Supplementary Analyses

Absolute change in pupil size

Mean absolute change in pupil size was calculated in addition to the relative change in pupil size to determine the absolute difference in pupil size before and after the game. Mean absolute change in pupil size was calculated by subtracting *pre-game* pupil from *post-game* pupil and averaging across the trials for each game type, player type, and block order per participant. We conducted the same analyses as done in the main manuscript on absolute change in pupil size to determine if our results substantially differed if we used absolute change in pupil size instead of relative change in pupil size.

Experiment 1. There was no significant three-way interaction among block order, game type, and player type ($F(1, 91) = 1.388$, $p = .242$, $\eta^2 = .014$). There was no significant interaction between game type and player type ($F(1, 91) = .021$, $p = .889$, $\eta^2 < .001$) or between player type and block order ($F(1, 91) = 1.105$, $p = .296$, $\eta^2 = .012$).

However, there was a significant interaction between game type and block order ($F(1, 91) = 4.892$, $p = .029$, $\eta^2 = .050$). Following up this significant interaction indicated that there was a significant main effect of game type ($F(1, 46) = 5.919$, $p = .019$, $\eta^2 = .113$) in Block 1. This finding indicates that participants showed less decrease from *pre-game* to *post-game* after exclusion games ($M = -7.01$, $SD = 227.47$) than inclusion games ($M = -95.43$, $SD = 240.94$). In contrast, there was no main effect of game type ($F(1, 45) = .731$, $p = .397$, $\eta^2 = .016$) in Block 2. There were no significant interactions between game type and player type ($ps > .397$) or main effects of players type ($ps < .300$) in both Block 1 and Block 2. There was a marginal main effect of block order overall, $F(1, 91) = 3.603$, $p = .061$, $\eta^2 = .038$), suggesting that there was a marginally greater increase from *pre-game* to *post-game* pupil size in Block 2 (M = 30.01, SD = 257.73) than Block 1 (M = -51.22, SD = 237.27).

These results closely mirror our findings on the relative change in pupil dilation reported in the main manuscript. There is an effect of game type in Block 1 with less decrease in post-game pupil size to exclusive players than inclusive players but no effects in Block 2.

Experiment 2. There was a marginal main effect of game type ($F(1, 44) = 3.933$, $p = .054$, $\eta^2 = .080$) indicating that there was marginally less decrease in pupil size after exclusion games (M = -51.09, SD = 219.2) than inclusion games (M = -119.9, SD = 198.1). There was no significant interaction ($p = .299$) and no main effect of player type ($p = .198$). Again, this result shows the same trend as the findings from relative change in pupil dilation.

Experiment 3. There was a significant main effect of game type ($F(1, 76) = 3.989$, $p = .049$, $\eta^2 = .049$) indicating that there was less decrease in pupil size after exclusion games (M = -72.15, SD = 167.9) than inclusion games (M = -146.1, SD= 161.8). There was no interaction ($p = .273$) and no main effect of player type ($p = .786$). Thus, across three experiments, the patterns observed in absolute change in pupil dilation closely mirror our findings from the average relative change in pupil dilation.

Time course analyses

We additionally examined the time course of the relative change in pupil size from *pre-game* to *post-game* according to game type (exclusion vs. inclusion) and player type (computer vs. human). The time interval in which pupil sizes were collected was 3 seconds long and we divided this interval into three equal segments: Time point 1 (1 ms to 1000 ms), time point 2 (1001 ms to 2000 ms), and time point 3 (2001 ms to 3000 ms). Figure 2 in the main manuscript shows the time course of the relative change in pupil size in each experiment.

Experiment 1 Block 1. A mixed ANOVA model was conducted with game type (exclusion vs. inclusion) and time point (1, 2, and 3) as within-subject factors, player type (computer vs. human) as a between-subject factor, and relative change in pupil size as the dependent variable. There was a significant main effect of game ($F(1, 46) = 4.176$, $p = .047$, $\eta^2 = .081$) as expected from the main analyses. There was also a significant main effect of time ($F(2, 92) = 21.779$, $p < .001$, $\eta^2 = .313$). Bonferroni-corrected post hoc tests indicated that relative change in pupil size in time point 1 ($M = -.01$; $SD = .07$) was significantly more negative than time points 2 ($M = .03$; $SD = .07$) and 3 ($M = .04$; $SD = .07$), $ps < .001$. However, time points 2 and 3 did not significantly differ ($p = .155$). There were no other significant main effects or interactions ($ps > .087$). This pattern indicates that *post-game* pupil size tended to increase from time point 1 to time point 2.

Experiment 2. The same mixed ANOVA model was conducted as Experiment 1. There was a significant main effect of game ($F(1, 44) = 5.531$, $p = .023$, $\eta^2 = .025$) as expected from the main analyses. There was also a significant main effect of time, ($F(2, 88) = 11.442$, $p < .001$, $\eta^2 = .046$). Bonferroni-corrected post hoc tests indicated that relative change in pupil size in time point 1 ($M = -.03$; $SD = .07$) was significantly more negative compared to time points 2 ($M = -.003$; $SD = .05$) and 3 ($M = .001$; $SD = .07$), $ps < .001$. Relative change in pupil size increased from time point 2 to 3 ($p = .035$). There were no other significant effects or interactions ($ps > .218$). Again, *post-game* pupil size tended to increase from time point 1 to later time points. This pattern echoes the pattern found in Experiment 1.

