Quantifying streams of thought during cognitive task performance using sequence analysis

Anthony P. Zanesco¹

Published online: 18 May 2020
© The Psychonomic Society, Inc. 2020

Abstract

Streams of thought vary in content from one moment to the next, and these temporal patterns have been argued to be critical to understanding the wandering mind. But few analytic methods have been proposed that can account for both the content and temporal ordering of categorical experience sampling thought probes over time. In the present study, I apply sequence analytic methods to quantify the dynamics of thought from time series sequences of categorical experience sampling thought probes delivered across five different cognitive tasks in the same individuals ($N = 545$). Analyses revealed some patterns of consistency in streams of thought within individuals, but also demonstrated considerable variability within and between task sessions. Hierarchical clustering of sequence dissimilarities further revealed common typologies of mind wandering across individuals. These findings demonstrate the application of sequence analytic methods for quantifying the dynamics of thought over the course of task performance and show that contextual task constraints are associated with how streams of thought unfold over time. More broadly, sequence analysis provides a valuable framework for investigation of time ordered cognitive and behavioral processes across psychological domains.

Keywords Dynamics · Markov model · Mind wandering · Multivariate distance matrix regression · Sequence analysis · Task-unrelated thought

Introduction

Thoughts alight from one topic to another in a seemingly unconstrained manner, likened famously to the many flights and perchings of a bird (James, 1890). As noted by Christoff et al. (2016), contemporary psychological research on mind wandering has focused almost entirely on understanding the many perchings of thought—the substantive content of thought (e.g., about the past or future, negative or positive) and how it might be defined with respect to one’s current goals or activities (e.g., whether thoughts are task-related or unrelated, deliberate or spontaneous). But investigations of its flights have been largely neglected. This is perhaps expected as experience sampling methods are widely used but rely on capturing random moments of cognition *in situ* to estimate the amount of time individuals engage in episodes of thought and cannot determine with precision when thoughts begin or end

(Smallwood, 2013). Rather, experience sampling can only catch thoughts at a moment in time, and their flights must be inferred from changes in content between sampled moments.

How reported content differs across sequential thought probes may nevertheless provide an estimate of how attention and thought varies from one moment to the next. Investigations of temporal patterns in sequences of thought probe reports may therefore be important for understanding how attention is maintained in a task-relevant manner and the cognitive mechanisms that contribute to continuity or variation in thought over time (Smallwood, 2013). Characterizing how thought streams unfold is also critical for understanding the many individual differences and contextual constraints that influence ongoing thought (Christoff et al., 2016, 2018; Smith, Mills, Paxton, & Christoff, 2018). Indeed, the dynamic and unconstrained manner in which streams of thought unfold is a central feature of subjective experience and has been argued to be fundamental to definitions of mind wandering (Christoff et al., 2016, 2018; Irving, 2016; Irving & Thompson, 2018; but see Seli et al., 2018b, 2018c). Nevertheless, few studies have proposed methods for directly quantifying the temporal ordering of thought probe reports. Experience sampling studies attempting to understand and

✉ Anthony P. Zanesco
apz13@miami.edu

¹ Department of Psychology, University of Miami, 5665 Ponce de Leon Blvd, Coral Gables, FL 33146, USA

characterize the dynamics of thought will have to account for temporal patterns within entire sets of successive probes.

Experience sampling studies of mind wandering have instead primarily characterized the content and amount of task-unrelated thought as a proportion or average of thought probe reports in cognitive tasks and daily life (Weinstein, 2018). Studies have nevertheless provided some insight into the factors that influence patterns of ongoing thought within individuals or across situations and cognitive task contexts, and the role attention plays in regulating thought during cognitive task performance (Smallwood & Schooler, 2015). Thought is often bound in its amount and content by contextual factors, habitual tendencies, or by deliberate attentional focus and goal-directed constraints. When cognition is successfully directed towards the accomplishment of some demanding goal, thought should be coerced in content towards relevant topic domains while the occurrence of task-unrelated thought is reduced (Smallwood & Schooler, 2015).

The amount of task-unrelated thought reported by individuals has been demonstrated to be constrained in the cognitive task context and depends in part on how attention is regulated over time in the service of task performance. Individuals strategically regulate instances of task-unrelated thought in response to dynamic situational demands (Seli et al., 2018a), and task-unrelated thought has been shown to gradually increase with time on task (e.g., Krimsky, Forster, Llabre, & Jha, 2017; Thomson, Seli, Besner, & Smilek, 2014). More frequent presentation of experience sampling probes also reduces the proportion of probes that catch individuals in episodes of task-unrelated thought (Schubert, Frischkorn, & Rummel, 2019; Seli, Carriere, Levene, & Smilek, 2013), as probes are likely used as cues by individuals to maintain attention on their current task. Furthermore, the content of thought is contextually constrained, in which individuals, unsurprisingly, are more likely to experience thoughts relating to their performance and the parameters of their current task (i.e., task-related interference; McVay & Kane, 2012; Robison & Unsworth, 2018; Welhaf et al., 2019).

Tendencies to engage in task-unrelated thought are moderately consistent within persons across sessions of different cognitive tasks (Kane et al., 2016; Welhaf et al., 2019). But the amount of task-unrelated thought experienced by individuals is inconsistently correlated between laboratory situations and daily life (Kane et al., 2017; Linz, Pauly, Smallwood, & Engert, 2019; McVay, Kane, & Kwapil, 2009). Turnbull et al. (2019) found that the features of cognitive tasks also interact with individual differences to influence patterns of ongoing thought. Specifically, when task demands are low, individuals engage in greater amounts of off-task thought. Furthermore, those with greater levels of fluid intelligence are even more likely to let their thoughts drift from the task with increasing time between target trials (Turnbull et al., 2019). Together, these findings suggest that tendencies for particular patterns

of ongoing thought are shared between individuals and that amounts of task-unrelated thought demonstrate a modicum of stability in the cognitive task context. This supports the notion that the contextual features of the current task, alongside one's cognitive capacities and other more habitual and automatic constraints, are associated with how streams of thought unfold. Individual difference studies of mind wandering have also provided support for this supposition, as the frequency and content of task-unrelated thought measured across different task contexts is reliably associated with individuals' cognitive abilities (e.g., Kane & McVay, 2012; Kane et al., 2007, 2016; Unsworth & McMillan, 2014; Welhaf et al., 2019).

That the frequency and content of task-unrelated thought displays these many contextual constraints and stable individual tendencies nevertheless appears hard to reconcile with the dynamic and seemingly unconstrained manner in which thought unfolds. How then do we understand the purposeful wandering of the mind (Irving, 2016), in which thoughts often appear aimless but also seem guided in line with our goals and current concerns? Past research is limited because estimating only the amount of task-unrelated thought ignores intra-individual variability in the temporal order of thought content and dependence between thoughts from one moment to the next. More than its overall amount, the within-person dynamics of thought are critically important for understanding the manner in which individuals self-generate and regulate its ongoing occurrence. Any comprehensive characterization must therefore consider approaches that can quantify the extent to which different thoughts systematically occur over time and from which thoughts they are likely to follow.

Recently, Welhaf et al. (2019) conducted secondary analyses of a large multi-measure dataset originally reported by Kane et al. (2016), in order to more directly quantify the temporal dynamics of thought content over successive experience sampling probes embedded in five cognitive tasks. Specifically, they explored whether systematic patterns of switching and stability between categorical reports of thought content predicted individual differences in cognitive ability. For example, individuals reported switches in task-unrelated thought content on roughly 40% of successive thought probe reports, which demonstrates both stability in the kinds of thoughts occurring from one probe to the next as well as a great deal of variability in content over successive probes. Importantly, this measure of thought switching, did not account for the specific content of thoughts across successive thought probes. That is, switches were assessed regardless of the kind of task-unrelated thought that followed particular thought probes. This is perhaps consequential as it implicitly assumes that transitions between thought probes do not depend on thought content or that probes have no bearing on what thoughts might be subsequently reported.

Instead, it appears critical that researchers consider both the temporal ordering of thoughts and the kinds of thoughts that

arise in sequence. One category of thought (e.g., thoughts about everyday things) repeatedly followed by the occurrence of another category of thought (e.g., personal worries) is a kind of temporal consistency that might be contrasted with more randomly occurring thought content. Likewise, task-unrelated thought might consistently follow more lengthy sequences of task-related focus. In this way, the content of thought does not appear independent of its dynamics. Ruby et al. (2013) provided support for this supposition by sampling the content of task-unrelated thought using a multidimensional experience sampling approach, in which individuals simultaneously rated their thoughts across several dimensions (e.g., on vs. off-task, about the past or future, negative or positive). They observed that affective dimensional components of thought predicted individuals' moods on the subsequent trial, particularly when affective thoughts were about the past or the future.

One approach that lends itself well to the analysis of successive categorical thought probes is the application of sequence analysis and related methods (Abbott & Tsay, 2000; Ritschard & Studer, 2018). Sequence analysis can be broadly defined as a set of techniques that take as their input entire sets of sequences of categorical data rather than individual data points (Abbott & Tsay, 2000). This provides for comparison of entire time series of categorical thought probe reports as whole units, accounting for both their content and dynamics over successive probes. Sequence analytic methods were originally applied in the domains of computer science and biology (e.g., Sankoff and Kruskal, 1983), but have also been continually developed for application in the social sciences (e.g., Cornwell, Gershuny, & Sullivan, 2020; Ritschard & Studer, 2018).

