Thought dynamics under task demands: Evaluating the influence of task difficulty on mind wandering and thought constraint

Nicholaus P. Brosowsky[1], Jonathan W. Schooler[2], & Paul Seli[1]

[1] Department of Psychology and Neuroscience, Duke University, Durham, USA

[1] Department of Psychological and Brain Sciences, University of California Santa Barbara, Santa Barbara, CA, USA

Author note

Correspondence concerning this article should be addressed to Nicholaus P. Brosowsky, Duke University, 417 Chapel Dr, Durham, NC 27708, USA. E-mail: nicholaus.brosowsky@duke.edu

Abstract

As research on mind wandering has accelerated, the construct's defining features have expanded and researchers have begun to appreciate different varieties of mind wandering. Recently, Christoff and colleagues argued for the importance of investigating a hitherto neglected variety of mind wandering: unconstrained thought. To date, with only a handful of studies investigating unconstrained thought, little is known about this intriguing type of mind wandering. Across two experiments, we examined, for the first time, whether changes in task difficulty influence rates of constrained versus unconstrained thoughts. In both experiments, participants completed either an easy (0-back) or hard (2-back) task and responded to intermittently presented thought probes that gauged thought constraint throughout the task. In Experiment 1, we found that participants completing the easy task engaged in unconstrained thoughts significantly more frequently than those completing the difficult task (~20% difference), and that, during both tasks, rates of unconstrained thoughts increased with time on task. In Experiment 2, we found that rates of unconstrained thought were manipulated independently of task relatedness. Finally, we found that unconstrained thought was not associated with a trait-level assessment of flow, or 'deep effortless concentration' but in exploratory analyses, such thoughts were found to be associated with age and intelligence. All data, analysis, manuscript, and experiment code can be found at
Keywords: mind wandering, freely moving thoughts, task-difficulty

Thought dynamics under task demands: Evaluating the influence of task difficulty on mind wandering and thought constraint

Over the past fifteen years, 'mind wandering' has become a prominent topic of interest (e.g., Christoff et al., 2016; Esterman et al., 2012; Mills et al., 2015; Schooler, 2004; Schooler et al., 2011; Wammes et al., 2016). Whereas researchers of mind wandering have tended to focus their investigations on task-unrelated thoughts, more recently, Christoff and colleagues have argued for the importance of studying a different, hitherto neglected variety of mind wandering: unconstrained thought (Christoff et al., 2016). Over the span of just four years, the seminal paper in which Christoff et al. first argued for the importance of unconstrained thought has become one of the most widely cited articles in the mind-wandering literature, with over 450 citations already[1]. While the field has undoubtedly been receptive to Christoff et al.'s argument, to date, little is actually known about unconstrained thought. Here, to shed some much needed light on this newly identified variety of mind wandering, we sought to determine whether unconstrained thought varies as a function of task difficulty.

From a methodological perspective, it is important to understand whether (and how) we can experimentally manipulate the various dimensions of mind wandering. Indeed, task-difficulty manipulations have been frequently employed in research on 'task-unrelated thought' (TUT)—the most commonly studied variety of mind wandering—which is defined as thought that is unrelated to one's ongoing task. Experiments manipulating TUT have contributed significantly to the study of mind wandering as they have allowed researchers to systematically push around TUT rates and observe the costs and benefits of doing so. For instance, task-difficulty manipulations have been used to examine the effect of different rates of TUT on creativity (Baird et al., 2012), reading comprehension (Forrin, Risko, & Smilek, 2019), and driving (Yanko & Spalek, 2013). To date, however, it is unclear whether rates of unconstrained thought are amenable to experimental manipulations; yet, in the same way that the literature on TUT has benefitted greatly from manipulations of TUT, the growing literature on unconstrained thought will surely benefit from the identification of a method that can be used to effectively manipulate thought constraint.

From a theoretical perspective, it is also important to understand the influence of task difficulty on thought constraint. Indeed, doing so could allow researchers to generate numerous testable hypotheses that could help to elucidate the nature of unconstrained thoughts. For instance, one prominent view in the mind-wandering literature is that engaging in TUT requires attentional resources (Smallwood & Schooler, 2006) and, consequently, task-difficulty effects are explained in terms of the availability of resources: difficult tasks require more attentional resources, leaving few resources available for the maintenance of TUTs. Considering thought constraint from this perspective, it seems plausible that the act of constraining our thoughts—as opposed to permitting them to roam free—is an effortful, resource-demanding process (Hypothesis A: Constraining thoughts is effortful and resource-demanding). On this view, people ought to engage in fewer constrained (and more unconstrained) thoughts when resources are in low supply. Such a prediction could be empirically tested by varying task demands (and hence, the resources required to perform those tasks) and indexing rates of (un)constrained thinking

[1] Google Scholar, 05/21/2020.

during task completion: Given that easy tasks require relatively fewer resources than hard tasks, this view would predict higher rates of constrained thoughts (or lower rates of unconstrained thoughts) during an easy versus a hard task.

Alternatively, it also seems plausible that failing to constrain our thoughts (i.e., allowing them to roam) may consume available resources (Hypothesis B: Unconstraint as effortful, exploratory thought). Indeed, to move from one thought to another, loosely associated thought might require executive resources to generate, evaluate, and select and/or reject thoughts (Beaty, Benedek Silvia, & Schacter, 2016). It is likewise possible that disengaging from constrained thoughts might be more resource-demanding than maintaining them. For instance, ruminative and obsessive thoughts are highly constrained, yet, to disengage from those thoughts to attend to a focal task, or to allow one's thoughts to move freely, would likely be effortful and resource demanding. On these bases, Hypothesis B would predict the opposite outcome of Hypothesis A: that is, we should observe lower rates of constrained thoughts (or higher rates of unconstrained thoughts) during an easy versus a hard task.

Finally, it is also possible that constraint is not influenced by task difficulty manipulations (Hypothesis C: constraint/unconstraint is automatic and resource-free). Some have argued that constraint might operate both in a controlled and automatic manner (e.g., Christoff, et al., 2016). If automatic, then it is possible that determining whether we constrain our thoughts or let them roam free is not dependent on the availability of resources (e.g., De Houwer et al.). As such, manipulations reducing the amount of resources available should have no influence on rates of thought constraint.

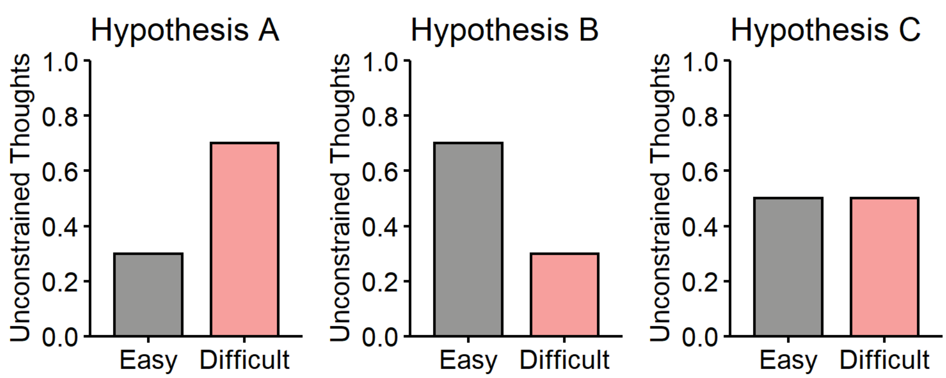

Figure 1

Predicted proportion of unconstrained thoughts as a function task difficulty. According to Hypothesis A, the act of constraining our thoughts is effortful and resource-demanding and predicts less constrained and more unconstrained thought when resources are reduced (i.e., in a difficult task). According to Hypothesis B, unconstrained thought is an effortful, exploratory process, and predicts less unconstrained and more constrained thought when resources are reduced (i.e., in a difficult task). Finally, according to Hypothesis C, constraint/unconstraint is automatic and resource-free and predicts no systematic change across task difficulty.

To the best of our knowledge, to date, no research has examined the influence of task-difficulty on (un)constrained thoughts. It is therefore unclear whether changes in task difficulty influence rates of thought constraint and, if so, in which direction. To shed light on this issue, across two experiments, we manipulated task difficulty via working-memory load using an n-back task (e.g., Baird et al., 2012; Konishi et al., 2015; Seli, Konishi, Risko, & Smilek, 2018; Smallwood et al., 2011).

