Supplementary Materials for 'Familiarity does not inhibit image-specific encoding of faces' by James D. Dunn, Kay L. Ritchie, Richard I. Kemp & David White

Corresponding author: david.white@unsw.edu.au

Supplementary analysis 1: Median split analysis by participant familiarity

In all experiments, participants included in the final analysis were familiar with at least 10 of the celebrities that were used as stimuli. However, there was considerable variability in the number of celebrities each participant was familiar with, which may have affected the strength of our familiarity manipulation. To investigate this, we performed a post-hoc 'median split' analysis that tested the effect of familiarity on image memory, separately for 'high familiarity' participants (who were familiar with more celebrities than average) and 'low familiarity participants' (familiar with less than average).

Full details of this analysis are reported below. For each experiment, data were reanalysed separately for high and low familiarity subsamples, using the same ANOVA model reported in the main paper. For brevity, only the critical comparisons of this analysis that involve familiarity are reported.

To summarise, across 5 experiments we did not observe a cost of familiarity in any of the image memory tasks, for either low or high familiarity participants. In Experiment 1, we did find a main effect of familiarity for the 'high familiarity' group but this was in the opposite direction to predicted. Based on this analysis, we conclude that differences in participants' familiarity with the celebrities cannot account for the lack of differences found between unfamiliar and familiar faces.

Experiment 1. Median familiarity with the 'familiar' celebrities was 75%. We reanalysed d-prime scores separately for the high familiarity group (n = 25, Mean familiarity = 90.1%, *SD = 7.1, min = 77.5%, max = 100%*) and the low familiarity group (n = 31, M = 58.9%, *SD = 14.2, min = 28%, max = 75%*). Six participants with the median familiarity score were assigned to low familiarity group.

For the high familiarity group, there was a significant main effect of Familiarity, F(1,

24) = 5.172, p = .032, η_p^2 = .177, but in to the opposite direction to predicted, with sensitivity being higher for familiar faces on this task than for unfamiliar faces. The interaction between Familiarity and Array Size was not significant, $F(1, 24) = 0.711$, p = .496, η_p^2 = .029. For the low familiarity group the main effect of Familiarity, $F(1, 30)$ = 0.089, p = .767, η_p^2 = .003, and interaction between factors, $F(1, 30) = 2.634$, p = .080, η_p^2 = .081, were not significant.

Experiment 2. Median familiarity with 'familiar' celebrities was 53.8%. Average familiarity for the high familiarity group was 72.9% (n = 28; *SD = 12.8, min = 55%, max = 100%*) and average familiarity for the low familiarity group was 41.5% ((n = 28; *SD = 6.4, min = 32.5%, max = 52.5%*).

For the high familiarity group, both the main effect of Familiarity, $F(1, 27) = 0.184$, p = .671, η_p^2 = .007, and interaction between factors, $F(1, 27) = 1.636$, p = .204, η_p^2 = .057, were not significant. For the low familiarity group, the main effect of Familiarity, $F(1, 27) = 0.918$, p = .346, η_p^2 = .033, and interaction between factors, $F(1, 27) = 0.313$, p = .732, η_p^2 = .011, were not significant.

Experiment 3. Median familiarity with 'familiar' celebrities was 68.8%. Average familiarity for the high familiarity group was 81.4% (n = 25, *SD = 8.1, min = 70%, max = 100%*) and average familiarity for the low familiarity group was 53.2% (n = 26, *SD = 9.7, min = 35%, max = 67.5%*).

For the high familiarity group, both the main effect of Familiarity, $F(1, 24) = 2.232$, p = .148, η_p^2 = .085, and interaction between factors, $F(1, 24) = 0.589$, p = .624, η_p^2 = .024, were not significant. For the low familiarity group, the main effect of Familiarity, $F(1, 25) = 0.417$, p = .524, η_p^2 = .016, and interaction between factors, $F(1, 25) = 0.625$, p = .601, η_p^2 = .024, were not significant.

Experiment 4. Median familiarity with 'familiar' celebrities was 67.5%. Average familiarity for the high familiarity group was 82% (n = 26, *SD = 7.8, min = 70%, max = 95%*) and average familiarity for the low familiarity group was 50.8% (n = 30, *SD = 10.8, min = 27.5%, max = 67.5%*).

For the high familiarity group, both the main effect of Familiarity, $F(1, 25) = 0.003$, $p = .956$, $\eta_p^2 = .000$, and interaction between factors, $F(1, 25) = 0.833$, $p = .370$, $\eta_p^2 = .032$, were not significant. For the low familiarity group, the main effect of Familiarity, $F(1, 29) = 0.836$, $p = .368$, $\eta_p^2 = .028$, and interaction between factors, $F(1, 29) = 2.594$, $p = .118$, $\eta_p^2 = .082$, were not significant.

