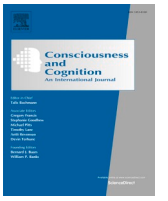
Consciousness and Cognition 128 (2025) 103815

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/yccog



Full Length Article

Opening the black box: Think Aloud as a method to study the spontaneous stream of consciousness

Anusha Garg [a,*], Shivang Shelat [a], Madeleine E. Gross [a], Jonathan Smallwood [b], Paul Seli [c], Aman Taxali [d], Chandra S. Sripada [d], Jonathan W. Schooler [a]

[a] Department of Psychological and Brain Sciences, University of California Santa Barbara, CA, USA
[b] Department of Psychology, Queen's University, Ontario, Canada
[c] Department of Psychology and Neuroscience, Duke University, North Carolina, USA
[d] Department of Psychiatry and Philosophy, Michigan Medicine, University of Michigan, MI, USA

ARTICLE INFO

Keywords:
Resting state
Think Aloud
Stream of consciousness
Mind wandering
Thinking
Text analytics
Natural language processing

ABSTRACT

Asking participants to Think Aloud is a common method for studying conscious experience, but it remains unclear whether this approach alters thought qualities—such as meta-awareness, rate of topic shifts, or the content of thoughts in task-absent conditions. To investigate this, we conducted two studies comparing thinking aloud to thinking silently. In Study 1, 111 participants alternated between 15-minute intervals of verbalizing and silently reflecting on their stream of consciousness in a counterbalanced design. A subset also reported topic shifts intermittently via self- and probe-catching methods. Results showed that the stream of consciousness was minimally reactive to the Think Aloud protocol, with no significant differences in meta-awareness and topic shifting rates. Moreover, among 21 thought qualities and 18 content topics analyzed, only three qualities (private thoughts, mind blanking, and session difficulty) and one topic (partner, intimacy, love, and sexual matters) differed between Think Aloud and Silent Think. In Study 2, 102 participants either did Think Aloud or Silent Think while responding to thought probes. Findings replicated the lack of differences in the frequency and meta-awareness of topic shifts between Think Aloud and Silent Think. Furthermore, no differences in reported cognitive load were observed between the two conditions. These results emphasize the value of the Think Aloud procedure for examining the stream of consciousness, demonstrating its reliability and minimal impact on the natural flow of thoughts. Thus, Think Aloud offers a robust model system for examining the otherwise unverbalized stream of consciousness in task-absent contexts.

1. Opening the black box: Think Aloud as a method to study the stream of consciousness

The stream of consciousness is a ubiquitous phenomenon, and there is substantial scientific interest in better understanding its nature, content, and qualities. The Think Aloud protocol, wherein participants verbalize their ongoing and continuous stream of consciousness, is a promising technique to assess the stream of consciousness. However, it is unclear if the natural flow of thoughts is interfered with and/or modulated in its structure, content, and qualities via the process of verbalizing the stream of consciousness—i.e., reactivity. Here, across two studies, we compare the Think Aloud methodology to the Silent Think method in naturalistic (i.e., task-

* Corresponding author at: Dept. of Psychological and Brain Sciences, 251, University of California Santa Barbara, Santa Barbara, USA.
E-mail address: anushagarg@ucsb.edu (A. Garg).

https://doi.org/10.1016/j.concog.2025.103815
Received 22 October 2024; Received in revised form 4 January 2025; Accepted 6 January 2025
Available online 15 January 2025

absent) conditions.

The Spontaneous Stream of Thought, hereafter referred to as SST, represents the images, ideas, memories, and thoughts that are spontaneously generated and generally follow an exploratory and meandering pattern of thinking (Christoff et al., 2011, 2016; Irving, 2016; Irving & Glasser, 2020; Mildner & Tamir, 2019; Schooler et al., 2011; Singer, 1975; Smallwood & Schooler, 2015). It has previously been studied using various techniques including thought probes, self-catching, experience sampling, and retrospective questionnaires.

Thought probes are administered by periodically interrupting a participant during a task to ask questions about the type of thoughts they had been experiencing in the past few seconds (Weinstein, 2017; see also triggered probes in Shelat et al., 2024). Self-catching, on the other hand, depends on participants' meta-awareness of their SST to determine whether they were on task/off task, and/or the types of thoughts they experienced (Seli et al., 2017). Experience sampling is similar to thought probes, however, participants are probed throughout their daily lives (e.g., via a smartphone app) rather than being constrained in a lab setting (McVay et al., 2009). Finally, retrospective questionnaires ask about qualities of thought once the task is completed (Weinstein, 2017). Each of these methods has offered valuable insights into SST, but all are constrained in their ability to capture its dynamic nature. Thought probes and experience sampling illuminate only brief moments of thoughts, while self-catching and retrospective questionnaires depend heavily on participants' memory and meta-awareness, which are prone to error. As a result, these methodologies provide only temporally isolated glimpses of thought content and qualities, limiting their capacity to fully capture the nature of SST.

Think Aloud (Ericsson & Simon, 1980, 1984, 1993) represents an alternative methodology that potentially offers a direct window into the content and qualities of spontaneous thoughts by having individuals verbalize their SST without requiring them to introspect (Ericsson and Fox, 2011). In Think Aloud, participants are encouraged to let their thoughts flow freely without interruptions, unlike the thought probes technique, which inherently disrupts this process. This uninterrupted flow more closely mirrors everyday thinking and better captures the continuous, serial structure of thoughts (Sripada & Taxali, 2020). As a result, Think Aloud addresses a key limitation of thought probes by avoiding interruptions to the thought processes. Think Aloud also potentially avoids the memory biases that may occur in retrospective questionnaires by capturing thoughts in real time. By not relying on participants' meta-reflections of their thoughts, Think Aloud may offer a more accurate characterization of the content and qualities of thoughts as they naturally occur.

Think Aloud has been especially useful in identifying thought processes associated with a concurrent experimenter-imposed task. The method has been used widely, including in studies examining the content of working memory, problem-solving, and human-technology interactions (e.g. Chein & Weisberg, 2014; de Groot, 1978; Fleck, 2008; Newell, 1972; Nielsen & Madsen, 2012; Paromita et al., 2023; Shelat, Marquez, Zheng, & Karasinski, 2024; Siddiq & Scherer, 2017). Within these contexts, researchers have examined the potential reactivity of the method, in particular whether Think Aloud interferes with cognitive processing or task performance.

In their initial conceptualization of Think Aloud, Ericsson and Simon (1980) distinguished between varying degrees of concurrent verbalizations and their impacts on cognitive processing. They postulated that vocalizing thoughts as they occur—i.e., without any interruptions to the thought process—would not impede cognitive processing. On the other hand, cognitive processing may be affected when participants are asked to reflect on their thought processes and explain their thoughts while performing a task (Ericsson & Simon, 1984, 1993). Reinforcing this dissociation, a meta-analysis showed minimal differences when participants merely vocalized thoughts without introspection. However, instructing participants on the specific content and thought processes to verbalize impacted task performance and completion times (Fox et al., 2011).

Collectively, research on the Think Aloud methodology suggests that it has minimal effects on the stream of consciousness (but see Schooler et al., 1993; Schooler, 2011). Although a potentially valuable technique for investigating SST, few studies have examined SST's reactivity to Think Aloud in contexts without a concurrent task. It is possible that thinking aloud without any primary task to engage in may introduce its own set of constraints. For example, as will be discussed, preliminary evidence by Sripada and Taxali (2020) suggests that thinking aloud may reduce topic shifting potentially by increasing participants' awareness of the sequential patterns in which thoughts emerge. If SST is reactive to Think Aloud, it highlights an important caveat when using verbalization to study the flow of thought in task-absent conditions. However, if SST is not reactive to Think Aloud, it instead validates the method's utility in investigating the nature of the stream of consciousness. Therefore, the overarching aim of the current work was to comprehensively examine the ways in which Think Aloud induces reactivity in the structure, content, and qualities of the stream of consciousness by comparing it to the naturally experienced silent thought flow.

2. Study 1

2.1. Introduction

SST may be reactive to Think Aloud on several different dimensions. For Study 1, we sought to compare Think Aloud to Silent Think on meta-awareness, topic shifts, and assorted thought qualities, and content topics.

2.1.1. Meta-awareness and topic shifts

Sripada and Taxali (2020) investigated the structure of the stream of consciousness using Think Aloud protocols, discovering a 'clump-and-jump' pattern where clusters of related thoughts are followed by jumps to new topics. Although the authors validated this pattern in a secondary pilot study using only the Silent Think protocol, participants reported higher rates of topic shifts in Silent Think compared to Think Aloud. This difference in topic shifting rates should be interpreted with caution for two reasons: a) they did not directly compare Think Aloud to the Silent Think protocol in the pilot study and b) the topic shifts were measured objectively via raters

in the Think Aloud study, and subsequently measured subjectively using self-catching in the pilot Silent Think study. Nonetheless, this potential difference in topic shifting rates, if genuine, could be a result of heightened meta-awareness—i.e., the ability to explicitly note the current contents of consciousness (Schooler, 2002; Schooler et al., 2011) during Think Aloud due to participants actively listening to themselves talk.

A subsequent investigation was also suggestive of differences in meta-awareness between Think Aloud and Silent Think. In Li et al. (2021), participants had a lower frequency of self-caught topic shifts in the Think Aloud condition compared to the Silent Think condition. These findings, contrary to Sripada and Taxali (2020), are suggestive of reduced meta-awareness when participants are asked to Think Aloud. Although Li et al's study offered important insights into the potential differences between Think Aloud and Silent Think, the authors noted critical limitations to their study design. Importantly, the study used a within-subjects design with three conditions, so participants may have been fatigued towards the end due to continuous verbalization. This concern is underscored by the absence of task order counterbalancing, which raises the possibility of order effects. Specifically, the authors were unable to test whether the increased workload of thinking aloud and self-catching topic shifts or the potential fatigue experienced by participants mediated the reduction in self-caught topic shifts. Our study, although conducted prior to the publication of Li et al, addressed this limitation by counterbalancing the conditions to account for participant fatigue.

Additionally, we employed a dual approach to measure meta-awareness and topic shifting using both self-caught and probe-caught measures (Sayette et al., 2009; Sayette et al., 2010). This combination has typically been used in mind wandering research, where thought probes serve as a way to measure the overall rate of mind wandering and self-caught monitoring captures episodes of mind wandering before the thought probes catch it, measuring meta-awareness. In the current study, we intermittently probed participants about their topic shifts and asked them to self-monitor for topic shifts. We hypothesized that thinking aloud might lead to more self-caught topic switches due to increased meta-awareness, while thinking silently might result in greater probe-caught topic shifting, reflecting a relative lack of meta-awareness.

2.1.2. Reactivity in qualities and content of thoughts

Our secondary exploratory analyses examined how Think Aloud compared to Silent Think might alter the stream of consciousness. In addition to comparing topic shifting rates, Li et al. (2021) also examined qualitative differences for the temporality, self- or other-focus, emotional valence, and mental experience form of thoughts in the stream of consciousness between the two modalities. They found no differences across these four measures. Our research extends this work by examining 21 thought qualities and 18 thought content categories differences between Think Aloud and Silent Think.

In expressive writing and talk therapy literature, studies have shown that thinking aloud or writing encourages participants to organize and integrate their thoughts (i.e., "narration creation") by converting them into language, which is highly structured (Singer, 2004; Smyth et al., 2001). Thinking silently, on the other hand, is more disorganized because it carries non-verbalizable emotions, images, and memories in addition to language (Lyubomirsky et al., 2006). It is important to note that these findings were established in conditions with no concurrent tasks, as opposed to the task-based Think Aloud paradigms that Ericsson and Simon (1980, 1984, 1993) initially assessed. Language—whether spoken or written—may enhance meta-awareness, as suggested by the finding that inner speech was associated with increased meta-awareness of mind wandering relative to visual and auditory imagery (Bastian et al., 2017). Accordingly, we speculated that Think Aloud might afford a degree of control over thoughts through language, such that thoughts would be more goal-directed, productive, and purposeful than silent thoughts.

However, we also considered that the controlled nature of Think Aloud might reduce divergent thinking. Mind wandering, an explorative state, can yield creative insights (Sripada, 2018) and enhance divergent thinking during incubation periods (Baird et al., 2012). Additionally, topic shifting during rest periods correlates positively with originality in divergent thinking tasks (Raffaelli et al., 2023; Schooler et al., 2011). Thinking aloud, on the other hand, has been shown to impair the solving of insight problems (Schooler et al., 1993). Despite some mixed evidence (Murray et al., 2021; Smeekens & Kane, 2016; Steindorf et al., 2021), we hypothesized that the controlled nature of Think Aloud might lead to fewer topic shifts, not reflecting the explorative nature of mind wandering as Silent Think would. Additionally, we hypothesized that participants might experience a decrease in curiosity-driven and creative thoughts in Think Aloud compared to Silent Think. Furthermore, we hypothesized that participants might report a more positive mood in Think Aloud than in Silent Think. Ruminative thought patterns—repetitive, intrusive, and negatively valenced—are linked to negative outcomes like dysphoric mood and depression (Lyubomirsky & Tkach, 2003; Nolen-Hoeksema, 2003; Watkins & Roberts, 2020). Similarly, mind wandering is associated with negative affect (Killingsworth & Gilbert, 2010; Smallwood et al., 2009; Stawarczyk et al., 2013), characterized by a discursive and unorganized nature of thoughts. Verbalizing thoughts may help individuals recognize, process and move on from negative thoughts (Lyubomirsky et al., 2006), thereby increasing positive mood compared to thinking silently.