In Experiment 1, participants viewed a series of numbers and were instructed to indicate whether a target was even or odd. In the “easy” task, participants were asked to indicate whether the currently presented number was even or odd (0-back), whereas in the “hard” task, participants were to indicate whether the number presented two trials ago was even or odd (2-back). Periodically throughout the task, we asked participants to indicate whether their thoughts were currently freely moving or not (Mills et al., 2018). Additionally, there has been some suggestion that constraint is related to the concept of a “flow state” (Csikszentmihalyi, 1988; Nakamura & Csikszentmihalyi, 2014) or “deep effortless concentration” (Marty-Dugas & Smilek, 2019). For exploratory purposes, we included a trait-level assessment of flow, or “deep effortless concentration” (Marty-Dugas & Smilek, 2019) to determine whether rates of unconstrained thought were associated with flow.

In Experiment 2, we adapted the thought probes to include a measure of task-relatedness in addition to thought constraint. Constraint and task-relatedness, although dissociable, are closely related in that off-task thought tends to be more unconstrained and on-task thought tends to be more constrained (Mills et al., 2018). Therefore, the aims of Experiment 2 were to determine whether constraint varies as a function of task difficulty independent of task-relatedness, further evaluate the relationship between flow and constraint, and conceptually replicate the results of Mills et al. (2018).

We supplemented the null hypothesis tests with corresponding Bayesian analyses (Rouder, Speckman, Sun, Morey, & Iverson, 2009). It is not possible to quantify the evidence for a null effect using conventional frequentist testing. A Bayes Factor, however, is a continuous measure of the relative strength of evidence and can quantify the degree to which the data are compatible with the null over the alternative hypothesis (Dienes, 2014; Rouder et al., 2009). All Bayesian analyses were performed using the R package BayesFactor and Bayes Factors (BFs) were calculated using its default settings (Morey & Rouder, 2018). BF10 indicates evidence in favor of the alternative hypothesis whereas BF01 indicates evidence in favor of the null hypothesis. To simplify the interpretation, we report the Bayes Factor in the direction the data supports (e.g., BF01 when there is more evidence in favor of the null over alternative hypothesis). As per previous recommendations, we refer to a BF > 3 as “moderate” and BF > 10 as “strong” evidence (Jeffreys, 1961; Rouder et al., 2009). To test for interaction effects, we contrasted the best model with the same model including/excluding the interaction effect.

Experiment 1

Method

Participants

Participants were 150 individuals who completed a Human Intelligence Task (HIT) posted on the Amazon Mechanical Turk (MTurk). All participants provided informed consent and were treated in accordance with guidelines approved by the ethics committee at Duke University. Participants were paid $2.00 (U.S. dollars) for completing the HIT, which lasted approximately

20 minutes. Only workers who had completed more than 5000 HITs with 98% approval rating could participate.

Materials

N-back task. Instructions and stimuli were displayed in the center of the screen. Stimuli were presented in a 72px black font on an off-white background Participants completed 20 practice n-back trials, 188 experimental n-back trials, and 12 thought probe trials. Stimuli consisted of the numbers 1 through 8 (4 even and 4 odd). Non-target stimuli were presented in black font and target stimuli were presented in red font. Target stimuli were randomly inserted between the second and 8th trial of each block of 8 n-back trials for a total 23 targets and 165 non-targets.

In both the 0-back and 2-back conditions, participants were instructed to withhold responses to black-colored digits and only respond when a red target appeared. In the 0-back condition, participants indicated whether the red digit was even or odd. In the 2-back condition, participants were presented a red question mark and indicated whether the digit presented two trials ago was even or odd.

Thought probes. Participants were presented 12 thought probes throughout the experiment. Thought probes were randomly inserted once every 15 n-back trials, between trials 5 and 11 (on average, every 60 seconds). Thought probes consisted of a single question (Mills et al., 2018) gauging thought constraint (“The thoughts I was experiencing were freely moving: YES/NO”). Verbatim instructions can be found in Appendix A.

Flow-state questionnaire. At the end of the experiment, participants completed a short questionnaire about their thoughts during the task. Participants reported their response on a 5-point likert scale (strongly disagree, disagree, neutral, agree, strongly agree):

1. I was able to focus on the task without straining to pay attention
2. I seemed to reach a level of deep focus almost effortlessly
3. I got in the zone and didn’t have to force myself to concentrate on the task
4. I easily paid deep attention to the task for extended periods of time.

Procedure

Participants were first given a general overview of the tasks and instructions on how to respond. In the 0-back group, participants were instructed to respond to the red digits (and not the black digits), and to indicate whether the red digits were even or odd, using the keyboard (‘e’ for even, ‘o’ for odd). In the 2-back group, participants were instructed to respond to the red question mark (“?”) and to indicate whether the digit presented two trials ago was even or odd using the keyboard. Participants had to complete a block of 20 practice trials with 4 targets and respond correctly to 3/4 of the targets before moving on to the experimental trials. If not, they repeated the practice trials. After completing the n-back task, participants completed the flow questionnaire.

Data analysis and manuscript preparation

This manuscript was prepared using R (R Core Team, 2019). A variety of notable R packages were used for data analysis (Bates et al., 2015; Fox & Weisberg, 2019; Kuznetsova et al., 2017; Singmann et al., 2019; Wickham et al., 2019; Wickham & Henry, 2019), data visualization (Fox & Weisberg, 2018; Kassambara, 2019; Wickham, 2016; Wilke, 2019), and general manuscript preparation (Aust & Barth, 2018).

Results

N-Back performance

First, we compared n-back performance using sensitivity (d-prime) and bias (c) scores across n-back groups. Target discrimination (d-prime) was significantly better in the 0-back versus 2-back groups, $\Delta M = 1.03$, 95% CI [0.55, 1.52], $t(148) = 4.20$, $p < .001$, $d = 0.69$, 95% CI [0.35, 1.02], $\mathrm{BF}_{10} = 432.55$. Additionally, there was significantly lower response bias (c) in the 0-back versus 2-back groups, $\Delta M = -0.29$, 95% CI [-0.54, -0.04], $t(148) = -2.27$, $p = .025$, $d = -0.37$, 95% CI [-0.7, -0.05], though there was only anecdotal evidence in favor of the alternative, $\mathrm{BF}_{10} = 1.84$.

Finally, we compared reaction times for accurate target responses across groups. We found no significant differences between group reaction time means, $\Delta M = -6.52$, 95% CI [-66.09, 53.05], $t(148) = -0.22$, $p = .829$, $d = -0.04$, 95% CI [-0.36, 0.29], $\mathrm{BF}_{01} = 5.57$. In addition to the standard reaction time analysis, we also performed an ex-Gaussian analysis. This approach estimates three parameters that correspond to different characteristics of the reaction time distribution: mu, sigma, and tau. Mu corresponds to the mean of the normal component of the distribution, sigma, to the standard deviation of the normal component of the distribution, and tau to the tail of the ex-Gaussian distribution (i.e., the mean and standard deviation of the exponential component of the distribution). These parameters can provide additional information regarding changes in reaction times that the estimated mean cannot provide.

Contrasting the 0-back and 2-back groups, we observed significantly larger mu in the 0-back compared to the 2-back, $\Delta M = 80.06$, 95% CI [19.50, 140.61], $t(148) = 2.61$, $p = .010$, $d = 0.43$, 95% CI [0.1, 0.75], $\mathrm{BF}_{10} = 3.88$. However, we also observed a significantly smaller sigma in the 0-back group ($\Delta M = -76.76$, 95% CI [-94.06, -59.47], $t(148) = -8.77$, $p < .001$, $d = -1.43$, 95% CI [-1.79, -1.07], $\mathrm{BF}_{10} = 1.30 \times 10^{12}$, and significantly smaller tau in the 0-back group, $\Delta M = -86.58$, 95% CI [-103.47, -69.68], $t(148) = -10.13$, $p < .001$, $d = -1.65$, 95% CI [-2.03, -1.28], $\mathrm{BF}_{10} = 3.17 \times 10^{15}$.

Counter-intuitively, we found participants in the 2-back responded faster in terms of mu, but were more variable (sigma), with higher frequency and magnitude of exceedingly long reaction times (tau). The difference in mu likely reflects the foreknowledge of the correct response in the 2-back group. That is, if participants were accurately tracking whether the digit presented two trials ago was even or odd, they will know the correct response prior to the target presentation, likely speeding reaction times on those trials. Differences in sigma and tau, in contrast, likely reflect performance when participants in the 2-back condition were uncertain of the correct response due to the increase in difficulty.

The results of the n-back analysis confirm that the 2-back task was indeed more difficult than the 0-back. Participant performance was notably worse in terms of target discrimination and bias. The reaction time analyses corroborate this interpretation increased variability, and higher frequency and magnitude of exceedingly long reaction times.

