Mind-wandering impedes response inhibition by affecting the triggering of the inhibitory process: Supplemental Materials

Supplemental Methods

Detailed study design

The task was designed using JavaScript, HTML, CSS, and PHP. The data and scripts were housed on the web server of the Department of Psychology, UCSD. The study had 6 sections and took ~1 hour to complete:

1) Sound test: Participants were asked to set the volume at 40% of their max system volume and maintain it throughout the study. Then they were presented with a sequence of 3-9 tones which they counted and reported. This ensured that the volume was set at an appropriate level.
2) Go practice: Participants then performed 20 Go trials. Each trial began with a black fixation square for 500±50 ms. After this, black left and right arrows were presented and they had to press the left and right arrow keys with their right hand, as quickly and accurately as they could, within a 1000 ms response deadline. To satisfy the criteria of this section, they had to achieve, *a*) a mean go reaction time (RT) of 300-600 ms, and, *b*) a go accuracy > 90%. They got 3 attempts for this. After that they were prompted to either exit the study and receive partial payment for their time, or keep trying to satisfy the criteria.
3) Stop practice: Participants then performed 40 trials of the auditory stop signal task (*Fig. 2A*). Each trial began with a fixation for 500±50 ms, followed by a left or right arrow, and they had to respond quickly with the corresponding left/right key press within a 1000 ms response deadline. In 25% trials, after a delay, called Stop Signal Delay (SSD), a tone was presented and, in these trials, they had to try to stop their response. The SSDs were tracked such that it reduced by 50 ms after a failed stop response, and increased by 50 ms after a successful stop response, so as to achieve successful stopping in roughly 50% of trials. The duration of each trial including the inter trial interval was 2500 ms. To satisfy the criteria of this section, they had to achieve, *a*) a mean go RT of 300-600 ms, *b*) a go accuracy > 90%, and *c*) a successful stop % of 25-75%. They got 3 attempts for this. After that they were prompted to either exit the study and receive partial payment for their time, or keep trying to satisfy the criteria.

4) Stop practice with mental-state probe: Same as the stop practice except that after every 16-22 trials (40-55 s), participants received a mental-state probe: "In the previous trial, was your mind: 1. Focused on the task, 2. Wandering or thinking of something else, 3. Blank or unfocused", where they responded by pressing the 1, 2, or 3 keys with their left hand (*Fig. 2B*). These corresponded to On-Task, MW, and Off-focus mental-states, respectively. They performed 45 trials, which included 2 such probes. To satisfy the criteria of this section, they had to achieve, *a*) a mean go RT of 300-600 ms, *b*) a go accuracy > 90%, and *c*) a successful stop of 25-75%. They got 3 attempts for this. After that they were prompted to either exit the study and receive partial payment for their time, or keep trying to satisfy the criteria.
5) Experiment: Same as stop practice with mental-state probe, except that there were 640 trials in 8 blocks (32 mental-state probes; 160 stop trials). Participants were given the option of taking breaks in between blocks.
6) Questionnaire: Participants completed the Mind-wandering questionnaire (Mrazek et al., 2013) and the Barrett Impulsivity Scale (BIS-11) (Stanford et al., 2009).

In study 1, 47 participants started the study, 15 stopped the study midway, 32 completed it, and 30 were selected for the analyses but only 11 met the number-of-MW-episodes criterion. In study 2, 356 participants started the study, 149 stopped the study midway, 207 completed the study, and 145 were selected for the analyses but only 40 met the-number-of-MW-episodes criterion.

Supplemental Results

Study 2: Off-focus episodes

There were 11 participants who reported >5 instances of Off-focus. We started by testing the effect of Off-focus on SSRT and mean go RT in these participants. $SSRT_{OFF}$ (345±17 ms [CI 307, 383 ms]) was significantly less than $SSRT_{ON}$ (315±12 ms [CI 288, 342 ms]; $t(10) = 2.6$, $p = 0.028$, $d = 0.8$, $BF_{10} = 2.7$, *Fig. S1A*). This was observed in the context of no significant difference in mean go RT between Off-focus (488±11 ms [CI 464, 513 ms]) and On-Task episodes (495±15 ms [CI 461, 529 ms]; $t(10) = 0.8$, $p = 0.417$, $d = 0.3$, $BF_{10} = 0.4$*; Fig. S1B*).

We then compared the BEESTS estimates. A paired signed rank test revealed no significant difference in the BEESTS estimate of mean SSRT between Off-focus episodes (383±34 ms [CI 307, 460 ms]) than On-Task (313±20 ms [CI 269 357 ms]) but Bayesian analysis suggested that there was weak evidence in favor of the alternate hypothesis ($z(11) = 1.8$, $p = 0.075$, $r = 0.5$, $BF_{10} = 1.3$; *Fig. S1C*). Further, the SD of the SSRT distribution was greater during Off-focus episodes (285±23 ms [CI 233, 337 ms]) compared to On-Task (208±30 ms [142, 274 ms]; $t(10) = 2.5$, $p = 0.032$, $d = 0.8$, $BF_{10} = 2.4$). Thus, stopping became slower and more variable in Off-focus episodes compared to On-Task. Further, there were greater pTF in the Off-focus episodes (25±3 % [CI 19, 32 %]) compared to On-Task (17±2 % [CI 12, 22 %]; $z(11) = 2.1$, $p = 0.033$, $r = 0.6$, $BF_{10} = 3.1$; *Fig. S1D*). The change in TF between Off-focus and On-Task episodes was well correlated with the change in mean SSRT ($r_p = 0.88$ [CI 0.47, 0.97], $p < 0.001$, $BF_{10} = 95.5$). Also, there was also a moderate but non-significant correlation between the change in pTF and the change in the SD of the SSRT distribution ($r_p = 0.87$ [CI 0.44, 0.96], $p < 0.001$, $BF_{10} = 65.8$). Thus, taken together, these results demonstrate that stopping becomes slower and more variable during Off-focus state and that this is mostly driven by the detrimental effect of Off-focus state on the trigger stage.