Experiment 5. Unlike the previous experiments, familiarity with the image set used in this experiment was much higher, with a median familiarity of 100%. Consequently, the median split analysis instead compared the results of participants with 100% familiarity (n = 27) against participants with lower familiarity (n = 24). Average familiarity for the low familiarity group was 93.8% (*SD = 4.3, min = 81.3%, max = 96.9%).*

For the high familiarity group, the main effect of Familiarity was significant, $F(1, 26) = 68.977$, $p < .001$, $\eta_p^2 = .726$, with higher sensitivity for familiar faces than unfamiliar faces. However, there was also a significant interaction between Familiarity and Task Type, $F(1, 26) = 82.364$, $p < .001$, $\eta_p^2 = .760$. Follow up comparisons show significantly higher sensitivity for familiar faces than unfamiliar faces on the identity task, $t(26) = 10.84$, $p < .001$, Cohen's $d = 4.254$, but no significant differences between the familiar and unfamiliar faces on the image task, $t(26) = 1.195$, $p = .178$, Cohen's $d = 0.468$. The two-way interaction between Familiarity and Instruction Timing, $F(1, 26) = 0.199$, $p = .659$, $\eta_p^2 = .008$, and the three way interaction, $F(1, 26) = 1.061$, $p = .313$, $\eta_p^2 = .039$, were both not significant.

For the low familiarity group, we found the same pattern of results. There was a significant main effect of Familiarity, $F(1, 23) = 36.883$, $p < .001$, $\eta_p^2 = .616$, with higher sensitivity for familiar than unfamiliar faces. Moreover, the interaction between Familiarity and Task Type was also significant, $F(1, 23) = 26.177$, $p < .001$, $\eta_p^2 = .532$. Follow up comparisons again show significantly higher sensitivity for familiar faces than unfamiliar faces on the identity task, $t(23) = 7.592$, $p < .001$, Cohen's $d = 3.168$,

but no significant differences between the familiar and unfamiliar faces on the image task, $t(23) < 1$. The two-way interaction between Familiarity and Instruction Timing, $F(1, 23) = 0.025$, $p = .877$, $\eta_p^2 = .001$, and the three way interaction, $F(1, 23) = 0.003$, $p = .954$, $\eta_p^2 = .000$, were both not significant.

Supplementary analysis 2: Response latency analysis

The dependent variables in the main paper were sensitivity (d-prime) and response bias (criterion). This choice was motivated primarily by our research question: we were interested in the extent to which perceptual representations in short term memory contained image-specific details. The perceptual sensitivity was taken as a measure of the extent to which image-specific details has been stored in memory. Secondly, we chose this analysis to be consistent with the critical analysis in Armann et al. (2016), which our experiments were designed to follow up. The one exception to this was the visual search task reported in Experiment 4 where response latency is an important consideration (as explained in the main paper, targets were always present in these arrays and so accuracy had to be interpreted together with response latency).

In response to reviewers' comments, we also conducted a post-hoc analysis of participant's response latencies in Experiments 1, 2, 3, and 5 and report the results below. This analysis was conducted to rule out speed-accuracy trade-offs as an account of the lack of differences between unfamiliar and familiar faces in image memory tasks.

Full details of this analysis are reported below. For each experiment, log transformations were applied to the median response latency for participants' correct trials in each condition, and then analysed using the same repeated measures ANOVA's used to analyse sensitivity and criterion in the main manuscript. For brevity, only the critical comparisons of this analysis that involve familiarity are reported.

To summarise, across 5 experiments we found one significant difference between unfamiliar and familiar faces in response latency to image memory tasks. This difference was found in Experiment 1, where participants took significantly longer to decide whether duplicated of familiar faces were present compared with unfamiliar faces. Consequently, we cannot rule out that the findings this experiment could be explained by a speed-accuracy trade-off. However, in Experiments 2 to 5 (note that Experiment 4 analysis reported in manuscript) we found no differences in response time

between unfamiliar and familiar faces. Overall, this analysis provides very little support for the possibility that null effects of familiarity on sensitivity can be attributed to a speed accuracy trade-off.

Experiment 1. There was a significant main effect of Familiarity, $F(1, 55) = 7.980$, $p = .007$, $\eta_p^2 = .127$, with participants taking longer on average with familiar faces than unfamiliar faces. The interaction between factors was not significant, $F(2, 110) = 1.203$, $p = .304$, $\eta_p^2 = .021$.

Experiment 2. The main effect of Familiarity, $F(1, 55) = 1.555$, $p = .218$, $\eta_p^2 = .027$, and interaction between factors, $F(2, 110) = 2.177$, $p = .118$, $\eta_p^2 = .038$, were both not significant.

Experiment 3. The main effect of Familiarity, $F(1, 50) = 0.010$, $p = .921$, $\eta_p^2 = .000$, and interaction between factors, $F(3, 150) = 1.413$, $p = .241$, $\eta_p^2 = .027$, were both not significant.

Experiment 5. There was a significant main effect of Familiarity, $F(1, 50) = 9.631$, $p = .003$, $\eta_p^2 = .162$, with follow up comparisons showing slower responses for unfamiliar faces than familiar faces.

Critically, this effect was qualified by a significant interaction between Familiarity and Trial Type, $F(1, 50) = 22.611$, $p < .001$, $\eta_p^2 = .311$. Follow up comparisons show significantly slower responses for unfamiliar faces than familiar faces on the identity task, $t(50) = 4.846$, $p < .001$, Cohen's $d = 1.372$, but no significant differences in latency between familiar faces and unfamiliar faces on the image task, $t(50) = 0.737$, $p = .465$, Cohen's $d = 0.211$. The interaction between Familiarity and Instruction Timing, $F(1, 50) = 1.403$, $p = .242$, $\eta_p^2 = .027$, and three-way interaction between all factors, $F(1, 50) = 0.081$, $p = .777$, $\eta_p^2 = .002$, were not significant.