Thought probes

To determine whether rates of unconstrained thoughts were influenced by task difficulty, we compared rates of self-reported thought constraint across n-back groups (see Figure 2). Here, we observed significantly higher rates of freely moving thoughts in the 0-back as compared to the 2-back group, $\Delta M = 0.22$, 95% CI [0.10, 0.34], $t(148) = 3.71$, $p < .001$, $d = 0.61$, 95% CI [0.28, 0.94], with strong evidence in favor of the alternative, $\mathrm{BF}_{10} = 80.93$.

Flow questionnaire

We found no differences between mean flow state responses across n-back groups, $\Delta M = 0.04$, 95% CI $[-0.32, 0.39]$, $t(148) = 0.20$, $p = .840$, $d = 0.03$, 95% CI $[-0.29, 0.36]$, $\mathrm{BF}_{01} = 5.59$ (see Figure 2). Given the non-normal distribution of responses, however, we also tested for difference using the Kruskal-Wallis Test, and found no differences between response distributions, $\chi^2(1) = 0.01$, $p = .922$.

Additionally, we were interested in how the response in the flow state questionnaire compared to the reported rates of unconstrained thoughts. Using a Pearson correlation found a significant, small to moderate, positive correlation, $r = .19$, 95% CI [.03, .34], $t(148) = 2.37$, $p = .019$, $\mathrm{BF}_{10} = 2.67$, indicating some overlap between these two constructs.

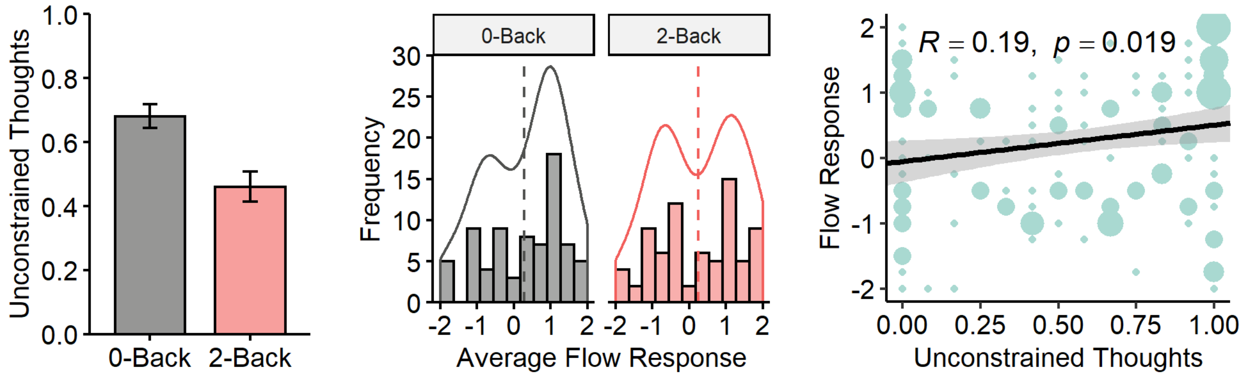

Figure 2

Results from Experiment 1. Proportion of unconstrained thoughts are plotted as a function of n-back task (left). The frequencies of average flow state responses are plotted as a function of n-back task (center) and the average flow state responses are plotted as a function of unconstrained thought proportions (right). Error bars represent 95% confidence intervals around the mean.

Discussion

In Experiment 1, we examined the influence of task difficulty on thought constraint. To manipulate task difficulty, we used two versions of the n-back task, throughout which we periodically presented thought probes to determine whether participants' current thoughts were constrained or unconstrained. Results across performance measures confirm that the 2-back was significantly more difficult than the 0-back. More important, however, we found that participants engaged in significantly higher rates of unconstrained thought in the easy versus the hard task and, for both groups, the proportion of unconstrained thoughts increased across trial blocks. Consistent with Hypothesis B, this preliminary evidence suggests that unconstrained thought is indeed resource dependent. That is, when the focal task requires fewer attentional resources, people tend to engage in more unconstrained thought. This suggests that such task difficulty could be a useful manipulation for studying (un)constrained thoughts.

We also compared self-reported rates of unconstrained thoughts with responses to the flow state questionnaire replicating prior work (Smith et al., under review). Although here, we found only a small-to-moderate correlation between measures. In any case, finding some correspondence between rates of unconstrained thoughts and flow is heartening. Indeed, on the surface, it seems sensible that these very similar constructs should overlap. However, the relatively low correspondence suggests that the constraint dimension of mind wandering is indeed distinct from the concept of a "flow state" (Csikszentmihalyi, 1988; Marty-Dugas & Smilek, 2019; Nakamura & Csikszentmihalyi, 2014).

One noteworthy limitation of Experiment 1 is that we only measured general rates of (un)constrained thoughts and did not measure the task-relatedness of those thoughts. Although prior work suggests that these two dimensions are dissociable (e.g., Mills et al., 2018, O'Neill et al., under review), it is possible that on- versus off-task thought constraint is differentially influenced by task difficulty. For instance, Mills et al. (2018) found off-task thoughts were more likely to be unconstrained than constrained and that on-task thoughts were more likely to be constrained than unconstrained. Moreover, task difficulty is known to influence rates of off task thought (e.g., Giambra, 1989; Smallwood et al., 2011; Teasdale et al., 1995; Thomson, Besner, & Smilek, 2013). What this suggests is the possibility that the task-difficulty manipulation, rather than influencing constraint, might simply be increasing the overall rate of off-task thoughts, which in turn, results in higher overall proportions of unconstrained thought when aggregated.

Furthermore, the relationship between flow and constraint might depend on the task-relatedness of thoughts. O'Neill et al. (under review) had participants complete a mind wandering clock task (e.g., Seli et al., 2018) and responded to intermittently presented thought probes gauging TUT and thought constraint. One interesting result from their study was the relatively high proportion of probe responses that were both on task and unconstrained (see Mills et al., 2018 for a similar result using an experience-sampling procedure). This is perhaps surprising given that constraint was initially conceived of as a measure of topical shifts in the content of one's thoughts, which is seemingly incompatible with being on task. To explain this surprising finding, O'Neill et al. proposed that perhaps the Mills et al. (2018) constraint probes do not measure topical shifts, but instead measure the effortfulness of thoughts (where "unconstrained" thoughts are those that feel effortless). As such, unconstrained thoughts—particularly on-task, unconstrained thoughts—might be closely associated with the concept of a "flow state" (Csikszentmihalyi, 1988; Nakamura & Csikszentmihalyi, 2014) or "deep effortless concentration" (Marty-Dugas & Smilek, 2019). It is possible then, that the low correspondence we observed between flow and constraint is the result of ignoring the task-relatedness of thoughts. If O'Neill et al.'s hypothesis is correct, then we should expect to observe a significant positive correlation between rates of on-task, unconstrained thoughts and scores on the flow questionnaire.

Experiment 2

Having established that rates of unconstrained thought can be manipulated via changes in task difficulty, the primary aim of Experiment 2 was to determine whether task difficulty influences rates of thought constraint for both on- and off-task thought. Task difficulty manipulations are known to influence rates of off task thought (e.g., Giambra, 1989; Smallwood et al., 2011; Teasdale et al., 1995; Thomson, Besner, & Smilek, 2013) and prior work has shown that off-task thought is more likely to be unconstrained and on-task thought is more likely to be constrained (Mills et al., 2018; Smith et al., under review).

Therefore, the primary aim of Experiment 2 was to determine how thought probe responses varied under task difficulty manipulations when both task-relatedness and thought constraint were measured. A secondary aim of Experiment 2 was to better evaluate the relationship between constraint and flow by testing O'Neill's hypothesis that the concept of a "flow state" or "deep effortless concentration" is closely associated with unconstrained on-task and not unconstrained thought in general.

The third and final aim of Experiment 2 was to conceptually replicate Mills et al. (2018) using a laboratory task. Mills et al. used a time-consuming and costly experience-sampling procedure. Methodologically, it would be important to know whether a simple online task would produce the same distribution of thought probe responses as the experience-sampling procedure. If so, one interested in testing hypotheses about thought constraint and task-relatedness could forgo the lengthy experience-sampling procedure and instead use the much simpler online task.

We again use the n-back task to manipulate task difficulty. However, we included thought probes that gauged both thought constraint (constrained, unconstrained) and task-relatedness (on-task, off-task). We also included a demographics survey and the Wordsum test, the latter of which is a brief vocabulary test that correlates well with full-scale measures of intelligence (Huang & Hauser, 1998). We included these measures to explore how rates of thought constraint might be related to age (e.g., Maillet & Schacter, 2016) and/or verbal intelligence (e.g., Mrazek et al., 2012), respectively. We ran two samples, both pre-registered prior to data collection. The first sample consisted of 150 participants, 75 per group, and the second consisted of 300 participants, 150 per group.

