IN REVISION: Behavior Research Methods

Attention in hindsight: Using stimulated recall to capture dynamic fluctuations in attentional engagement

Effie J. Pereira*[1], Samantha Ayers-Glassey*[1], Jeffrey D. Wammes[2], & Daniel Smilek[1]

[1] *Department of Psychology, University of Waterloo*

[2] *Department of Psychology, Queen's University*

* equal contribution to the manuscript

Abstract count: 211

Word count: 9,843

Correspondence to:

Effie J. Pereira
Department of Psychology, University of Waterloo
PAS Building, 200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1-519-888-4567 ext. 42813
Email: effie.pereira@uwaterloo.ca

Abstract

Attentional engagement is known to vary on a moment-to-moment basis. However, few self-report methods can effectively capture dynamic fluctuations in attentional engagement over time. In the current paper, we evaluated the utility of stimulated recall, a method wherein individuals are asked to remember their subjective states while using a mnemonic cue, for the measurement of temporal changes in attentional engagement. Across three experiments, we asked participants to watch video lectures and assessed their in-the-moment levels of attentional engagement using intermittent thought probes. Then, we used stimulated recall by cueing participants with short video clips from the lectures to retrospectively assess the levels of attentional engagement they had experienced when they first watched those clips within the video lectures. Experiment 1 tested the statistical overlap between in-the-moment and video-stimulated ratings, Experiment 2 tested the generalizability of stimulated recall across different types of videos, and Experiment 3 tested order effects by presenting the video-stimulated probe clips out of sequential order. Across all experiments, we found statistically robust correspondence between in-the-moment and video-stimulated ratings of attentional engagement, illustrating a strong convergence between these two methods of assessment. Taken together, our findings indicate that stimulated recall provides a new and practical methodological approach that can accurately capture dynamic fluctuations in subjective attentional states over time.

Keywords: attentional engagement, retrospective attention, probe methods, stimulated recall

Attention is responsible for shaping our experience of the world. By flexibly controlling and guiding our limited computational resources over time, this neurocognitive mechanism can dynamically shift and vary to bring different sources of information into our consciousness (Anderson et al., 2011; Carrasco, 2011; Desimone & Duncan, 1995; Itti & Koch, 2001; Luck & Vecera, 2002; Pashler, 1998; Posner & Petersen, 1990; Wolfe & Horowitz, 2004). As time unfolds, attentional engagement can fluctuate on a moment-to-moment basis between states of attentiveness when we are focused on the external environment and states of mind wandering when we are focused on our internal thoughts. Prior studies have demonstrated that attentional engagement can vary across diverse timescales (e.g., over the course of a day, Busch & VanRullen, 2010; Hylan, 1898; over the span of an hour, Smallwood et al., 2008; Terhune et al., 2017; over a period of seconds, Cheyne et al., 2009; Liddell, 1919) and across various activities (e.g., during work, Gable et al., 2019; Risko et al., 2013; Smallwood et al., 2011; Szpunar, Moulton, et al., 2013; during leisure activities, Feng et al., 2013; Killingsworth & Gilbert, 2010; Kopp et al., 2016; during periods of rest, Christoff et al., 2016; Delorme & Brandmeyer, 2019; Hasenkamp et al., 2012; Tusche et al., 2014). Shifts in engagement have also been associated with wide-ranging behavioural outcomes (e.g., changes in reading and comprehension, McVay & Kane, 2012; Schad et al., 2012; Unsworth & McMillan, 2013; fluctuations in learning and retention, Kane et al., 2007; Risko et al., 2012; differences in overall performance, Bastian & Sackur, 2013; Cheyne et al., 2009; Smallwood et al., 2007). For these reasons, sophisticated methods are needed that can comprehensively capture dynamic fluctuations in attentional engagement.

Currently, two self-report approaches are commonly used in the literature when assessing changes in attentional engagement over time. *Thought probes* (Giambra, 1995; Killingsworth &

Gilbert, 2010; Larson & Csikszentmihalyi, 1978; Wammes et al., 2016) are experience-sampling methods that rely on immediate assessments by interrupting participants throughout their task to ask about their current engagement at that moment in time. For example, Smallwood et al. (2004) randomly presented thought probes during a sustained attention task to demonstrate that this method can accurately distinguish when states of mind wandering were related or unrelated to the task at hand. In contrast, *retrospective reports* (Kahneman et al., 2004; Matthews et al., 1999) are aggregate methods that rely on delayed recall by waiting until the completion of the task to ask participants about their overall engagement throughout the task. For instance, Baird et al. (2012) used retrospective reports to assess mind wandering during demanding or undemanding tasks to show that easy and simple tasks that allow the mind to freely wander can facilitate creative problem solving.