Since the impact on MW and Off-focus states (compared to On-Task) on stopping performance was largely similar, we contrasted the behavior between MW and Off-Focus episodes. Unfortunately, there were very few participants who had both >5 episodes of MW and Off-focus so we could not perform a paired comparison. Instead, we compared the performance in the MW episodes in the MW-participants to that in the Off-focus episodes in the Off-focus-participants (i.e. unpaired comparison). There was no significant difference for any of the following comparisons (*Fig. S1E-H*): 1) SSRT (MW: 332±10 ms [CI 312, 352 ms]; Off-focus:

345±17 ms [CI 307, 383 ms]; $t(49) = 0.6$, $p = 0.537$, $d = 0.2$, $BF_{10} = 0.4$; *Fig. S1E*); 2) mean go RT (MW: 531±14 ms [CI 503, 558 ms]; Off-focus: 495±15 ms [CI 461, 529 ms]; $t(49) = 1.3$, $p = 0.197$, $d = 0.4$, $BF_{10} = 0.6$; *Fig. S1F*); 3) the BEESTS estimate of mean SSRT (Off-focus: 383±34 ms [CI 307, 460 ms]; MW: 339±15 ms [CI 310, 369 ms]; $z = 1.1$, $p = 0.277$, $r = 0.2$, $BF_{10} = 0.6$; *Fig. S1G*); 4) TF% (Off-focus: 25±3 % [CI 19, 32 ms]; MW: 21±1 % [CI 18, 24 ms]; $z = 1.2$, $p = 0.221$, $r = 0.2$, $BF_{10} = 0.8$; *Fig. S1H*). Thus, behaviorally there did not seem to be a difference between the MW and Off-focus episodes, albeit the metrics of action-stopping tended to be worse in the Off-focus episodes and the percentage of Off-focus reports were far fewer than MW reports (*Table 2*).

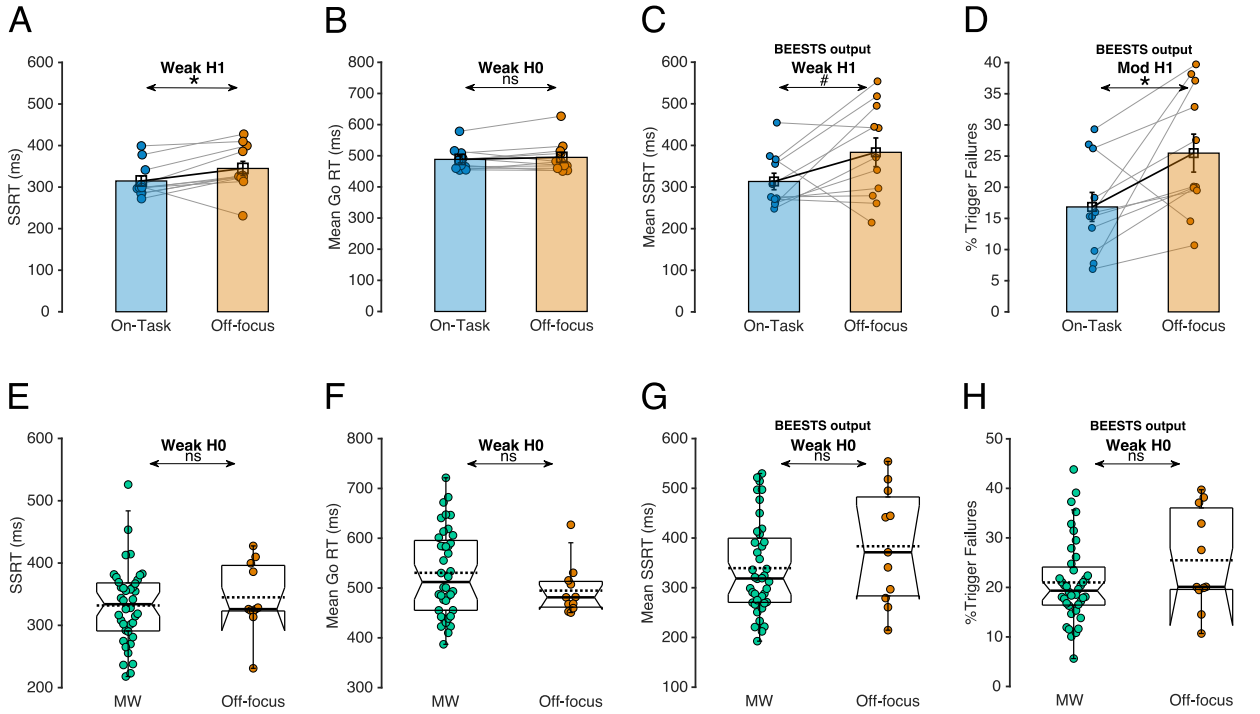

Figure S1 | Behavior and BEESTS estimates in the Off-focus episodes. ***A.*** *Comparison of behavioral estimate of SSRT during the On-Task (blue) and Off-focus (orange) episodes in Off-focus-participants. Other details same as Fig 4B.* ***B.*** *Same as A but for mean go RT.* ***C.*** *Same as A but for the mean SSRT estimated from BEESTS.* ***D.*** *Same as A but for the % trigger failures estimated from BEESTS.* ***E.*** *Beeswarm plot of the behavioral estimate of SSRT between the MW-participants in the MW episodes (green) and the Off-focus-participants in the Off-focus episodes (orange). Each dot represents a participant, overlaid by the box-plot. The notch represents the 95% confidence interval of the median, and the dotted line represents the mean. F. Same as E but*

*for mean go RT. G. Same as E but for the mean SSRT estimated from BEESTS. H. Same as E but for % trigger failures estimated from BEESTS. p>0.1: ns, p<=0.1 & p>0.05: #, p<=0.05 & p>0.01: *. BF_{10}>0.3 & BF_{10}<=1: Weak H0, BF_{10}>1 & BF_{10}<=3: Weak H1, BF_{10}>3 & BF_{10}<=10: Mod H1.*

Supplemental Tables

Table S1*: Behavioral responses for the MW-participants in the MW and On-Task episodes in study 1 and 2 (Values represent mean±s.e.m.)*

