*Experiment 1 Supplementary Methods: Participant Instructions*

Below is an example crime scenario used in Experiment 1. The bold text below was changed from participant to participant, depending on whether the participant was randomly assigned to the Christian or Atheist condition of the Suspect factor (See Experiment 1 Method). The text was not in bold in the study.

*“It is a dark and stormy night. You are a detective called to the scene of a crime. A local restaurant has been set on fire. Fortunately, the culprit was observed by a crowd of people as he ran from the scene, and a suspect has already been apprehended.  He is local businessman Matthew Johnson. Mr Johnson is a married father of three children, a member of Rotary International, and a volunteer for the local branch of* ***American Atheists / the Christian church****. The witnesses, however, don't all agree about whether Mr Johnson is the man they saw running from the scene. Your task now is to determine whether to charge him with the crime. You will hear from ten of the witnesses, and after each witness you need to rate the likelihood that Mr Johnson is the guilty party. In your experience, on a dark and stormy night like tonight, 60% of the witnesses to a crime are accurate about what they report, and 40% are mistaken.”*

*Experiment 1 Supplementary Results*

Here, we also report our pre-registered analyses using as the dependent variable *Ground Truth Deviation* - the subtraction of participant probability ratings from their corresponding ground truth mathematical probabilities. Our pre-registered hypothesis #1 was that participants would deviate from the ground truth most for guilty claims about atheist Suspects. Using an ANOVA with factors *Suspect*, *Claim* and *Affiliation,* this hypothesis was not confirmed. There was no Suspect main effect *F*(1, 597) = .002, *p* = .95, *ηp2* = <.001 or Suspect\*Claim interaction *F*(1, 597) = .041, *p* = .84, *ηp2* = <.001. Our pre-registered hypothesis #3 was that religiously affiliated participants would show larger effects of Suspect. When we added *Affiliation* to the ANOVA reported above, we found no main effect of affiliation *F*(1, 597) = .002, *p* = .95, *ηp2* = <.001 nor any interactions involving affiliation ps > .39. Incidentally, one effect was significant in this ANOVA, the main effect of claim *F*(1, 451) = 7.63, *p* = <.001, *ηp2* = <.02, which confirms that participants were more conservative (tended to increase their probabilities estimates less across the sequence), compared to the ground truth model (Figure 1).

*Experiment 2 Supplementary Methods: Participant Instructions*

At the beginning of the experiment, participants viewed the following instructions:

*“You have been assigned the role of a detective and will investigate 8 crimes, for which a different suspect has been identified in each case. Based on your own opinion, you will make an initial rating about the innocence or guilt of the suspect in each case before interviewing witnesses. In each crime scenario you will interview 10 witnesses, each of whom will state an opinion about the guilt or innocence of the suspect. The witnesses will be shown in a random order. Before and after each witness, you will be required to rate the probability of the suspect’s ‘innocence’ or ‘guilt’ based on the evidence presented to you.”*

Scenario summaries and their associated sequences are shown in Supplementary Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1** |  |  |  |
| *Characteristics of the scenarios* |  |  |  |
| Scenario | Suspect | Claimsa | Guilt Sequence |
| It was a foggy day when a car was damaged and broken into | male | IGGIGGIGGG | Guilty |
| It was an early summer’s morning when a member of the public was assaulted on their way to work | male | GGIGGGIGGI | Guilty |
| It was a very late night when a member of the public was attacked in the park | male | IGIIIIGIGI | Innocent |
| It was a busy and early morning when a member of the public was mugged at a bus stop | male | IIIIGIGGII | Innocent |
| It was a dark and frosty evening when a family home and business was robbed | female | IGGGGIIGGG | Guilty |
| It was a foggy afternoon when a child was abducted on the street | female | GIGIGGIGGG | Guilty |
| It was a late winter’s night when a robbery was committed | female | IIGIGIIGII | Innocent |
| It was a frosty and misty morning when a handbag was stolen from a shopper | female | IIGIGIIIGI | Innocent |
| aG’s represent guilty witness claims. I’s represent innocent claims | | | |

Before each sequence commenced, participants read the following. Bold text changed depending on scenario (Supplementary Table 1) but was not bold in the study.

*“The suspect in question is* ***male****. 10 witnesses came forward to explain what they had seen and whether they believe the suspect in question was guilty or not guilty of committing the crime. All the witnesses are only 70% sure of what they witnessed, as* ***it was a dark night****. You will now be shown the evidence presented by the 10 witnesses, one at a time, in a sequence. Before and after each witness, you need to make a new rating of the probability to which you believe the* ***male*** *suspect is ‘innocent’ or ‘guilty’ using the sliding tool.”*

**A group of blue and black bars

Description automatically generatedSupplementary Figure S1**

*Analysis of residuals for the delta model*

*Note.* Upper row shows delta model probability estimates for all evidence samples (left) and for participant-specific mean probability values, averaged over evidence samples (right). Lower row shows residuals for probability prediction for all evidence samples (left) and for participant-specific mean residual values, averaged over evidence samples (right). The normally shaped distributions suggest that BIC assumptions are satisfied.