Most commonly, sequence analysis involves defining the dissimilarity between categorical sequences using an edit distance based on the number of edit operations required to transform one sequence into another. The pairwise dissimilarities of entire streams of thought can be defined in this way, allowing for tests of differences between whole sets of sequences or the identification of clusters of similar sequences (Studer, Ritschard, Gabadinho, & Müller, 2011). While researchers have increasingly identified data-driven structure in the content of multidimensional thought probes (e.g., Linz et al., 2019; Ruby et al., 2013; Turnbull et al., 2019), clusters of temporal patterns in sequences (i.e., typologies of ongoing thought) have not previously been described. In addition to analysis and clustering based on sequence dissimilarities, information theoretic measures that quantify the randomness of thought content and the temporal variability of thought probe sequences can be calculated from categorical sequences (i.e., Elzinga & Liefbroer, 2007; Elzinga, 2010; Shannon, 1948). Finally, Markov chain modeling of transitions can be used to quantify the systematic tendencies of thought content to precede and follow other content categories. Together, these

measures may prove useful for characterizing the dynamics of categorical thought probe sequences.

In the present study, I conducted secondary analyses of data originally reported by Kane et al. (2016) and Welhaf et al. (2019), with the goal of demonstrating the utility of sequence analysis and related methods for quantifying the dynamics of thought. In addition, I had the aim of examining variability and consistency in streams of thought across successive thought probe reports and between cognitive task contexts. I had several specific research questions which guided my investigation. Specifically, I sought to examine whether systematic similarities and differences could be identified in the frequency of and transitions between kinds of reported thought content between cognitive tasks. To this end, I characterized sources of variability in thought probe sequences, and directly compared temporal sequences and thought probe transitions between tasks. Finally, to the degree that some amount of systematicity in thought probe sequences might be observed within or between individuals, I had the exploratory goal of identifying data-driven typologies of mind wandering through clustering of sequence dissimilarities to evaluate common thought streams across individuals and cognitive tasks.

Method

Deidentified participant data were acquired from an online data-sharing repository via the Open Science Framework (<https://osf.io/guhw7/>) provided by Welhaf et al. (2019). Secondary analyses conducted using these data applied the same participant exclusions and outlier treatment as reported in Kane et al. (2016) and Welhaf et al. (2019).

Participants

Kane et al. (2016) recruited and assessed 545 undergraduates at the University of North Carolina at Greensboro (UNCG). Complete details regarding participant recruitment and demographic characteristics are reported in Kane et al. (2016). 545 participants completed the first of three assessments, 492 returned to complete the second assessment, and 472 completed all three assessments. All participants gave informed consent and the study was approved by the Institutional Review Board of UNCG.

Procedure

Participants completed a battery of cognitive-behavioral tasks during three separate assessment sessions. The complete description of tasks and their order over assessments is described by Kane et al. (2016). Importantly, five tasks included experience sampling thought probes presented randomly during task performance. These tasks were administered over

different days of assessment. At the first assessment, a semantic version of the Sustained Attention to Response Task (SART; Robertson et al., 1997) was administered, followed later by a Letter Flanker task. At the second assessment, an Arrow Flanker task was administered followed later by a Number Stroop task. Finally, a 2-Back working memory task was administered at the third assessment.

Experience sampling of thought content

Forty-five thought probes occurred randomly for participants throughout the SART (7% of all trials), whereas 20 occurred in the Arrow Flanker (4.2% of block-1 trials and 16.6% of block-2 trials), 12 in the Letter Flanker (8.3% of trials), 20 in the Number Stroop (13% of block-2 trials), and 15 in the 2-Back task (6.3% of trials). At the presentation of each probe, participants selected among 8 possible options to describe whether their attention was currently focused “on-task” (option 1); whether they experienced thoughts directed towards “task performance and evaluation” (option 2); thoughts about normal life concerns and “everyday things” (option 3); thoughts about one’s current physical, cognitive or emotional “state of being” (option 4); thoughts about “personal worries” (option 5); fantastical thoughts and “daydreams” (option 6); thoughts about stimuli in the “external environment” (option 7); and “other” thoughts that were not described by other categories (option 8).

Sequence analyses of thought probes

Sequence analysis of participants’ thought content considered each response to thought probes as a discrete categorical state that occurred in sequence across the duration of cognitive tasks. The entire range of categorical responses were considered (i.e., from 1 “on-task” to 8 “other”) for all analyses and calculation of measures of sequence dynamics. Successive categorical responses to thought probes thus formed whole

sequences of thought content for each participant and cognitive task. Variability in the timing between probes was ignored. Thought probe sequences were manipulated and visualized in *R* using package *TraMineR* (Gabadinho, Ritschard, Müller, & Studer, 2011) and *seqHMM* (Helske & Helske, 2019).

Information theoretic analyses of entropy, complexity, and turbulence Several dependent measures were calculated from the categorical thought probe sequences using package *TraMineR* (Gabadinho et al., 2011). These measures quantified variability in the temporal ordering and content of sequences of thought probe reports. Specifically, I calculated the average number of unique kinds of thought probe categories reported, and the proportion of transitions between content categories out of the total possible transitions (i.e., length of sequence - 1). I also calculated three information theoretic measures: sequence *entropy* (Shannon, 1948), *complexity* (Elzinga, 2010), and *turbulence* (Elzinga & Liefbroer, 2007). Figure 1 depicts a schematic showing three sequences of thought probes and their corresponding sequence entropy, complexity, and turbulence.

Entropy in this context reflects how unpredictable one’s thought probe responses are based on the size of the “alphabet” of content categories (i.e., the range of possible thought probe responses from 1 to 8) and their proportion of occurrence. Entropy was standardized to range from 0 to 1. Entropy is 0 when the same thought content occurs for all experience sampling probes. In contrast, the maximum entropy would be reached when the occurrence of each content category was equiprobable (i.e., occurred at random). Low entropy (close to 0) therefore indicates that thought probe responses display some systematic tendencies, whereas high entropy (close to 1) suggests that probe responses are unpredictable in their occurrence. Entropy is calculated from the distribution of thought content occurrences only and does not depend on the length of the sequence or the sequential ordering of content.

Fig. 1 Sequences of 45 experience sampling thought probe responses (rated according to 8 content categories) during the SART are depicted on rows for three participants (A, B, and C). Pairwise dissimilarities between sequences are calculated based on optimal matching (OM) with substitution costs derived from transition probabilities of the entire set of SART sequences ($n = 526$) and insertion / deletion (indel) costs set as half

the maximum substitution cost. Sequences A and B are more alike (dissimilarity = 37.9) than they are to sequence C (dissimilarities > 60). Sequence A has more entropy (0.75), complexity (0.65), and turbulence (0.58) than sequences B (entropy = 0.54, complexity = 0.44, turbulence = 0.36) and C (entropy = 0.60, complexity = 0.55, turbulence = 0.48)

In contrast, the measure of complexity considers both the ordering of thought probes in sequence as well as their frequency distribution (i.e., sequence entropy). Complexity is standardized (from 0 to 1) with respect to the sequence length. Complexity is 0 when a sequence contains only a single content category for all probes (i.e., 0 transitions and 0 sequence entropy), and 1 when the maximum number of transitions occurs (i.e., length - 1) and the duration of each content occurrence across probes are uniform. Finally, turbulence is calculated based on the number of distinct subsequences found in each sequence and the variance of the duration of each subsequence. Turbulence is also standardized to range from 0 to 1 based on the maximum possible turbulence for a sequence of the same length. When turbulence is low, few unique kinds of thought content categories occur in a sequence, and when they do occur, the same category occurs many times in succession. In contrast, when turbulence is high, different thought content frequently occurs but rarely lasts across subsequent thought probes, instead transitioning to the occurrence of a different kind of content category.

I compared these five dependent measures between tasks using mixed effects models in *R* with package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). That is, data for each separate dependent measure (e.g., entropy) from the five cognitive tasks were nested within individuals in the model and compared among tasks. The SART served as the reference condition for fixed effects of task. Random subject intercepts were included to allow for within-person dependencies. Type III tests of fixed effects are reported for omnibus tests, and parameter estimates are given for model effects. Model parameters were estimated using restricted maximum likelihood and degrees of freedom were calculated by Satterthwaite approximation. The proportion of variance explained (R^2) by the fixed effects was also calculated (Nakagawa, Johnson, & Schielzeth, 2017).

To evaluate the consistency of these measures between- and within-participants, I also calculated the intraclass correlation (ICC) on null (i.e., unconditional means) models using the *R* package *sjstats* (Lüdtke, 2019). For repeated measures data, the ICC describes the proportion of variance attributable to differences in the means of individuals versus the share of variance attributable to within-person variation and measurement error. Often, ICC is used as an index of consistency and measurement reliability over multiple measurements. This is because the ICC will be large for highly reliable measures that are stable over repeated assessments when within-person variation and measurement error is low. ICC above 0.75 is commonly interpreted as a good level of reliability, whereas from 0.50 to 0.75 is moderate, and 0.50 and below is poor (Koo & Li, 2016).

Markov chain transition analyses Markov chain transition probabilities were calculated based on the entire sample of

thought probe sequences for each task using package *seqHMM* (Helske & Helske, 2019). Transition probabilities reflect the probability of thought probes transitioning from one particular thought probe content to the next. To test whether transitions were more likely to occur than chance, I compared transition probabilities to a null distribution of probabilities occurring by chance using non-parametric permutation-based statistical comparisons. Thought probe sequences were shuffled randomly in their ordering for each individual for 5000 iterations and an empirical null distribution of transition probabilities occurring by chance were calculated. The actual transition probabilities were compared to the distribution of permutations to identify whether the observed probability was sufficiently extreme (two-tailed $\alpha = .010$, i.e., permutations greater than or less than 99.5% of all permutations) with respect to the null distribution of probabilities occurring by chance.

Observed differences in transition probabilities between cognitive tasks were also compared to a null distribution of differences occurring by chance using non-parametric permutation-based statistical comparisons. I shuffled sequences randomly between pairs of cognitive tasks for 5000 iterations and calculated an empirical null distribution of transitions probabilities. I then compared the actual difference in transition probabilities for each transition-pair to the empirical null-distribution to identify whether the observed probability was sufficiently extreme (greater than or less than 99.5% of all permutations) with respect to the null distribution of differences occurring by chance.