Parameters	**Study 1**		**Study 2**	
Group	**On-Task**	**MW**	**On-Task**	**MW**
Mean Correct Go RT	514±21 ms	502±21 ms	522±13 ms	531±14 ms
Mean Failed Stop RT	475±18 ms	464±20 ms	474±11 ms	488±11 ms
Correct Go	99±0 %	98±0 %	99±0 %	98±0 %
Error Go	1±0 %	1±0 %	1±0 %	2±0 %
Successful Stop	57±3 %	38±3 %	50±2 %	45±2 %
SSRT	290±18 ms	335±19 ms	306±9 ms	332±10 ms
Mean SSD	194±17 ms	183±18 ms	202±14 ms	199±14 ms

Supplemental Figures

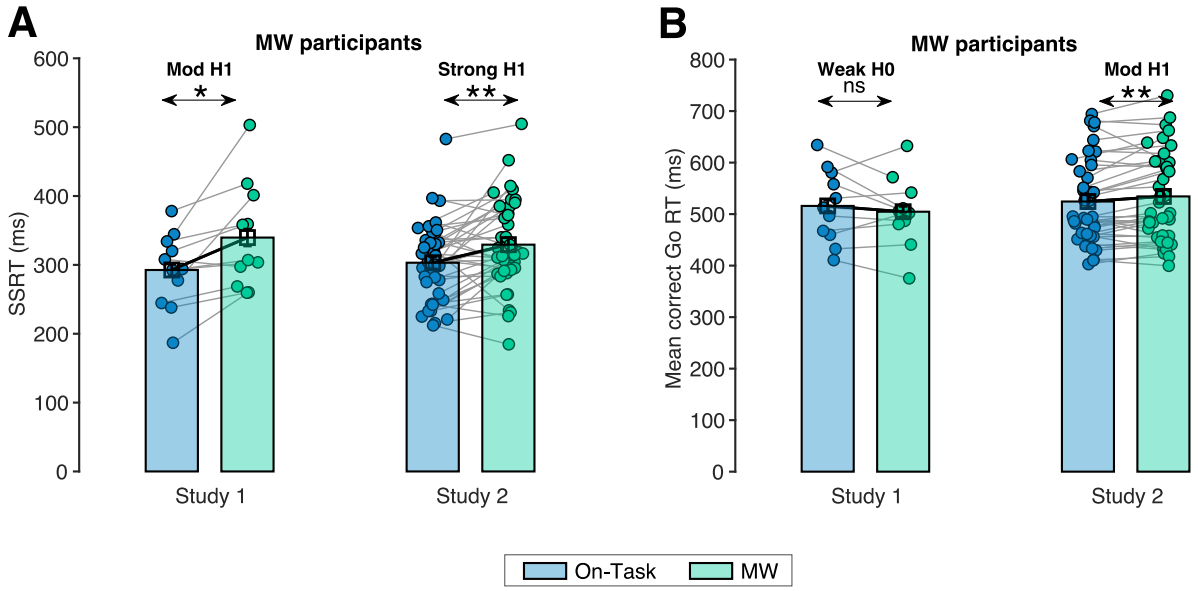

Figure S2 | Behavior when 10 s (4 trials) prior to the probe report is used instead of 15 s (6 trials). This shows that the effect of slower SSRT during MW remained irrespective of the episode duration used. ***A.*** *Behavioral estimate of SSRT in the MW participants in the On-Task (blue) and MW (green) episodes in Study 1 and 2. Each dot represents a participant and the bar and crosshairs represent the mean±s.e.m. across the sample.* $SSRT_{MW}$ *was significantly slower than* $SSRT_{ON}$ *in both studies (Study 1:* $SSRT_{ON}$ *= 291±16 ms [CI 255, 327 ms],* $SSRT_{MW}$ *= 340±23 ms [CI 288, 391 ms]; t(10) = 3.1, p = 0.012, d = 0.9,* BF_{10} *= 5.2; Study 2:* $SSRT_{ON}$ *= 304±9 ms [CI 286, 322 ms],* $SSRT_{MW}$ *= 331±10 ms [CI 310, 352 ms]; t(39) = 3.5, p = 0.001, d = 0.6,* BF_{10} *= 26.4).* ***B****. Same as A but for mean go RT. Mean Go RT was not significantly different during MW compared to On-Task in study 1 but was significantly slower in study 2 (Study 1: Go* RT_{ON} *= 519±21 ms [CI 472, 565 ms], Go* RT_{MW} *= 507±20 ms [CI 461, 552 ms]; t(10) = 095, p = 0.408, d = 0.3,* BF_{10} *= 0.4; Study 2: Go* RT_{ON} *= 524±13 ms [CI 498, 551 ms], Go* RT_{MW} *= 535±14 ms [CI 507, 564 ms]; z(40) = 2.6, p = 0.008, r = 0.4,* BF_{10} *= 5.0). p>0.1: ns, p<=0.05 & p>0.01: *, p<=0.01 & p>0.001: **.* BF_{10}*>=0.3 &* BF_{10}*<1: Weak H0,* BF_{10}*>1 &* BF_{10}*<=3: Weak H1,* BF_{10}*>3 &* BF_{10}*<=10: Mod H1,* BF_{10}*>10: Strong H1.*