2.1.3. Potential impact of probes

With the inclusion of thought probes in the current study, we created an experiment group that was given the combination of the probes (self- and probe-caught topic shifts) and a control group that was not given any probes. This was done to assess the reactivity of thought probes on thought content and qualities, allowing us to investigate whether the inclusion of probes changes the nature of mind wandering. Given that past research has indicated little reactivity upon the inclusion of thought probes (Kane et al., 2021; Robinson et al., 2019; Wiemers & Redick, 2019) we did not anticipate substantial effects here but sought to rule it out in the present context. Similarly, we included counterbalancing, wherein half the participants were asked to Think Aloud first, and the other half were asked to engage in Silent Mind Wandering first. While we did not make any specific predictions for this dimension of task order, we included this condition to rule out the possibility of order effects.

2.2. Method

2.2.1. Participants and design

122 undergraduate students were recruited from the University of California, Santa Barbara via the Psychological and Brain Sciences Department Subject Pool. All participants received course credit for completing the study. Participants were excluded if they had missing data (N = 5) or failed to follow instructions (N = 6), resulting in a total of 111 participants. They were divided into 4 experimental groups: Think Aloud first With Probes (N = 23), Silent Think first With Probes (N = 35), Think Aloud first Without Probes (N = 31), and Silent Think first Without Probes (N = 22).

2.2.2. Materials

2.2.2.1. Thought qualities Inventory. A list of 21 questions were created based on mind wandering studies that frequently probe participants about thought qualities during tasks that induce a mind wandering state (e.g., Smallwood et al., 2011; Stawarczyk et al., 2011; Stawarczyk et al., 2013). Some example items include questions about goal-directed thoughts, "How often were your thoughts goal-directed?", boredom, "How bored were you during the session?", and creative thoughts, "How often did you have new and/or creative thoughts?". All items were measured on a five-point Likert scale, with responses ranging from never/not at all (1) to a great deal (5). Two items deviated from the 5-point Likert scale: The first item asked participants about their perceived reason for shifting between topics, such as due to boredom. The second item inquired about the medium in which participants experienced their thoughts, such as inner speech or as images. Finally, participants who completed the Think Aloud session responded to three additional items, referencing their experience of thinking aloud on their stream of consciousness— "How natural did you find it to talk out loud?", "How filtered were you in your verbalization of the protocol?", and "To what degree were you unable to vocalize your thoughts?". For the full list of items, please refer to Appendix A.

2.2.2.2. Positive and negative affect Schedule. The 20-item PANAS measures mood and feelings. It includes two subscales, namely the positive and negative attribute subscales (Watson et al., 1988). Each subscale contains ten positive and negative markers of emotion, measured on a five-point Likert scale (1 = very slightly/not at all – 5 = extremely), in accordance with items that represent affect like "Active", and "Upset".

2.2.2.3. 18 Categories Coding Scheme. Developed as an adaptation of the Personal Concerns Inventory (Klinger and Cox, 2004), the 18 Categories Coding Scheme consists of a comprehensive list of the 'categories' of thought content experienced in the stream of consciousness. This was created using the dataset of Think Aloud transcripts (Sripada & Taxali, 2020, Garg et al., *in preparation*). Within each transcript, human raters assigned "jumps" whenever they thought the participant had switched topics, creating blocks of thoughts that related to one topic each. These blocks were then given categories, thereby forming 18 such categories of the most frequently recurring topics in the stream of consciousness across all participants. Each category consists of a definition, along with keywords that represent the category. An example item is "Education – Relating to college, and classes. Keywords – slacking off, homework, major, essay, exam, quiz, teacher, professor, minor, course, grades, deadlines, attendance, syllabus, lecture, scholarship, awards, PhD, Masters, undergraduate, freshman, sophomore, junior, senior, seminar, presentation, conference." (see Appendix B).

2.2.2.4. Keywords. Participants were asked to recall and type keywords for all the topics they thought of during the Think Aloud and Silent Think sessions—one keyword each representing the topic. These words were allowed to be whatever the participants felt best represented the topic. This data was analyzed and coded by human raters, who assigned each of the words to one of the categories in the 18 Categories Coding Scheme. An example of this would be the keyword "test" being assigned the category of "Education". The purpose of the keywords was to assess if there were any content topics that were not included in the 18 Categories Coding Scheme.

2.2.2.5. Individual difference measures. All participants completed a battery of self-report measures, namely the Adult Self Report Scale-Screener (ASRS-Screener, Kessler et al., 2005), NEO Five Factor Inventory (NEO-FFI, McCrae & Costa, 2004), Daydreaming Frequency Scale (DDFS, Giambra, 1993), Short Boredom Proneness Scale (SBPS, Struk et al., 2017), Mind Wandering Questionnaire (MW-S and MW-D, Carriere et al., 2013), Five-Dimensional Curiosity Scale (FDC, Kashdan et al., 2018), Five Facet Mindfulness Questionnaire (FFMQ, Baer et al., 2006), Aberrant Salience Inventory (ASI, Cicero et al., 2010), Self Talk Scale (STS, Brinthaupt et al., 2009), and Jourard Self Disclosure Questionnaire (JSDQ, Jourard & Lasakow, 1958). The scope of the current paper does not include a discussion of the results from these measures and will instead be reported elsewhere in a different paper.

2.2.3. Procedure

Due to COVID-19 restrictions, the study was conducted over Zoom. Participants assured the experimenter they were in a quiet, secluded space before the study began. After the participants completed informed consent, the experimenter explained the session instructions as adapted from Sripada & Taxali (2020). They emphasized the need for participants to let their thoughts flow naturally, similar to situations like a long, boring bus ride or a shower. Participants were instructed to express their thoughts exactly as they occurred, whether in short sentences, words, or random interruptions.

In the Think Aloud condition, participants verbalized their stream of consciousness for 15 min. Those in the With Probes condition responded to periodic thought probes and self-caught topic shifting. Thought probes, administered every minute, asked participants to

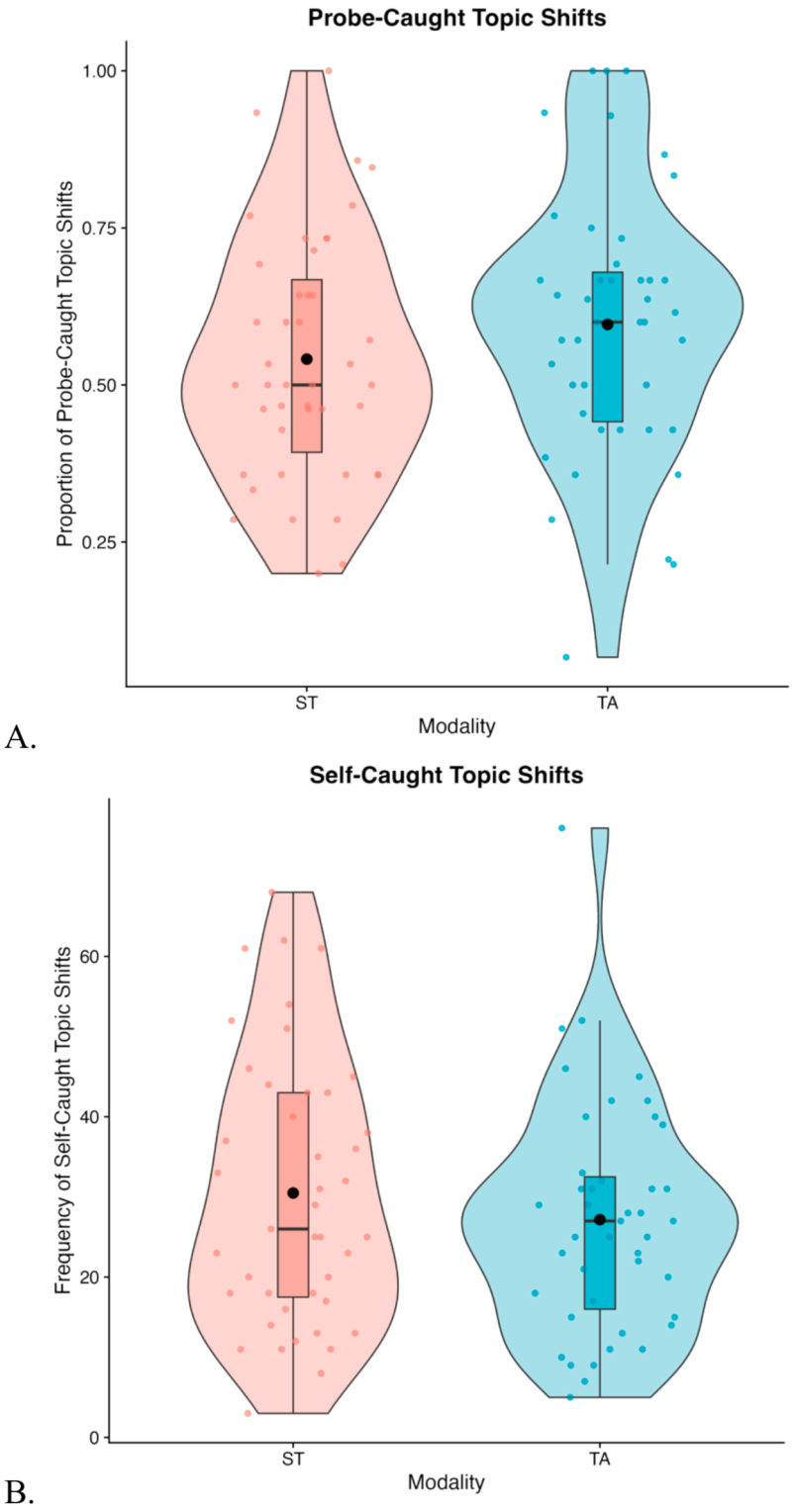


Fig. 1. Raincloud plot including half violin plots for each modality, boxplots displaying the interquartile range and median (horizontal line), jittered points representing individual data points, and mean (black circles). Here, ST = Silent Think and TA = Think Aloud. **a.** Distribution of probe-caught topic shifts across modality. **b.** Distribution of self-caught topic shifts across modality in Study 1.

state whether they had shifted topics in the past 5seconds with a simple verbal yes or no response. Participants were given examples, such as shifting from thinking about friends and family to school and grades. Self-caught topic shifts required participants to type *'1'* whenever they noticed a topic shift, independent of the probes. Participants in the Without Probes condition completed the Think Aloud session without probe- and self-catching topic shifts.

Participants were informed that they would be recorded during the session and that all identifying details would be removed in transcription. They were shown the experimenter's empty room and assured of privacy. The experimenter left the room during the session and returned afterward. Following the Think Aloud session, participants completed the Thought Qualities Inventory, PANAS, 18 Categories Coding Scheme, Keywords, and the State Daydreaming Scale.

The same instructions were then given for the Silent Think condition, where participants silently engaged in their stream of consciousness. After completing the Silent Think session, participants answered the same questionnaires regarding their thought content and qualities. The order of sessions (Silent Think first or Think Aloud first) was counterbalanced. Finally, all participants completed a battery of individual difference measures (see Methods) and were debriefed.

2.3. Results

2.3.1. Analysis strategy

To mitigate the impact of order effects, we isolated our analyses to the first time block (Time 1: Think Aloud first and Silent Think first) when Order interacted with either the Modality or Probe conditions. Conversely, when no Order interactions were detected, we reported the collapsed results over both time blocks (Time 1 and Time 2: Think Aloud second and Silent Think second).

Additionally, to assess evidence in favor of the null over the alternative hypotheses, we included results from Bayes Factors (BF_{10}). It is typically used as a continuous measure to compare the relative strength of two models rather than making a binary decision based on p-values, calculated using the *BayesFactor* package in R (Morey & Rouder, 2022) to supplement traditional frequentist hypothesis testing (Dienes, 2014; Rouder et al., 2009). Bayes Factor is a likelihood ratio, comparing how well the null and alternative hypotheses predict the observed data. Prior research has delineated thresholds for determining the extent to which the evidence supports the null hypothesis, such that "a $BF_{10} < 0.01$ provides "extreme" evidence in favor of the null, a $BF_{10} = 0.01$–0.03 provides "very strong" evidence in favor of the null, a $BF_{10} = 0.03$–0.10 provides "strong" evidence in favor of the null, a $BF_{10} = 0.10$–0.33 provides "moderate" evidence in favor of the null, a $BF_{10} = 0.33$–1 provides "anecdotal" evidence in favor of the null, a $BF_{10} = 1$–3 provides "anecdotal" evidence in favor of the alternative, a $BF_{10} = 3$–10 provides "moderate" evidence in favor of the alternative, a $BF_{10} = 10$–30 provides "strong" evidence in favor of the alternative, a $BF_{10} = 30$–100 provides "very strong" evidence in favor of the alternative, a $BF_{10} > 100$ provides "extreme" evidence in favor of the alternative" (Welhaf & Bugg, 2024, p.5; Jeffreys, 1961; Rouder et al., 2009; Lee & Wagenmaker, 2013).