Optimal matching and analyses of sequence dissimilarities

Pairwise sequence dissimilarities were calculated for thought probe sequences in *R* using package *TraMineR*. Dissimilarities were based on optimal matching (OM) with substitution costs derived from transition probabilities and insertion / deletion (indel) costs set as half the maximum substitution cost. OM algorithms function by defining an edit distance based on the number of substitutions, insertions, and deletions required to transform one sequence of categorical states into another with corresponding costs associated with these different operations (Abbot & Tsay, 2000; Studer & Ritschard, 2016). For example, a stream of thought defined by the sequence “on-task” (probe option 1) → “task-related interference” (option 2) → “on-task” (option 1) → “task-related interference” (option 2) would require two substitutions to transform the sequence into “on-task” (option 1) → “daydreams” (option 6) → “on-task” (option 1) → “daydreams” (option 6) by replacing the second and fourth items in the sequence. The edit distance dissimilarity between those sequences would therefore reflect the cost of two substitutions (or four indels). Figure 1 depicts a schematic showing three sequences of thought probes and their corresponding sequence dissimilarities for illustration.

Importantly, a number of statistical tests can be conducted based on generalization of partitions of sum of squares ($SS_{total} = \frac{1}{2n} \times \sum d_{ij}$) from the matrices of pairwise dissimilarities (d_{ij}) calculated through OM. In the context of OM dissimilarities based on edit distances, these methods have been known as “discrepancy analysis” (Studer et al., 2011). More broadly, these techniques are known as multivariate distance matrix regression (MDMR; McArdle & Anderson, 2001; McArtor, Lubke, & Bergeman, 2017; Zapala & Schork, 2012), and allow for tests of statistical associations between matrices of dissimilarities and categorical or continuous predictors. MDMR has also recently been extended to the mixed effects modeling framework, allowing for analyses to account for hierarchical or dependent data, implemented in package *MDMR* in *R* (McArtor et al., 2017; McArtor, 2018). MDMR of Euclidean-based distances commonly use the squared dissimilarities in partitioning variance in dissimilarity matrices. It has instead been argued that the unsquared dissimilarities are preferable when dissimilarity is based on edit distances (Studer et al., 2011). Unsquared dissimilarities were therefore used here. Typically, significance is evaluated based on permutation testing of a pseudo- F statistic comparing observed statistics to a distribution under the null hypothesis. But analytic p -values have now been validated for these tests (McArtor et al., 2017). A pseudo- R^2 can also be defined as the sum of squares between groups ($SS_{between}$) over SS_{total} and represents the proportion of discrepancy among sequences accounted for by covariates in the fixed effects MDMR model.

Ward hierarchical agglomerative clustering of sequence dissimilarities was used to identify unique clusters of individuals’ thought probe sequences for each task, and for all individuals’ sequences combined across tasks. Clustering was conducted using the *WeightedCluster* package in *R* (Studer, 2013). I chose the optimal number of clusters based on examination of three cluster selection criteria (see Studer, 2013, for further discussion): including the point biserial correlation, Hubert’s gamma, and the weighted average silhouette width (AWS). The lowest number of clusters was considered optimal for the cluster selection identified by the local maxima of all three criteria. The weighted AWS reflects the homogeneity of the cluster and can be interpreted as a measure of similarity between the sequences comprising a cluster. Clusters with very similar sequences will have a high AWS, whereas clusters of more heterogeneous sequences will have a lower AWS. Kaufman and Rousseeuw (2005) suggested that $AWS < 0.25$ reflected little cluster homogeneity, whereas 0.26 to 0.50 suggests weak homogeneity, and $AWS > 0.51$ reasonable homogeneity. The proportion of discrepancy explained (pseudo- R^2) in sequences was also examined for each cluster solution based on partitions of dissimilarities according to cluster assignment as a categorical covariate. It is important to note that

more clusters will necessarily explain more discrepancy until pseudo- R^2 approaches 1, with a cluster for each individual sequence. Finally, to evaluate whether participants were organized into the same clusters together across tasks, individuals’ cluster assignments identified from one cognitive task were used as a categorical predictor of the pairwise sequence dissimilarities of each other task.

Results

Sequences of thought probe responses are depicted in Figure 2 for all participants in each of five cognitive tasks. There were 526 thought probe sequences in total for SART, 479 for Arrow Flanker, 462 for Letter Flanker, 478 for Number Stroop, and 461 for 2-Back tasks. The proportion of thought probes categorized according to on-task focus or categories of task-related interference or task-unrelated thoughts are provided in Table 1 for each task.

Kolmogorov–Smirnov tests demonstrated that the response distributions of thought probes differed significantly between tasks (all $ps < .001$). Most notably, proportions of on-task reports appeared lower in the SART, whereas reports of task-related interference appeared greater in the SART, compared to other tasks. Despite these differences, proportions of thought probes were highly similar. Across all tasks, reports of being on-task, experiencing task-related interference, or thoughts about one’s current state were the most prevalent. Daydreams, worries, and thoughts about everyday things were the next most frequent categories.

Dependent measures of thought probe sequence dynamics

Descriptive statistics are provided in Table 2 for dependent measures of thought probe sequence dynamics for each of the five cognitive tasks. Comparisons between tasks are provided below for each measure, and parameter estimates from mixed effects models are given in Table 3.

Number of unique thought content categories reported

Individuals reported between 3.7 to 5.5 unique kinds of thought content categories, on average, during task performance for each of the five cognitive tasks (see Table 2). The number of thought contents reported differed significantly between the five cognitive tasks in a systematic manner, $F(4, 1901) = 175.00$, $p < .001$. The number of content categories was significantly lower in the Arrow Flanker ($b = -1.443$, $p < .001$, 95% CI [-1.593, -1.293]), the Letter Flanker ($b = -1.278$, $p < .001$, 95% CI [-1.429, -1.126]), the Number Stroop ($b = -1.535$, $p < .001$, 95% CI [-1.685, -1.384]), and the 2-Back ($b = -1.848$, $p < .001$, 95% CI [-1.999, -1.696]), compared to the SART (intercept $b = 5.523$, 95% CI [5.390, 5.656]).

Fig. 2 Sequences of experience sampling thought probe responses (rated according to 8 categories) during cognitive tasks are depicted (left) on rows for each participant. The distribution of thought probe responses for each thought probe are also shown (right). 45 thought probes occurred

during the SART ($n = 526$), whereas 20 occurred in the Arrow Flanker ($n = 479$), 12 in the Letter Flanker ($n = 462$), 20 in the Number Stroop ($n = 478$), and 15 in the 2-Back task ($n = 461$)

Individuals therefore reported more kinds of thought content categories in the SART compared to other tasks. The intra-

class correlation (ICC) for the number of reported thought contents was 0.311, suggesting that 31.1% of variance was

Table 1 Sample size and the proportion of thought probes for each task

Task	<i>N</i>	1	2	3	4	5	6	7	8
SART	526	0.216	0.274	0.051	0.232	0.047	0.096	0.036	0.047
Arrow Flanker	479	0.340	0.170	0.067	0.207	0.064	0.098	0.022	0.031
Letter Flanker	462	0.283	0.131	0.076	0.268	0.064	0.095	0.045	0.038
Number Stroop	478	0.332	0.217	0.063	0.213	0.053	0.071	0.022	0.027
2-Back	461	0.369	0.207	0.045	0.200	0.054	0.084	0.019	0.021

Note: The sample size (N) and the proportion of all probes classified as each thought probe response (1 “on-task”; 2 “task-related thoughts”; 3 “everyday things”; 4 “current state”; 5 “worries”; 6 “daydreams”; 7 “environment”; and 8 “other”) are provided for each cognitive task

Table 2 Descriptive statistics of dependent measures of thought probe sequences

Task	Length	No. Reported	Prop. Transitions	Entropy	Complexity	Turbulence
SART	45	5.529 (1.641)	0.499 (0.212)	0.610 (0.191)	0.546 (0.198)	0.492 (0.193)
Arrow Flanker	20	4.090 (1.600)	0.468 (0.253)	0.494 (0.234)	0.476 (0.237)	0.491 (0.230)
Letter Flanker	12	4.242 (1.481)	0.598 (0.257)	0.559 (0.217)	0.574 (0.228)	0.613 (0.223)
Number Stroop	20	3.994 (1.538)	0.471 (0.241)	0.484 (0.218)	0.473 (0.221)	0.488 (0.218)
2-Back	15	3.683 (1.524)	0.476 (0.261)	0.463 (0.224)	0.464 (0.235)	0.494 (0.233)

Note: Means and standard deviations are given for the number of unique thought content categories reported, the proportion of total transitions, and thought probe sequence entropy, complexity, and turbulence for each of the five cognitive tasks. The length of thought probe sequences is also provided

attributable to differences between individuals, but that the majority of variance (~69%) was attributable to within-person variation and measurement error. Task context explained 15% of the observed variance ($R^2 = .150$) in the number of unique categories reported.

Proportion of total transitions Participants reported different thought content on subsequent probes for 47% to 60% of all possible transitions on average for each of the different cognitive tasks (see Table 2). The proportion of transitions differed significantly between cognitive tasks, $F(4, 1896.5) = 43.94$, $p < .001$. Transitions were significantly lower in the Arrow Flanker ($b = -0.032$, $p = .006$, 95% CI [-0.054, -0.009]), the Number Stroop ($b = -0.028$, $p = .013$, 95% CI [-0.051, -0.006]), and the 2-Back ($b = -0.025$, $p = .029$, 95% CI [-0.048, -0.003]), compared to the SART (intercept $b = 5.523$, 95% CI [0.477, 0.519]). In contrast, transitions were more prevalent in the Letter Flanker ($b = 0.101$, $p < .001$, 95% CI [0.079, 0.124]), compared to the SART. Individuals therefore

had more transitions between different thought content categories in the SART and Letter Flanker compared to other tasks. The ICC was .469 and task context explained 4% of the observed variance ($R^2 = .039$) in the proportion of transitions.