*Experiment 2 Supplementary Methods: Delta Model Parameter Recovery*

We tested whether the delta model (i.e., the best-fitting model, see Results) was associated with acceptable parameter recovery. For each of the 104 participants, we used that participants’ fitted parameters and the set of eight sequences to simulate probability estimates. We then fitted the delta model to these simulated data using the same procedures for fitting as used for the participant data. Figure S2 shows the correlations between simulated and fitted parameters, which show highly accurate parameter recovery.

**Supplementary Figure S2.**

*Parameter recovery for delta model*

*A screenshot of a graph

Description automatically generated*

*Note:* Numbers and colour scale represent Pearson’s r computed between simulated and recovered free parameters.

*Experiment 3 Supplementary Methods: Participant Instructions*

Participants began the study with the following instructions:

“*You will read two scenarios about fictional crimes that are being processed in your local court, for which you have been selected to be part of the jury. After each scenario you will complete a task during which you will be able to see up to 10 witness statements. The witness statements will say either “Guilty” or “Innocent”. The ratio of statements may differ between the two tasks as the witnesses can only be 60% certain that they saw the defendant due to the circumstances surrounding the two crimes. After each statement, you can choose to see another witness statement by choosing ‘More information’. Or you can decide on a verdict by choosing ‘Verdict’ which will be followed by a screen where you will be able to choose either ‘Guilty’ or ‘Innocent’. The task will finish once a verdict has been made and you will move onto the next task*.”

The beginning of each sequence of witnesses began with a scenario:

“*You have been chosen to be part of a jury at your local court. The defendant is accused of burgling a home during the night. The street lights were off so visibility was minimal and witnesses found it difficult to make out the intruder’s face. Due to these circumstances the witnesses can only be 60% certain whether they saw the defendant*”.

*Experiment 3 Supplementary Results: Ideal observer analysis*

We calculated the number of draws to decision for the two sequences in Experiment 3 using an ideal observer. This model is well-documented (Averbeck, 2015; Furl & Averbeck, 2011; Hauser et al., 2018; Moutoussis et al., 2011). At sequence position *t* in state *s* (The current witness claim of guilty or innocent), one action *a* (guilty verdict, innocent verdict or sample again) will be chosen when it has a maximal action value *Q*(*a*;*s*). We used the same equation as in Experiment 1, derived from Bayes’ rule, to compute the probability of being guilty , or innocent , based on the current number of guilty claims , the total number of claims so far and the probability of the majority claim *q* = 0.6.

The action value for deciding in favour of guilt is and, similarly, the action value for deciding in favour of innocence is . We fixed cost of being wrong to = 1000 (as per participants’ instructions). The action value for sampling another witness claim incorporates the cost to sample = £10 (as per instructions) and requires a backwards induction algorithm to consider expected values of successive states on future samples. The ideal observer renders a correct verdict for GGIGIIGGIG after three samples (cost = £30). For GGGIIGGIGI, the model renders a correct verdict after eight samples (cost = £80). Unfortunately, the two unique data points of ideal observer performance precluded statistical comparisons with human participants. A visual comparison of participants with this ideal observer standard shows slight evidence for undersampling, consistent with our expectations from previous research (Furl & Averbeck, 2011), although ideal observer performance is within the 95% confidence intervals of participants’ performance.

*Experiment 3 Supplementary Results: Exploratory analyses*

We report exploratory correlations among *Draws to Decision*, *Accuracy*, *Court Costs*, *Proportion Guilty Verdicts* (i.e., the number of guilty verdicts out of the two sequences per participant), the four Likert scales participants answered after each sequence (*Confidence*, *Responsibility*, *Danger Others*, *Danger Self*) and the total score of the PDI (*PDItotal*). In Table 1, we report exploratory correlations among Draws to Decision, Accuracy, Court Costs, Proportion Guilty Verdicts, our Likert measures and PDItotal. Not surprisingly, Confidence is highly intercorrelated with the two performance measures Accuracy and Court Costs. We further found that Proportion Guilty Verdicts, participants’ ratings of the defendants’ responsibility for their behaviour, and participants’ ratings of the defendant’s danger to others all positively intercorrelated with each other. This danger to others variable was additionally positively correlated with participants’ ratings of the defendants’ danger to self.

**Supplementary Table 2**