A power analysis prior to data collection was not possible given that no other studies have examined on- and off-task (un)constrained thought under task-difficulty manipulations. We collected data from the first sample and found the critical three-way interaction between n-back, thought constraint, and task relatedness to be non-significant ($p > .05$) and the critical two-way interaction between thought constraint and n-back condition to be significant ($p = .02$). To improve our confidence in this result, we decided to collect a second, larger sample to augment the first. Below, we report the results of an analysis combining data from both samples; however, individual sample analyses can be found in the supplementary material (critically, both samples converged on the same general conclusions). As in Experiment 1, we supplemented the null hypothesis tests with corresponding Bayesian analyses.

Method

Participants

Sample 1 consisted of 150 individuals who completed a HIT on Amazon Mechanical Turk; Sample 2 consisted of 300 individuals who likewise completed a HIT on Amazon Mechanical Turk. All participants provided informed consent and were treated in accordance with guidelines approved by the ethics committee at Duke University. Participants were paid $2.50 (U.S. dollars) for completing the HIT, which lasted approximately 25 minutes. Only workers who had completed more than 5000 HITs with 98% approval rating could participate.

Materials

N-back task. The n-back tasks were identical to those used in Experiment 1.

Thought probes. Instructions and stimuli were identical to Experiment 1. However, in addition to asking participants to report whether their thoughts were freely moving or not (identical to Experiment 1), we also asked whether their current thoughts were on- or off-task ("Just prior to seeing this screen, I was: (a) on-task (b) thinking task-unrelated thoughts"). Verbatim instructions are reported in Appendix B.

Wordsum. The Wordsum (Huang & Hauser, 1998) is a verbal intelligence test for which participants are asked to identify which of five words comes closes in meaning to a target word. For example, one of the target words is *accustom* and the five options are *disappoint*, *customary*, *encounter*, *get used to*, or *business*. The traditional version consists of 10 target words. In the version presented here, we included two additional target words, *audacious* (options: *smart*, *daring*, *brave*, *loud*, or *outgoing*) and *encumber* (options: *oppress*, *gather*, *impede*, *press*, or *encompass*). The Wordsum test correlates well with other, full-scale, measures of intelligence such as the WAIS-R (Huang & Hauser, 1998). As such, it is commonly used as a proxy for intelligence in many psychological, sociological, and political science studies (e.g., Malhotra et al., 2007; Pennycook et al., 2012) and included in the 16 General Social Surveys (Davies & Smith, 1994).

Flow state and demographics questionnaires. Following the Wordsum, participants completed the same flow state questionnaire completed in Experiment 1, and an optional demographics questionnaire.

Procedure

The procedure was identical to Experiment 1. After the n-back task participants completed the flow questionnaire, followed by the Wordsum, then the demographics survey.

Results

Target discrimination (d-prime) was significantly better in the 0-back versus 2-back groups, $\Delta M = 0.58$, 95% CI [0.34, 0.83], $t(448) = 4.62$, $p < .001$, $d = 0.44$, 95% CI [0.25, 0.62], $\mathrm{BF}_{10} = 2.71 \times 10^3$. Additionally, there was significantly lower response bias (c) in the 0-back versus 2-back groups, $\Delta M = -0.38$, 95% CI $[-0.49, -0.26]$, $t(448) = -6.51$, $p < .001$, $d = -0.61$, 95% CI $[-0.8, -0.42]$, $\mathrm{BF}_{10} = 3.93 \times 10^7$.

We also found no significant differences between group reaction time means, $\Delta M = 8.16$, 95% CI $[-25.96, 42.28]$, $t(448) = 0.47$, $p = .639$, $d = 0.04$, 95% CI $[-0.14, 0.23]$, $\mathrm{BF}_{01} = 8.59$. However, we did observe significant differences in the mu ($\Delta M = 78.02$, 95% CI $[42.50, 113.55]$, $t(448) = 4.32$, $p < .001$, $d = 0.41$, 95% CI $[0.22, 0.59]$, $\mathrm{BF}_{10} = 749.84$), sigma ($\Delta M = -69.03$, 95% CI $[-79.18, -58.89]$, $t(448) = -13.37$, $p < .001$, $d = -1.26$, 95% CI $[-1.46, -1.06]$, $\mathrm{BF}_{10} = 1.44 \times 10^{31}$), and tau parameters ($\Delta M = -69.87$, 95% CI $[-80.26, -59.47]$, $t(448) = -13.21$, $p < .001$, $d = -1.25$, 95% CI $[-1.45, -1.04]$, $\mathrm{BF}_{10} = 3.10 \times 10^{30}$).

These results are consistent with those from Experiment 1: the 2-back group was quicker in terms of mu but showed greater variability (sigma) and a longer reaction-time tail (tau). The differences in mu again likely reflect the foreknowledge of the correct response prior to the target stimulus, whereas the differences in sigma and tau likely reflect the increase in general task difficulty.

Thought Probes

We analyzed response proportions using 2 x 2 x 2 mixed analysis of variance (ANOVA) with task relatedness (on versus off task) and constraint (constrained versus unconstrained) as within-subjects factors and n-back task (0-back versus 2-back) as the between-subjects factor (see Figure 3). It is also worth noting that the dependent measure in our design is ipsative (i.e., proportions sum to 1) and likely violate assumptions typically required for an ANOVA. However, the ANOVA has been shown to be robust against violations of assumptions resulting from ipsativity (Greer and Dunlap 1997).

First, the critical three-way interaction was non-significant, $F(1{,}448) = 0.17$, $MSE = 0.12$, $p = .683$, $\hat{\eta}_p^2 < .001$, 90% CI [0, 0.01]. According to the Bayesian analyses, the model containing all main effects and all two-way interactions ($\mathrm{BF}_{10} = 3.74 \times 10^{83}$) was preferred over the full model including the three-way interaction by a factor of 7.53. Therefore, we continued to analyze and interpret each of the following two-way interactions.

N-back condition and task relatedness. The two-way interaction between n-back task and task relatedness was significant, $F(1{,}448) = 46.31$, $MSE = 0.06$, $p < .001$, $\hat{\eta}_p^2 = .094$, 90% CI [0.05, 0.14], with strong evidence in favor of including this interaction effect, $\mathrm{BF}_{10} = 1.03 \times 10^7$. In a follow-up analysis, we found the proportion off task thoughts was significantly lower in the 2-back versus 0-back condition, $\Delta M = -0.16$, 95% CI $[-0.21, -0.11]$, $t(448) = -6.81$, $p < .001$, $d = -0.64$, 95% CI $[-0.83, -0.45]$, $\mathrm{BF}_{10} = 2.30 \times 10^8$.

Thought constraint and task relatedness. The two-way interaction between task relatedness and constraint was also significant, $F(1{,}448) = 98.51$, $MSE = 0.12$, $p < .001$, $\hat{\eta}_p^2 = .180$, 90% CI [0.13, 0.23], with strong evidence in favor of including the interaction effect, $\mathrm{BF}_{10} = 2.43 \times 10^{31}$. Following up the interaction, we found off task thoughts were more likely to be unconstrained than constrained, $M_d = -0.17$, 95% CI $[-0.20, -0.15]$, $t(449) = -13.86$, $p < .001$, $d = -0.65$, 95% CI $[-0.79, -0.52]$, $\mathrm{BF}_{10} = 1.88 \times 10^{33}$, and on task thoughts were more likely to be constrained than unconstrained, $M_d = 0.15$, 95% CI [0.09, 0.21], $t(449) = 4.94$, $p < .001$, $d = 0.23$, 95% CI [0.1, 0.36], $\mathrm{BF}_{10} = 7.22 \times 10^3$.

Thought constraint and N-back condition. Finally, the two-way interaction between n-back and constraint was also significant, $F(1{,}448) = 10.43$, $MSE = 0.13$, $p = .001$, $\hat{\eta}_p^2 = .023$, 90% CI [0.01, 0.05], with strong evidence in favor of including the interaction effect, $\mathrm{BF}_{10} =$

152.68. Following up the interaction, we found off higher rates of unconstrained thought in the 2- versus 0-back condition $\Delta M = 0.11$, 95% CI [0.04, 0.17], $t(448) = 3.23$, $p = .001$, $d = 0.3$, 95% CI [0.12, 0.49], $\mathrm{BF}_{10} = 15.65$.

Summary. To summarize, we found evidence *against* a three-way interaction between n-back task, task-relatedness, and constraint, and could therefore meaningfully interpret the significant two-way interactions. The two-way interaction between n-back and task-relatedness shows higher rates of off-task and lower rates of on-task in the 0- versus 2-back condition. The two-way interaction between constraint and task-relatedness shows that off-task thoughts were more often unconstrained than constrained and on-task thoughts were more often constrained than unconstrained. Finally, the two-way interaction between n-back and constraint show lower rates of constrained thoughts and higher rates of unconstrained thoughts in the 0- versus 2-back condition.

