Supplementary Material
Title: Changes in Pupil Size Track Self-Control Failure
Authors: Sean R. O'Bryan, Mindi M. Price, Jessica L. Alquist, Tyler Davis and Miranda Scolari

Corresponding author:
Sean R. O'Bryan
Department of Cognitive, Linguistic & Psychological Sciences, Brown University
sean_obryan@brown.edu

Index:

Participant demographics. Among the sample of 89 collected participants, reported age and racial identification were as follows: M_{age} = 19.57, SD_{age} = 1.93; 31.46% Hispanic or Latino, 62.92% Not Hispanic or Latino, 5.62% not reported; 70.79% White, 7.87% Black or African American, 3.37% Native American, 2.25% Asian, 1.12% Native Hawaiian or Other Pacific Islander, 4.49% more than once race, 10.11% unknown or not reported.

Ishihara color test. Because color is a critical feature in the Stroop task, participants completed a computerized version of the Ishihara color test (Ishihara, 1917) before the start of the experiment to ensure all had normal color vision. Each stimulus consisted of colored dots clustered together in the shape of a plate. Within the plate was a set of uniquely colored dots that formed a number. People with normal color vision can detect the number easily, whereas those with color deficiencies find it difficult or impossible. Participants viewed 5 dot plates and verbally indicated, without time pressure, the number they saw. All participants identified all numbers correctly, so none were excluded based on their color vision.

Self-reported effort and motivation. Participants completed a series of self-report measures related to their subjective experiences at the end of the experiment, including questions about their effort and motivation during each task. As described in the main text, participants assigned to the depletion group reported exerting more effort during the attention control video than those in the control group, suggesting that our manipulation was effective. Alternatively, when asked about the Stroop task (where instructions were identical for both groups), reports of subjective effort did not differ between the groups (depletion: $M = 4.7$, $SD = 1.7$; control: $M = 4.6$, $SD = 1.7$), $t(84) = 0.13$, $p = 0.90$, $d = 0.03$. Likewise, with respect to subjective motivation, participants assigned to the depletion condition reported being more motivated to perform well on the initial attention control task ($M = 5.2$, $SD = 2.0$) than those assigned to the control condition ($M = 4.2$, $SD = 1.7$), $t(85) = 2.71$, $p = .008$, $d = 0.58$. In contrast, both groups reported experiencing similar levels of subjective motivation during the Stroop task (depletion: $M = 5.3$, $SD = 1.8$; control: $M = 4.7$, $SD = 1.8$), $t(83) = 1.43$, $p = 0.16$, $d = 0.31$. This pattern of results is consistent with our observations of self-reported effort reported in the main text.

Correlations between self-report data, Stroop accuracy, and PD. As an exploratory analysis, we were interested in testing whether 1) any measures of self-reported effort or motivation collected following the experiment were reliably associated with participants' Stroop performance, and 2) whether these self-reports tracked either sustained (i.e., low-frequency) or task-evoked (i.e., high-frequency) PD during each respective task. The results of these correlation analyses, collapsed across groups, are detailed in Supplementary Table 2.

Additionally, we tested whether each of the reported associations differed significantly based on group assignment (depletion condition vs. control condition).

To summarize, first, the analyses suggest that self-reported effort and motivation across both task phases were generally weak predictors of Stroop accuracy on incongruent trials; only reported motivation to perform well on the Stroop task approached significance, $r = .19$, 95% CI [-.02, .39], $t(81) = 1.77$, $p = .08$. Interestingly, with respect to the associations between self-reports and each PD measure, we found that reported Stroop motivation – which did not differ between groups – was reliably associated with several PD measures. Specifically, we found that higher motivation to perform well on the Stroop task was negatively correlated with both phasic PD responses during the attention control video, $r = -.28$, 95% CI [-.47, -.07], $t(81) = -2.65$, $p = .009$, and tonic PD responses during the Stroop task itself, $r = -.28$, 95% CI [-.47, -.07], $t(81) = -2.65$, $p = .009$. These associations were of a similar magnitude among both groups. Accordingly, it is possible that participants whose attention was captured more frequently by the salient stimuli during the video phase, regardless of condition assignment, felt less motivated to engage with the subsequent Stroop task. Likewise, patterns of PD suggesting an exploratory state (high tonic PD) during the Stroop task may indicate states of lower subjective motivation among participants.