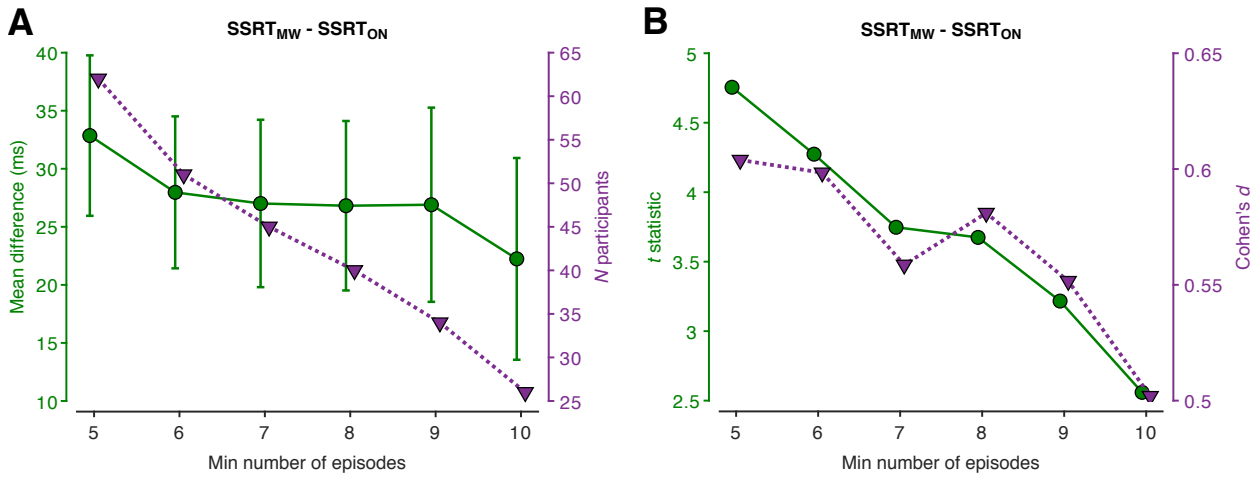

Fig. S3 | Difference between behavioral $SSRT_{MW}$ and $SSRT_{ON}$ for different number of MW episodes. This shows that the effect of slower SSRT during MW remained irrespective of the number-of-MW-episodes cutoff used. ***A.*** *(Left) Mean difference between behavioral SSRT between the MW episodes and On-Task episodes for different (minimum number of MW episodes) cutoffs, pooled across both studies. The dot and cross-hairs represent the mean±s.e.m. of the difference. (Right) Number of participants (purple triangle) selected when that MW episode cutoff is used.* ***B.*** *(Left) Same as A but for the t-statistic of the difference between $SSRT_{MW}$ and $SSRT_{ON}$. The p-value was significant for all the values. (Right) Same as A but for Cohen's d.*

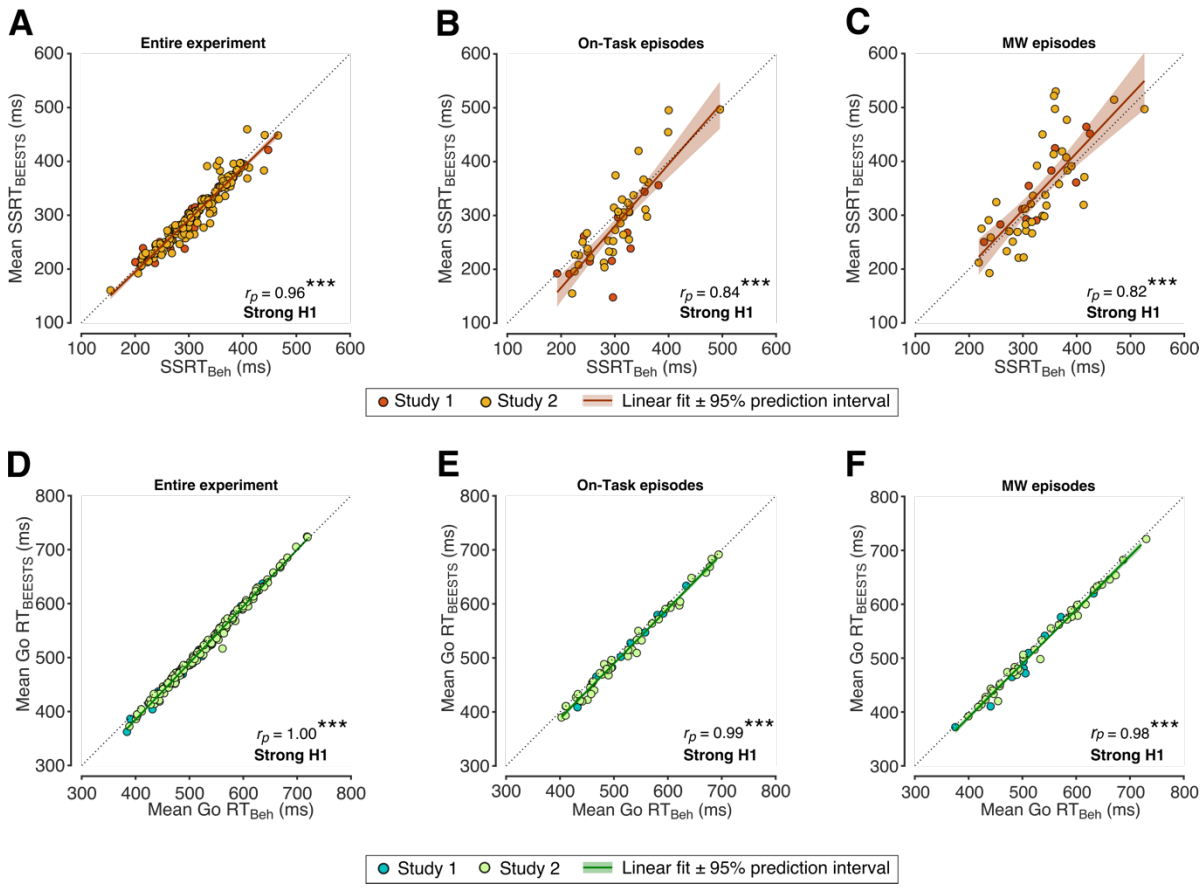

Fig. S4 | Correlation between the behavioral and BEESTS estimate of mean SSRT and mean go RT. This shows reliable BEESTS estimation as there was a strong correlation between behavior and BEESTS estimates. ***A.*** *Correlation between SSRT estimate from the entire experiment across all participants (Study 1 – red, Study 2 - yellow) (r_p = 0.96 [CI 0.94, 0.97], p < 0.001, BF_{10} > 100). Each dot represents a subject while the line and shaded region represents the linear regression fit and the 95% confidence interval, respectively.* ***B.*** *Same as A but for the mean SSRT in the On-Task episodes estimated from the behavior and BEESTS in the MW-participants (r_p = 0.83 [CI 0.71, 0.90], p < 0.001, BF_{10} > 100).* ***C.*** *Same as A but for mean SSRT in the MW episodes estimated from the behavior and BEESTS in the MW-participants (r_p = 0.76 [CI 0.60, 0.85], p < 0.001, BF_{10} > 100).* ***D.*** *Same as A but for mean correct Go RT from the entire experiment across all participants (Study 1 – dark green, Study 2 – light green) (r_p = 1.00 [CI 1.00, 1.00], p < 0.001, BF_{10} > 100).* ***E.*** *Same as B but for mean correct Go RT in the On-Task episodes (r_p = 1.00 [CI 0.99, 1.00], p < 0.001, BF_{10} > 100).* ***F.*** *Same as C but for mean correct Go RT in the MW episodes (r_p = 0.99 [CI 0.99, 1.00], p < 0.001, BF_{10} > 100). Thus, the BEESTS estimates were well correlated with that observed from behavior. p<=0.001: ***. BF_{10}>10: Strong H1.*