Sequence entropy Mean thought probe sequence entropy ranged from 0.46 to 0.61 among tasks, which indicated that individuals' sequences demonstrated a fair amount of systematicity in their responding (see Table 2). Entropy differed significantly between cognitive tasks, $F(4, 1900.7) = 68.12$, $p < .001$. Entropy was significantly lower in the Arrow Flanker ($b = -0.117$, $p < .001$, 95% CI [-0.137, -0.097]), the Letter Flanker ($b = -0.048$, $p < .001$, 95% CI [-0.069, -0.028]), the Number Stroop ($b = -0.126$, $p < .001$, 95% CI [-0.146, -0.106]), and the 2-Back ($b = -0.147$, $p < .001$, 95% CI [-0.168, -0.127]), compared to the SART (intercept $b = 0.609$, 95% CI [0.591, 0.628]). Individuals' sequences had greater entropy in the SART compared to other tasks. Figure 3

Table 3 Parameter estimates from mixed effects models of dependent measures of thought probe sequences

Measure	No. Reported	Prop. Transitions	Entropy	Complexity	Turbulence
Fixed Effects					
SART (Intercept)	5.523 (0.068)***	0.498 (0.011)***	0.609 (0.009)***	0.545 (0.010)***	0.491 (0.010)***
Arrow Flanker	-1.443 (0.077)***	-0.032 (0.011)**	-0.117 (0.010)***	-0.070 (0.010)***	-0.002 (0.010)
Letter Flanker	-1.277 (0.077)***	0.101 (0.011)***	-0.048 (0.011)***	0.031 (0.010)**	0.123 (0.010)***
Number Stroop	-1.535 (0.077)***	-0.028 (0.011)*	-0.126 (0.010)***	-0.073 (0.010)***	-0.004 (0.010)
2-Back	-1.848 (0.077)***	-0.025 (0.012)*	-0.147 (0.011)***	-0.083 (0.010)***	0.001 (0.010)
Random Effects					
Intercept variance	0.995 (0.997)	0.028 (0.168)	0.021 (0.143)	0.024 (0.155)	0.023 (0.151)
Residual variance	1.444 (1.202)	0.032 (0.178)	0.027 (0.163)	0.026 (0.162)	0.025 (0.159)
Subjects	537	537	537	537	537
Observations	2406	2406	2406	2406	2406

Note: Restricted maximum likelihood estimates are given from mixed effects models of the number of unique thought content categories reported, the proportion of total transitions, and thought probe sequence entropy, complexity, and turbulence for fixed effects of cognitive task. The SART serves as the reference condition. The number of subjects (N) and observations contributing to the analyses are provided. Standard errors are reported in parentheses for fixed effects and standard deviations for random effects. * $p < .05$, ** $p < .01$, *** $p < .001$

Fig. 3 Model estimated means and 95% confidence intervals are provided from mixed effects models for sequence entropy, complexity, and turbulence derived from thought probe sequences from the SART ($n = 526$), the Arrow Flanker ($n = 479$), the Letter Flanker ($n = 462$), the Number Stroop ($n = 478$), and the 2-Back ($n = 461$). Observed subject data are plotted from individuals' sequences as dots in gray

shows individual participant data for these comparisons, along with the model estimated means for each task. The ICC was 0.437 and task context explained 6% of the observed variance ($R^2 = .062$) in entropy.

Sequence complexity Mean sequence complexity ranged from 0.46 to 0.57 among tasks (see Table 2). Complexity differed significantly between cognitive tasks, $F(4, 1896.3) = 44.89, p < .001$. Complexity was significantly lower in the

Arrow Flanker ($b = -0.070, p < .001, 95\% \text{ CI } [-0.091, -0.050]$), the Number Stroop ($b = -0.073, p < .001, 95\% \text{ CI } [-0.093, -0.053]$), and the 2-Back ($b = -0.083, p < .001, 95\% \text{ CI } [-0.103, -0.062]$), compared to the SART (intercept $b = 0.545, 95\% \text{ CI } [0.526, 0.564]$). But complexity was greater in the Letter Flanker ($b = 0.031, p = .003, 95\% \text{ CI } [0.010, 0.051]$), compared to the SART. Individuals' sequences were therefore more complex in the SART and Letter Flanker compared to other tasks (see Figure 3). The ICC was 0.478, and task context explained 4% of the observed variance ($R^2 = .039$) in complexity.

Sequence turbulence Mean sequence turbulence ranged from 0.49 to 0.61 (see Table 2). Turbulence differed significantly between cognitive tasks, $F(4, 1895.6) = 54.84, p < .001$. Turbulence was significantly higher in the Letter Flanker ($b = 0.123, p < .001, 95\% \text{ CI } [0.103, 0.143]$), but did not differ for the Arrow Flanker ($b = -0.002, p = .853, 95\% \text{ CI } [-0.022, 0.018]$), the Number Stroop ($b = -0.004, p = .696, 95\% \text{ CI } [-0.024, 0.016]$), and the 2-Back ($b = 0.001, p = .901, 95\% \text{ CI } [-0.019, 0.021]$), compared to the SART (intercept $b = 0.491, 95\% \text{ CI } [0.472, 0.509]$). Turbulence was therefore higher in the Letter Flanker than in other tasks (see Figure 3). The ICC was 0.473 and task context explained 5% of the observed variance ($R^2 = .048$) in turbulence.

Transition probabilities of experience sampling thought probes

An examination of the Markov chain transition probabilities for thought probe sequences revealed similar patterns of transitions for tasks overall. Bivariate correlations between sets of transition probabilities were high among tasks (r range = .926 – .983). Figure 4 depicts the transition probability matrices of thought probes. Each row of the transition probability matrix reflects the total probability (i.e., each row sums to 1) of thought probes transitioning to other probes.

Overall, individuals were highly likely to report the same thought content that they last reported. This was indicated by high transition probabilities along the diagonal of matrices. Accordingly, patterns of transitions appeared more ordered than random, and fluctuations in thought content in the tasks overall behaved with some amount of systematicity. This was reflected in a number of transition probabilities that were more or less likely to occur than if they had been ordered at random (i.e., transition probabilities more extreme than 99.5% of 5000 random permutations of thought probes). Figure 5 depicts transitions that were significantly different from those expected by chance.

Despite overall similarities, many transition probabilities differed significantly between cognitive tasks (differences between observed transition probabilities that were more extreme than 99.5% of permuted differences). Figure 6 depicts the observed difference for significant transition pairs. Of

Fig. 4 Transition probabilities indicate the total probability (rows sum to 1) of all transitions from each thought content category to each other category based on sequences for each cognitive task

note, transitions from $1 \rightarrow 1$, $2 \rightarrow 1$, $4 \rightarrow 1$, and $8 \rightarrow 1$ were generally less likely to occur in the SART compared to other cognitive tasks. In contrast, transitions from $2 \rightarrow 2$, $4 \rightarrow 2$, and $6 \rightarrow 2$ were more likely to occur. These patterns of transitions suggest that during the SART, individuals were less likely to return to an on-task report after experiencing some amount of task-related interference or task-unrelated thoughts about their current state. Furthermore, they were more likely to continue experiencing task-related interference over time and report additional task-related interference after experiencing thoughts about their current state or daydreams. Another noteworthy pattern of transitions occurred for the Letter Flanker, in which individuals were less likely to remain in an on-task state (i.e., transitions from $1 \rightarrow 1$), and less likely to transition from $1 \rightarrow 2$, $2 \rightarrow 2$, and $3 \rightarrow 2$, compared to other tasks. This suggests that in the Letter Flanker, individuals had a harder time regulating their attention between successive thought probes and were less likely to experience recurrent task-related interference.

Sequence analysis and clustering of experience sampling thought probes

Sequence discrepancy analysis Pairwise sequence dissimilarities based on OM were calculated between participants'

thought probe sequences of cognitive tasks. Mixed effects multivariate distance matrix regression (MDMR) of sequence dissimilarities, with 2406 sequences hierarchically nested within individuals, demonstrated that tasks significantly differentiated sequences, $F(4, 2401) = 457.90$, $p < .001$, and explained about 8% of total discrepancy (pseudo- $R^2 = .084$). Pairwise comparisons indicated that sequences from all tasks differed significantly from one another (all $ps < .001$). Finally, a generalization of the ICC from partitions of sums of squares of dissimilarities indicated that 31.6% of variance among sequences was attributable to variance between individuals, but that the majority of variance (68.4%) was attributable to intra-individual variation between and within sessions.

The length of thought probe sequences also contributes in some part to discrepancy between sequences. This is because thought probes must be inserted / deleted to match pairs of sequences in length. Given that thought probe sequences differed in length between tasks, sequence comparisons were also constrained to only the first 12 successive thought probes. Tasks differentiated sequences to a smaller degree when considering the first 12 thought probes, $F(4, 2401) = 120.04$, $p < .001$, accounting for only 3% of discrepancy (pseudo- $R^2 = .031$). Again, all tasks differed significantly from one another (all $ps < .001$). The ICC (0.343) was slightly larger when considering only the first 12 thought probes.