Finally, although correlations between self-report measures and other indices of effort/performance were generally well-matched between groups, we observed one relationship that was significantly modulated by assigned task demands: the association between Stroop motivation and sustained PD during the attention control video. Specifically, depletion group participants who had high sustained PD during the attention control task reported *less* motivation to perform well on the Stroop task, $r = -.30$, 95% CI [-.55, .01], $t(41) = -2.00$, $p = .05$, while those assigned to the control group tended to report *more* motivation, $r = .18$, 95% CI [-.14, .47], t(38) = 1.14, $p = .26$, and this interaction was significant, $F(1, 79) = 5.02$, $p = .03$. Thus, additional demands for sustained attention faced by the depletion group may have led to less subjective motivation during the later Stroop task, while more sustained attention among controls who were not faced with this requirement appears to have a weak positive effect (if any) on later reports of motivation. We emphasize that because these analyses are exploratory, future research should seek to further clarify the associations between subjective report data, PD, and behavioral indices of ego-depletion.

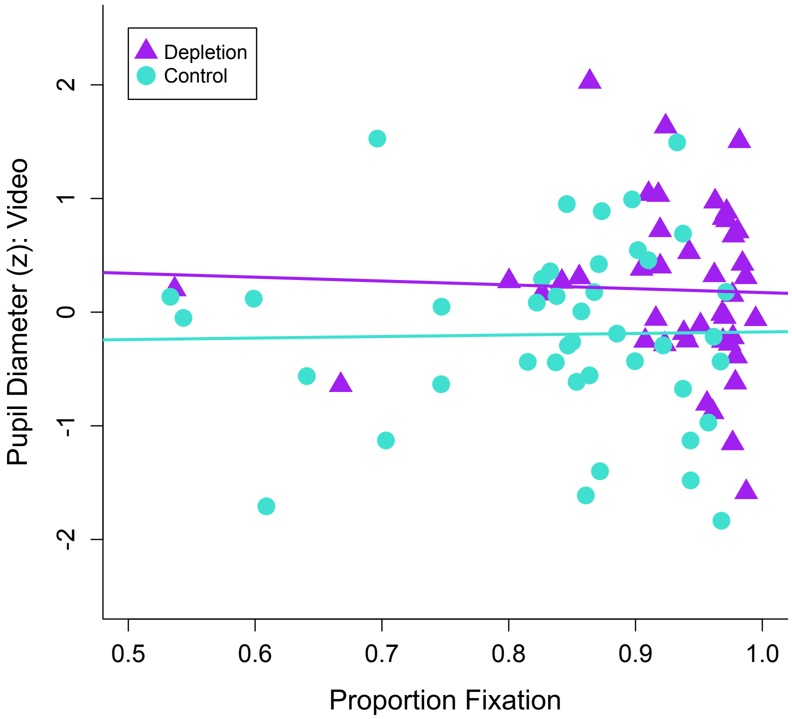
Stroop task: Speed-accuracy tradeoff. As reported in the main text, across groups, accuracy was significantly higher on congruent trials than incongruent trials (classic Stroop effect). We confirmed this behavioral pattern did not reflect a speed-accuracy tradeoff by computing subject-level Pearson correlations between mean accuracy and log-transformed RT for correct trials (Ratcliff, 1993). In direct opposition to a speed-accuracy trade-off, where higher accuracy is achieved by executing slower responses, the mean correlation coefficients for both groups were negative and significant, (depletion: $r = -0.20$, 95% CI [-0.31, -0.09], t(43) = -3.73, $p < .001$; control: $r = -0.15$, 95% CI [-0.25, -0.05], t(44) = -3.12, $p = .003$). This suggests that higher accuracy was associated with faster RTs on average across participants.

References

Ishihara, S. (1917). *Tests for Color Blindness*. Handaya, Tokyo: Hongo Harukicho.