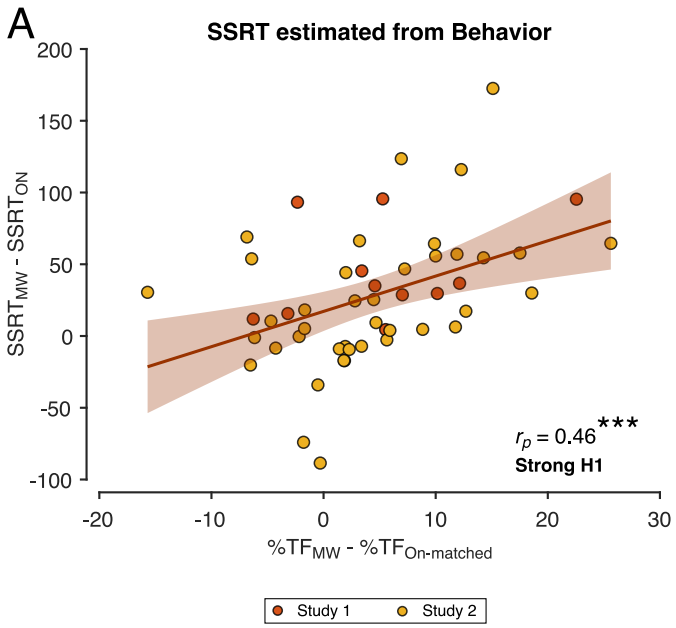

Fig. S5 | Correlation between the change in %TF and change in SSRT estimated from behavior. This shows that the correlation between the change in pTF and the change in the BEESTS estimate of SSRT between the MW and On-matched episodes is unlikely to be a simulation artifact as the change in pTF was also correlated to the change in the behavioral estimate of SSRT. A. *Each dot represents a participant (Study 1 – red, Study 2 – yellow). The line and shaded region represent the linear regression and its 95% confidence interval. There is a significant positive correlation between the change in %TF (MW – On-matched) and the change in* $SSRT_{Beh}$ *(MW – On-Task), replicated that seen in the BEESTS results (*$r_p = 0.43$ *[CI 0.17, 0.62],* $p = 0.002$*,* $BF_{10} = 20.5$*). p<=0.001: ***.* BF_{10}*>10: Strong H1.*

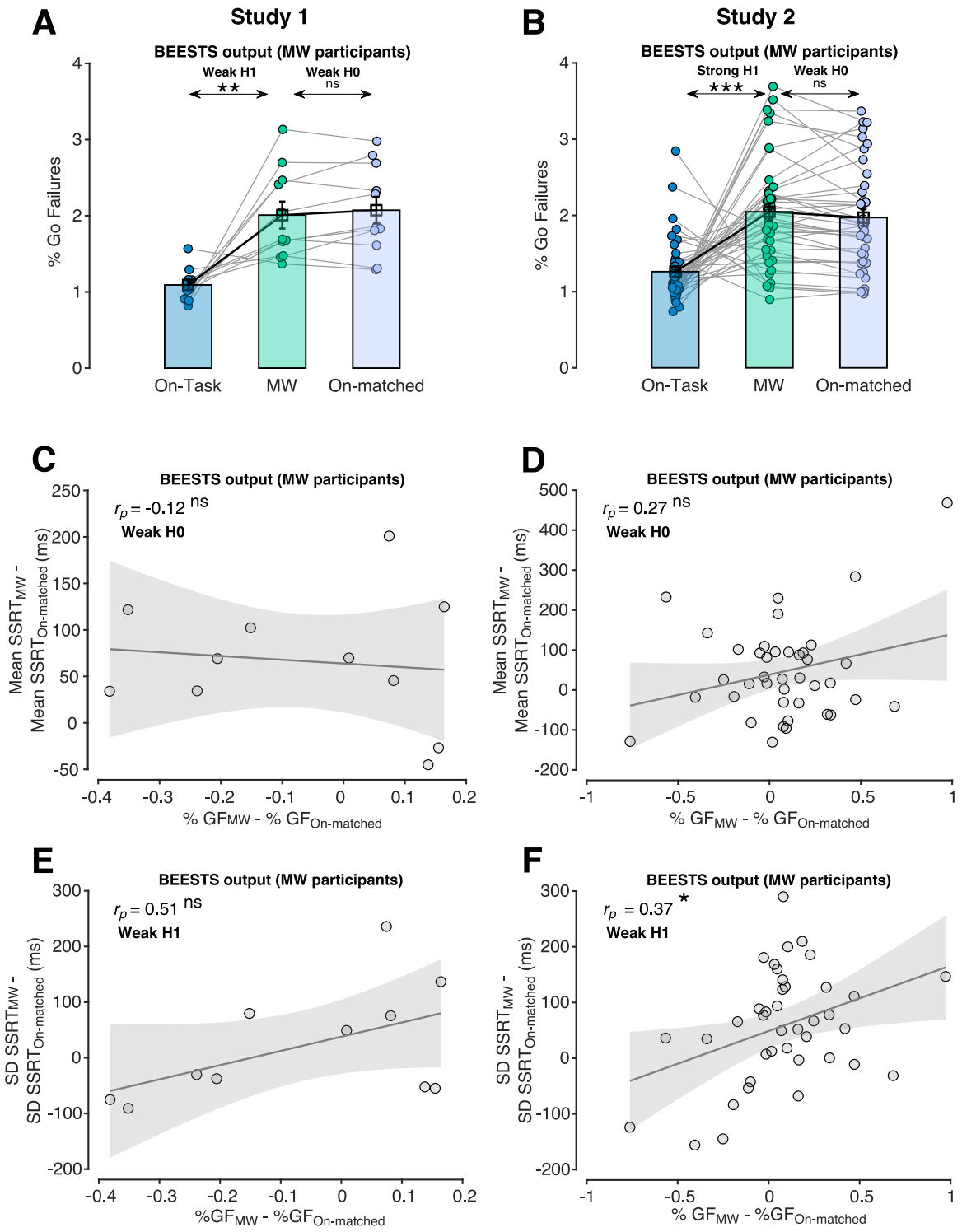

Fig. S6 | Comparison of percentage go failure estimates among different mental-states. This shows that go failures were less affected by mental-states than were trigger failures. A. Study 1. *A 1-way repeated measures ANOVA with episode (On-Task, MW, On-matched) as the independent variable and percentage of go failures (pGF) as the dependent variable revealed that there was a*