Fig. 5 Transition probabilities for each cognitive task are shown for pairs of transitions that were significantly extreme with respect to the null distribution of permuted transitions. Blue cells indicate transitions less

likely to occur than if thought probes had been ordered at random, and red cells indicate transitions more likely than at random

Clustering of sequences for cognitive tasks Hierarchical agglomerative clustering of sequence dissimilarities revealed several typologies of thought probe sequences for each cognitive task. The optimal clusters were chosen based on the local maxima among three model assessment criteria, with preference for the solution with the smallest number of clusters. The model selection criteria are shown in Figure 7 for cluster solutions 1 through 20, and the optimal clusters are depicted in Figure 8 for each task. Clusters of individuals identified in one cognitive task were then evaluated to see how well they explain discrepancy in other cognitive tasks. Table 4 provides the pseudo- R^2 for each cluster solution as a measure of discrepancy explained. Optimal clusters of sequences from each task explained between 17% to 32% of dissimilarity between the sequences of thought probes from that same task.

Four clusters were identified as optimal in the SART that explained 17.45% of discrepancy among individuals' sequences. Clusters were generally ill-defined as the weighted ASW ranged from -0.011 to 0.307, with sequences belonging to Cluster 1 being the most homogenous and sequences in Cluster 4 being the least homogenous. Nevertheless, cluster typologies were visually distinct and appeared to be differentiated based on different proportions of thought content (see Figure 8). Cluster 1 contained individuals who reported more

on-task probes, whereas Cluster 2 contained individuals who reported more task-related interference. Cluster 3 contained individuals who reported more thoughts about their current state of being. Finally, Cluster 4 contained individuals whose sequences contained more daydreams but also appeared more heterogeneous overall. Individuals' cluster membership was then applied to other task sequences to examine whether thought probe sequence typologies from the SART consistently identified the same groupings of individuals' sequences from other tasks. The 4 clusters only explained between 3.6% to 5.4% of discrepancy in other tasks (see Table 4).

Six clusters were optimal in the Arrow Flanker that explained 29.5% of discrepancy. Clusters 1 (ASW = 0.517), 2 (ASW = 0.654), and 4 (ASW = 0.538), were adequately defined and showed a moderate level of homogeneity between sequences. Cluster typologies 1 through 4 were similar in thought content prevalence to those identified in other tasks. When applied to other tasks, the 6 clusters explained between 4.5% to 10.7% of discrepancy in the other tasks (see Table 4). Interestingly, clusters of sequences identified in the Arrow Flanker explained a noteworthy amount of discrepancy in the Number Stroop (pseudo- R^2 = 0.107).

Six clusters were optimal in the Letter Flanker that explained 22.9% of discrepancy. Only Cluster 2 (ASW = 0.611) showed reasonable homogeneity between sequences.

Fig. 6 Observed differences between transition probabilities of cognitive tasks in columns compared to those in rows are shown for pairs of transitions that were significantly extreme with respect to the null distribution of permuted differences

Clusters 1 (ASW = 0.347) and 3 (ASW = 0.346) were less defined but showed more homogeneity than Clusters 4 (ASW = 0.234), 5 (ASW = -0.070), and 6 (ASW = -0.030). Cluster typologies 1 through 4 were similar to those identified in other tasks. The 6 clusters explained between 3.7% to 5.8% of discrepancy in other tasks.

Eight clusters were optimal in the Number Stroop that explained 32.3% of discrepancy. Only Clusters 3 (ASW = 0.642), 4 (ASW = 0.651), and 8 (ASW = 0.599) were reasonably defined. Clusters 1 (ASW = 0.362) and 2 (ASW = 0.203) were less defined but were more specified than the remaining clusters (ASW range = -0.066 to -0.025). Again, cluster typologies 1 through 4 were similar to those in other tasks. The 8 clusters explained between 4.4% to 11.1% of discrepancy in other tasks. Notably, the Number Stroop clusters explained (pseudo- R^2 = 0.111) a moderate amount of discrepancy among sequences from the Arrow Flanker.

Three clusters were optimal in the 2-Back that explained 19.8% of discrepancy. Clusters were generally ill-defined, with Clusters 1 (ASW = 0.295) and 2 (ASW = 0.335) being the most homogenous, whereas Cluster 3 had poor levels of homogeneity (ASW = 0.084). Finally, the 3 clusters were poor

predictors of sequences in other tasks, as they explained only 1.3% to 2.7% of discrepancy.

Clustering of all sequences combined Importantly, hierarchical clustering across all 2406 thought probe sequences combined across tasks suggested that 4 global clusters were optimal. Figure 9 depicts each global cluster typology of sequences and the corresponding proportion of thought probes for clusters. Global clusters together explained 17.1% of the discrepancy among all 2406 sequences. While the structure of clusters was poorly defined (ASW range = 0.057 to 0.179), they each appeared visually distinct and were differentiated based on their proportions of thought probes (Kolmogorov–Smirnov p s < .001). The typology of sequences comprising Cluster 1 reflected sequences with a high proportion of on-task reports. Cluster 2 reflected sequences with a high proportion of task-related interference. Cluster 3 reflected sequences with a high proportion of thoughts about their current physical, cognitive or emotional state of being. Finally, Cluster 4 reflected sequences with a high proportion of worries, daydreams, and other kinds of task-unrelated thoughts. Cluster 4 also appeared more heterogeneous than other clusters. A χ^2

test of independence indicated that the assignment of sequences to clusters were dependent on cognitive task, $\chi^2(12) = 565.94$, $p < .001$. Table 5 provides the number of sequences from each cognitive task that were assigned to clusters. Cluster 3 occurred most frequently in the SART, whereas Cluster 4 occurred most frequently in the Letter Flanker and least frequently in the SART.

Global cluster typologies differed in terms of the five dependent measures characterizing the overall dynamics of sequences. Comparisons between cluster typologies are described below for each measure, and parameter estimates from mixed effects models are given in Table 6. The number of unique thought content categories reported by individuals differed significantly between global clusters, $F(3, 2318.5) = 3.23$, $p = .022$. More unique categories were reported in

Cluster 2 ($b = 0.209$, $p = .031$, 95% CI [0.019, 0.399]), Cluster 3 ($b = 0.313$, $p = .004$, 95% CI [0.097, 0.528]), and Cluster 4 ($b = 0.170$, $p = .036$, 95% CI [0.010, 0.330]), compared to Cluster 1 (intercept $b = 4.202$, 95% CI [4.060, 4.344]). Cluster 1, which reflected those sequences with a high proportion of on-task probe reports, was therefore more homogeneous in thought content than other clusters. Clusters explained little observed variance ($R^2 = .004$) in unique thought content.

The proportion of transitions differed significantly between global clusters, $F(3, 2231.9) = 60.15$, $p < .001$. The proportion of transitions was significantly lower in Cluster 3 ($b = -0.056$, $p < .001$, 95% CI [-0.084, -0.027]) but higher in Cluster 2 ($b = 0.051$, $p < .001$, 95% CI [0.025, 0.076]) and Cluster 4 ($b = 0.099$, $p < .001$, 95% CI [0.078, 0.120]), compared to Cluster 1 (intercept $b = 0.462$, 95% CI [0.442, 0.483]). There were thus fewer transitions among sequences belonging to Cluster 1 and Cluster 3 than the remaining clusters. Clusters explained 5% of the observed variance ($R^2 = .051$) in the proportion of transitions.

Entropy differed significantly between global clusters, $F(3, 2284.9) = 35.53$, $p < .001$. Entropy was significantly higher in Cluster 2 ($b = 0.058$, $p < .001$, 95% CI [0.034, 0.081]) and Cluster 4 ($b = 0.094$, $p < .001$, 95% CI [0.074, 0.114]) compared to Cluster 1 (intercept $b = 0.478$, 95% CI [0.459, 0.496]). Cluster 3 did not significantly differ from Cluster 1 ($b = 0.010$, $p = .455$, 95% CI [-0.017, 0.037]). Thought probe content was therefore more unpredictable in Clusters 2 and 4 compared to other clusters. Figure 10 shows individual participant data for these comparisons, along with the model estimated means for each task. Clusters explained 4% of observed variance ($R^2 = .035$) in entropy.

Complexity differed significantly between global clusters, $F(3, 2229.5) = 49.68$, $p < .001$. Complexity was significantly higher in Cluster 2 ($b = 0.052$, $p < .001$, 95% CI [0.029, 0.075]) and Cluster 4 ($b = 0.094$, $p < .001$, 95% CI [0.074, 0.113]) compared to Cluster 1 (intercept $b = 0.466$, 95% CI [0.448, 0.485]). Cluster 3 did not significantly differ from Cluster 1 ($b = -0.025$, $p = .065$, 95% CI [-0.051, 0.001]). Sequences had more complexity in Clusters 2 and 4 compared to other clusters (see Figure 10). Clusters explained 4% of observed variance ($R^2 = .042$) in complexity.

Finally, turbulence differed significantly between global clusters, $F(3, 2231.5) = 87.91$, $p < .001$. Turbulence was significantly higher in Cluster 2 ($b = 0.042$, $p < .001$, 95% CI [0.019, 0.064]) and Cluster 4 ($b = 0.104$, $p < .001$, 95% CI [0.085, 0.122]) compared to Cluster 1 (intercept $b = 0.476$, 95% CI [0.458, 0.494]). In contrast, Cluster 3 had significantly lower turbulence ($b = -0.066$, $p < .001$, 95% CI [-0.091, -0.040]). Compared to Cluster 1, sequences had more turbulence in Clusters 2 and 4, and less turbulence in Cluster 3 (see Figure 10). Clusters explained 7% of observed variance ($R^2 = .072$) in turbulence.