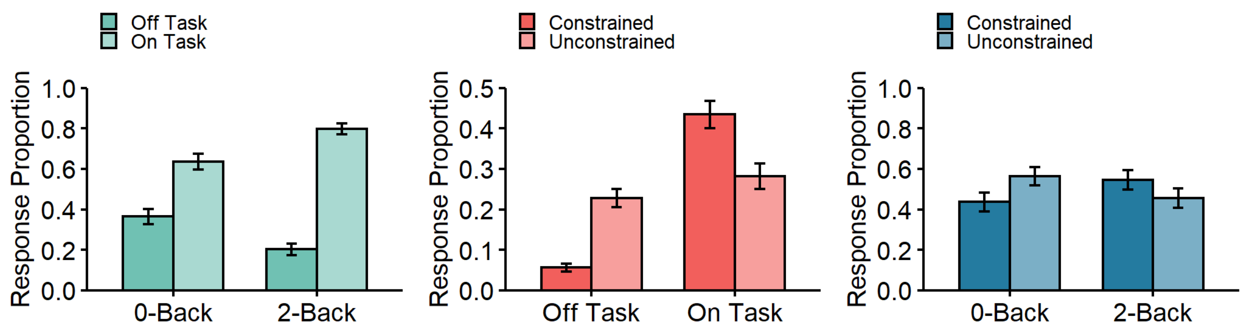

Figure 3

Results of the thought-probe trials from Experiment 2. Plotted are the two-way interactions between task relatedness and n-back (left), task relatedness and constraint (center), and constraint and n-back (right). Error bars represent 95% confidence intervals around the mean.

Flow Questionnaire

We analyzed the flow state questionnaire responses and compared them to the thought probe responses (see Figure 4). We found no significant difference between mean flow state responses across n-back groups using a *t*-test, $\Delta M = 0.12$, 95% CI $[-0.08, 0.31]$, $t(448) = 1.19$, $p = .235$, $d = 0.11$, 95% CI $[-0.07, 0.3]$ and no significant difference using the Kruskal-Wallis Test, $\chi^2(1) = 1.89$, $p = .169$.

We were also interested in how the response in the flow state questionnaire compared to each of the response types. We compared each using a Pearson correlation and found a small, non-significant correlation between rates of constrained on-task thought and flow responses, $r = .08$, 95% CI $[-.01, .17]$, $t(448) = 1.66$, $p = .099$, $\mathrm{BF}_{01} = 2.36$; a significant positive correlation between unconstrained on-task thought and flow responses, $r = .21$, 95% CI $[.12, .29]$, $t(448) = 4.50$, $p < .001$, $\mathrm{BF}_{10} = 1.92 \times 10^3$; a significant negative correlation between constrained off-task thought and flow responses, $r = -.21$, 95% CI $[-.30, -.12]$, $t(448) = -4.54$, $p < .001$, $\mathrm{BF}_{10} = 2.32 \times 10^3$; and a significant negative correlation between unconstrained thought and flow responses, $r = -.32$, 95% CI $[-.40, -.23]$, $t(448) = -7.13$, $p < .001$, $\mathrm{BF}_{10} = 2.35 \times 10^9$.

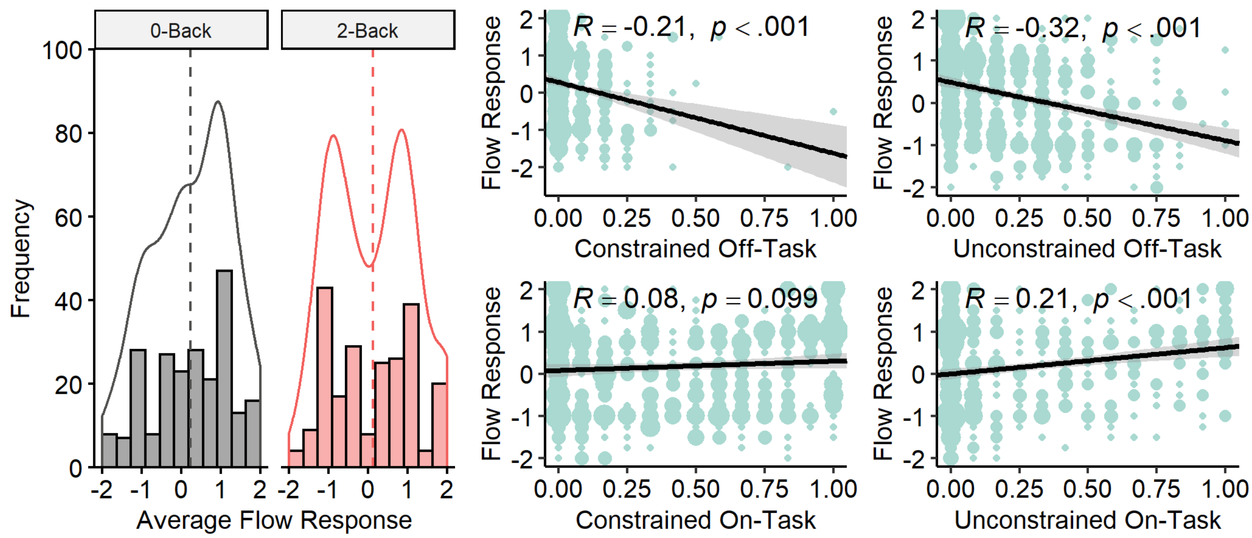

Figure 4

Results from the flow questionnaire in Experiment 2 (both samples). Frequency of average flow responses are plotted as a function of n-back task.

Exploratory Analyses

Thought probe responses and age. First, we correlated age with each of the thought probe response proportions (see Figure 5). Out of the 450 participants, 377 completed the optional demographics survey. We found that age was not significantly correlated with constrained off-task thought, $r = -.09$, 95% CI $[-.19, .01]$, $t(375) = -1.83$, $p = .068$, $\mathrm{BF}_{01} = 1.60$, but was negatively correlated with unconstrained off-task thought, $r = -.15$, 95% CI $[-.25, -.05]$, $t(375) = -3.00$, $p = .003$, $\mathrm{BF}_{10} = 9.63$. Additionally, age was positively correlated with constrained on-task thought, $r = .25$, 95% CI $[.15, .34]$, $t(375) = 4.96$, $p < .001$, $\mathrm{BF}_{10} = 1.51 \times 10^4$ and negatively correlated with unconstrained on-task thought, $r = -.13$, 95% CI $[-.23, -.03]$, $t(375) = -2.55$, $p = .011$, $\mathrm{BF}_{10} = 2.86$.

Thought probe responses and verbal intelligence. Next, we correlated Wordsum test accuracy scores with the thought probe response proportions (see Figure 5). Here, we found that Wordsum scores did not significantly correlate with either constrained off-task thought, $r = .03$, 95% CI $[-.06, .12]$, $t(448) = 0.62$, $p = .535$, $\mathrm{BF}_{01} = 7.50$. However, it did positively correlate with constrained on-task thought, $r = .15$, 95% CI $[.06, .24]$, $t(448) = 3.27$, $p = .001$, $\mathrm{BF}_{10} = 20.13$ and negatively correlate with unconstrained on-task thought, $r = -.25$, 95% CI $[-.33, -.16]$, $t(448) = -5.37$, $p < .001$, $\mathrm{BF}_{10} = 1.10 \times 10^5$, and unconstrained off-task thought, $r = .11$, 95% CI $[.02, .20]$, $t(448) = 2.32$, $p = .021$, $\mathrm{BF}_{10} = 1.53$.

Using the subset of participants who completed the demographics questionnaire (N = 377), we tested whether the previous associations were present after controlling for age. The results of each linear regression model produced the same directional effects as the previous analyses: We found Wordsum scores were not associated with constrained off-task thought (b = -0.03, 95% CI [-0.19, 0.13], t(374) = -0.40, p = .688), but were positively associated with constrained on-task thought (b = 0.05, 95% CI [0.00, 0.11], t(374) = 2.10, p = .036), negatively associated with unconstrained on-task thought (b = -0.10, 95% CI [-0.15, -0.04], t(374) = -3.53, p < .001), and positively associated with unconstrained off-task thought (b = 0.08, 95% CI [0.00, 0.16], t(374) = 2.08, p = .038).