2.3.2. Topic shifting and meta-awareness

2.3.2.1. Probes. First, to assess differences in rates of topic shifting and meta-awareness between modalities, we compared probe-caught and self-caught topic shifts, before standardizing and averaging the two to develop a combined-probe topic shifting metric. Since we found no interaction effects of Order with Modality in a 2-way ANOVA: 2 (Order: Think Aloud first and Silent Think first) x 2 (Modality: Think Aloud and Silent Think), we report the results of the paired-samples t-tests on the differences between Think Aloud and Silent Think for the rate of topic shifting and meta-awareness as assessed by self-caught, probe-caught, and combined-probes.

While 58 participants were given probes, the probe-caught and self-caught data was analyzed from 43 participants. We excluded 15 participants due to experimenter error during recording, participants failing to follow instructions, or data corruption.

2.3.2.1.1. Probe-caught. Due to malfunctions in the presentation of the probes, some participants did not receive an equal number of probes (min = 9, max = 15 probes). Hence, we standardized the rate of probe-caught topic shifts by calculating the proportion of times a participant answered *'yes'* to the probes in relation to the number of probes presented. There were no significant differences in probe-caught topic shifting rates between Think Aloud ($M = 0.59$, $SD = 0.21$) and Silent Think ($M = 0.54$, $SD = 0.19$), $t(42) = -1.91$, $p = 0.062$, $d = -0.26$ (Fig. 1a). A Bayesian paired-sample *t*-test yielded a *Bayes Factor* (BF_{10}) = 0.87, suggesting that the results were only 1.14 ($BF_{01} = 1/BF_{10}$) times more likely under the null hypothesis than under the alternate hypothesis. In other words, the test yielded anecdotal evidence that there is no difference in probe-caught topic shifts across modalities, as the data slightly favored the null hypothesis but did not provide strong evidence in either direction.

2.3.2.1.2. Self-caught. Participants self-reported topic shifting during the session by pressing a button; their total number of button presses reflected the number of times they caught themselves switching between topics. There were no significant differences in self-caught topic shifting rates between Think Aloud ($M = 27.16$, $SD = 14.37$) and Silent Think ($M = 30.48$, $SD = 16.77$), $t(42) = 1.77$, $p = 0.083$, $d = 0.21$ (Fig. 1b). Again, a Bayesian paired-sample *t*-test was conducted on our measure of self-caught topic shifting, yielding $BF_{10} = 0.69$. This suggested that the results were only 1.56 times more likely under the null hypothesis. The test yielded anecdotal evidence in favor of the hypothesis that there is no difference in probe-caught topic shifts across modalities.

2.3.2.1.3. Combined probes. The probe-caught and self-caught topic shifts were then standardized to z-scores and averaged to compute an overall session-intermittent topic shifting score. There were no significant differences between Silent Think ($M = -0.01$, $SD = 0.82$) and Think Aloud ($M = 0.01$, $SD = 0.75$) for the combined probes topic shifting, $t(42) = -0.26$, $p = 0.79$, $d = -0.03$. The $BF_{10} = 0.17$, suggesting that the data are 5.88 times more likely under the null hypothesis compared to the alternate hypothesis, providing moderate evidence in favor of the null hypothesis.

2.3.2.1.4. Retrospective topic shifting item in thought qualities Inventory. The rate of topic shifting was also examined using a single item question upon completion of each session: "How often did you topic shift?" scored on a 5-point Likert scale *(1 = 5 times – 5 = More than 20 times)*. A three-way mixed ANOVA: 2 (Probes: with and without) x 2 (Order: Think Aloud first and Silent Think first) x 2 (Modality: Think Aloud and Silent Think) was conducted. There were no significant differences, $F(1,107) = 0.31$, $p = 0.58$, $\eta^2_G = 0.001$, between Think Aloud ($M = 3.40$, $SD = 1.41$) and Silent Think ($M = 3.49$, $SD = 1.42$) for topic shifting reported retrospectively (the manuscript will refer to the single item question on topic shifting in the thought qualities inventory as 'retrospective topic shifting/ shifts').

2.3.2.1.5. Combined probes and retrospective topic shifting. The probe-caught, self-caught, and retrospective topic shifts were standardized to z-scores, followed by averaging the three to compute an overall topic shifting score for the combined probed and retrospectively reported topic shifts. There were no significant differences between Think Aloud ($M = 0.004$, $SD = 0.74$) and Silent Think ($M = 0.008$, $SD = 0.77$) for overall topic shifting, $t(40) = 0.03$, $p = 0.97$, $d = 0.004$.

Finally, we examined the reliability of the self-reports of topic shifting for probes and for the retrospective topic shifting question by correlating them with rater-assigned topic shifts within the Think Aloud transcripts. Three raters independently assigned 'jumps' to all Think Aloud transcripts whenever they noticed a shift in topics, following a procedure used by Sripada and Taxali (2020). Percent agreement averaged across the three raters was 89.1 % and Fleiss' *kappa* = 0.645, close to the original findings by Sripada and Taxali (2020) at *percent agreement* = 91.3 % and *Fleiss' kappa* = 0.69. The agreement was substantially better than chance, $z = 130$, $p < 0.001$ (Cohen, 1960). The sample size for the reliability tests is 38 participants, due to indescribable audio recordings where participants were speaking too quietly, or missing sentences.

Rater-assigned topic shifts were strongly positively correlated with self-caught topic shifts, $r(36) = 0.76$, $p < 0.001$ and moderately correlated with both combined-probes, $r(36) = 0.52$, $p < 0.001$ and combined probes and retrospective topic shifts, $r(36) = 0.63$, $p < 0.001$. They were not correlated with probe-caught topic shifts, $r(36) = 0.006$, $p = 0.97$. These moderate to strong correlations with rater-assigned topic switches indicate that the participant-reported topic shifts for self-caught, combined-probes, and combined probes and retrospective topic shifts during Think Aloud are validated by a more objective scoring approach.

2.3.3. Retrospective reports on the reactivity of Think Aloud

Here, we first present the effects of Think Aloud on the 21 thought qualities collected retrospectively using the Thought Qualities Inventory. We discuss the main effects of Modality in the three-way ANOVAs – 2 (Probes: with and without) x 2 (Order: Think Aloud

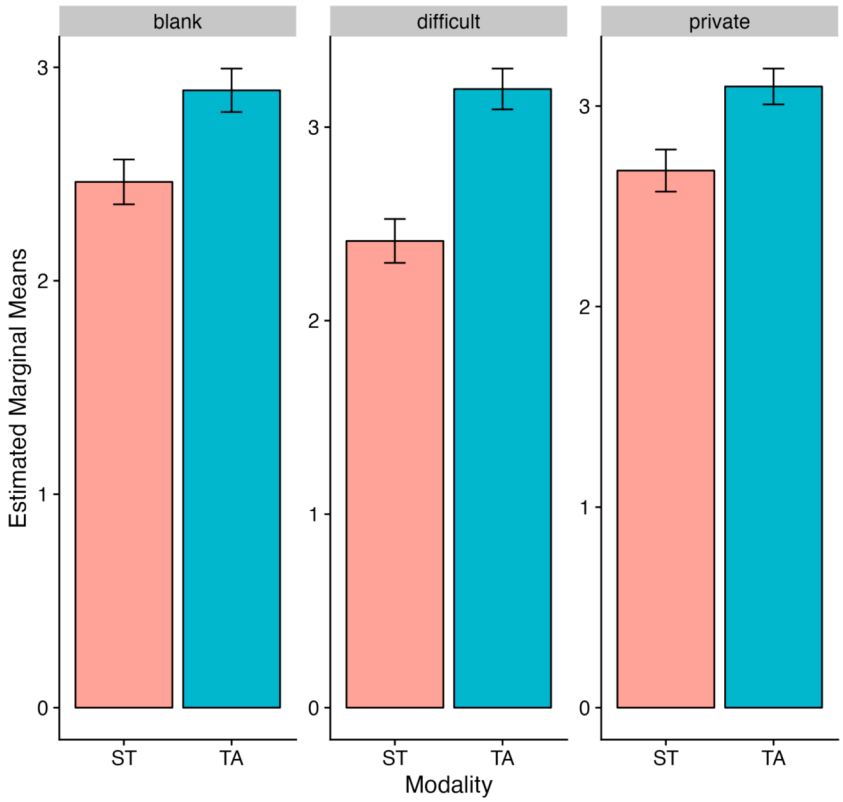


Fig. 2. Estimated marginal means for the main effects of modality on thought qualities assessing amount of mind blanking (blank), perceived session difficulty (difficult), and number of private thoughts (private) at $\alpha = 0.0022$. Bars represent the estimated marginal means (EMMs) for each modality, with error bars indicating the standard errors. Here, ST = Silent Think and TA = Think Aloud.

first and Silent Think first) x 2 (Modality: Think Aloud and Silent Think) for each variable. As 21 three-way ANOVAs were conducted, we corrected for multiple comparisons using Bonferroni corrections. For the phenomenological differences results, we report only the statistically significant effects after Bonferroni corrections at $\alpha = 0.0022$ (See supplementary materials section 1 for significant effects at $\alpha = 0.05$ and $\alpha = 0.1$). Secondly, we discuss the differences in thought content collected using the 18 Categories Coding Scheme between Think Aloud and Silent Think.

2.3.3.1. Phenomenological differences. Of the 21 thought qualities, only 3 qualities were significantly different between Think Aloud and Silent Think. Participants reported greater difficulty in doing the Think Aloud session ($M = 3.25$, $SD = 1.12$) than the Silent Think session ($M = 2.46$, $SD = 1.21$), $F(1,107) = 27.11$, $p < 0.001$, $\eta^2_G = 0.107$. They also reported more mind blanking in the Think Aloud condition ($M = 2.94$, $SD = 1.07$) than in the Silent Think condition ($M = 2.50$, $SD = 1.10$), $F(1,107) = 9.79$, $p = 0.002$, $\eta^2_G = 0.039$. Finally, participants reported having more private thoughts in Think Aloud ($M = 3.13$, $SD = 0.90$) than in Silent Think ($M = 2.75$, $SD = 1.10$), $F(1,107) = 15.65$, $p < 0.001$, $\eta^2_G = 0.041$ (Fig. 2). To see the results of all 21 qualities, please refer to Table 1.

As for the interactions between Order and Modality, only difficulty from the Thought Qualities Inventory was significant. However, there were no significant differences at Time 1.

Additionally, a chi-square test of independence was performed to examine the relation between Modality and the medium within which participants experienced their thoughts. Participants had the option to choose multiple responses amongst images, words, inner monologue, TV show or movie, and strong and consistent personal narrative. We found no significant differences between mediums based on modalities. Overall, participants chose 'inner monologue' most often in both Think Aloud ($M = 0.66$) and Silent Think ($M = 0.68$).

2.3.3.2. Think Aloud only. Three questions pertaining solely to the Think Aloud session were asked in the Thought Qualities Inventory. The results indicated that participants found it only a little bit natural to Think Aloud ($Mdn = 2$, $M = 2.63$, $SD = 1.37$). However, they were rarely unable to vocalize their thoughts ($Mdn = 3$, $M = 2.65$, $SD = 1.04$) and they also were only a little bit filtered in their verbalization of their thoughts ($Mdn = 2$, $M = 2.54$, $SD = 1.02$). There were also no differences between participants' experience of the session dependent upon whether they did Think Aloud first or second, nor were there any differences based on whether they were assigned to the probes condition or not.

2.3.3.3. Content differences. When assessing for differences in thought content, participants reported whether they experienced a certain topic more using the 18 Categories Coding Scheme (Garg et al., *in prep*). All chi-sq results indicated only one significant difference amongst the 18 categories—Partner, intimacy, love, and sexual matters, wherein participants in the Silent Think condition ($M = 0.64$, $SD = 0.48$) had more thoughts related to the category than participants in the Think Aloud condition ($M = 0.50$, $SD = 0.50$): $X^2(1) = 4.26$, $p = 0.036$, $\phi = 0.196$.

Overall, the categories most frequently thought of were Education (Silent Think: $M = 0.74$, SD = 0.43; Think Aloud: M = 0.81, SD = 0.39) and Friends, family, relatives, and acquaintances (Silent Think: $M = 0.84$, SD = 0.36; Think Aloud: M = 0.83, SD = 0.38) (Fig. 3). Additionally, participants were also asked to write topics that they thought of during the sessions independent of the 18 Categories Coding Scheme (refer to Keywords in Method). There were no new themes that emerged beyond the 18 Categories that we provided the

Table 1
Means, standard deviations, and results of the simple main effects of Modality from the three-way ANOVA.