significant effect of episode ($F(2,20) = 16.2$, $p = 0.002$, $\eta_p^2 = 0.62$, $BF_{10} > 100$). Post-hoc multiple comparisons revealed that pGF was significantly greater in MW episodes (2.0±0.2 % [CI 1.6, 2.4 %]) compared to that in On-Task episodes (1.1±0.1 % [CI 1.0, 1.2 %]; $t(10) = 4.0$, $p_{BH} = 0.005$, $BF_{10} = 18.5$). However, there was no significant difference in pGF between MW and On-matched episodes (2.1±0.2 % [CI 1.7, 2.5 %]; $t(10) = 1.0$, $p_{BH} = 0.327$, $BF_{10} = 0.5$). ***B. Study 2.*** *A Friedman test with episode (On-Task, MW, On-matched) as the independent variable and pGF as the dependent variable revealed that the factor episode had a significant effect ($\chi^2(2,20) = 27.2$, $p < 0.001$, Kendall's $W = 0.2$, $BF_M > 100$). Post-hoc multiple comparisons revealed that pGF was significantly greater in MW episodes (2.0±0.2 % [CI 1.8, 2.3 %]) compared to that in On-Task episodes (1.3±0.1 % [CI 1.0, 1.4 %]; $z(40) = 4.1$, $p_{BH} < 0.001$, $BF_{10} > 100$) but not significantly different from that in On-matched episodes (2.0±0.2 % [CI 1.8, 2.2 %]; $t(39) = 1.6$, $p_{BH} = 0.129$, $BF_{10} = 0.5$).* ***C. Study 1.*** *The correlation between the change in mean SSRT between MW and On-matched episodes and the change in pGF between MW and On-matched episodes was not significant ($r_p = -0.12$ [CI -0.60, 0.46], $p = 0.728$, $BF_{10} = 0.4$) (the line and shaded region represents the linear fit and the 95% confidence interval).* ***D. Study 2.*** *The correlation between the change in mean SSRT between MW and On-matched episodes and the change in pGF between MW and On-matched episodes was not significant ($r_p = 0.27$ [CI -0.05, 0.52], $p = 0.097$, $BF_{10} = 0.7$).* ***E. Study 1.*** *The correlation between the change in the SD of the SSRT distribution between MW and On-matched episodes and the change in pGF between MW and On-matched episodes was not significant using frequentist approach but there was a weak evidence for the alternate hypothesis using Bayesian approach ($r_p = 0.52$ [CI -0.13, 0.81], $p = 0.105$, $BF_{10} = 1.2$).* ***F. Study 2.*** *There was a weak but significant correlation between the change in the SD of the SSRT distribution between MW and On-matched episodes and the change in pGF between MW and On-matched episodes was not significant ($r_p = 0.37$ [CI 0.06, 0.59], $p = 0.021$, $BF_{10} = 2.6$). p>=0.1: ns, p<=0.05 & p>0.01: *, p<0.01 & p<=0.001: **, p<0.001: ***. BF_{10}>=0.3 & BF_{10}<1: Weak H0, BF_{10}>1 & BF_{10}<=3: Weak H1, BF_{10}>10: Strong H1.*

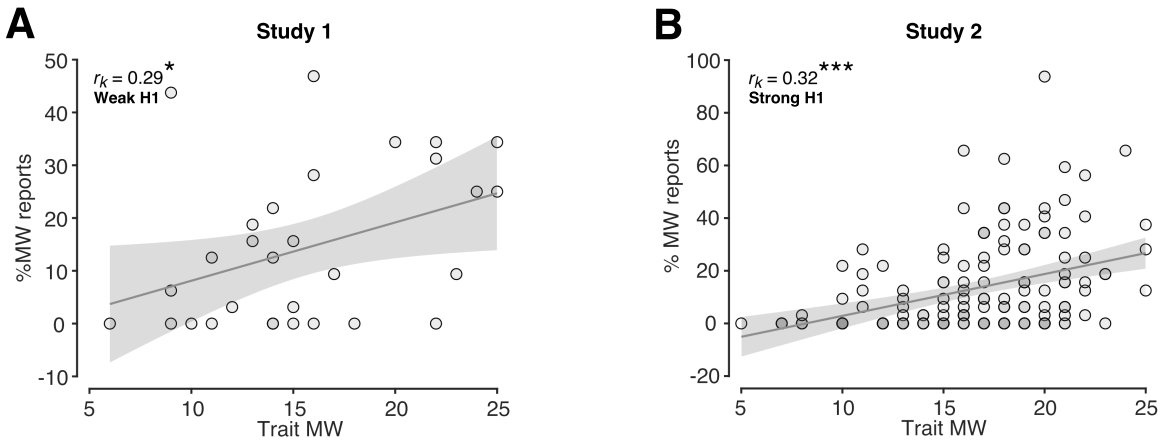

Fig. S7 | Correlation between state-level and trait mind-wandering. This shows that participants reported their mental-state accurately and that the probe reports could be trusted as there was a reliable correlation between state-level and trait mind-wandering across participants. ***A.*** *Across-participants correlation between trait MW based on the MW questionnaire and % MW reports during the entire experiment in Study 1 (r_k = 0.29 [CI 0.03, 0.49], p = 0.037, BF_{10} = 2.5). Each dot represents a subject. The line represents a linear fit and the shaded region represents the 95% confidence interval.* ***B.*** *Same as A but for Study 2 (r_k = 0.32 [CI 0.21, 0.42], p < 0.001, BF_{10} > 100). p<=0.05 & p>0.01: *, p<=0.001: ***. BF_{10}>1 & BF_{10}<=3: Weak H1, BF_{10}>10: Strong H1.*