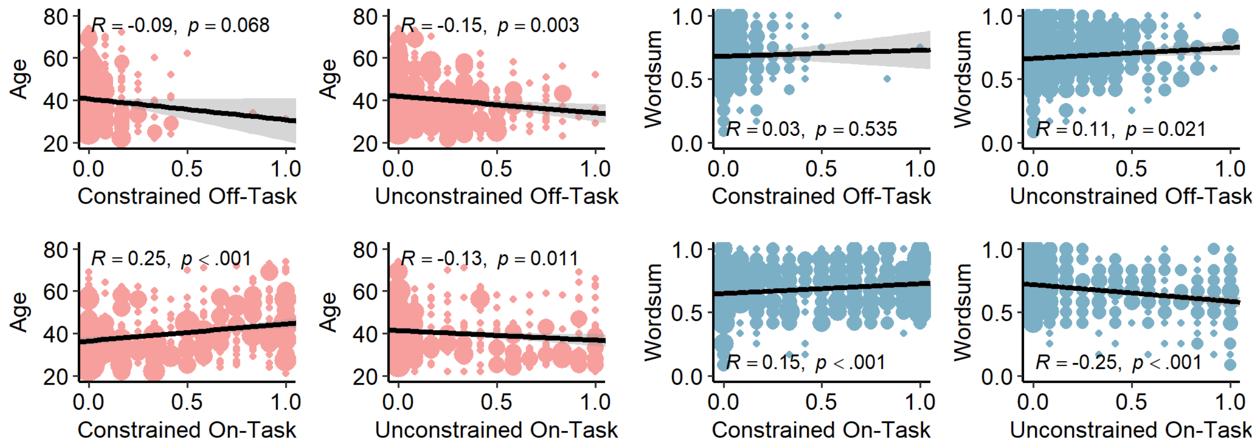

Figure 5

Results from Experiment 2 (both samples). Though probe responses are plotted as a function of age (left) and Wordsum scores (right).

Flow and age. Next, we examined the relationship between flow state responses and age (see Figure 6) and found no significant correlation, r < .01, 95% CI [-.10, .10], t(375) = -0.03, p = .976.

Flow and verbal intelligence. Finally, we examined the relationship between flow state responses and Wordsum test accuracy scores (see Figure 6). First we used a Pearson correlation and found a significant negative association, r = -.11, 95% CI [-.20, -.02], t(448) = -2.38, p = .018. Using a linear regression model controlling for age, we also find an significant negative association, b = -0.62, 95% CI [-1.19, -0.04], t(374) = -2.11, p = .035.

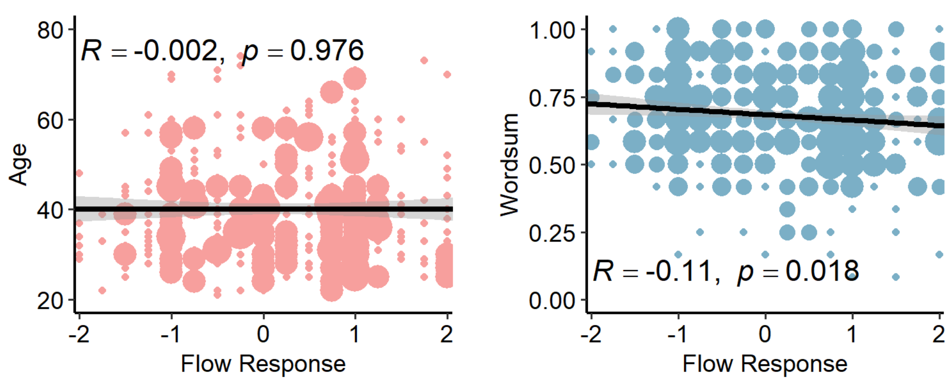

Figure 6

Results from Experiment 2 (both samples). Flow state questionnaire responses are plotted as a function of age (left) and Wordsum accuracy scores (right).

Discussion

In Experiment 2, we examined the influence of task difficulty on thought probe responses when both thought constraint and task relatedness were measured. To manipulate difficulty, we used two versions of the n-back task and periodically presented thought probes to determine whether participants' current thoughts were constrained or unconstrained. Results across performance measures confirm that the 2-back was significantly more difficult than the 0-back.

More importantly, however, we found that participants engaged in significantly higher rates of unconstrained thought in the easy versus the hard task, independent of whether those thoughts were on or off task. We did find, however, the change in thought constraint across n-back groups was considerably smaller in Experiment 2 compared to Experiment 1. This is likely due to the additive effects of manipulating both off-task and unconstrained thoughts. Replicating prior work, we found lower rates of off task thought in the hard task as compared to the easy task (e.g., Giambra, 1989; Smallwood et al., 2011; Teasdale et al., 1995; Thomson, Besner, & Smilek, 2013). Similarly, we found that whereas off-task thoughts were more likely to unconstrained than constrained, on-task thoughts were more likely to be constrained (Mills et al., 2018; Smith et al., under review). Taken together, these results suggest that future work using task difficulty to manipulate constraint should also measure task-relatedness to determine how much of the change is due a shift in off-task thought.

We also failed to replicate the positive association between flow state responses and rates of unconstrained thought we observed in Experiment 1. Whereas unconstrained on-task thought was positively correlated, unconstrained off-task thought was *negatively* correlated. The inconsistent result from Experiment 1 can likely be explained as a result of aggregating over on- and off-task thoughts (not measured in Experiment 1) and the fact that most thought probe responses are on-task. The result of the aggregation would be a small, overall positive relationship. Again, this points to the importance of measuring both the constraint and task-relatedness dimensions. However, the overall pattern of results suggests that the flow state concept (Csikszentmihalyi, 1988; Marty-Dugas & Smilek, 2019; Nakamura & Csikszentmihalyi, 2014) does not align well with constraint dimension of mind wandering, but instead seems to be better captured by the task-relatedness dimension.

Turning to our exploratory analyses, we found age was positively associated with constrained on-task thought, but negatively associated with unconstrained off-task thought. This aligns with prior work showing rates of task-unrelated thought tends to decrease with age (). It also suggests that "constrained on task" reports may be capturing traditional measures of task-related thought and "unconstrained off task" traditional measures of task-unrelated thoughts. However, we also found that age was negatively associated with unconstrained on-task thought. This result is particularly interesting as it suggests that "unconstrained on-task thought" is more like mind wandering than focused on task thought.

We also examined the relationship between verbal intelligence, as indexed by the WordSum scores, and thought probe responses. To our knowledge, there is only one study that used the WordSum test and they found no association between rates of task-unrelated thought and verbal intelligence (Phillips, Mills, D'Mello & Risko, 2016). However, more generally, task-unrelated thought (i.e., mind wandering) has been shown to be negatively associated with measures of general intelligence (Mooneyham and Schooler, 2013; Mrazek et al. 2012). In this regard, our result is mixed. On the one hand, constrained on task thought was positively associated with verbal intelligence. This is consistent with the prior claim. On the other hand, we found a small, positive association between unconstrained off-task thought and verbal intelligence and a negative association between unconstrained on-task thought and verbal intelligence. Again, this suggests that unconstrained on-task thought more closely resembles traditional conceptualizations of mind wandering than it does focused, on task thought.

Finally, we examined the relationship between flow state responses, age, and verbal intelligence. Previous work has indicated that the tendency to experience flow may actually increase across the lifespan (Mary-Dugas and Smilek, 2018; Ullén et al., 2016). However, despite a large sample and wide distribution of ages, here we find no association between age and flow state responses. Looking at verbal intelligence we find a small negative association between flow state responses and WordSum scores. Prior work has found small but inconsistent associations between intelligence and flow state using much larger samples and concluded that they were essentially unrelated (Ullén, et al., 2012).

Thought Constraint and Time on Task

The primary aim of the current study was to examine the influence of task difficulty on rates of (un)constrained thought. From a resource-theory perspective we hypothesized that either constrained or unconstrained thoughts might be resource dependent resulting in diverging predictions about how task difficulty should influence the rate of (un)constrained thought. Across the two experiments we found that when the task was made more difficult, and resources become more limited, participants reported lower rates of unconstrained thought. However, we did find that combinations of task-relatedness and constraint produced differing associations across our exploratory measures (i.e., flow, age, verbal intelligence). To further test the resource-dependency hypotheses, we performed to two additional exploratory analyses to examine how thought probe responses changed as a function of time on task.

These exploratory analyses are motivated by prior work showing that with increasing time on task, performance tends to decrease (Helton & Russell, 2011, 2012; Mackworth, 1948; Parasuraman et al., 2009) and mind wandering tends to increase (e.g., Cunningham, Scerbo, & Freeman, 2000; McVay & Kane, 2012; Thomson, Seli, Besner, & Smilek, 2014). These observations are thought to result from an attentional resource trade-off: As participants become familiar with the task, they may automatize parts of their performance freeing up more resources for mind wandering (e.g., Brosowsky et al. under review) or changes in motivation and/or boredom may cause participants to devote more attentional resources to mind wandering at the cost of their performance in the task (Brosowsky et al., under review; Thomson, Besner, & Smilek, 2015; for similar perspectives, see Esterman, Grosso, et al., 2016; Hancock, 2013; Kurzban, Duckworth, Kable, & Myers, 2013).