Measure	Silent Think		Think Aloud		$F(1, 112)$	*p-value*	η^2_G
	M	*SD*	*M*	*SD*			
Blank	2.47	1.10	2.89	1.05	9.79	0.002	0.039
Boredom	2.59	1.21	2.41	1.22	1.87	0.174	0.005
Control	3.13	1.11	3.41	0.98	5.55	0.020	0.016
Creative	2.64	1.09	2.43	1.01	3.31	0.072	0.011
Curious	2.96	1.04	2.95	0.95	0.07	0.796	<0.001
Difficult	2.44	1.23	3.20	1.09	27.11	<0.001	0.107
Distress	2.11	0.99	2.34	1.04	7.39	0.008	0.015
Enjoyable	3.00	1.19	2.85	1.15	2.23	0.139	0.006
Fragmented	2.84	1.22	2.57	1.03	3.18	0.077	0.012
Goal directed	3.01	1.07	2.83	1.14	2.23	0.138	0.005
Interesting	3.08	1.24	2.89	1.22	2.05	0.155	0.005
Meaningful	3.65	0.93	3.74	1.02	0.48	0.489	0.001
Memory	3.62	0.92	3.71	0.94	0.78	0.380	0.002
Negative PANAS	17.1	6.43	17.4	6.57	0.88	0.350	<0.001
Positive PANAS	22.7	8.45	22.2	8.06	0.72	0.399	<0.001
Private	2.71	1.10	3.12	0.92	15.65	<0.001	0.041
Productive	2.75	1.11	2.82	1.05	0.56	0.458	0.001
Purpose	2.80	1.10	2.82	1.20	0.03	0.855	<0.001
Repetition	3.44	1.03	3.14	0.89	6.36	0.013	0.018
Topic shift	3.48	1.43	3.41	1.42	0.31	0.581	<0.001
Vague	2.41	1.08	2.35	0.98	0.51	0.475	0.002
Why topic shift	2.59	0.92	2.56	0.84	0.38	0.538	0.001

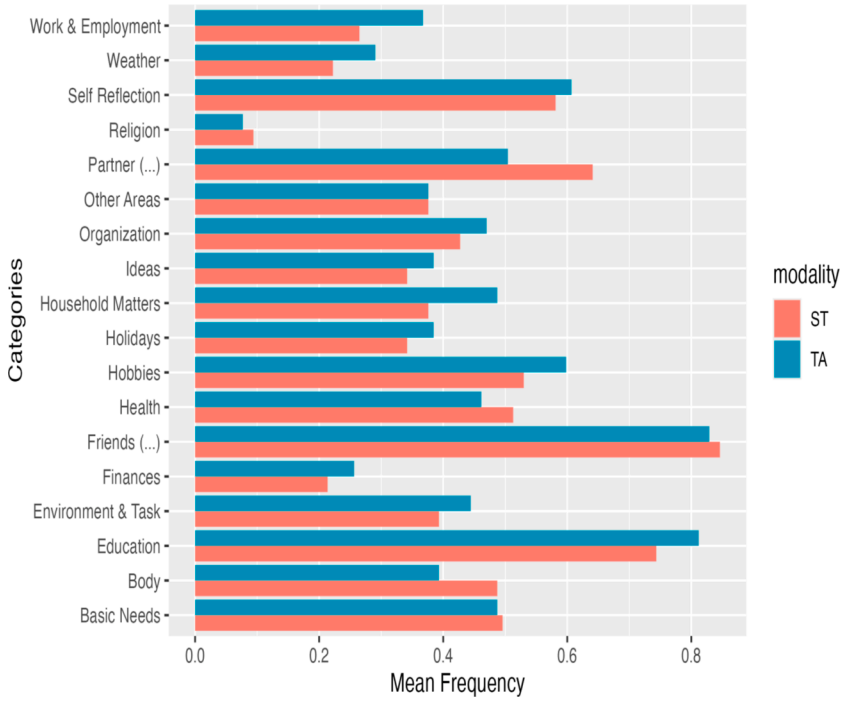


Fig. 3. Mean frequency of 18 topics in the 18 Categories Coding Scheme by modality. Bars represent the mean frequency of topics for each category, with different colors indicating different modalities. The x-axis shows the mean frequency, while the y-axis shows the thought topics from the 18 Categories Coding Scheme. Here, ST = Silent Think and TA = Think Aloud. Only the Partner, Intimacy, Love, and Sexual Matters topic was significantly different between Think Aloud and Silent Think.

participants, except for frequent mentioning of social media.

2.3.4. *Reactivity of probes*

The study also addresses the question of whether the introduction of thought probes during a session affects thought qualities and the nature of the stream of consciousness. Amongst the 21 three-way ANOVAs conducted, the only statistically significant main effect of the Probe condition was for the rate of topic shifting, $F(1,107) = 40.01$, $p < 0.001$, $\eta^2_G = 0.212$. Participants in the With-Probes ($M = 4.09$, $SD = 1.28$) condition retrospectively reported topic shifting more often than participants in the Without-Probes ($M = 2.74$, $SD = 1.22$) condition (Fig. 4).

2.4. *Discussion*

Does thinking aloud alter the qualities and content in the stream of consciousness? We investigated this question by comparing the Think Aloud and Silent Think protocols, where participants let their minds wander naturally for 15 minutes while either verbalizing their stream of consciousness or thinking silently. Specifically, we assessed differences in the rate of topic shifting and meta-awareness during the sessions, and the reactivity of Think Aloud on SST by comparing retrospective reports on the experienced qualities and content at the end of each session.

The results of Study 1 largely reinforce Think Aloud's validity as a method for studying the stream of consciousness. There were no significant differences in the amount of topic shifting and meta-awareness, examined through rates of self-caught shifts, probe-caught shifts, and a retrospective report of topic shift frequency. Composite measures derived from these different metrics again did not vary between Think Aloud and Silent Think, and supplementary Bayesian analyses even showed evidence in favor of their similarity. We further validated these self-reports with rater-assigned topic shifts in the Think Aloud transcripts, finding moderate to strong correlations with participant-provided rates.

Of the 21 different phenomenological qualities of the stream of consciousness that we measured, only 3 qualities were significantly different between Think Aloud and Silent Think: the number of private thoughts, amount of mind blanking, and difficulty in doing the session. As anticipated, participants preferred to keep more thoughts private rather than speak them aloud. This is evidenced by the fact that of the 18 content topics analyzed, the only difference between modalities for thought content was in the category of partner, intimacy, love and sexual matters. Indeed, in the Silent Think condition, the most frequently occurring category of thoughts pertained to Partner, Intimacy, Love, and Sexual Matters, suggesting that participants might have hesitated to verbalize such personal and intimate thoughts due to self-consciousness or discomfort. The Think Aloud method had minimal impact on the topics participants thought about during the session, indicating that it did not substantially alter the distribution of thought topics.

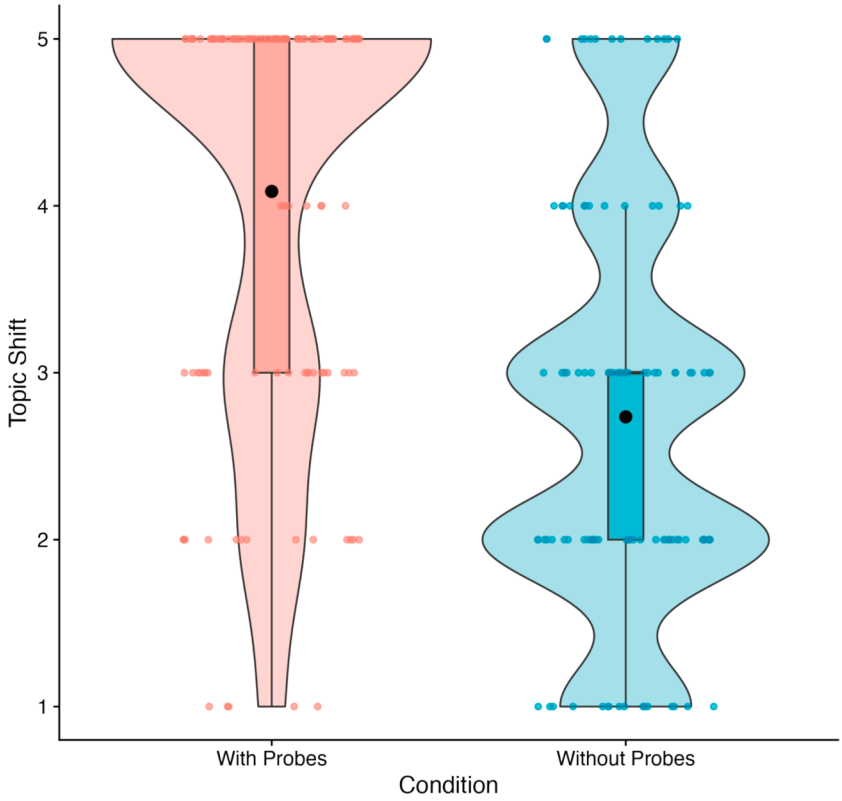


Fig. 4. Raincloud plot illustrating the distribution of topic shifts across different conditions. It includes half violin plots for each condition, boxplots displaying the interquartile range and median (horizontal line), jittered points representing individual data points, and mean (black circles).

Participants also displayed greater instances of mind blanking during Think Aloud than Silent Think. Two models of mind blanking have been proposed: a lack of mental content and a lack of linguistically or conceptually determinable mental content (Kaufmann et al., 2024). Potentially, in this study, both mechanisms could have been at play. Participants might have experienced moments during Think Aloud when they simply lacked sufficient mental content to report, while at other times, they could have encountered a linguistic fog, making the translation of thoughts into verbalizable content difficult. On a similar note, Baumeister (1984) proposed a model of choking under pressure, whereby pressure increases conscious attention to one's performance and impairs automatic or overlearned procedures. In the current study, the pressure to verbalize thoughts immediately may have impacted people's ability to come up with something to say, leading to choking via increased mind-blanking.

Curiously, our results trended against what we had originally hypothesized. We expected that a reduction in topic shifting in the Think Aloud condition would be driven by an increase in meta-awareness associated with verbalization. Accordingly, we hypothesized that self-catching (previously used as a measure of meta-awareness) would be particularly pronounced in the Think Aloud condition relative to Silent Think. Contrary to this prediction, we found a marginally significant trend in the opposite direction, with numerically more self-caught topic switches in the Silent Think condition than in Think Aloud. Similarly, we had hypothesized that Silent Think would involve less meta-awareness, resulting in more probe-caught topic shifts in the Silent Think condition. However, here too we found a marginally significant trend indicating more probe-caught topic switches in the Think Aloud condition than the Silent Think condition.

There are three possible reasons why self-caught/probe-caught patterns trended in the opposite direction as hypothesized. First, it is possible that the self-caught/probe-caught methodology, which has been effectively used to identify aware and unaware mind wandering (Chu et al., 2023; Sayette et al., 2009; Sayette et al., 2010), is less effective for assessing meta-awareness in the context of topic shifting. When participants attend to both simultaneously, either self-caught or probe-caught topic shift rates could be inflated by the other. Additionally, while self-caught methods are intended to be the 'true' indicator of meta-awareness, participants may have responded to topic shifting probes with or without meta-awareness (Chu et al., 2023). Thus, the comparison of self-caught to probe-caught topic shifting may not be an accurate reflection of differences in meta-awareness.

Second, thinking aloud may increase cognitive load, thereby reducing meta-awareness compared to thinking silently. One of the thought qualities that differed between Think Aloud and Silent Think pertained to the difficulty experienced during sessions. Participants indicated experiencing greater difficulty thinking aloud than silently. Indeed, Li et al. (2021) also showed that when participants simultaneously Think Aloud and self-catch topic shifts, the overall rate of self-caught topic shifts was lower than rater-assigned topic shifts. When asked why this may be, their participants mentioned forgetting to self-catch topic shifts while also

trying to Think Aloud.

Third, given that the effect was only marginally significant, we did not rule out the possibility of it simply being a statistical deviation. Thus, we were also interested in exploring whether the marginally significant trends represented real differences that might show up when investigated using a different measure in a new sample. To address these alternative possibilities, we conducted Study 2.

3. Study 2

3.1. Introduction

The results of Study 1 suggested that SST is minimally reactive to the Think Aloud protocol, exhibiting no differences in the rate of topic shifting and meta-awareness of topic shifts. However, there were marginal effects in the self-caught reports of topic shifting between conditions which were in the opposite direction to our predictions; specifically, the Think Aloud condition had marginally fewer self-caught topic shifts—a finding that is consistent with Li et al., (2021), where subjects reported significantly lower self-caught topic shifts in the Think Aloud condition compared to the Silent Think condition. The reduction in self-caught probes may suggest decreased meta-awareness during Think Aloud compared to Silent Think. However, participants also reported greater difficulty thinking aloud than silently which raises the possibility that a greater cognitive load in the Think Aloud procedure explains the marginally reduced self-caught reports. In other words, participants may have had a harder time self-monitoring given the more demanding nature of thinking aloud. Similarly, Li et al. (2021) conjectured that the difference in the amount of self-caught topic shifts between Silent Think and Think Aloud in their study may be due to the increased cognitive burden of maintaining meta-awareness while thinking aloud.