Fig. 7 Three cluster selection criteria including the point biserial correlation (red), Huberts gamma (light blue), and the weighted average silhouette width (dark blue) are depicted for each cognitive task for cluster solutions 1 through 20. Selection of optimal clusters was determined based on the local maxima of criteria with the smallest number of clusters

Fig. 8 Sequences of experience sampling thought probe responses (rated according to 8 categories) are depicted for optimal clusters of each cognitive task. Clusters were labeled and organized in order according to their visual similarity between tasks

Discussion

The stream of thought rarely follows the same meandering path twice for an individual, but nevertheless flows with some predictability in the cognitive task context. Sequence analytic methods applied to categorical experience sampling thought probes revealed a modicum of order and consistency in streams of thought for individuals based on systematic patterns of transitions and information theoretic measures of entropy, complexity, and turbulence. Individuals were likely to report the same cognitive state or kind of task-unrelated thought on probes which followed their current probe report,

as transition pairs between the same kinds of thought content were more likely to occur than if sequences had been ordered at random. That the same thought probe content was highly likely to re-occur over time also meant that other transition pairs were less likely to systematically occur.

Sequences of thought probes varied more within individuals across different sessions of cognitive tasks than they did between individuals. Participants completed five different cognitive tasks with embedded thought probes across three different cognitive-behavioral assessment sessions. The intra-class correlation coefficient (ICC) calculated from mixed effects models of sequence entropy, complexity, and

Table 4 Proportion of sequence dissimilarity explained by sequence cluster solutions

Clusters	SART	Arrow Flanker	Letter Flanker	Number Stroop	2-Back
SART (<i>N</i>)	526	470	454	469	455
4 Clusters	0.175	0.043	0.054	0.046	0.036
Arrow Flanker (<i>N</i>)	470	479	414	474	456
6 Clusters	0.046	0.295	0.045	0.107	0.049
Letter Flanker (<i>N</i>)	454	414	462	410	399
6 Clusters	0.058	0.052	0.229	0.047	0.037
Number Stroop (<i>N</i>)	469	474	410	478	454
8 Clusters	0.056	0.111	0.044	0.323	0.049
2-Back (<i>N</i>)	455	456	399	454	461
3 Clusters	0.019	0.025	0.013	0.027	0.198

Note: The proportion of discrepancy explained (pseudo- R^2) from analyses of sequence dissimilarity is provided for optimal clusters of sequences from each of five cognitive tasks, as well as the discrepancy explained when cluster membership from one task is applied to other tasks. The sample size (*N*) for the set of individuals with complete pairwise data is also provided

turbulence demonstrated that a minority of variance (< 50%) was explained by variability between individuals, whereas a majority could be attributable to unexplained within-person variability. This was consistent with low ICC calculated from pairwise dissimilarities of entire thought probe sequences. Together, these values indicate poor to fair levels of measurement reliability when considering common standards for interpreting ICC in relation to test–retest reliability (Koo & Li, 2016). These findings also suggest that streams of thought as assessed by sequences of thought probe responses demonstrate considerable variability within and between different task contexts, even minutes later when engaged in another cognitive task.

There were also systematic differences between cognitive tasks in thought probe sequences that contributed to within-person variability between sessions. Differences between tasks explained about 8% of the overall discrepancy in pairwise sequence dissimilarities. Indeed, tasks also differed overall in the frequency distribution of thought probes, the number of unique categories reported, patterns of transition probabilities, and sequence entropy, complexity, and turbulence. The temporal dynamics of thought therefore differed on average between cognitive tasks. These findings are in line with the notion that the demands and features of cognitive tasks are associated with not only the frequency and content of task-unrelated thought, but also *how* entire sequences of thought probe responses unfold over the course of task performance.

Differences between tasks may result from specific features of the cognitive task context, such as the demand and kind of cognitive processes engaged during performance, the timing and frequency of experience sampling probes presented, the total duration of the task in question, and features of stimuli and stimulus presentation. Tasks in the current study varied

along each of these dimensions. Prior research has suggested that the overall frequency of task-unrelated thought is reduced when probes are presented more frequently (Schubert et al., 2019; Seli et al., 2013). There is also some evidence that task demand and probe presentation rates influence the variety of reported thought content and the dynamics of thought probe sequences (Turnbull et al., 2019). At least in the present study, the SART was the longest duration task with the most probes presented (7% of trials). The 45 probes of the SART exhibited the greatest variety of reported thought content and corresponding sequence entropy. Conversely, the Letter Flanker was a shorter task that had the fewest probes (8% of trials), but had content that was the most temporally volatile, demonstrating the greatest proportion of transitions, and greatest sequence complexity and turbulence among tasks. It is also possible that the ordering of tasks over assessments may have contributed to patterns of thought probes within and between tasks. Future studies should continue to examine the influence of probe presentation rate and other task or assessment factors on the dynamics and content of thought beyond simply assessing the overall frequency of probe responses.

In contrast to patterns of within-person variability, there were also a great many similarities between participants overall. Despite having different proportions of specific thought probe content, the distributions of thought content among tasks were also highly similar: on-task reports and task-related interference being the most prevalent. Thoughts about one's own current state, followed by daydreams, were the next most frequent categories of thought. Similarly, the overall transition matrices of transition probabilities between thought probes were highly correlated between tasks. Finally, a number of sequence typologies could be identified through hierarchical clustering of thought probe sequences that explained between 17% to 32% of discrepancy between sequences.

Fig. 9 Sequences of experience sampling thought probe responses (rated according to 8 categories) are depicted based on 2406 sequences in total from all five cognitive tasks. The distribution of thought probe responses

for clusters is also shown. Sequences differ in length because more or less thought probes were presented in different cognitive tasks

Clusters shared a large degree of visual and conceptual similarity between different cognitive tasks.

Table 5 Number of sequences for global cluster typologies by task

Task	1	2	3	4
SART	168	125	177	56
Arrow Flanker	175	84	47	173
Letter Flanker	73	43	15	331
Number Stroop	161	127	56	134
2-Back	133	92	29	207

Note: The number of sequences assigned to global cluster typologies 1 through 4 are provided for each of five cognitive tasks

Four clusters were prevalent throughout tasks: a cluster of individuals who reported a high degree of task-related focus, individuals who reported a high degree of task-related interference, individuals who primarily engaged in mind wandering about their current state, and those who primarily daydreamed. But consistent with broader patterns of within-person variability, individuals tended to be grouped according to different typologies in each cognitive task. Clusters of individuals' sequences in one cognitive task were therefore poor predictors of groupings of sequence dissimilarities for all the other tasks, explaining roughly 1% to 11% of discrepancy in other tasks. Clustering of sequences from all tasks combined also identified common typologies of mind wandering. Seventeen percent of the discrepancy among all thought probe sequences was explained by these broader profiles. Notably,

Table 6 Parameter estimates from mixed effects models for dependent measures of global cluster typologies

Measure	No. Reported	Prop. Transitions	Entropy	Complexity	Turbulence
Fixed Effects					
Cluster 1 (Intercept)	4.202 (0.073)***	0.462 (0.010)***	0.478 (0.009)***	0.466 (0.010)***	0.476 (0.009)***
Cluster 2	0.209 (0.097)*	0.051 (0.013)***	0.058 (0.012)***	0.052 (0.012)***	0.042 (0.011)***
Cluster 3	0.313 (0.110)**	-0.056 (0.015)***	0.010 (0.014)	-0.025 (0.013)	-0.066 (0.013)***
Cluster 4	0.170 (0.081)*	0.099 (0.011)***	0.094 (0.010)***	0.094 (0.010)***	0.104 (0.009)***
Random Effects					
Intercept variance	0.893 (0.945)	0.025 (0.157)	0.017 (0.132)	0.021 (0.145)	0.020 (0.140)
Residual variance	1.967 (1.403)	0.033 (0.181)	0.030 (0.172)	0.027 (0.165)	0.026 (0.160)
Subjects	537	537	537	537	537
Observations	2406	2406	2406	2406	2406

Note: Restricted maximum likelihood estimates are given from mixed effects models of the number of unique thought content categories reported, the proportion of total transitions, and thought probe sequence entropy, complexity, and turbulence for fixed effects of global cluster typology. Cluster 1 serves as the reference condition. The number of subjects (*N*) and observations contributing to the analyses are provided. Standard errors are reported in parentheses for fixed effects and standard deviations for random effects. * $p < .05$, ** $p < .01$, *** $p < .001$

Cluster 1, representing those individuals who reported high levels of task-related focus, were more homogeneous in thought content than other clusters, and tended to display less temporal variability. Cluster 4, however, demonstrated the greatest levels of temporal variability, as sequences had more entropy, complexity, and turbulence than other typologies. Similar typologies of mind wandering therefore emerge on the whole across cognitive task contexts, whereas individuals are highly variable in their patterns of ongoing thought between different cognitive tasks.

In total, the present analyses of the dynamics of thought probe sequences illustrate the conceptual tension between spontaneity and consistency in patterns of thought over time. One's stream of thought, represented by sequences of experience sampling thought probes, appears to show a modicum of consistency in the cognitive task context. But even minutes later these dynamic patterns are likely to differ between task sessions. Yet, on the whole, consistent typologies of mind wandering were identified from individuals according to how their mind wandering unfolds over time. Linz et al. (2019) reported conceptually similar findings when comparing multivariate thought probe ratings between cognitive tasks and daily life. While the same components of thought content appeared invariant across contexts, amounts of task-unrelated thought were inconsistently correlated within individuals (Linz et al., 2019). It is thus possible that there are common temporal structures to ongoing thought that nevertheless vary within individuals from one moment to the next. Given the large degree of variability within and between thought probe sequences across different cognitive tasks, it seems likely that appropriately characterizing the range of contextual constraints on ongoing thought will be challenging for theoretical definitions that define mind wandering according to their

degree of influence (i.e., Christoff et al., 2016; cf. Seli et al., 2018b).