Experiment 3. A mixed ANOVA model was conducted with time point (1, 2, and 3) as a within-subject factor, game type (exclusion vs. inclusion) and player type (computer vs. human) as between-subject factors, and relative change in pupil size as the dependent variable. There was a significant main effect of game ($F(1, 76) = 4.280$ $p = .042$, $\eta^2 = .053$) as expected from the main analyses. There was also a significant main effect of time ($F(2, 152) = 8.369$, $p < .001$, $\eta^2 = .098$). Bonferroni-corrected post hoc tests indicated that relative change in pupil size in time point 1 ($M = -.04$; $SD = .07$) was significantly more negative compared to time points 2 ($M = -.01$, $SD = .06$) and 3 ($M = -.003$; $SD = .07$), $ps = .002$, but time points 2 and 3 did not significantly differ ($p = 1.00$). There were no other significant effects or interactions ($ps > .452$). This pattern indicates again that *post-game* pupil size increased from time point 1 to time point 2.

Across the three experiments, we observe the same pattern in pupil dilation. *Post-game* pupil size increases from time point 1 to time point 2. This consistent increase from the first 1 second to the later time points are also observed in Bradley and Lang (2015), Snowden et al. (2016), and many other works; this pattern probably reflects a simple light reflex (i.e., initial pupil size constriction to the stimuli coming on screen). Bradley and Lang (2015) found that pupil dilation in response to affective differences (i.e., whether image is pleasant or unpleasant) are found at both the early time window of initial constriction and later time window (i.e., after the first 1 second of picture onset). Our results are in line with these findings.

Bayesian analyses

We conducted Bayesian analyses using JASP (JASP Team, 2016) and followed the guidelines from van Doorn et al (2019). A Bayesian analysis compares how likely different statistical models fit the observed data by providing a Bayes Factor (BF) for each statistical model. The BF_{01} values of each models can be compared to indicate evidence for how much more likely one model is than another model given the observed data. $BF_{inclusion}$ is a value generated from Bayesian model averaging. $BF_{inclusion}$ indicates how well the effect predicts the data by comparing the performance of all models that include the effect to the performance of all the models that do not include the effect. For BFs, values between .03 to .1 indicate strong evidence for H_0, values between .33 to .1 indicate moderate evidence for H_0, values in between 0.33 and 1 indicate weak to inconclusive evidence for H_0, values between 1 to 3 indicate weak to inconclusive evidence for H_1, values between 3 to 10 indicate moderate evidence for H_1 , and values between 10 to 30 indicate strong evidence for H_1. See OSF (https://osf.io/rnmpa/?view_only=3d3fd6a939b6458eaec33e063bb9da2f) for JASP files that contain the Bayesian analyses.

Experiment 1 Block 1. A Bayesian mixed ANOVA model was conducted with game type (exclusion vs. inclusion) as a within-subject factor, player type (computer vs. human) as a between-subject factor, and relative change in pupil size as the dependent variable. We used the default prior options for the effects (i.e., r = 0.5 for the fixed effects) as the main analysis. The model including game as the main effect was 2.486 more likely than the null model and 8.128 times more likely than the full model including the game type and player type interaction. $BF_{inclusion}$ for game type was 1.82 (weak to inconclusive evidence for the effect of game type), player type was .29 (moderate evidence for no effect of player type), and interaction of game type and player type was .26 (moderate evidence for no effect of interaction). To assess the robustness of the result, we also repeated the analysis for two different prior specifications (i.e., r = 1 and r = .2 for the fixed effects). Across all three prior specifications, the model with the main effect of game type was the best performing model compared to all other models (BF_{01} values ranged between 1.789 to 26.692), indicating that the main effect of game best explains the observed data.

Experiment 2. The same Bayesian mixed ANOVA model was conducted as in Experiment 1. With the prior specification of r = .5, the model including game as the main effect was 1.842 more likely than the null model and 4.496 times more likely than the full model including the game type and player type interaction. The $BF_{inclusion}$ for game type was 1.40 (weak to inconclusive evidence for the effect of game type), player type was .43 (weak to inconclusive evidence for no effect of player type), and interaction of game type and player type was .39 (weak to inconclusive evidence for no effect of interaction). Across all three prior specifications, the model with the main effect of game type was the best performing model compared to all other models (BF_{01} values ranged between 1.332 and 11.861), indicating that the main effect of game best explains the observed data.

Experiment 3. A Bayesian between-subject ANOVA model was conducted with game type (exclusion vs. inclusion) and player type (computer vs. human) as between-subject factors, and relative change in pupil size as the dependent variable. With the prior specification of r = .5, the model including game as the main effect was 1.620 more likely than the null model and 13.409 times more likely than the full model including the game type and player type interaction. The $BF_{inclusion}$ for game type was 1.15 (weak to inconclusive evidence for the effect of game type), player type was .2 (moderate evidence for no effect of player type), and interaction of game type and player type was .15 (moderate evidence for no effect of interaction). Across all three prior specifications, the model with the main effect of game type was the best performing model compared to all other models (BF_{01} values between 1.040 and 44.670), indicating that the main effect of game best explains the observed data.

References

JASP Team (2018). JASP (Version 0.10.0)[Computer software].

van Doorn, J., van den Bergh, D., Boehm, U., Dablander, F., Derks, K., Draws, T., Evans, N. J., Gronau, Q. F., Hinne, M., Kucharsky, S., Ly, A., Marsman, M., Matzke, D., Komarlu Narendra Gupta, A. R., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J.

(2019). The JASP guidelines for conducting and reporting a Bayesian analysis. Manuscript submitted for publication. Preprint available at https://psyarxiv.com/yqxfr.