In order to tease apart these explanations for the observed marginal reduction in self-catches—i.e., decreased meta-awareness and increased cognitive load (Li et al., 2021)—we aimed to mitigate the potentially confounding effects of self-catches by employing a probe-caught measure of meta-awareness in Study 2. Specifically, participants were periodically asked if they had switched topics in the past few seconds. If they responded yes, they would answer another thought probe assessing whether they had been aware of the topic shift prior to the probe (Smallwood et al., 2007). This eliminated the confound of participants monitoring for shifts and provided a clearer metric for meta-awareness of topic shifting. To further examine the potential confound that Think Aloud increased cognitive demand, we incorporated the NASA Task Load Index as a robust, multidimensional measure of workload–both as a dependent variable and as a possible mediator for the relationship between modality and meta-awareness of topic shifting. This allowed us to expand on Study 1′s difficulty finding, which was just based on a single item. Additionally, as our finding of the rate of self-caught topic shifts was marginally significant, we wished to assess whether these differences would be observed using a different approach—i.e., using only thought probes to catch meta-aware topic shifts. We also wanted to confirm whether the absence of differences in the rate of topic shifting was observed in a new sample, given that the Bayes Factor indicated only “anecdotal evidence” in favor of the null hypothesis for probe-caught topic shifts across the two modalities.

In Study 2, our goal was to replicate the key findings of Study 1 while addressing whether the marginal effects regarding self-caught and probe caught topic shifts observed were real or spurious. This question is critical because if these marginal effects are real, it suggests potential reactivity in the Think Aloud condition, warranting further exploration of alternative explanations for the difference in participants’ reported topic shifts. For instance, the marginal reduction in self-caught topic shifts could indicate a decrease in meta-awareness or an increase in cognitive load while thinking aloud. Importantly, if cognitive load is shown to increase in Think Aloud, this would suggest that the procedure is reactive in certain conditions, even if the reactivity was not consistently observed here. Conversely, if cognitive load does not differ between conditions, it would reinforce the non-reactive nature of Think Aloud on SST, indicating that any observed differences in topic shifts are not attributable to reactivity but to other factors.

Our primary focus for analyses were whether Think Aloud differed in the rate of topic shifting, amount of meta-awareness of topic shifting, and cognitive load to Silent Think in a between-subjects design. Specifically, we hypothesized that participants would indicate no differences in the rate of topic shifting between modalities but would experience greater cognitive load in the Think Aloud condition compared to the Silent Think condition. Given the ambiguous findings from the self- and probe-caught measures in Study 1, we propose potential outcomes regarding the relationship between modality and meta-awareness of topic shifts—if cognitive load is higher in Think Aloud, then meta-awareness of topic shifts will be decreased compared to the Silent Think condition; if cognitive load does not differ between conditions, meta-awareness of topic shifts will be higher in the Think Aloud condition; and finally, meta-awareness will be the same between conditions as repeated presentation of the thought probes will increase the baseline meta-awareness of topic shifts.

The study was pre-registered on the Open Science Foundation. In the Preregistrations Deviations Table (Willroth & Atherton, 2024), we have documented deviations from the original hypotheses based on new knowledge gained from a reanalysis of Study 1′s findings (see supplementary materials section 2).

3.2. Method

3.2.1. Participants and design

A total of 118 participants were recruited from the same subject pool as Study 1. After examining the data, we excluded 16 participants for failing to follow instructions (N = 8), or data corruption (N = 8). This left a final sample of 102 participants, which was slightly above our preregistered sample size of $N = 100$. Study 2 was a between-subjects design comparing two groups: Think Aloud and Silent Think.

3.2.2. Materials

3.2.2.1. NASA task load Index (NASA-TLX). This 6-item weighted scale measures various dimensions of workload after a task. It has six subscales: mental demand, physical demand, frustration level, own performance, effort, and temporal demand (Hart, 1988). For example, the item for mental demand asks "How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?". The NASA-TLX follows a two-step approach—the first level entails participants' weightings of the subdimensions' contribution to overall workload, and the second step asks participants to rate the subdimensions themselves. Each subdimension is rated on a continuous scale from 0 (lowest) to 100 (highest). This enables us to analyze both the overall weighted score and also each subscale individually (for calculation procedures, see NASA Task Load Index (TLX) v.1.0 manual).

3.2.2.2. Covariates. Three surveys measuring trait mindfulness, trait *meta*-awareness, and social desirability were included as covariate measures in Study 2. All participants completed the Five-Factor Mindfulness Questionnaire-15 (FFMQ-15, Baer et al., 2012), Multidimensional Awareness Scale (MAS, DeMarree & Naragon-Gainey, 2022), and the Marlowe-Crowne Social Desirability Scale-Short Form (MCSDS-SF, Strahan & Gerbasi, 1972). We have included a description of each survey and the corresponding results in the supplementary materials section 3.

3.2.3. Procedure

After participants completed informed consent, the experimenter led them to a dimly lit room in the laboratory. Participants received the same instructions as Study 1 for either the Think Aloud or Silent Think session, depending on their assigned condition. However, participants engaged in the session for 20 minutes instead of 15 minutes.

During the 20 minutes, all participants, regardless of condition, were interrupted periodically with one thought probe every 2-minute subblock. The probes were presented at randomly selected intervals of 60, 75, 90, or 105 seconds. The first part of each probe inquired about topic shifting: "In the past 5 seconds, did you shift between topics? Please press the button 'Y' for yes or 'N' for no on the keyboard." If the participant responded yes to the topic shift thought probe, they then saw a meta-awareness question: "Were you aware of having topic shifted prior to the administration of the probe?" with the same method to respond yes or no (Smallwood et al., 2007). Participants who responded no to the topic shift thought probe did not receive the meta-awareness thought probe.

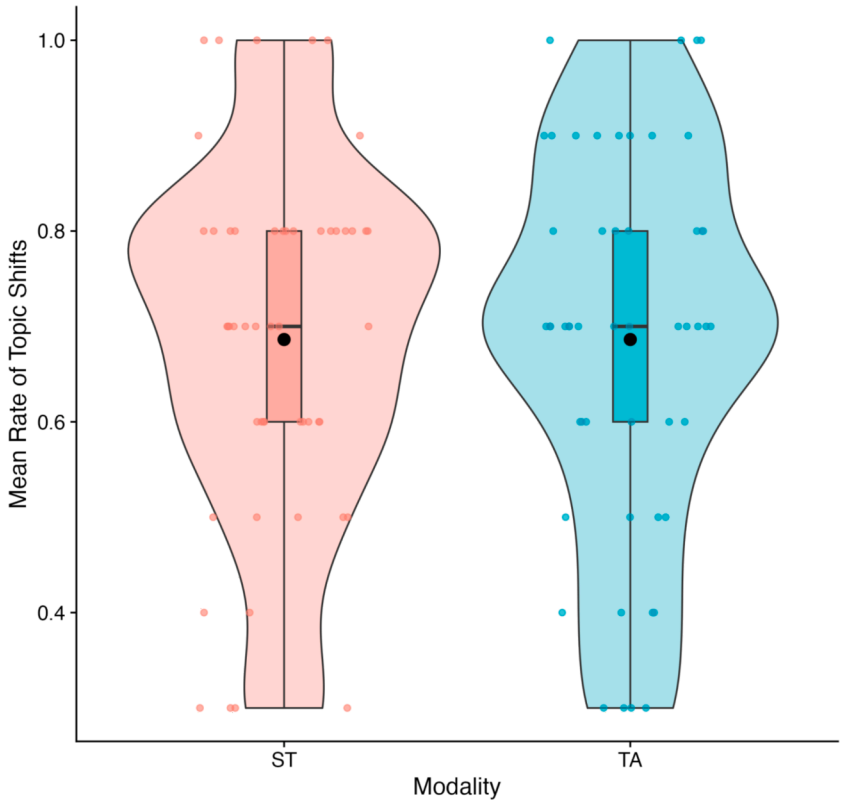


Fig. 5. Raincloud plot illustrating the mean rate of topic shifts across Silent Think (ST) and Think Aloud (TA) in Study 2. It includes half violin plots to display the density distribution of the averaged topic shift rates for each modality, boxplots displaying the interquartile range and median (horizontal line), jittered points representing individual data points, and mean (black circles).

After informing the participant that they would be recorded during the session, providing examples of each scenario, and ensuring the participants understood the instructions, the experimenter left the room and re-entered once the session was completed. Upon completion of the session, participants completed the NASA-TLX, FFMQ-15, MAS, and MCSD-SF questionnaires. They were then debriefed and given course credit.

3.3. Results

3.3.1. Topic shifting and meta-awareness of topic shifting

Like Study 1, Study 2′s main aim was to test whether there were differences in topic shifting rates and meta-awareness between Think Aloud and Silent Think. The 10 binary responses of yes or no to the topic-shifting probes were averaged to create a metric for the rate of topic shifting for each participant. An independent samples *t*-test showed that the proportion of yeses to the topic shifting probes were remarkably similar across Think Aloud ($M = 0.68$, $SD = 0.18$) and Silent Think ($M = 0.68$, $SD = 0.19$), $t(100) = 0$, $p = 1$, $d = 0$ (Fig. 5). A follow-up Bayesian *t*-test was conducted, giving $BF_{10} = 0.209$; this suggested that the results were 4.78 times more likely under the null hypothesis, providing moderate evidence in support of the null hypothesis over the alternative.

Prior to comparing meta-aware topic shift rates, we dropped any instances where the participant responded no to the initial topic shift probe. The final metric for the meta-awareness of topic shifting was calculated as the proportion of yeses to the *meta*-awareness probe. Through an independent samples *t*-test, we found no significant differences between Think Aloud ($M = 0.61$, $SD = 0.25$) and Silent Think ($M = 0.55$, $SD = 0.30$) for the proportion of times a participant was meta-aware of having switched between topics in their stream of consciousness, $t(100) = -1.093$, $p = 0.276$, $d = -0.216$ (Fig. 6). A Bayesian *t*-test was conducted, yielding $BF_{10} = 0.356$. This suggested that the results are 2.81 times more likely under the null hypothesis, providing anecdotal evidence in favor of the null hypothesis over the alternative.

3.3.2. Cognitive load

Next, we examined the differences in cognitive load between Think Aloud and Silent Think using the NASA-TLX scale. An independent samples *t*-test was conducted between Think Aloud ($M = 42.11$, $SD = 16.65$) and Silent Think ($M = 40.21$, $SD = 19.60$) on the participants overall weighted TLX scores for each session, resulting in no significant differences between the two modalities, $t(100) = -0.526$, $p = 0.599$, $d = -0.104$. A Bayesian *t*-test was conducted, yielding $BF_{10} = 0.236$, suggesting that the results are 4.23 times more

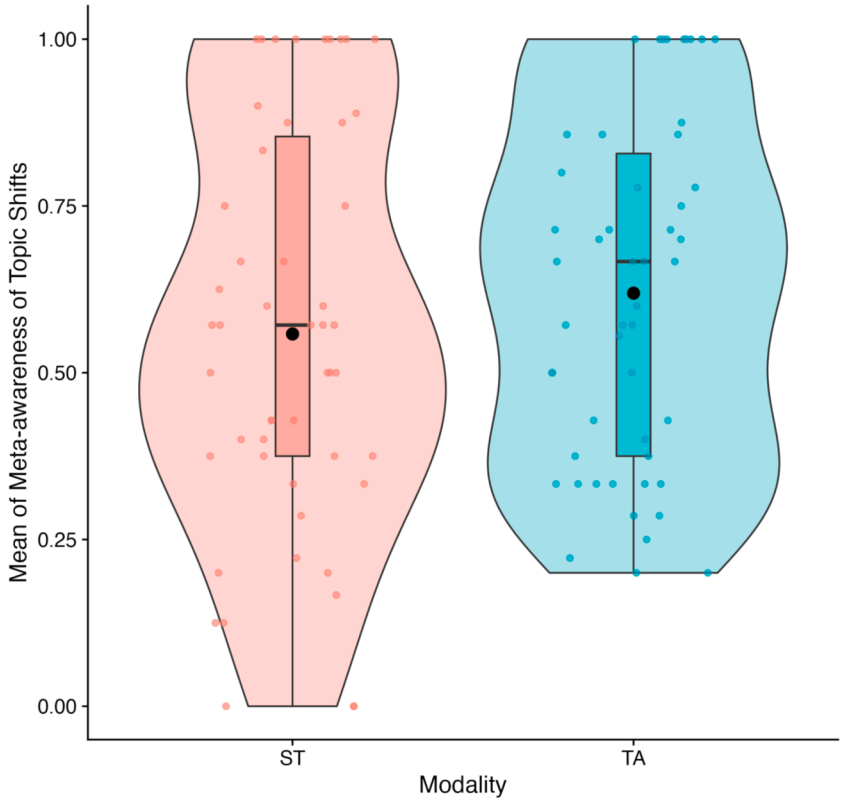


Fig. 6. Raincloud plot illustrating the mean amount of *meta*-awareness of topic shifts across Silent Think (ST) and Think Aloud (TA) in Study 2. It includes half violin plots to display the density distribution of the averaged topic shift rates for each modality, boxplots displaying the interquartile range and median (horizontal line), jittered points representing individual data points, and mean (black circles).

likely under the null hypothesis and indicating that there is moderate evidence to support the null hypothesis over the alternative. Similarly, there were no significant differences for any of the subscales of the NASA-TLX between Think Aloud and Silent Think either.