Studies considering only the amount of off-task thought ignore the large range of reported thought content and the considerable dynamics evinced during task performance. One open question is whether thought probe variability is behaviorally meaningful. Studies should continue to explore how mind wandering dynamics predict task performance and relate to individual differences in cognitive abilities over and above the amount of reported mind wandering. Accordingly, several studies have already provided some evidence that patterns of ongoing thought are associated with stable individual differences in cognition (i.e., Turnbull et al., 2019; Welhaf et al., 2019). Given the importance of thought dynamics for understanding how and why mind wandering occurs, it is my hope to have demonstrated the value of sequence analysis and related methods for quantifying and analyzing dynamics of categorical experience sampling thought probe sequences. It is clear to me that researchers ought to consider not only whether individuals experience spontaneous, deliberate, or task-unrelated thoughts, but also the content and qualities of those thoughts, and the kinds of cognitive states that precede and follow.

One limitation of dissimilarity-based sequence analyses and clustering are the many possible choices for quantifying sequence dissimilarity. For dissimilarity based on edit distances, such as optimal matching (OM), there is also the challenge of specifying appropriate indel and substitution costs, which determines how sequences are matched, and can influence sequence dissimilarity by emphasizing certain properties of the data over others (Elzinga & Studer, 2015). Future studies utilizing OM to quantify sequence dissimilarities in thought probes ought to consider edit costs that might be

Fig. 10 Model estimated means and 95% confidence intervals are provided from mixed effects models for sequence entropy, complexity, and turbulence for global clusters of individuals' thought probe sequences for Cluster 1 ($n = 710$), Cluster 2 ($n = 471$), Cluster 3 ($n = 324$), and Cluster 4 ($n = 901$). Observed subject data are plotted from individuals' sequences as dots in gray

interpreted and defined based on explicit theories from the cognitive science of mind wandering. That is, in determining whether two sequences are dissimilar, should the ordering of thoughts in sequence take precedence over the kind of

thoughts reported? OM dissimilarities in general are less sensitive to the fixed order of sequence states in time (Elzinga & Studer, 2015; Lesnard, 2010), and, by determining substitution costs using transition probabilities, sequences are considered more similar if substitutions involve states that were more likely to transition to and from one another. A stream of thought probe responses from $1 \rightarrow 2 \rightarrow 1 \rightarrow 2$, indicating transitions between on-task focus and task-related interference, seems conceptually more similar in the present case to $2 \rightarrow 1 \rightarrow 2 \rightarrow 1$ than it does for a sequence from $1 \rightarrow 6 \rightarrow 1 \rightarrow 6$. Likewise, this is the conclusion based on OM, but alternative specification of indel and substitution costs can change how dissimilarities might be calculated.

It is also true that many experience sampling paradigms employ continuous rating procedures that preclude the use of categorical sequence methods. Indeed, the way thoughts are measured using experience sampling ought to influence emergent temporal patterns and constrain analytic methods. One extension of the general analytic principles employed herein might involve the use of multivariate distance matrix regression (MDMR) and hierarchical clustering of Euclidean distances for time series of continuous thought probe ratings. Markov chain modeling can also be applied to ordinal thought probe ratings to quantify patterns of systematic transitions (see Zanesco, Denkova, Witkin, & Jha, 2019). Acknowledging that the content of thought is not only dynamic but multidimensional, many studies have also begun to collect multivariate ratings from several dimensions of thought for single experience sampling probes (e.g., Kane et al., 2017; Linz et al., 2019; Ruby et al., 2013; Sormaz et al., 2018; Turnbull et al., 2019; Wang et al., 2018). These studies have commonly employed principle components analysis to reduce multivariate ratings into dimensional components for analysis in a data-driven manner. One alternative is to use mixed effects MDMR to analyze entire sets of multivariate ratings without first reducing ratings into components, or by categorizing sets into distinct profiles of thought according to clustering algorithms. Thought profiles, in turn, might be analyzed as a time series sequence of categorical states. Multichannel sequence analysis (e.g., Gauthier, Widmer, Bucher, & Notredame, 2010) and Markov modeling (e.g., Helske, Helske, & Eerola, 2018) may also be useful when applied to multivariate thought probe ratings.

More broadly, there are many viable applications of sequence analysis for investigating time ordered neurocognitive and behavioral categorical processes within other domains of psychological research. There is great interest in applying functional neuroimaging and electroencephalography techniques to understand brain dynamics and their links to spontaneous thought (Andrews-Hanna, Smallwood, & Spreng, 2014; Christoff et al., 2016, 2018; Turnbull et al., 2019; Wang et al., 2018). OM-based sequence analysis might be used to examine the links between ongoing thought and

dynamic patterns in time series of categorical brain states identified in time-varying functional magnetic resonance imaging data (i.e., dynamic functional connectivity; Lurie et al., 2020), or to analyze more complex sequences of topographically defined brain electric microstates (e.g., Zanesco, King, Skwara, & Saron, 2020). OM-based sequence analysis is also well suited to analyzing other kinds of categorical data collected in experience sampling studies, such as behavioral sequences in daily life. Finally, these techniques might be applied in clinically oriented research to identify and characterize patterns of rumination and repetitive thinking in individuals' thought streams (Smith & Alloy, 2009).

Other methods that can account for the time-varying nature of thoughts should continue to be used in studies of mind wandering. Past studies have employed latent growth curve models to examine systematic trends in dimensions of thought over time (Smith et al., 2018), examined time-lagged associations between thought reports (Ruby et al., 2013), quantified the semantic distance between one instance of thought to another (Gray et al., 2018), reconstructed thought trajectories over time (Nakatani, Ganschow, & van Leeuwen, 2019), or employed hidden Markov modeling methods to characterize mind wandering dynamics and the response-generating processes underlying thought probe sequences (Bastian & Sackur, 2013; Zanesco et al., 2019). Even more straightforward, but no less important, are approaches quantifying counts of transitions of different thought content (Welhaf et al., 2019), or those that add new dimensions of responses to thought probes such as by asking individuals to reflect and report on their own thought dynamics, such as whether thoughts are experienced as freely-moving (Mills, Raffaelli, Irving, Stan, & Christoff, 2018) or evolving (Sormaz et al., 2018; Wang et al., 2018). Alongside sequence analysis, these methods have much to offer the cognitive science of mind wandering.

Acknowledgements I thank Matthew S. Welhaf, Michael J. Kane, and their colleagues, for providing their data online (<https://osf.io/guhw7/>) and encouraging secondary analyses by interested researchers. I also thank Brandon G. King and Michael J. Kane for thoughtful comments on an earlier version of this manuscript.

Open practices statement The data that support the findings of this study are openly available in the OSF repository and can be found at: <https://osf.io/mjk3t/>

References

- Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological Methods & Research*, 29(1), 3–33. <https://doi.org/10.1177/0049124100029001001>.
- Andrews-Hanna, J. R., Smallwood, J., & Spreng, N. R. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, 1316(1), 29–52. <https://doi.org/10.1111/nyas.12360>
- Bastian, M., & Sackur, J. (2013). Mind wandering at the fingertips: automatic parsing of subjective states based on response time variability. *Frontiers in Psychology*, 4, 573. <https://doi.org/10.3389/fpsyg.2013.00573>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, N. R., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews Neuroscience*, 17, 718–731. <https://doi.org/10.1038/nrn.2016.113>
- Christoff, K., Mills, C., Andrews-Hanna, J. R., Irving, Z. C., Thompson, E., Fox, K. C. R., & Kam, J. W. Y. (2018). Mind-wandering a scientific concept: Cutting through the definitional haze. *Trends in Cognitive Sciences*, 22(11), 957–959. <https://doi.org/10.1016/j.tics.2018.07.004>
- Cornwell, B., Gershuny, J., & Sullivan, O. (2020). The social structure of time: Emerging trends and new directions. *Annual Review of Sociology*, 45, 301–320. <https://doi.org/10.1146/annurev-soc-073018-022416>
- Elzinga, C. H. (2010). Complexity of categorical time series. *Sociological Methods & Research*, 38(3), 463–481. <https://doi.org/10.1177/0049124109357535>
- Elzinga, C. H., & Liefbroer, A. C. (2007). De-standardization of family-life trajectories of young adults: A cross-national comparison using sequence analysis. *European Journal of Population*, 23, 225–250. <https://doi.org/10.1007/s10680-007-9133-7>
- Elzinga, C. H., & Studer, M. (2015). Spell sequences, state proximities, and distance metrics. *Sociological Methods & Research*, 44(1), 3–47. <https://doi.org/10.1177/0049124114540707>
- Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software* 40(4), 1–37. <https://doi.org/10.18637/jss.v040.i04>
- Gauthier, J. A., Widmer, E. D., Bucher, P., & Notredame, C. (2010). Multichannel sequence analysis applied to social science data. *Sociological Methodology*, 40(1), 1–38. <https://doi.org/10.1111/j.1467-9531.2010.01227.x>
- Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., ... Lewis, K. (2018). “Forward flow”: A new measure to quantify free thought and predict creativity. *American Psychologist*, 74(5), 539–554. <https://doi.org/10.1037/amp0000391>
- Helske, S., & Helske, J. (2019). Mixture hidden Markov models for sequence data: The seqHMM package in R. *Journal of Statistical Software*, 88(3), 1–32. <https://doi.org/10.18637/jss.v088.i03>
- Helske, S., Helske, J., & Eerola, M. (2018). Combining sequence analysis and hidden Markov models in the analysis of complex life sequence data. In G. Ritschard & M. Studer. (Eds.). *Sequence analysis and related methods: Innovative methods and applications* (pp 185–200). Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-319-95420-2>
- Irving, Z. C. (2016). Mind-wandering is unguided attention: Accounting for the “purposeful” wanderer. *Philosophical Studies*, 173(2), 547–571. <https://doi.org/10.1007/s11098-015-0506-1>
- Irving, Z. C., & Thompson, E. (2018). The philosophy of mind-wandering. In K. C. R. Fox & K. Christoff (Eds.). *The Oxford handbook of spontaneous thought: Mind-wandering, creativity, and dreaming* (pp. 87–96). New York, NY: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190464745.013.19>
- James, W. (1890). *The Principles of Psychology*. New York: Henry Holt and Company.
- Kane, M. J., & McVay, J. C. (2012). What mind wandering reveals about executive-control abilities and failures. *Current Directions in*