Experiment 1

To determine if rates of unconstrained thoughts varied as a function of time on task, we submitted the proportion of unconstrained thoughts to a mixed ANOVA with n-back task (0-

Back versus 2-Back) as the between-subjects factor and trial block (1, 2, 3, and 4) as the within-subjects factor (see Figure 7).

The results of this analysis showed a significant main effect of n-back task, $F(1{,}148) = 13.75$, $MSE = 0.53$, $p < .001$, $\hat{\eta}_p^2 = .085$, 90% CI [0.03, 0.16], and a significant effect of block, $F(2.63{,}389.57) = 10.59$, $MSE = 0.06$, $p < .001$, $\hat{\eta}_p^2 = .067$, 90% CI [0.03, 0.11], but no significant interaction between n-back task and block, $F(2.63{,}389.57) = 0.96$, $MSE = 0.06$, $p = .402$, $\hat{\eta}_p^2 = .006$, 90% CI [0, 0.02]. Additionally, the model containing only main effects was preferred over the null, $\mathrm{BF}_{10} = 8.68 \times 10^5$, and over the full model, $\mathrm{BF}_{10} = 20.05$.

Experiment 2

To determine how probe response proportions varied with time on task we submitted each probe response to a mixed ANOVA with n-back task (0-Back versus 2-Back) as the between-subjects factor and trial block (1, 2, 3, and 4) as the within-subjects factor (see Figure 7).

Constrained on-task. We found a significant main effect of n-back task, $F(1{,}448) = 14.48$, $MSE = 0.51$, $p < .001$, $\hat{\eta}_p^2 = .031$, 90% CI [0.01, 0.06]; significant main effect of block, $F(2.85{,}1277.19) = 24.28$, $MSE = 0.05$, $p < .001$, $\hat{\eta}_p^2 = .051$, 90% CI [0.03, 0.07]; but no interaction effect, $F(2.85{,}1277.19) = 0.17$, $MSE = 0.05$, $p = .906$, $\hat{\eta}_p^2 < .001$, 90% CI [0, 0]. The model containing both main effects was preferred over the null, $\mathrm{BF}_{10} = 2.00 \times 10^{14}$, and over the full model, $\mathrm{BF}_{10} = 238.59$.

Constrained off-task. We found no effect of n-back task, $F(1{,}448) = 3.49$, $MSE = 0.05$, $p = .062$, $\hat{\eta}_p^2 = .008$, 90% CI [0, 0.03]; a significant main effect of block, $F(2.81{,}1260.72) = 5.05$, $MSE = 0.02$, $p = .002$, $\hat{\eta}_p^2 = .011$, 90% CI [0, 0.02]; and no interaction effect, $F(2.81{,}1260.72) = 1.48$, $MSE = 0.02$, $p = .220$, $\hat{\eta}_p^2 = .003$, 90% CI [0, 0.01]. Though the model containing only the main effect of block was preferred over the full model, $\mathrm{BF}_{10} = 34.81$, there was only anecdotal evidence in favor of the block-only model over the null, $\mathrm{BF}_{10} = 1.87$. Therefore, there was little evidence to suggest that constrained off-task thought varied as a function of time on task.

Unconstrained on-task. We found no effect of n-back task, $F(1{,}448) = 1.05$, $MSE = 0.47$, $p = .305$, $\hat{\eta}_p^2 = .002$, 90% CI [0, 0.02]; a significant main effect of block, $F(2.82{,}1262.36) = 3.01$, $MSE = 0.04$, $p = .032$, $\hat{\eta}_p^2 = .007$, 90% CI [0, 0.01]; and no interaction effect, $F(2.82{,}1262.36) = 0.57$, $MSE = 0.04$, $p = .621$, $\hat{\eta}_p^2 = .001$, 90% CI [0, < .001]. However, the null model was preferred over all other models (BFs > 3). In particular, the null was preferred over the model containing only the main effect of block, $\mathrm{BF}_{01} = 10.05$. This suggests that the rate of unconstrained on-task thought did not vary across blocks.

Unconstrained off-task. We found a significant effect of n-back task, $F(1{,}448) = 43.69$, $MSE = 0.21$, $p < .001$, $\hat{\eta}_p^2 = .089$, 90% CI [0.05, 0.13]; a significant main effect of block, $F(2.93{,}1310.67) = 25.74$, $MSE = 0.05$, $p < .001$, $\hat{\eta}_p^2 = .054$, 90% CI [0.03, 0.07]; but no interaction effect, $F(2.93{,}1310.67) = 1.43$, $MSE = 0.05$, $p = .232$, $\hat{\eta}_p^2 = .003$, 90% CI [0, 0.01]. Additionally, model containing both main effects was preferred over the null, $\mathrm{BF}_{10} = 7.08 \times 10^{20}$, and the full model, $\mathrm{BF}_{10} = 39.27$.

Summary

To briefly summarize the results of the time-on-task analyses (see Figure 6): In Experiment 1, we found that rates of unconstrained thought increased with time on task. The results of Experiment 2, however, showed that constraint did not consistently increase with time on task but instead differed depending on task-relatedness. Here, we found that constrained on-task thoughts decreased with time on task, but unconstrained on-task thought did not vary with time on task. We also found that unconstrained off-task thoughts increased with time on task, but no evidence that constrained off-task thoughts varied.

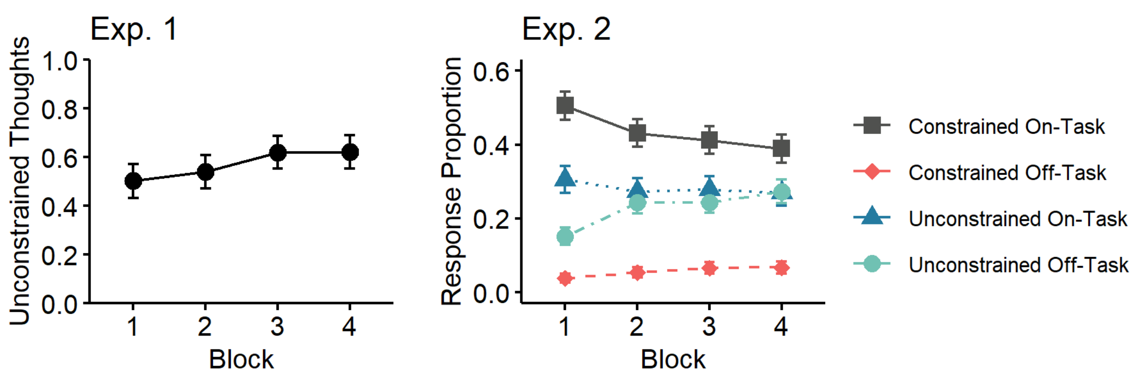

Figure 7

Results from Experiments 1 and 2. Thought probe responses are plotted as a function block (collapsed over n-back task). Error bars represent 95% confidence intervals around the means.

General Discussion

Of primary interest was whether manipulating task difficulty would influence reported rates of thought constraint (Christoff et al., 2016; Mills et al., 2018; Smith et al., under review). In Experiment 1, we found lower rates of thought constraint in the more difficulty 2-back condition as compared to the 0-back condition suggesting that indeed, constraint can be manipulated. In Experiment 2, we sought to determine how task difficulty influences thought probe responses when both task relatedness and thought constraint were measured. Replicating and extending Experiment 1, we found that, independent of whether thoughts were on or off task, participants in the more difficult 2-back condition reported lower rates of thought constraint than participants in the 0-back condition. Furthermore, we replicated prior work that (a) participants engage in more off-task thought during easier tasks (e.g., Giambra, 1989; Smallwood et al., 2011; Teasdale et al., 1995; Thomson, Besner, & Smilek, 2013), and (b) off-task thought was more likely to be unconstrained than constrained and on-task thought was more likely to be constrained than unconstrained (Mills et al., 2018; Smith et al., under review). Together, these results suggest that unconstrained thought is indeed resource dependent and when the focal task requires less attentional resources, people tend to engage in more unconstrained thought.