Although both thought probes and retrospective reports have been widely used for the measurement of attentional engagement, they both have a serious shortcoming in that they are not suited for capturing temporal precision in the subtle fluctuations of attentional engagement over time. Specifically, when utilizing thought probes, measuring attentional engagement on a frequent moment-to-moment basis is theoretically possible but not practically feasible since probes interrupt the flow of the task and can disrupt ongoing attentional states. Past studies have established that thought probes appearing several minutes apart allow sufficient time for individuals to re-engage with their task (Leroy et al., 2020; Marty-Dugas et al., 2021; Welhaf et al., 2022). This finding implies that probing at a higher rate could likely disrupt natural engagement and/or task performance, an idea supported by prior work showing that the frequency of thought probes throughout a task influences the relative proportion of reported mind wandering (Greve & Was, 2022; Schubert et al., 2020; Seli, Carriere, et al., 2013).

Although retrospective reports avoid this issue since they do not interrupt task performance, they suffer from other limitations. Namely, retrospective reports typically require participants to aggregate (i.e., sum or average) their experiences over a more extended period to time (sometimes even across the whole task), and as commonly implemented, they fail to direct participants to report their engagement for specific points in time (Ellison et al., 2020; Karapanos et al., 2010). Thus, for different reasons, thought probes and retrospective reports lack temporal precision because they under-sample the underlying attentional engagement one is attempting to quantify.

Given the need for approaches that can track dynamic fluctuations in attentional engagement over time with a high degree of precision, one viable approach would be to utilize methods that can allow for greater temporal specificity when measuring attentional states. An existing method that could be adapted for these purposes is *stimulated recall*, wherein participants' experiences are retrospectively assessed upon completion of a task through the use of a mnemonic cue (e.g., audio, images, video) that involves re-presenting segments of the original event (Bloom, 1953; Calderhead, 1981; Gass & Mackey, 2016; Lyle, 2003). For example, stimulated recall has been previously used by asking individuals to rate the moment-to-moment affective states they experienced when giving a presentation, while watching a video recording of their presentation as a recall cue (Gregersen et al., 2014). In this manner, stimulated recall preserves the advantages of thought probes and retrospective reports in that it allows for the measurement of subjective states over time; however, this method additionally allows for the measurement of as many specific time points as necessary via the recall cues. This provides the opportunity for a high degree of temporal precision when retrospectively probing dynamic and free-flowing experiences.

For these reasons, stimulated recall has proven advantageous within numerous fields of research to study a wide range of cognitive and affective experiences. For example, within behavioural research, stimulated recall has been used to examine constraints on working memory due to cognitive load (Beers et al., 2006); within communication, it has been used to study individuals' second-language experiences (MacIntyre & Legatto, 2010) and musicians' techniques for successful collaboration (Dempsey, 2010); within education, it has lent insight into students' thoughts during classroom lectures (O'Brien, 1993), children's perceptions of their learning (Morgan, 2007), and teachers' pedagogical beliefs (Meade & McMeniman, 1992); and within clinical domains, this method has accurately captured physicians' memory for clinical encounters (Sinnott et al., 2017) and experiences of the physician-patient relationship (Paskins et al., 2014). Given the efficacy of this method across numerous domains, it seems likely that this approach could transfer successfully to measuring attentional engagement as well.

Here, we assessed the viability of stimulated recall for capturing dynamic fluctuations in attention over time by evaluating the convergence between in-the-moment and retrospective ratings of attentional engagement. To do so, participants completed two experimental phases of the study. In Phase 1, participants completed a video-viewing task, which included thought probes to capture participants' in-the-moment ratings of attentional engagement (immediate reports). Following this, in Phase 2, we implemented video-stimulated recall by presenting participants with short clips from the video they had just viewed and asking them to retrospectively estimate their level of attentional engagement when they first viewed that section of the video (video-stimulated reports). Video-stimulated reports were then compared against immediate reports to assess for statistical overlap. Experiment 1 assessed the overlap for a 30-minute video when video-stimulated reports were measured in the same sequential runtime order as immediate reports. Experiments 2 and 3 then assessed the generality of our findings by using two 15-minute videos that varied on different engagement metrics (Experiment 2) and querying whether video-stimulated reports remained accurate when presented out of sequential order (Experiment 3)[1]. Across all three experiments, our findings indicated that attentional engagement as measured via stimulated recall statistically overlapped with in-the-moment probe responses, validating that this methodological approach can be used to provide accurate and dynamic fluctuations of attentional engagement over time.