Additionally, a mediation analysis was conducted to analyze whether cognitive load mediated the relationship between meta-awareness of topic shifting and modality. The mediation was not statistically significant, $b = 0.002$ (95 % bootstrapped CI: −0.01, 0.02), $z = 0.279$, $p = 0.781$. For a detailed description of the mediation analysis, see supplementary materials section 4.

3.4. Discussion

We conducted Study 2 to replicate our results from Study 1, focusing on differences in the rate of topic shifting, amount of meta-awareness of shifts, and difficulty levels between Think Aloud and Silent Think. Study 2 also explored potential reasons for the marginal differences in self-catching between the two modalities observed in Study 1. Specifically, the results of Study 1 left open three possible interpretations of the marginal decrease in self-caught probes in the Think Aloud condition: reduced meta-awareness, increased cognitive load, or a statistical deviation. To examine these alternatives both meta-awareness and cognitive load were measured, the latter by measuring task difficulty across several validated subscales in the NASA-TLX. We found no differences between Think Aloud and Silent Think for both meta-awareness and difficulty, as well as observing no differences in topic-switching between the two modalities. Collectively, this suggests that the differences in difficulty observed between the modalities in Study 1 were either because of the *combination* of Think Aloud and self-monitoring, as thinking aloud alone in Study 2 was not more demanding than Silent Think, or a statistical deviation that was not replicated in a new sample. Finally, to rule out an explanation that workload mediated any relationship between Think Aloud and meta-awareness, we conducted a set of follow-up mediation analyses and found no evidence to support that conjecture (see supplementary materials section 4). In sum, the results of Study 2 reinforced the null findings from Study 1 in a new sample and using a between-subjects design, finding no significant differences in the rate of topic shifting and the amount of meta-awareness of topic shifts between the Think Aloud and Silent Think modalities. Furthermore, the lack of difference in cognitive load between the Think Aloud and Silent Think conditions, Study 2 provides additional evidence of the non-reactivity of the Think Aloud procedure on the spontaneous stream of consciousness.

4. General discussion

In recent times, research on the spontaneous stream of consciousness has witnessed a significant surge, bringing to the forefront a compelling and ever-curious inquiry into the content and structure of thoughts experienced during idle moments (Christoff et al., 2016; Schooler et al., 2011; Smallwood & Schooler, 2015). Among the methodologies used, Think Aloud has emerged as a valuable technique offering valuable insights into the content of SST (Sripada & Taxali, 2020). However, the reactivity of SST to the Think Aloud protocol remains relatively understudied. Our studies sought to examine how the protocol compares with the Silent Think state, as thinking aloud might alter the qualities and content of thoughts due to increased meta-awareness during verbalization. We therefore conducted two experiments to assess Think Aloud's impact on the rate of topic shifting, meta-awareness, cognitive load, and the qualities and content of the stream of consciousness. In general, our results suggested that the stream of consciousness is minimally reactive to the Think Aloud protocol. The current studies help define the limits of Think Aloud in researching the stream of consciousness, but also pave the way for future studies to use the technique when examining the content of unguided and unconstrained thought.

4.1. Meta-awareness

We hypothesized that thinking aloud could impact the spontaneous stream of consciousness by increasing meta-awareness, as verbalizing thoughts could bring them to the forefront of attention. This self-monitoring of internal thoughts might inadvertently alter the qualities, content, and structure of SST. To first assess meta-awareness differences between Think Aloud and Silent Think, Study 1 had participants self-catch topic shifts as a measure of meta-awareness and respond to probes about topic shifts during 15-minute sessions (Chu et al., 2023; Sayette et al., 2009; Sayette et al., 2010). The results suggested that thinking aloud did not lead to an increase in the amount of meta-awareness compared to thinking silently. However, marginally significant trends raised the possibility that thinking aloud might reduce meta-awareness with indications of more self-caught and fewer probe-caught shifts when thinking silently, and vice-versa when thinking aloud.

Study 2 explored this possibility by using a different method of measuring meta-awareness of topic shifts. Rather than comparing the relative frequency of self- and probe-caught topic shifts, in Study 2 we exclusively used thought probes to assess topic shifts, but after every probe we additionally asked participants about their awareness of topic shifts. Here, we replicated our primary results and augmented our findings with a set of Bayesian analyses: there were no differences in the amount of meta-awareness of topic shifts between thinking silently and thinking aloud.

4.2. Topic shifts

In Study 1, we hypothesized that greater meta-awareness during Think Aloud would decrease the frequency of topic shifts, as hearing oneself talk might help maintain focus on a single topic. This conjecture was based on preliminary findings from a pilot study wherein Silent Think had a higher rate of topic shifting than Think Aloud (Sripada & Taxali, 2020). Study 1 used multiple measures to assess the impact of Think Aloud on the rate of topic shifts via self-caught shifts, probe-caught shifts, retrospective report of topic shift frequency, and composite measures. In Study 2, topic shifts were measured using only the probe-caught method. Across both studies,

the results indicated no differences in the rate of topic shifting between thinking aloud and silently.

Our studies replicated the null findings by Li et al. (2021), who also found no differences in topic shifting rates between Think Aloud and Silent Think in an eastern population in China. However, they compared self-caught topic shifts in a Silent Think condition to rater-assigned topic shifts in two Think Aloud conditions. Our study addressed this discrepancy by instead comparing the rate of topic shifts across the same "unit of measurement" between the modalities—for example, directly comparing self-caught topic shifts in Silent Think to self-caught topic shifts in Think Aloud. Li et al. had also reported lowered self-caught topic shifts in the Think Aloud + self-catch condition compared to the Silent Think + self-catch condition. Although our results from Study 1 did not replicate the finding of differing self-caught topic shifts between Think Aloud and Silent Think, we did find differences in the amount of perceived difficulty between the modalities.

4.3. Cognitive load

Across two studies, we found somewhat discrepant evidence regarding the nature of the relationship between cognitive load and verbalizing thoughts. In Study 1, participants reported experiencing greater difficulty in verbalizing thoughts than thinking through them silently, while in Study 2, we found no significant difference in the amount of cognitive load across the Think Aloud and Silent Think conditions. There are two alternate explanations for this discrepancy.

First, the reports of thinking aloud being more difficult to do than thinking silently can shed some light on the ambiguity in Li et al.'s findings, which indicated differences in self-caught topic shifts between modalities, but no differences in topic shifts when comparing the Silent Think + self-catch condition to rater-assigned topic shifts in the Think Aloud condition. The findings from our and Li et al.'s studies suggest that requiring participants to self-monitor while concurrently thinking aloud is especially challenging. The added burden of self-monitoring may be similar to asking participants to introspect on their thoughts, which has previously been reactive to the task (Fox et al., 2011). Reinforcing this, Ericsson and Fox (2011) indicate that "requests for explanations and descriptions—the defining attributes of introspection—are associated with significant reactivity" in the Think Aloud protocol (p. 351). As such, self-catching or self-monitoring thoughts may be a form of introspection in the context of the Think Aloud task. Therefore, one possible explanation for the increased difficulty of thinking aloud in Study 1 compared to the null finding of cognitive load in Study 2 might be that Think Aloud alone does not incite greater cognitive load, but the Think Aloud + self-catch combination does. Given the possibility of increased load when thinking aloud and self-catching topic shifts, we recommend using only thought probes for measuring meta-awareness when using the Think Aloud paradigm. More generally, we encourage future researchers to explore the potential cognitive load demands of the self-catching method as it might apply in other paradigms such as mind wandering, where it has been used extensively.

Second, it is important to consider that the observed differences in difficulty between the modalities in Study 1 may have resulted from sampling variability, which was not replicated in a follow-up experiment. Since Study 2 measured cognitive load differently and replaced self-caught with probe-caught topic shifts, it is difficult to attribute the increased difficulty in Study 1 solely to the Think Aloud + self-catch combination. Replicating and extending both Study 1 and Study 2 would help clarify these alternative explanations.

If the observation of a lack of differences in cognitive load between the Think Aloud and Silent Think conditions in Study 2 proves reliable, it would have important implications for the potential reactivity of Think Aloud on the stream of consciousness. In principle, the added burden of verbalizing thoughts could have impacted the structure, content, and/or qualities of thoughts as they are naturally experienced in the stream of consciousness. The finding that Think Aloud does not induce greater cognitive load reduces concerns about this potentially major source of reactivity of the Think Aloud procedure on SST.

4.4. Reactivity of content and qualities

In Study 1, participants were asked to retrospectively report on their stream of consciousness for a comprehensive set of 21 thought qualities and 18 categories of thought content. Our findings suggest that the stream of consciousness was minimally reactive to Think Aloud, with only 3 qualities (number of private thoughts, amount of mind blanking, and session difficulty) and 1 thought topic (partner, intimacy, love, and sexual matters) differing between Think Aloud and Silent Think.

Despite these few differences, Study 1 suggested that SST is largely non-reactive to the Think Aloud protocol. This underscores Think Aloud's value as a method for studying the stream of consciousness, showing that it has minimal impact on thought content and qualities. Understanding the content of spontaneous thought is crucial because of its impact on human flourishing and well-being (Smallwood & Andrews-Hanna, 2013; Li et al., 2021). For instance, past research on mind wandering has found that ruminative and/or past related thought content can contribute to emotional distress (Poerio et al., 2013; Smallwood & O'Connor, 2011; Stawarczyk et al., 2013; Watkins, 2008). On the other hand, mind wandering is associated with greater productivity, fulfillment, and positive affect when individuals think goal-directed, future, and planning-related thoughts (Klinger, 1971, 1999, 2009; Franklin et al., 2013; Poerio et al., 2015). The Think Aloud protocol can extend this research by providing a direct glimpse into the content of thoughts and furthering our understanding of the impact of thought content on well-being.

4.5. Exploratory results

4.5.1. Thought probes reactivity

Although self-reported topic shifting was not affected by thinking aloud, it was impacted by the inclusion of periodic thought probes querying participants regarding their topic shifts. In Study 1, participants, when asked post-session, reported a higher frequency

of topic switches in the 'with-probes' condition compared to the 'without-probes' condition. This outcome may be attributed to the nature of the probe- and self-caught topic shifts, whereby the frequent presentation of thought probes and the implicit nature of asking participants to be vigilant of their own topic shifts might have led the participants to be more attentive to their topic shifts in the 'with-probes' condition. It is also plausible that the interruption due to the probes increased the likelihood of topic changes as well. This is important for future researchers to consider when using thought probes, as the presentation of the probe may inflate the reports of the phenomenon being measured by the probe.

Notably, the inclusion of the probes did not demonstrate any other reactivity in the structure of the stream of consciousness beyond the aforementioned effect. This finding supports the utility of thought probes, as previous studies have also shown that administering thought probes during concurrent tasks does not significantly impact the overall cognitive processes (Kane et al., 2021; Wiemers & Redick, 2019).

4.5.2. Exploratory considerations

Given the exploratory nature of Study 1, here, we briefly touch on the findings for the thought qualities that were significant at $\alpha = 0.05$ and at $\alpha = 0.1$. The results hinted that Think Aloud during the stream of consciousness provided more structure and direction to thoughts, resulting in reduced fragmentation and repetition of ideas when spoken aloud. However, this controlled nature of Think Aloud came at a cost, as the results suggested that Think Aloud restricted instances of creative thinking (Bastian et al., 2017; Schooler et al., 1993; Schooler & Melcher, 1995; Sripada, 2018). We also observed a notable rise in positive mood during Think Aloud compared to Silent Think. Earlier research has highlighted the discomfort that individuals may experience when left alone with their thoughts (Wilson et al., 2014). However, Think Aloud may serve as a deterrent to this negative effect. These findings are exploratory and must be interpreted with caution, but they warrant further investigation, nonetheless. We urge researchers to replicate Study 1 with focused hypotheses keeping this data in mind.

4.6. Limitations

The current studies are not without their limitations. First, thoughts that occur during the stream of consciousness can be personal, and some thoughts may be difficult to verbalize if they occur in another medium such as visual imagery (Amit et al., 2017). Because participants may experience discomfort when vocalizing their thoughts, more Think Aloud sessions may make participants feel more comfortable in reflecting their thoughts as they come. Previous research shows that thinking aloud becomes easier after practice (Van Someren et al., 1994). Accordingly, participants may have become comfortable with the protocol during the first few minutes of their 15-minute session in the current studies. Nonetheless, we recommend that future researchers include practice Think Aloud sessions prior to administration of the task.

Second, in Study 1 we collected data on thought content and qualities retrospectively rather than embedded within the thought probes. We chose to administer our measures after the SST session because we anticipated that probing participants in the middle of the session to assess the many variables we were measuring would have significantly interrupted the task, potentially disrupting the stream of consciousness. However, it is important to note that the accuracy of the retrospective reports may also have suffered from memory biases. We encourage researchers to correlate self-reports of a select few thought qualities with either rater-assigned qualities or with results from large language models to verify the validity of self-reports in future studies.

Finally, as Study 1 was conducted for the first time online during the pandemic, at-home internet and Zoom malfunctions led to a significant loss of data. Study 1 would have benefited from an increase in sample size and with less disturbances during the sessions themselves. While Study 2 attempted to address these concerns by conducting the study in person in a laboratory setting, it did not measure differences in the same thought qualities and content as Study 1.

