Supplemental Materials

Signal detection theory (SDT) analysis – Experiment 1

Another approach to investigate participants' awareness is to use the well-established framework of signal detection theory (SDT). Type 2 SDT provides valuable insights into participants' metacognitive sensitivity and wagering strategies by measuring sensitivity and bias independently (Clifford, Arabzadeh, & Harris, 2008; Fleming & Dolan, 2010; Higham, 2007; Kunimoto, Miller, & Pashler, 2001). In applying SDT to the IGT and post-decision wagering, a *hit* is a high wager after a good deck selection and a *false alarm* a high wager after a bad deck selection. A constant of 0.5 was added to the counts of hits, false alarms, misses and correct rejections in order to prevent infinite values for the calculation of *d'* (metacognitive sensitivity) and ln β (metacognitive bias).

A 2 (group [control, questionnaire]; between) × 10 (block: 10 trials each; within) mixed ANOVA was performed to investigate whether there were any differences in *d'* between the two conditions (Figure S1A). The analysis revealed no main effect of condition, $F(1, 28) < 1$, $MSE = 1.73$, $p = .90$. Figure 2 shows a tendency for *d'* to increase across blocks, resulting in a significant main effect of block, $F(9, 252) = 14.26$, $MSE = 0.60$, $p<.001$, $\eta_G^2 = 0.28$. As expected, the group × block interaction was not significant indicating that the questionnaire did not increase participants' metacognitive sensitivity, $F(9, 252) = < 1$, $MSE = 0.60$, $p = .94$.

Moreover, we can obtain useful insights into participants' wagering strategies by examining the bias measure ln β (ln β = 1 if no bias; ln β > 1 if conservative; ln β < 1 if liberal) (Higham, 2007; Macmillan & Creelman, 2005). Analysis of variance revealed that neither the main effect of condition, $F(1, 28) < 1$, $MSE = 1.00$, $p = .84$, nor the interaction (group × block), $F(9, 252) = 0.38$, $MSE = 0.29$, $p = .94$, were significant. There was a significant main effect of block, $F(9, 252) = 7.63$, $MSE = 0.29$, $p <.001$, $\eta_G^2 = 0.17$, as participants became more liberal across blocks (mean ln β ranged from 0.04 (block 1) to -0.58 (block 10) in the control group and from 0.26 (block 1) to -0.54 (block 10) in the questionnaire group).

SDT analysis – Experiment 2

A 2 (group [simple wagering, modified wagering]; between) × 10 (block: 10 trials each; within) mixed ANOVA was computed (see Figure S1B). The analysis revealed a

significant main effect of group, $F(1, 58) = 4.69$, $MSE = 2.43$, $p = .034$, $\eta^2_G = 0.02$, indicating that the modified pay-off matrix was more sensitive in assessing participants' task knowledge (simple wagering: $M = 0.20$, $SEM = 0.06$; modified wagering: $M = 0.47$, $SEM = 0.06$). Also, there was a significant effect of block, $F(6.81, 394.91) = 12.52$, $MSE = 0.79$, $p < .001$, $\eta^2_G = 0.14$. The interaction between block and group was significant, $F(6.81, 394.91) = 2.77$, $MSE = 0.79$, $p = .009$, $\eta^2_G = 0.03$. Simple effects analysis revealed significant differences between the two groups in blocks 4 and 5 (block 4: $F(1, 58) = 10.51$, $MSE = 0.74$, $p = .002$; block 5: $F(1, 58) = 9.13$, $MSE = 1.33$, $p = .004$), a pattern of results which resembles the differences found in advantageous wagering between the two groups.

We also investigated the mean bias ($\ln \beta$) in the two groups; in terms of loss aversion, we can ask whether the type of wagering matrix caused participants to develop a liberal or a conservative strategy about the wagers they placed. A 2 × 10 (group [simple wagering, modified wagering] × block) mixed ANOVA revealed that neither the group × block (main effect) interaction, $F(9, 522) = 0.96$, $MSE = 0.22$, $p = .47$, nor the main effect of group, $F(1, 58) = 1.60$, $MSE = 0.41$, $p = .21$, reached significance indicating that, in general, the different pay-off matrices did not affect participants' wagering strategy. However, there was a significant effect of block, $F(9, 522) = 2.53$, $MSE = 0.22$, $p = .008$, $\eta^2_G = 0.04$.

SDT analysis – Experiment 3

The confidence-accuracy relationship was examined using Type 2 SDT (see Figure S1C). The mean *d'* exceeded chance in block 5 for both the confidence ratings scales (2pts: $M = 0.45$, $t(39) = 2.80$, $p = .007$, 4pts: $M = 0.57$, $t(37) = 4.17$, $p < .001$) although it was not significantly above chance for the 2-point scale in block 6. In contrast, the mean *d'* for wagering was only marginally above chance in block 6 ($M = 0.34$, $t(39) = 1.99$, $p = .05$), and never reliably exceeded chance for the rest of the task. A 3 (group) × 10 (block) mixed ANOVA on the mean *d'* confirmed a significant main effect of group, $F(2, 115) = 7.08$, $MSE = 1.83$, $p = .001$, $\eta^2_G = 0.03$, due to significant differences between wagering and the confidence scales (M Wagering = 0.02, M 2pts = 0.31, M 4pts = 0.34) based on pairwise comparisons between wagering and confidence ratings using Tukey HSD, $p = .005$ (2pts) and $p = .002$ (4pts). There was a significant effect of block, $F(18, 1035) = 18.89$, $MSE = 0.78$, $p < .001$. The interaction between group and the block main effect, however, did not reach significance, $F(18, 1035) = 1.11$, $MSE = 0.78$, $p = .34$.