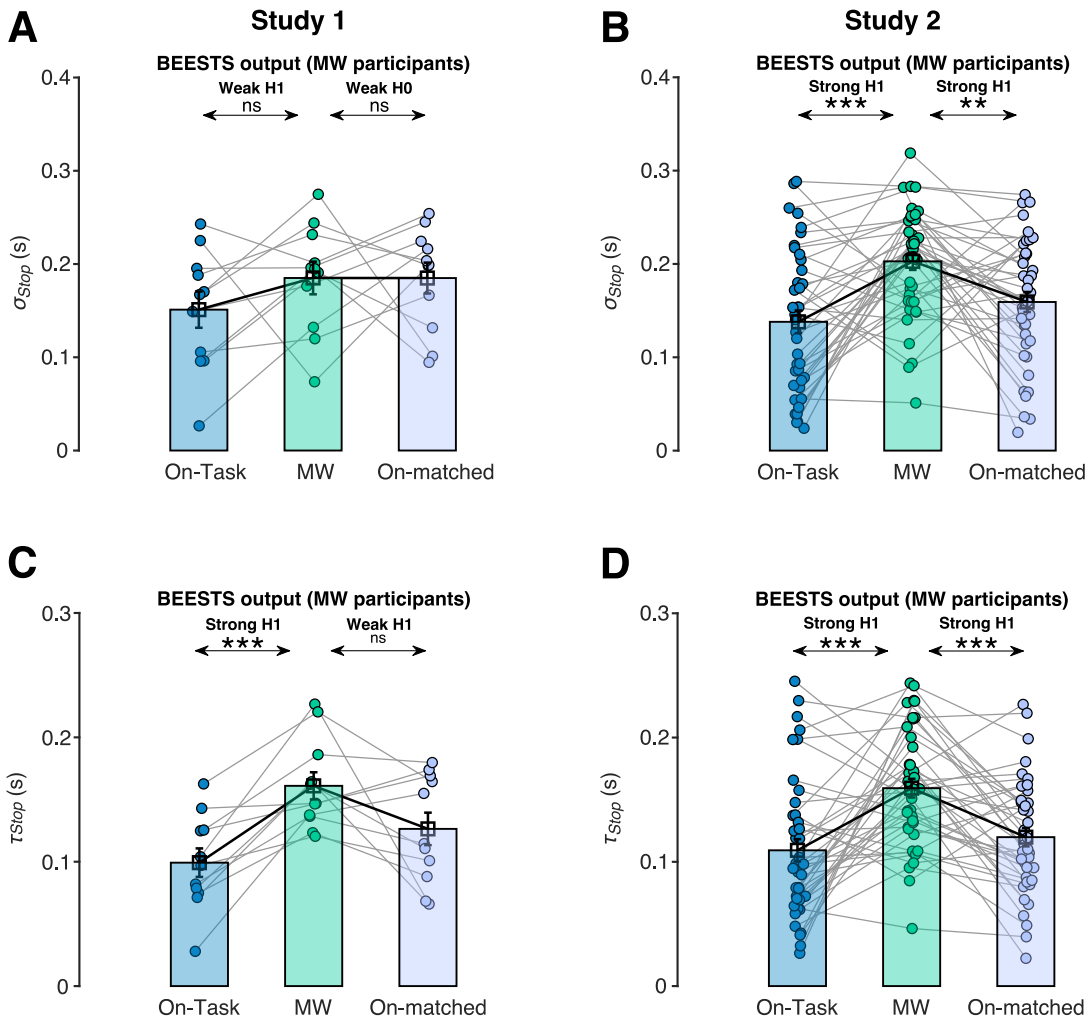

Figure S8 | Comparison of BEESTS parameter estimates among different mental-states. This shows that both the σ_{Stop} and τ_{Stop} parameters increased during MW compared to both On-Task and On-matched episodes. ***A.*** *σ_{Stop} in the On-task (deep blue), MW (green), and On-matched (light blue) episodes. Each dot represents a participant while the bar and cross-hairs represent the mean±s.e.m. ($F(2,20) = 1.4$, $p = 0.276$, $\eta_p^2 = 0.12$, $BF_M = 1.8$; On: 0.15±0.01 s [CI 0.11, 0.19 s], MW: 0.19±0.01 s [CI 0.15, 0.22 s], On-matched: 0.19±0.01 s [CI 0.15, 0.22 s]. On vs. MW: $t(10) = 1.4$, $p_{BH} = 0.375$, $d = 0.4$, $BF_{10} = 0.7$; MW vs. On-matched: $t(10) = 0.02$, $p_{BH} = 0.986$, $d = 0.0$, $BF_{10} = 0.3$).* ***B.*** *Same as A but for study 2 ($F(2,20) = 12.5$, $p < 0.001$, $\eta_p^2 = 0.2$, $BF_M > 100$; On: 0.14±0.01 s [CI 0.11 0.16 s], MW: 0.20±0.01 s [CI 0.18 0.22 s], On-matched: 0.16±0.01 s [CI 0.15, 0.18 s]. On vs. MW: $t(39) = 4.4$, $p_{BH} < 0.001$, $d = 0.7$, $BF_{10} > 100$; MW vs. On-matched: $t(39) = 4.7$, $p_{BH} = 0.004$, $d = 0.7$, $BF_{10} > 100$).* ***C.*** *Same as A but for τ_{Stop} in study 1 ($F(2,20) =$*