4.7. Implications and concluding remarks

The mind has often been metaphorically described as a "black box," an opaque container whose inner workings are mysterious and inaccessible. Our research demonstrated that Think Aloud can serve as a model system for unverbalized SST, allowing us to open this black box and gain valuable insights into the content, qualities, and structure of the spontaneous stream of consciousness. The results from the current studies empirically validate and support the minimal reactivity of Think Aloud on the stream of consciousness. This is crucial for ensuring the accuracy and reliability of data collected through Think Aloud, enabling researchers to explore the landscape of human thought with confidence.

The evidence for minimal reactivity is particularly emphasized by the absence of significant differences in meta-awareness, topic shifting, and cognitive load between the Think Aloud and Silent Think conditions. This is especially important because if verbalizing the stream of consciousness had imposed a greater cognitive load, it would have signaled potential reactivity of thinking aloud on qualities of thought that were not measured here. The lack of cognitive load differences supports the idea that Think Aloud does not interfere with the natural flow of thought. However, while our studies largely affirm that SST is non-reactive to Think Aloud, some subtle signs of reactivity were observed in areas such as thought control, creativity, and well-being, suggesting these domains may be particularly sensitive and warrant further exploration. Indeed, if these differences prove to be robust, they might be effectively harnessed for interventions, such as for improving emotion regulation, increasing goal directed thinking, and decreasing the frequency of negative thoughts. For example, if thinking aloud reduces repetitive thoughts and is associated with a more positive mood, it might be helpful for targeting rumination.

Additionally, it is important to consider the inherent challenges in generalizing the applicability of Think Aloud protocols given the

linguistic limitations imposed by the method. Although the current manuscript does not examine individual differences, variability in verbalization abilities across individuals may determine the extent to which Think Aloud can be utilized to examine the stream of consciousness. Several factors, such as the presence of inner speech, personal meaning, cultural contexts, and linguistic abilities can influence the results generated using Think Aloud. For example, not everyone experiences inner speech to the same extent; some may primarily rely on abstract ideas or mental imagery rather than linguistic formats. Indeed, individuals with anendophasia often report thinking without words, and only translating images into words when necessary, a process that can also impose a higher cognitive load (Nendergaard & Lupyan, 2024). Moreover, some individuals may not experience thought in the language in which SST is being asked to be verbalized, which could lead to the loss of essential meaning, less topic shifting, greater cognitive load, and skewness in thought representations. Such variability suggests that thinking aloud might differentially impact participants and that Think Aloud protocols might be more representative of the inner thoughts of some individuals relative to others. Future research should explore how individual differences may moderate the impact of thinking aloud on the stream of thought, and people's perceptions of the degree to which their Think Aloud protocols reflect their inner thoughts.

Our current studies build on the initial explorations by Sripada and Taxali (2020) and Li et al. (2021) on the use of Think Aloud as a method to study the stream of consciousness. Previously, the exploration of the stream of consciousness has primarily relied on methods that involved self-reflection and answering questions while engaging in concurrent tasks. Although thought probes and experience sampling have been valuable tools in this area of research, there is a need to further investigate the potential of Think Aloud as it may help to overcome many of the limitations posed by the traditional methods, such as memory biases, the impact of introspection, and interruptions to thought flow. Additionally, our studies also identified an important limitation of the use of thought probes, where the presentation of probes altered the variable being measured. In contrast, the Think Aloud protocol encourages a continuous, non-introspective flow of thought, minimizing reliance on memory and providing information on thought content and qualities without interrupting participants or increasing meta-awareness.

Further validation of the Think Aloud protocol is crucial, considering the extent of its practical implications. If the stream of consciousness is minimally reactive to the Think Aloud protocol, it can provide a unique window into the real-time cognitive processes in areas such as problem-solving, creative ideation, mind wandering, usability research, emotion regulation, and more. While it may be an uncomfortable process, or impose a cognitive burden on certain individuals, the protocol's flexibility, cost-effectiveness, and ability to provide rich, real-time data on the inner workings of the mind may outweigh its potential pitfalls.

In conclusion, we have demonstrated the applicability of Think Aloud in a new domain of research by assessing whether and how people's unconstrained thoughts are altered by being spoken. The largely non-reactive impact of thinking aloud on SST suggests that it can enrich our understanding of the manner in which spontaneous thoughts meander down the stream of consciousness.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, A.G. used ChatGPT in order to correct the language and better the readability of the manuscript. The draft has since undergone 9 revisions, and the authors have reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Anusha Garg: Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shivang Shelat:** Writing – review & editing, Validation, Software, Project administration, Methodology, Conceptualization. **Madeleine E. Gross:** Writing – review & editing. **Jonathan Smallwood:** Writing – review & editing, Methodology, Conceptualization. **Paul Seli:** Writing – review & editing, Methodology, Conceptualization. **Aman Taxali:** Software, Resources, Formal analysis. **Chandra S. Sripada:** Writing – review & editing. **Jonathan W. Schooler:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Templeton Religious Trust grant (TRT0487 to M.E.G. and J.W.S.) and National Science Foundation Graduate Research Fellowship (2139319 to S.S.). We thank all our research assistants for their hard work in data collection, data cleaning, and partial analyses.

Appendix A

Thought Qualities Inventory.

Not at all (1) *A little bit* (2) *Somewhat* (3) *A moderate amount* (4) *A great deal* (5)

1. How bored were you doing the session?
2. How distressed were you during the session?
3. How much guidance/control did you have over your thoughts?
4. How natural did you find it to talk aloud (Think Aloud only)?
5. How often, if at all, did you experience a feeling of having no reportable content (mind blanking)?
6. How enjoyable did you find the session?
7. To what degree were you unable to vocalize your thoughts (Think Aloud only)?
8. My thoughts had a clear sense of purpose.
9. My thoughts were vague and non-specific.
10. My thoughts were fragmented and disjointed.
11. How interesting were the topics you were thinking about?

Not at all (1) *A little bit* (2) *Somewhat* (3) *A moderate amount* (4) *A great deal* (5)

12. How often were your thoughts goal-directed?
13. How often did you have new and/or creative ideas?
14. How productive were your thoughts?
15. To what degree did you repeat the same thoughts?
16. To what degree do you remember the thoughts from the session?
17. How often were your thoughts private?
18. To what degree were your thoughts related to things you were curious about?
19. To what extent were your thoughts related to topics that are meaningful to you?

5 times (1) *10 times* (2) *15 times* (3) *20 times* (4) *More than 20 times* (5)

20. How often did you topic shift?

Boredom (1) *Intrusion from the environment* (2) *Association to previous topic* (3) *Other* (4)

21. Of the following reasons, why do you think you topic switched the most often?

Images (1) *Words* (2) *Inner monologue*(3) *TV show or movie* (4) *Strong and consistent personal narrative* (5)

22. How best would you describe your thoughts in terms of the medium within which you experienced them? (You can choose more than 1).

Appendix B

18 Categories Coding Scheme items.
Please indicate from the following topics which you thought of during the task (you can choose more than 1) −.

1. Body: Anything related to the body and face that is not a concern related to health.
2. Health: Relating to physical and mental health, disorders and diseases.
3. Education: Relating to college, classes.
4. Work and Employment – Current or Future: Job related concerns.
5. Finances: Monetary concerns, paying bills, looking for a job for the purpose of earning money.
6. Friends, family, relatives, and acquaintances: Thoughts related to other people who have an established non-romantic relationship with the subject. (For romantic relationships, see Partner, Intimacy, Love, and Sexual Matters, for random people, see other areas).
7. Partner, Intimacy, Love, and Sexual Matters: Romantic relationships relating to a partner, or a romantic interest.
8. Self-Reflections and Evaluations: Anything to do with reflecting about oneself. This does not include evaluating anything else, just self.
9. Home and Household Matters: Fixing things, groceries.
10. Organization (Plans and To-Do list): Trying to map out their day, or organize certain things in order to help them.
11. Hobbies, Activities, and Recreation: Talking about enjoying certain things in life, activities, and hobbies.
12. Religion, Faith, and Spiritual Matters: Relating to faith or a spiritual pathway.
13. Holidays: Talking about holidays, and everything that relates to it.
14. Weather: Relating to concerns about the weather.
15. Immediate Basic Needs: Talking about what they want at the moment, relating to needs such as food, thirst, rest.
16. Environment and the Task: Environmental/Stimulus driven thoughts, thoughts about the task.
17. Ideas: Thoughts about topics in science/philosophy/history/literature/art/technology/nature/politics or any other topic of interest that is not immediately related to your work or coursework.
18. Other Areas: Random thoughts about random things (song stuck in head, wonder woman is awesome), evaluation of random things (not current environment).

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2025.103815.

Data availability

Data will be made available on request.

References

Amit, E., Hoeflin, C., Hamzah, N., & Fedorenko, E. (2017). An asymmetrical relationship between verbal and visual thinking: Converging evidence from behavior and fMRI. *NeuroImage, 152*, 619–627. https://doi.org/10.1016/j.neuroimage.2017.03.029

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment, 13*(1), 27–45. https://doi.org/10.1177/1073191105283504

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. https://doi.org/10.1177/0956797612446024

Bastian, M., Lerique, S., Adam, V., Franklin, M. S., Schooler, J. W., & Sackur, J. (2017). Language facilitates introspection: Verbal mind-wandering has privileged access to consciousness. *Consciousness and Cognition, 49*, 86–97. https://doi.org/10.1016/j.concog.2017.01.002

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology, 46*(3), 610–620. https://doi.org/10.1037/0022-3514.46.3.610

Brinthaupt, T. M., Hein, M. B., & Kramer, T. E. (2009). The self-talk scale: Development, factor analysis, and validation. *Journal of Personality Assessment, 91*(1), 82–92. https://doi.org/10.1080/00223890802484498

Carriere, J. S., Seli, P., & Smilek, D. (2013). Wandering in both mind and body: Individual differences in mind wandering and inattention predict fidgeting. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 67*(1), 19–31. https://doi.org/10.1037/a0031438

Chein, J. M., & Weisberg, R. W. (2014). Working memory and insight in verbal problems: Analysis of compound remote associates. *Memory & Cognition, 42*, 67–83. https://doi.org/10.3758/s13421-013-0343-4

Christoff, K., Gordon, A., & Smith, R. (2011). The role of spontaneous thought in human cognition. In *Neuroscience of decision making* (pp. 259–284). Psychology Press.

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), 718–731. https://doi.org/10.1038/nrn.2016.113

Chu, M. T., Marks, E., Smith, C. L., & Chadwick, P. (2023). Self-caught methodologies for measuring mind wandering with meta-awareness: A systematic review. *Consciousness and Cognition, 108*, Article 103463. https://doi.org/10.1016/j.concog.2022.103463

Cicero, D. C., Kerns, J. G., & McCarthy, D. M. (2010). The aberrant salience inventory: A new measure of psychosis proneness. *Psychological Assessment, 22*(3), 688–701. https://doi.org/10.1037/a0019913

de Groot, A. (1978). Thought and choice in chess. *Mouton*. https://doi.org/10.1515/9783110816877-008

DeMarree, K. G., & Naragon-Gainey, K. (2022). Individual differences in the contents and form of present-moment awareness: The Multidimensional Awareness Scale. *Assessment, 29*(3), 583–602. https://doi.org/10.1177/1073191120986605

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Ericsson, K. A., & Fox, M. C. (2011). Thinking aloud is not a form of introspection but a qualitatively different methodology: Reply to Schooler (2011). *Psychological Review*. https://doi.org/10.1037/a0022388

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Ericsson, K. A., & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. *Bradford Books/MIT Press*. https://doi.org/10.7551/mitpress/5657.001.0001