SDT analysis – Experiments 4A and 4B

Participants' confidence-accuracy levels in Experiment 4A were examined using Type 2 SDT. As shown in Figure S1D (circle markers), meta-cognitive sensitivity as measured by *d'* was significantly above chance (*d'* = 0) even in the first 10 trials indicating that even a few deck selections sufficed to acquire the advantageous strategy. In other words, participants were able to discriminate between good and bad decks and make an appropriate wager. Figure S1D shows a tendency for *d'* to increase across blocks, although the main effect of block was not significant $F(9, 180) = 1.57$, $MSE = 0.43$, $p = .13$.

Figure S1D (square markers) shows the mean *d'* across blocks in Experiment 4B, which exceeded chance on block 5, $M = 0.78$, $t(18) = 3.97$, $p < .001$. Also, there was a significant effect of block, $F(9, 162) = 8.47$, $MSE = 0.57$, $p < .001$, $\eta_G^2 = 0.28$, as metacognitive discrimination gradually increased over time.

References

Clifford, C. W. G., Arabzadeh, E., & Harris, J. A. (2008). Getting technical about awareness. *Trends in Cognitive Sciences, 12*, 54-58. doi:10.1016/j.tics.2007.11.009

Fleming, S. M., & Dolan, R. J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Consciousness and Cognition, 19*, 352-363. doi:10.1016/j.concog.2009.11.002

Higham, P. A. (2007). No special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General, 136*, 1-22. doi:10.1037/0096-3445.136.1.1

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition, 10*, 294-340. doi:10.1006/ccog.2000.0494

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

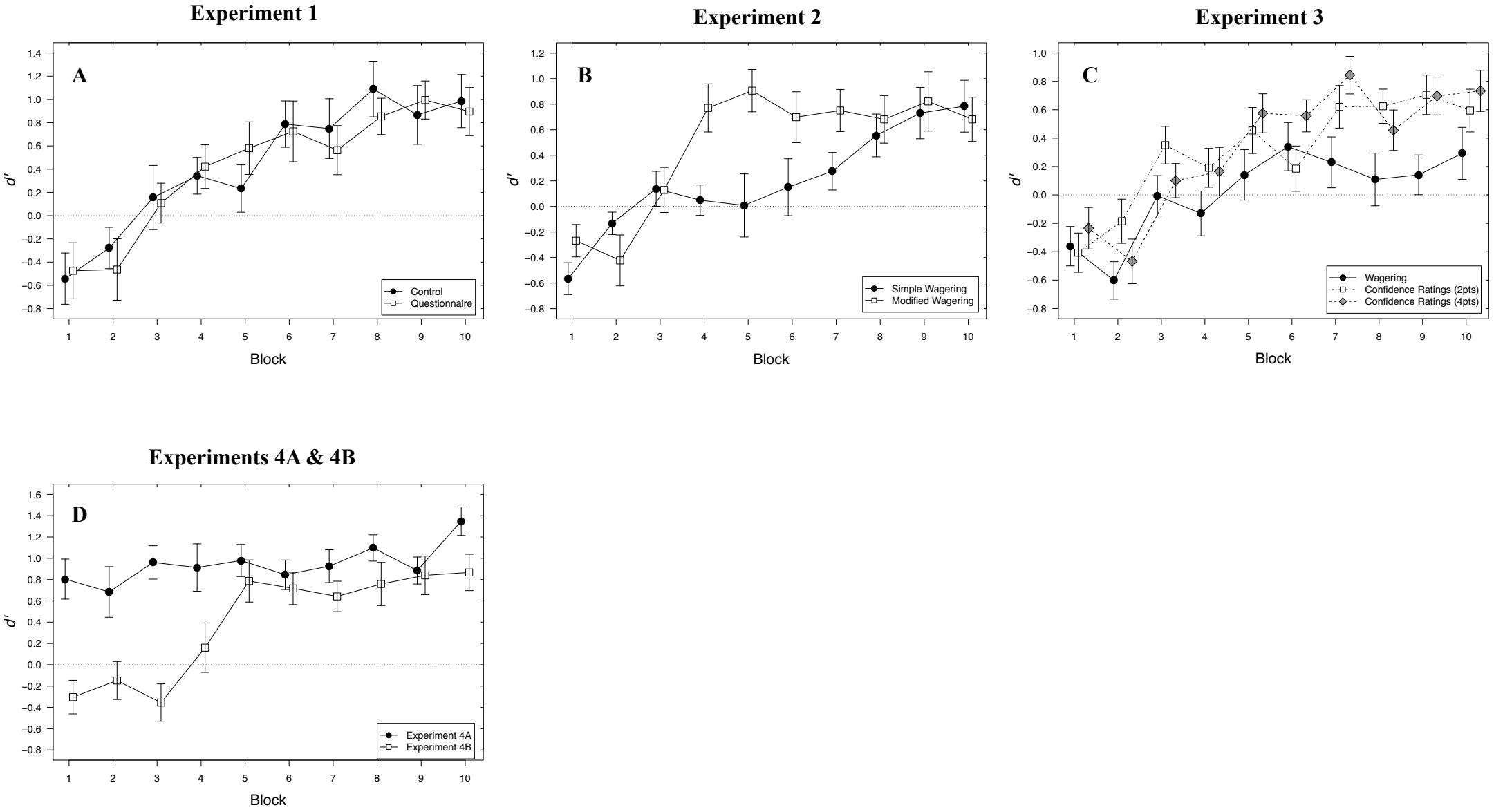

Figure S1. Participants' metacognitive sensitivity (mean *d'*) across experiments. Error bars represent ±1 SEM.