- Psychological Science*, 21(5), 348–354. <https://doi.org/10.1177/0963721412454875>
- Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science*, 18(7), 614–621. <https://doi.org/10.1111/j.1467-9280.2007.01948.x>
- Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General*, 145(8), 1017–1048. <https://doi.org/10.1037/xge0000184>
- Kane, M. J., Gross, G. M., Chun, C. A., Smeekens, B. A., Meier, M. E., Silvia, P. J., & Kwapil, T. R. (2017). For whom the mind wanders, and when, varies across laboratory and daily-life settings. *Psychological Science*, 28(9), 1271–1289. <https://doi.org/10.1177/0956797617706086>
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Hoboken, New Jersey: John Wiley & Sons Inc.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Krimsky, M., Forster, D. E., Llabre, M. M., & Jha, A. P. (2017). The influence of time on task on mind wandering and visual working memory. *Cognition*, 169, 84–90. <https://doi.org/10.1016/j.cognition.2017.08.006>
- Lesnard, L. (2010). Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns. *Sociological Methods & Research*, 38(3), 389–419. <https://doi.org/10.1177/0049124110362526>
- Linz, R., Pauly, R., Smallwood, J., & Engert, V. (2019). Mind-wandering content differentially translates from lab to daily life and relates to subjective stress experience. *Psychological Research*. <https://doi.org/10.1007/s00426-019-01275-2>
- Lüdtke, D. (2019). sjstats: Statistical functions for regression models. R package version 0.17.4. <https://CRAN.R-project.org/package=sjstats>
- Lurie, D. J., Kessler, D., Bassett, D. S., Betzel, R. F., Breakspear, M., Kheiholz, S., ... Calhoun, V. D. (2020). Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Human Brain Mapping*, 41(1), 30–69. https://doi.org/10.1162/netn_a_00116
- McArdle, B. H., & Anderson, M. J. (2001). Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology*, 82(1), 290–297. [https://doi.org/10.1890/0012-9658\(2001\)082\[0290:FMMTCJ\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2001)082[0290:FMMTCJ]2.0.CO;2)
- McArtor, D. B. (2018). MDMR: Multivariate Distance Matrix Regression. R package version 0.5.1. <https://CRAN.R-project.org/package=MDMR>
- McArtor, D. B., Lubke, G. H., & Bergeman, C. S. (2017). Extending multivariate distance matrix regression with an effect size measure and the asymptotic null distribution of the test statistic. *Psychometrika*, 82(4), 1052–1077. <https://doi.org/10.1007/s11336-016-9527-8>
- McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, 141(2), 302–320. <https://doi.org/10.1037/a0025250>
- McVay, J. C., Kane, M. J., & Kwapil, T. R. (2009). Tracking the train of thought from the laboratory into everyday life: An experience-sampling study of mind wandering across controlled and ecological contexts. *Psychonomic Bulletin & Review*, 16(5), 857–863. <https://doi.org/10.3758/PBR.16.5.857>
- 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. <https://doi.org/10.1016/j.concog.2017.10.003>
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14, 20170213. <https://doi.org/10.1098/rsif.2017.0213>
- Nakatani, C., Ganschow, B., & van Leeuwen, C. (2019). Long-term dynamics of mind wandering: ultradian rhythms in thought generation. *Neuroscience of Consciousness*, 2019(1), niz007. <https://doi.org/10.1093/nc/niz007>
- Ritschard, G., & Studer, M. (Eds.) (2018). *Sequence analysis and related methods: Innovative methods and applications*. Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-319-95420-2>
- Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). ‘Oops!’: Performance correlations of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35(6), 747–758. [https://doi.org/10.1016/S0028-3932\(97\)00015-8](https://doi.org/10.1016/S0028-3932(97)00015-8)
- Robison, M. K., & Unsworth, N. (2018). Cognitive and contextual correlates of spontaneous and deliberate mind-wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(1), 85–98. <https://doi.org/10.1037/xlm0000444>
- Ruby, F. J. M., Smallwood, J., Engen, H., & Singer, T. (2013). How self-generated thought shapes mood—The relation between mind-wandering and mood depends on the socio-temporal content of thoughts. *PLoS ONE* 8(10): e77554. <https://doi.org/10.1371/journal.pone.0077554>
- Sankoff, D., & Kruskal, J. B. (1983). *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. London, England: Addison-Wesley.
- Schubert, A. L., Frischkorn, G. T., & Rummel, J. (2019). The validity of the online thought-probing procedure of mind wandering is not threatened by variations of probe rate and probe framing. *Psychological Research*. <https://doi.org/10.1007/s00426-019-01194-2>
- Seli, P., Carriere, J. S., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology*, 4:430. <https://doi.org/10.3389/fpsyg.2013.00430>
- Seli, P., Carriere, J. S., Wammes, J. D., Risko, E. F., Schacter, D. L., & Smilek, D. (2018a). On the clock: Evidence for the rapid and strategic modulation of mind wandering. *Psychological Science*, 29(8), 1247–1256. <https://doi.org/10.1177/0956797618761039>
- Seli, P., Kane, M. J., Metzinger, T., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., & Smilek, D. (2018b). The family-resemblances framework for mind-wandering remains well clad. *Trends in Cognitive Sciences*, 22(11), 959–961. <https://doi.org/10.1016/j.tics.2018.07.007>
- Seli, P., Kane, M. J., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., & Smilek, D. (2018c). Mind-wandering as a natural kind: A family-resemblances view. *Trends in Cognitive Sciences*, 22(6), 479–490. <https://doi.org/10.1016/j.tics.2018.03.010>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.
- Smallwood, J. (2013). Distinguishing how from why the mind wanders: A process-occurrence framework for self-generated mental activity. *Psychological Bulletin*, 139(3), 519–535. <https://doi.org/10.1037/a0030010>
- Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, 66, 487–518. <https://doi.org/10.1146/annurev-psych-010814-015331>

- Smith, G. K., Mills, C., Paxton, A., & Christoff, K. (2018). Mind-wandering rates fluctuate across the day: Evidence from an experience-sampling study. *Cognitive Research: Principles and Implications*, 3:54. <https://doi.org/10.1186/s41235-018-0141-4>
- Smith, J. M., & Alloy, L. B. (2009). A roadmap to rumination: A review of the definition, assessment, and conceptualization of this multifaceted construct. *Clinical Psychology Review*, 29(2), 116–128. <https://doi.org/10.1016/j.cpr.2008.10.003>
- Sormaz, M., Murphy, C., Wang, H., Hymers, M., Karapanagiotidis, T., Poerio, G., ... Smallwood, J. (2018). Default mode network can support the level of detail in experience during active tasks states. *Proceedings of the National Academy of Sciences of the United States of America*, 115(37), 9318–9323. <https://doi.org/10.1073/pnas.1721259115>
- Studer, M. (2013). WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. *LIVES Working Papers*, 24. <https://doi.org/10.12682/lives.2296-1658.2013.24>
- Studer, M., & Ritschard, G. (2016). What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society*, 179(2), 481–511. <https://doi.org/10.1111/rssa.12125>
- Studer, M., Ritschard, G., Gabadinho, A., Müller, N. S. (2011). Discrepancy analysis of state sequences. *Sociological Methods & Research*, 40(3), 471–510. <https://doi.org/10.1177/0049124111415372>
- Thomson, D. R., Seli, P., Besner, D., & Smilek, D. (2014). On the link between mind wandering and task performance over time. *Consciousness and Cognition*, 27, 14–26. <https://doi.org/10.1016/j.concog.2014.04.001>
- Tumbull, A., Wang, H., Schooler, J. W., Jefferies, E., Margulies, D. S., & Smallwood, J. (2019). The ebb and flow of attention: Between-subject variation in intrinsic connectivity and cognition associated with the dynamics of ongoing experience. *Neuroimage*, 185, 286–299. <https://doi.org/10.1016/j.neuroimage.2018.09.069>
- Unsworth, N., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica*, 150, 14–25. <https://doi.org/10.1016/j.actpsy.2014.04.001>
- Wang, H., Poerio, G., Murphy, C., Bzdok, D., Jefferies, E., & Smallwood, J. (2018). Dimensions of experience: Exploring the heterogeneity of the wandering mind. *Psychological Science*, 29(1), 56–71. <https://doi.org/10.1177/0956797617728727>
- Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior Research Methods*, 50, 642–661. <https://doi.org/10.3758/s13428-017-0891-9>
- Welhaf, M. S., Smeeckens, B. A., Gazzia, N. C., Perkins, J. B., Silvia, P. J., Meier, M. E., ... Kane, M. J. (2019). An exploratory analysis of individual differences in mind wandering content and consistency. *Psychology of Consciousness: Theory, Research, and Practice*. <https://doi.org/10.1037/cns0000180>
- Zanesco, A. P., Denkova, E., Witkin, J. E., & Jha, A. P. (2019). Experience sampling of the degree of mind wandering distinguishes hidden attentional states. *PsyArXiv*. <https://doi.org/10.31234/osf.io/sxdrw>
- Zanesco, A. P., King, B. G., Skwara, A. C., & Saron, C. D. (2020). Within and between-person correlates of the temporal dynamics of resting EEG microstates. *NeuroImage*, 211, 116631. <https://doi.org/10.1016/j.neuroimage.2020.116631>
- Zapala, M. A., & Schork, N. J. (2012). Statistical properties of multivariate distance matrix regression for high-dimensional data analysis. *Frontiers in Genetics*, 3:190. <https://doi.org/10.3389/fgene.2012.00190>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.