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Bradford Books/MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001.
Fleck, J. I. (2008). Working memory demands in insight versus analytic problem solving. *European Journal of Cognitive Psychology, 20*(1), 139–176. https://doi.org/10.1080/09541440601016954
Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*(2), 316–344. https://doi.org/10.1037/a0021663
Franklin, M. S., Mrazek, M. D., Anderson, C. L., Smallwood, J., Kingstone, A., & Schooler, J. W. (2013). The silver lining of a mind in the clouds: Interesting musings are associated with positive mood while mind-wandering. *Frontiers in Psychology, 4*, 583. https://doi.org/10.3389/fpsyg.2013.00583
Giambra, L. M. (1993). The influence of aging on spontaneous shifts of attention from external stimuli to the contents of consciousness. *Experimental Gerontology, 28*(4–5), 485–492. https://doi.org/10.1016/0531-5565(93)90073-M
Hart, S. G. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human mental workload*. Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9.
Irving, Z. C. (2016). Mind-wandering is unguided attention: Accounting for the "purposeful" wanderer. *Philosophical Studies, 173*, 547–571. https://doi.org/10.1007/s11098-015-0506-1
Irving, Z. C., & Glasser, A. (2020). Mind-wandering: A philosophical guide. *Philosophy Compass, 15*(1). https://doi.org/10.1111/phc3.12644. e12644.
Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
Jourard, S. M., & Lasakow, P. (1958). Some factors in self-disclosure. *The Journal of Abnormal and Social Psychology, 56*(1), 91–98. https://doi.org/10.1037/h0043357
Kane, M. J., Smeekens, B. A., Meier, M. E., Welhaf, M. S., & Phillips, N. E. (2021). Testing the construct validity of competing measurement approaches to probed mind-wandering reports. *Behavior Research Methods, 1–40*. https://doi.org/10.3758/s13428-021-01557-x
Kashdan, T. B., Stiksma, M. C., Disabato, D. J., McKnight, P. E., Bekier, J., Kaji, J., & Lazarus, R. (2018). The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality, 73*, 130–149. https://doi.org/10.1016/j.jrp.2017.11.011
Kaufmann, A., Parmigiani, S., Kawagoe, T., Zabaroff, E., & Wells, B. (2024). Two models of mind blanking. *European Journal of Neuroscience, 59*(5), 786–795. https://doi.org/10.1111/ejn.16164
Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E. V., & Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine, 35*(2), 245–256. https://doi.org/10.1017/S0033291704002892
Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science, 330*(6006), 932. https://doi.org/10.1126/science.1192439
Klinger, E. (1971). *Structure and functions of fantasy*. John Wiley & Sons.
Klinger, E. (1999). Thought flow: Properties and mechanisms underlying shifts in content. In J. A. Singer, & P. Salovey (Eds.), *At play in the fields of consciousness: Essays in honor of Jerome L. Singer* (pp. 29–50). Erlbaum.
Klinger, E. (2009). Daydreaming and fantasizing: Thought flow and motivation. In K. D. Markman, W. M. P. Klein, & J. A. Suhr (Eds.), *Handbook of imagination and mental simulation* (pp. 225–239). Psychology Press. https://doi.org/10.4324/9780203809846.ch15.
Klinger, E., & Cox, W. M. (2004). Motivation and the theory of current concerns. In W. M. Cox, & E. Klinger (Eds.), *Handbook of motivational counseling* (pp. 3–27). John Wiley & Sons. https://doi.org/10.1002/9780470713129.
Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. *Cambridge University Press*. https://doi.org/10.1017/CBO9781139087759
Li, H. X., Lu, B., Chen, X., Li, X. Y., Castellanos, F. X., & Yan, C. G. (2021). Exploring self-generated thoughts in a resting state with natural language processing. *Behavior Research Methods, 1–19*. https://doi.org/10.3758/s13428-021-01710-6
Lyubomirsky, S., & Tkach, C. (2003). The consequences of dysphoric rumination. In C. Papageorgiou, & A. Wells (Eds.), *Depressive rumination: Nature, theory and treatment* (pp. 21–41). John Wiley & Sons. https://doi.org/10.1002/9780470713853.ch2.
Lyubomirsky, S., Sousa, L., & Dickerhoof, R. (2006). The costs and benefits of writing, talking, and thinking about life's triumphs and defeats. *Journal of Personality and Social Psychology, 90*(4), 692–708. https://doi.org/10.1037/0022-3514.90.4.692
McCrae, R. R., & Costa, P. T., Jr. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences, 36*(3), 587–596. https://doi.org/10.1016/S0191-8869(03)00118-1
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
Mildner, J. N., & Tamir, D. I. (2019). Spontaneous thought as an unconstrained memory process. *Trends in Neurosciences, 42*(11), 763–777. https://doi.org/10.1016/j.tins.2019.09.001
Morey, R. D., Rouder, J. N. (2022). *BayesFactor: Computation of Bayes factors for common designs* (R package version 0.9.12-4.4). https://cran.r-project.org/package=BayesFactor.
Murray, S., Liang, N., Brosowsky, N., & Seli, P. (2021). What are the benefits of mind wandering to creativity? *Psychology of Aesthetics, Creativity, and the Arts, 18*(3), 403–414. https://doi.org/10.1037/aca0000420
Nedergaard, J. S., & Lupyan, G. (2024). Not everybody has an inner voice: Behavioral consequences of anendophasia. *Psychological Science, 35*(7). https://doi.org/10.1177/09567976241243004
Newell, A. (1972). *Human problem solving*. Prentice Hall.
Nielsen, L., & Madsen, S. (2012). October). The usability expert's fear of agility: An empirical study of global trends and emerging practices. In *In Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (pp. 261–264). https://doi.org/10.1145/2399016.2399057
Nolen-Hoeksema, S. (2003). The response styles theory. In C. Papageorgiou, & A. Wells (Eds.), *Depressive rumination: Nature, theory and treatment* (pp. 105–123). John Wiley & Sons. https://doi.org/10.1002/9780470713853.
Paromita, P., Mundnich, K., Nadarajan, A., Booth, B. M., Narayanan, S. S., & Chaspari, T. (2023). Modeling inter-individual differences in ambulatory-based multimodal signals via metric learning: A case study of personalized well-being estimation of healthcare workers. *Frontiers in Digital Health, 5*, Article 1195795. https://doi.org/10.3389/fdgth.2023.1195795
Poerio, G. L., Totterdell, P., & Miles, E. (2013). Mind-wandering and negative mood: Does one thing really lead to another? *Consciousness and Cognition, 22*(4), 1412–1421. https://doi.org/10.1016/j.concog.2013.09.012
Poerio, G. L., Totterdell, P., Emerson, L. M., & Miles, E. (2015). Love is the triumph of the imagination: Daydreams about significant others are associated with increased happiness, love and connection. *Consciousness and Cognition, 33*, 135–144. https://doi.org/10.1016/j.concog.2014.12.011
Raffaelli, Q., Malusa, R., de Stefano, N. A., Andrews, E., Grilli, M. D., Mills, C., & Andrews-Hanna, J. R. (2023). Creative minds at rest: Creative individuals are more associative and engaged with their idle thoughts. *Creativity Research Journal, 36*(3), 396–412. https://doi.org/10.1080/10400419.2023.2227477
Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225
Sayette, M. A., Reichle, E. D., & Schooler, J. W. (2009). Lost in the sauce: The effects of alcohol on mind wandering. *Psychological Science, 20*(6), 747–752. https://doi.org/10.1111/j.1467-9280.2009.02351.x
Sayette, M. A., Schooler, J. W., & Reichle, E. D. (2010). Out for a smoke: The impact of cigarette craving on zoning out during reading. *Psychological Science, 21*(1), 26–30. https://doi.org/10.1177/0956797609354059
Schooler, J. W. (2002). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences, 6*(8), 339–344. https://doi.org/10.1016/S1364-6613(02)01949-6
Schooler, J. W. (2011). Introspecting in the spirit of William James: Comment on fox, ericsson, and best. *Psychological Bulletin, 137*(2), 345–350. https://doi.org/10.1037/a0022390
Schooler, J. W., & Melcher, J. (1995). The ineffability of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 97–133). MIT Press.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*(2), 166–183. https://doi.org/10.1037/0096-3445.122.2.166
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. https://doi.org/10.1016/j.tics.2011.05.006
Seli, P., Ralph, B. C., Risko, E. F., Schooler, J. W., Schacter, D. L., & Smilek, D. (2017). Intentionality and meta-awareness of mind wandering: Are they one and the same, or distinct dimensions? *Psychonomic Bulletin & Review, 24*, 1808–1818. https://doi.org/10.3758/s13423-017-1249-0
Shelat, S., Marquez, J. J., Zheng, J., & Karasinski, J. A. (2024). Collaborative system usability in spaceflight analog environments through remote observations. *Applied Sciences, 14*(5), 2005. https://doi.org/10.3390/app14052005
Shelat, S., Schooler, J. W., & Giesbrecht, B. (2024). Predicting attentional lapses using response time speed in continuous performance tasks. *Frontiers in Cognition, 3*, Article 1460349. https://doi.org/10.3389/fcogn.2024.1460349
Siddiq, F., & Scherer, R. (2017). Revealing the processes of students' interaction with a novel collaborative problem solving task: An in-depth analysis of think-aloud protocols. *Computers in Human Behavior, 76*, 509–525. https://doi.org/10.1016/j.chb.2017.08.007
Singer, J. A. (2004). Narrative identity and meaning making across the adult lifespan: An introduction. *Journal of Personality, 72*, 437–459. https://doi.org/10.1111/j.0022-3506.2004.00268.x
Singer, J. L. (1975). Navigating the stream of consciousness: Research in daydreaming and related inner experience. *American Psychologist, 30*(7), 727–738. https://doi.org/10.1037/h0076928
Smallwood, J., & Andrews-Hanna, J. (2013). Not all minds that wander are lost: The importance of a balanced perspective on the mind-wandering state. *Frontiers in Psychology, 4*, 441. https://doi.org/10.3389/fpsyg.2013.00441
Smallwood, J., & O'Connor, R. C. (2011). Imprisoned by the past: Unhappy moods lead to a retrospective bias to mind wandering. *Cognition & Emotion, 25*(8), 1481–1490. https://doi.org/10.1080/02699931.2010.545263
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
Smallwood, J., Fitzgerald, A., Miles, L. K., & Phillips, L. H. (2009). Shifting moods, wandering minds: Negative moods lead the mind to wander. *Emotion, 9*(2), 271–276. https://doi.org/10.1037/a0014855
Smallwood, J., McSpadden, M., & Schooler, J. W. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin & Review, 14*, 527–533. https://doi.org/10.3758/BF03194102
Smallwood, J., Schooler, J. W., Turk, D. J., Cunningham, S. J., Burns, P., & Macrae, C. N. (2011). Self-reflection and the temporal focus of the wandering mind. *Consciousness and Cognition, 20*(4), 1120–1126. https://doi.org/10.1016/j.concog.2010.12.017
Smeekens, B. A., & Kane, M. J. (2016). Working memory capacity, mind wandering, and creative cognition: An individual-differences investigation into the benefits of controlled versus spontaneous thought. *Psychology of Aesthetics, Creativity, and the Arts, 10*(4), 389–400. https://doi.org/10.1037/aca0000046
Smyth, J., True, N., & Souto, J. (2001). Effects of writing about traumatic experiences: The necessity for narrative structuring. *Journal of Social and Clinical Psychology, 20*(2), 161–172. https://doi.org/10.1521/jscp.20.2.161.22266
Sripada, C. S. (2018). An exploration/exploitation trade-off between mind wandering and goal-directed thinking. In K. C. Fox, & K. Christoff (Eds.), *The Oxford handbook of spontaneous thought: Mind-wandering, creativity, and dreaming* (pp. 23–34). Oxford University Press.
Sripada, C., & Taxali, A. (2020). Structure in the stream of consciousness: Evidence from a verbalized thought protocol and automated text analytic methods. *Consciousness and Cognition, 85*, Article 103007. https://doi.org/10.1016/j.concog.2020.103007
Stawarczyk, D., Majerus, S., & D'Argembeau, A. (2013). Concern-induced negative affect is associated with the occurrence and content of mind-wandering. *Consciousness and Cognition, 22*(2), 442–448. https://doi.org/10.1016/j.concog.2013.01.012
Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D'Argembeau, A. (2011). Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica, 136*(3), 370–381. https://doi.org/10.1016/j.actpsy.2011.01.002
Steindorf, L., Hammerton, H. A., & Rummel, J. (2021). Mind wandering outside the box—About the role of off-task thoughts and their assessment during creative incubation. *Psychology of Aesthetics, Creativity, and the Arts, 15*(4), 584–597. https://doi.org/10.1037/aca0000373
Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 28*(2), 191–193. https://doi.org/10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0.CO;2-G
Struk, A. A., Carriere, J. S., Cheyne, J. A., & Danckert, J. (2017). A short boredom proneness scale: Development and psychometric properties. *Assessment, 24*(3), 346–359. https://doi.org/10.1177/1073191115609996
Van Someren, M., Barnard, Y. F., & Sandberg, J. (1994). *The Think Aloud method: A practical approach to modelling cognitive processes.* Academic Press.
Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin, 134*(2), 163–206. https://doi.org/10.1037/0033-2909.134.2.163
Watkins, E. R., & Roberts, H. (2020). Reflecting on rumination: Consequences, causes, mechanisms and treatment of rumination. *Behaviour Research and Therapy, 127*, Article 103573. https://doi.org/10.1016/j.brat.2020.103573
Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063
Weinstein, Y. (2017). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior Research Methods, 50*(2), 642–661. https://doi.org/10.3758/s13428-017-0891-9
Welhaf, M. S., & Bugg, J. M. (2024). Positively framing mind wandering does not increase mind wandering in older adults. *Psychological Research Psychologische Forschung, 1–11*. https://doi.org/10.1007/s00426-024-01983-4
Wiemers, E. A., & Redick, T. S. (2019). The influence of thought probes on performance: Does the mind wander more if you ask it? *Psychonomic Bulletin & Review, 26*(2), 367–373. https://doi.org/10.3758/s13423-018-1529-3
Willroth, E. C., & Atherton, O. E. (2024). Best laid plans: A guide to reporting preregistration deviations. *Advances in Methods and Practices in Psychological Science, 7*(1). https://doi.org/10.1177/25152459231213802 (25152459231213802).
Wilson, T. D., Reinhard, D. A., Westgate, E. C., Gilbert, D. T., Ellerbeck, N., Hahn, C., & Shaked, A. (2014). Just think: The challenges of the disengaged mind. *Science, 345*(6192), 75–77. https://doi.org/10.1126/science.1250830