condition ( $M = 30.27$ ,  $SD = 23.87$ ) were significantly less likely to conclude that an error could have occurred during analysis than participants who received a likelihood ratio with weak diagnostic information ( $M = 45.78$ ,  $SD = 20.49$ ),  $p < .001$ , and participants who only received weak diagnostic information ( $M = 46.51$ ,  $SD = 18.99$ ),  $p < .001$ . However, participants in the Likelihood Ratio Only condition were significantly more likely to conclude that an error could have occurred during analysis than participants who received a likelihood ratio with strong diagnostic information ( $M = 17.25$ ,  $SD = 26.76$ ),  $p = .001$ , or participants who received only strong diagnostic information ( $M = 17.35$ ,  $SD = 25.46$ ),  $p = .001$ . This finding demonstrates that participants' conclusions about the likelihood that an error could have occurred during the analysis is driven largely by the diagnostic information and not the likelihood ratio.