Our results, however, are somewhat at odds with O'Neill et al. (2019). They found that the frequency of unconstrained thoughts did not vary with changes in task demands. Specifically, they found intentional mind wandering varied according to task demands, but unconstrained mind wandering largely remained the same regardless of task demands. The difference in timescales, however, could help explain this discrepancy. In their study, they manipulated task demands within 20-second trial windows–with the assumption that task demands are higher closer to the start/end of the trial. In our study we manipulated the overall level of task demands across groups. As such, it may be the case the rates of unconstrained thoughts fluctuate at a slower rate than the other dimensions of mind wandering. Indeed, the overall percentage of unconstrained thoughts in O'Neill et al. (2019) was 72%, closely resembling the 68% we observed for the 0-back group in Experiment 1. This would suggest that overall task-wide difficulty was, unsurprisingly, quite low in the clock task. The differences in fluctuation timescales would also reaffirm the claim that constraint is a theoretically important and dissociable dimension of mind wandering.

From a methodological perspective our results demonstrate that task difficulty can be used to manipulate rates of thought constraint. However, the results of Experiment 2 demonstrate the need to measure task relatedness. Since both task relatedness and thought constraint are influenced by task difficulty it is possible that changes in task relatedness could appear as changes in constraint. In Experiment 2, we found constraint was manipulated independent of task relatedness, but with a diminished effect. Furthermore, our distribution of thought probe responses was similar to, but not identical to Mills et al. (2018) who used experience-sampling. The main discrepancy being the rates of constrained off-task thought. Whereas Mills et al. found ~20% of reports to be constrained off-task, participants in our task only reported ~5%. Therefore, our laboratory task should be sufficient for those interested in the other thought probe responses, but inappropriate if one were interested in reports of constrained off-task thought.

Finally, we reported a number of exploratory analyses examining the relationships between thought probe responses and flow, age, and verbal intelligence. Of those results, there were two key takeaways: first, thought constraint and task relatedness do not appear to be redundant. That is, relationships between thought probe responses and other measures did not converge consistently along one dimension or the other as one might expect if they were redundant. Second, unconstrained thought does not appear to reflect the concept of a "flow state" (Csikszentmihalyi, 1988; Nakamura & Csikszentmihalyi, 2014) or "deep effortless concentration" (Marty-Dugas & Smilek, 2019). Although in Experiment 1 we found a small positive relationship, in Experiment 2 we found flow state responses did not consistently map onto the constraint dimension. However, unconstrained on-task thought was positively associated with flow state responses. It is possible that experiencing a flow state is not reflective of unconstrained thought in general, but a particular kind of "on task" unconstrained thought which fits well with the idea that a flow state is effortless concentration (Marty-Dugas & Smilek, 2019). More work is needed however, to determine whether this is the case.

References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*.

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science*, *23*(10), 1117–1122.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative Cognition and Brain Network Dynamics. *Trends in Cognitive Sciences*, *20*(2), 87–95. https://doi.org/10.1016/j.tics.2015.10.004

Christoff, K., Irving, Z. C., Fox, K. C., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews Neuroscience*, *17*(11), 718.

Csikszentmihalyi, M. (1988). *The flow experience and its significance for human psychology*.

Davies, J. A., & Smith, T. W. (1994). *General social surveys 1972–1994: Cumulative codebook*. National Opinion Research Center.

Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2012). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*, *23*(11), 2712–2723.

Fox, J., & Weisberg, S. (2018). Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. *Journal of Statistical Software*, *87*(9), 1–27. https://doi.org/10.18637/jss.v087.i09

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Huang, M.-H., & Hauser, R. (1998). Trends in Black-White test-score differentials: II. The WORDSUM vocabulary test. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures*. American Psychological Association.

Kassambara, A. (2019). *Ggpubr: 'Ggplot2' based publication ready plots*. https://CRAN.R-project.org/package=ggpubr

Konishi, M., McLaren, D. G., Engen, H., & Smallwood, J. (2015). Shaped by the past: The default mode network supports cognition that is independent of immediate perceptual input. *PloS One*, *10*(6), e0132209.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Maillet, D., & Schacter, D. L. (2016). From mind wandering to involuntary retrieval: Age-related differences in spontaneous cognitive processes. *Neuropsychologia*, *80*, 142–156. https://doi.org/10.1016/j.neuropsychologia.2015.11.017

Malhotra, N., Krosnick, J. A., & Haertel, E. (2007). The psychometric properties of the GSS Wordsum vocabulary test. *GSS Methodological Report*, *11*.

Marty-Dugas, J., & Smilek, D. (2019). Deep, effortless concentration: Re-examining the flow concept and exploring relations with inattention, absorption, and personality. *Psychological Research*, *83*(8), 1760–1777.

McVay, J. C., Meier, M. E., Touron, D. R., & Kane, M. J. (2013). Aging ebbs the flow of thought: Adult age differences in mind wandering, executive control, and self-evaluation. *Acta Psychologica*, *142*(1), 136–147.

Mills, C., D'Mello, S. K., & Kopp, K. (2015). The influence of consequence value and text difficulty on affect, attention, and learning while reading instructional texts. *Learning and Instruction*, *40*, 9–20.

Mills, C., Raffaelli, Q., Irving, Z. C., Stan, D., & Christoff, K. (2018). Is an off-task mind a freely-moving mind? Examining the relationship between different dimensions of thought. *Consciousness and Cognition*, *58*, 20–33.

Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General*, *141*(4), 788–798. https://doi.org/10.1037/a0027968

Nakamura, J., & Csikszentmihalyi, M. (2014). The Concept of Flow. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 239–263). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_16

O'Neill, K., Smith, A. P., Smilek, D., & Seli, P. (under review). *Dissociating the freely-moving thought dimension of mind-wandering from the intentionality and task-unrelated thought dimensions* [Paper submitted for publication].

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin & Review*, *21*(5), 1309–1315.

Schooler, J. W. (2004). Zoning Out while Reading: Evidence for Dissociations between Experience and Metaconsciousness Jonathan W. Schooler, Erik D. Reichle, and David V. Halpern. *Thinking and Seeing: Visual Metacognition in Adults and Children*, *203*.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, *15*(7), 319–326.

Seli, P., Carriere, J. S., Wammes, J. D., Risko, E. F., Schacter, D. L., & Smilek, D. (2018). On the clock: Evidence for the rapid and strategic modulation of mind wandering. *Psychological Science*, *29*(8), 1247–1256.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2019). *Afex: Analysis of factorial experiments*. https://CRAN.R-project.org/package=afex

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., Carlson, J. M., & Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PloS One*, *6*(3), e18298.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946.

Wammes, J. D., Boucher, P. O., Seli, P., Cheyne, J. A., & Smilek, D. (2016). Mind wandering during lectures I: Changes in rates across an entire semester. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 13.

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *Dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2019). *Tidyr: Tidy messy data*. https://CRAN.R-project.org/package=tidyr

Wilke, C. O. (2019). *Cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. https://CRAN.R-project.org/package=cowplot

Appendix A: Experiment 1 Instructions

Thought probe instructions

"*Your thoughts are considered "freely moving" when they seem to wander around on their own, flowing from one thing to the other, when there is no overarching purpose or direction to them, when it seems like your thoughts could land on pretty much anything, and when images and memories easily come to mind*

You will be presented the prompt: The thoughts I was experiencing were moving freely:" and respond by clicking (a) YES (b) NO".

Appendix B: Experiment 2 Instructions

"While you are completing the task, you may find yourself thinking about something other than the task. These thoughts are referred to as "task-unrelated thoughts." Experiencing task-unrelated thoughts is perfectly normal, especially when one has to do the same thing for a long period of time.

We would like to determine how frequently you are focused on the task and how frequently you are thinking about thoughts that are unrelated to the task. To do this, throughout the task, you will be periodically presented a "thought-sampling screen" asking you to report whether you were ON TASK or experiencing TASK-UNRELATED THOUGHTS (TUTs) just before the thought-sampling screen was presented.

Being ON TASK means that, just before the thought-sampling screen appeared, you were focused on task. For instance, if you were thinking about your performance on the task or thinking about when you should make a button press, these thoughts would count as being ON TASK. On the other hand, experiencing TASK-UNRELATED THOUGHTS (TUTs) means that you were thinking about something completely unrelated to the task. Some examples of TUTs include thoughts about what to eat for dinner, an upcoming event, or something that happened to you earlier in the day. Any thoughts that you have that are not related to the task count as TASK-UNRELATED THOUGHTS.

When the thought-sampling screen is presented, we would like you to indicate whether, just prior to seeing the screen, your thoughts were ON TASK or TASK-UNRELATED. The thought-sampling screen will look like this: Just prior to seeing this screen, I was: (a) on task (b) thinking task-unrelated thoughts.

We will also ask whether your thoughts are freely moving or not. Your thoughts are considered "freely moving" when they seem to wander around on their own, flowing from one thing to the other, when there is no overarching purpose or direction to them, when it seems like your thoughts could land on pretty much anything, and when images and memories easily come to mind. You will be presented the prompt: The thoughts I was experiencing were moving freely:" and respond by clicking (a) YES (b) NO".