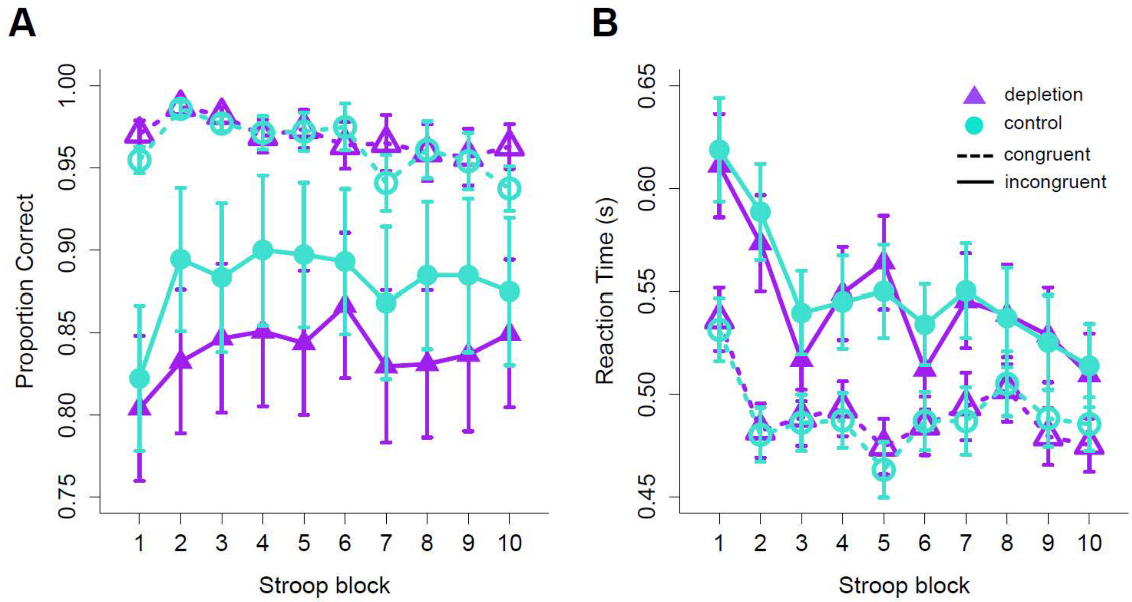
Ratcliff, R. (1993). Methods of dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.

Fig. S1



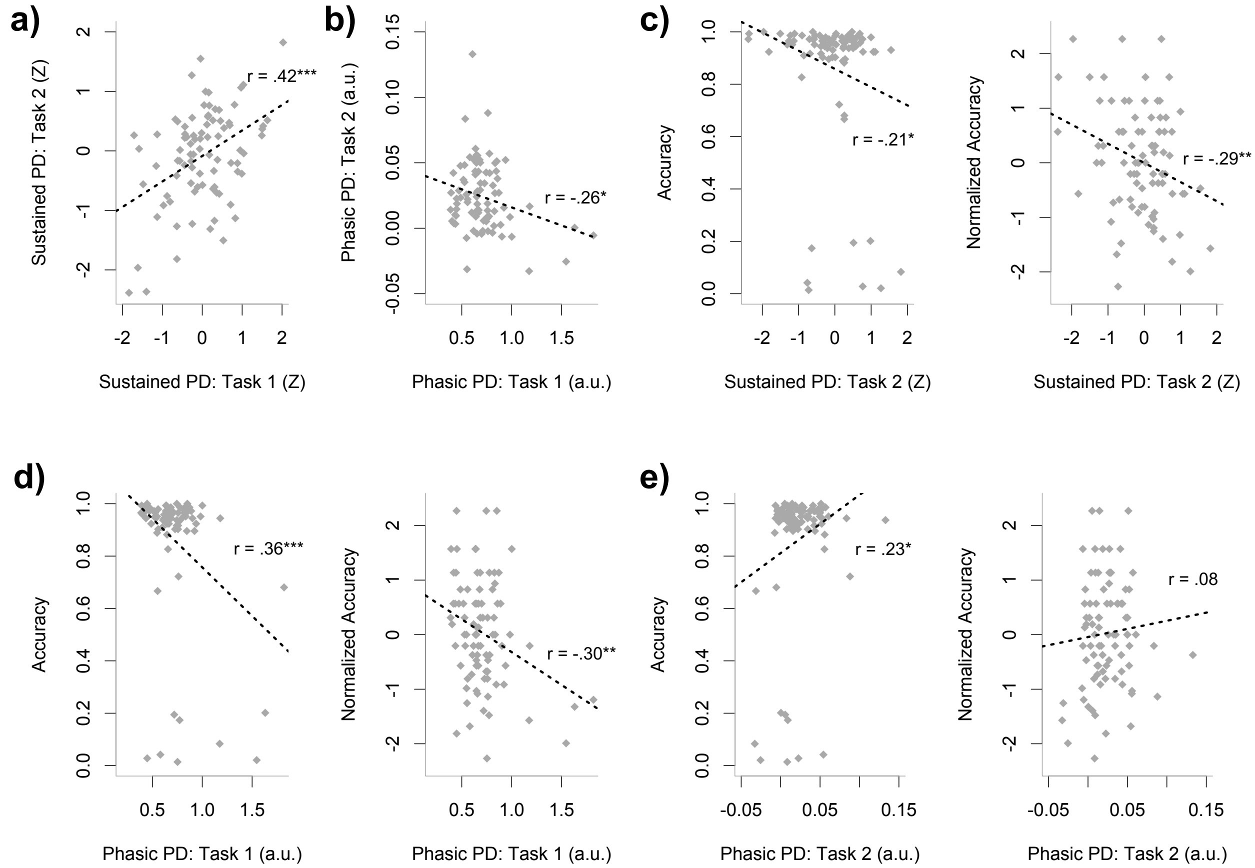
Supplementary Figure 1. Correlations between sustained pupil diameter collected during the video task and the proportion of time spent maintaining fixation within the interest area for the depletion (purple triangles) and control (turquoise circles) groups.