10.0, $p < 0.001$, $\eta_p^2 = 0.5$, $BF_M = 69.2$; On: 0.10±0.01 s [CI 0.07, 0.13 s], MW: 0.16±0.01 s [CI 0.14, 0.19 s], On-matched: 0.13±0.01 s [CI 0.10, 0.16 s]. On vs. MW: $t(10) = 5.5$, $p_{BH} < 0.001$, $d = 1.6$, $BF_{10} > 100$; MW vs. On-matched: $t(10) = 2.0$, $p_{BH} = 0.077$, $d = 0.6$, $BF_{10} = 1.4$). ***D.*** *Same as C but for study 2 ($\chi^2(2,20) = 13.9$, $p < 0.001$, $BF_M = 31.4$, Kendall's $W = 0.5$; $BF_M > 100$; On: 0.11±0.08 s [CI 0.09, 0.13 s], MW: 0.16±0.01 s [CI 0.14, 0.18 s], On-matched: 0.12±0.01 s [CI 0.11, 0.14 s]. On vs. MW: $W(40) = 139$, $z(40) = 3.6$, $p_{BH} = 0.001$; $r = 0.6$, $BF_{10} > 100$; MW vs. On-matched: $t(39) = 3.2$, $p_{BH} = 0.004$, $d = 0.5$, $BF_{10} = 14.0$). p>=0.1: ns, p<0.01 & p<=0.001: **, p<=0.001: ***. BF_{10}>=0.3 & BF_{10}<1: Weak H0, BF_{10}>1 & BF_{10}<=3: Weak H1, BF_{10}>10: Strong H1.*

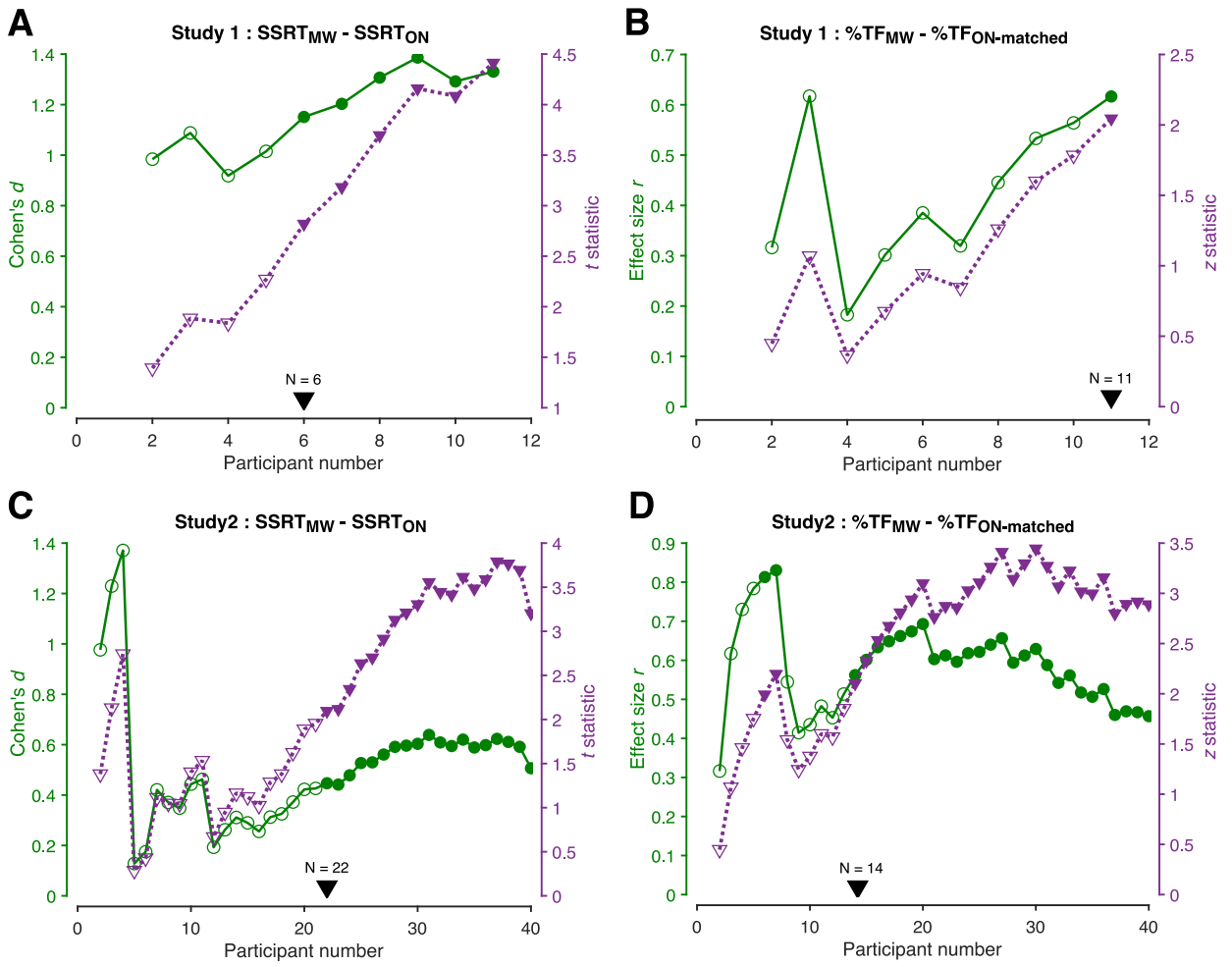

Fig. S9 | Reliability of the main results. This shows that the effect of slower SSRT and greater pTF during MW was reliable as effect sizes and test statistic remained significant once sufficient number of participants were added. ***A.*** *(Left) Effect size of the difference between the behavioral estimate of SSRT in the MW and On-Task episodes (green line and circles) as a function of participant number in study 1. Filled circles represent significant difference in SSRT. The black triangle at the bottom represents the subject from which the p-values become consistently significant. (Right) The t-statistic of the difference (purple dotted line and triangles). Filled triangles represent significant difference in SSRT.* ***B.*** *Same as A but for the difference between %TF between the MW and On-matched episodes. Other details same as A.* ***C.*** *Same as A but for study 2.* ***D.*** *Same as B but for study 2. Note that the p-value becomes significant for N = 6 and 7 but becomes non-significant again and then becomes consistently significant from N = 14 onwards.*

References

Mrazek MD, Phillips DT, Franklin MS, Broadway JM, Schooler JW. 2013. Young and restless: validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Front Psychol* **4**:560. doi:10.3389/fpsyg.2013.00560

Stanford MS, Mathias CW, Dougherty DM, Lake SL, Anderson NE, Patton JH. 2009. Fifty years of the Barratt Impulsiveness Scale: An update and review. *Pers Individ Dif* **47**:385–395. doi:10.1016/j.paid.2009.04.008