Fig. S2



Supplementary Figure 2. Behavioral results. Mean (A) accuracy and (B) reaction time for the depletion (purple triangles) and control (turquoise circles) groups over 10 Stroop task blocks. Dashed lines/open markers correspond to congruent trials; solid lines/filled markers correspond to incongruent trials. Error bars depict between-subject SEM.

Fig. S3.



Supplementary Figure 3. Bivariate correlations between PD measures, raw accuracy, and normalized accuracy. Normalized accuracy was computed by transforming accuracy values to percentiles and multiplying this uniform distribution by the inverse normal CDF. Note that all analyses in the main text were computed using raw accuracy. (A) Correlation between sustained PD during the attention control task and tonic PD during Stroop (path d_{21}). (B) Correlation between phasic PD responses during the attention control task and phasic PD evoked for incongruent vs. congruent trials during Stroop. (C) Correlations between tonic PD during Stroop, raw accuracy, and normalized accuracy. (D) Correlations between phasic PD during the attention control task, raw accuracy, and normalized accuracy. (E) Correlations between phasic PD during the Stroop task, raw accuracy, and normalized accuracy. * = $p < .05$, ** = $p < .01$, *** = $p <.001$.

Table S1

Supplementary Table 1. Parameter estimates and 95% bootstrapped confidence intervals for all effects in the serial mediation model.

Model Path	**Estimate**	**95% Confidence Interval**
a_1	**0.40**	**[0.07, 0.72] ***
a_2	-0.09	[-0.39, 0.21]
d_{21}	**0.54**	**[0.20, 0.90] ***
b_1	0.03	[-0.05, 0.12]
b_2	**-0.19**	**[-0.36, -0.03] ***
Indirect Effect 1 ($a_1 * b_1$)	0.01	[-0.02, 0.06]
Indirect Effect 2 ($a_2 * b_2$)	0.01	[-0.04, 0.10]
Serial Indirect Effect ($a_1 * d_{21} * b_2$)	**-0.03**	**[-0.13, -0.001] ***
Total Indirect Effect	-0.01	[-0.10, 0.07]
Direct Effect (c')	-0.03	[-0.14, 0.07]
Total Effect	-0.04	[-0.18, 0.09]

Table S2

Supplementary Table 2. Correlations between self-report measures, Stroop accuracy, and PD. -S = sustained pupil diameter -P = phasic pupil diameter ** = $p < .01$, uncorrected.

	Attn. Task Effort	**Attn. Task Motivation**	**Stroop Effort**	**Stroop Motivation**
Attn. Task PD - S	.05	-.05	.06	-.02
Attn. Task PD - P	-.18	-.17	-.14	**-.28****
Stroop PD - S	-.17	-.17	-.01	**-.28****
Stroop PD - P	.02	.05	.01	-.16
Stroop Accuracy	.10	.10	-.05	.19