Experiment 1

Methods

Participants

Prior to data collection, an a-priori power analysis was conducted using G*Power (test family: exact, statistical test: correlation; Faul et al., 2009) to determine the required sample size needed to detect a small effect size (ρ = 0.1; Cohen, 1988). Based on an alpha of .05 and power of .95, it was determined that approximately 103 participants would be needed. Because participants were recruited through online settings where the possibility of data loss due to standards of quality control is higher (Peer et al., 2014), we strategically overshot this sample size. As such, 112 participants (64 women, 45 men, 1 non-binary, 2 non-response; M_{age} = 20.9 years, SD_{age} = 1.9 years) were recruited for the experiment through the university's online participant pool. Informed written consent was obtained from all, and course credit was provided

[1] Although the data were collected in order of Experiment 2 (Jan to Aug 2021), Experiment 3 (Apr to Aug 2021), and Experiment 1 (May to Oct 2021), they are presented here in the current order to aid with theoretical and methodological comparisons.

as compensation for their time. The study was conducted in accordance with the Declaration of Helsinki, and all protocol and procedures were approved by the University of Waterloo's research ethics board (#42793).

Materials and Stimuli

Video Lectures. A single 30-minute video lecture was used as the stimulus for Experiment 1: a graduation speech featured by TED Talk on evolutionary psychology titled *The Uniqueness of Humans* (Sapolsky, 2009). The video contained both audio and visual elements and was displayed at full-screen resolution, embedded in the browser, with no playback controls visible. Prior to the video-viewing task, each participant was asked to report on their familiarity with the video topic on a scale from 1 (not at all familiar) to 7 (very familiar). Overall, the topic was reported as having a mean familiarity of 3.7 (SD = 1.6).

Immediate Thought Probes. Participants' in-the-moment attentional engagement was assessed using 20 immediate thought probes. Each immediate probe asked, "*Just before this screen appeared, how engaged were you with the video lecture?*", with response options presented as a multiple-choice Likert-style scale ranging from 1 (not at all) to 7 (very). Immediate probe timestamps were programmed to occur randomly throughout the lecture, with the stipulation that probes could not occur within 60 seconds of one another. On average, immediate probes appeared every 90 seconds throughout the lecture.

Video-Stimulated Probes. Participants' retrospective attentional engagement was measured using twenty 15-second video clips from the lecture, followed by a video-stimulated probe. Each video-stimulated probe asked, "*How engaged were you when you first watched this* *section of the video lecture?*”, with response options presented as a multiple-choice Likert-style scale ranging from 1 (not at all) to 7 (very).

Video-stimulated probe timestamps were programmed to occur at least 10 seconds outside of any immediate probe timestamps, such that video-stimulated probes never overlapped with any immediate probes for the same participant, i.e., participants’ attentional engagement was never probed twice for the exact same content. This was done to ensure that video-stimulated responses were not biased by participants’ initial immediate probe responses, and instead based on their actual remembered attentional states from their initial viewing of the lecture content. All clips associated with the video-stimulated probes were presented in sequential runtime order, i.e., in the same chronological order as they had originally appeared in the video lecture.

Other Materials. There were several additional measures employed within the study design that were not directly related to the aims of the current paper and were thus not included in the current analyses. We summarize them here in the interest of full methodological disclosure. Specifically, after completing both phases of the study (i.e., the video-viewing task and video-simulated recall), participants completed a multiple-choice quiz pertaining to the content of the video lecture, a self-report measure of their level of motivation in completing the study, and a self-report measure of attentional control (Derryberry & Reed, 2002).

Procedure

Figure 1 illustrates the task sequence for the study. After providing informed consent, participants were asked to indicate their demographic information (i.e., gender, age). Participants then completed Phase 1 of the study. Here, they completed the video-viewing task, where they were asked to view a video lecture and respond to immediate thought probes that would occur at random intervals. Participants were also instructed that they would be tested on their knowledge of the video content afterwards and were asked not to take notes or multitask while viewing the lecture. Once Phase 1 of the study was complete, participants moved on to Phase 2. During Phase 2, participants completed the video-stimulated recall task, where they were asked to view short clips from the video lectures that they had just watched and then respond to each with a video-stimulated probe.

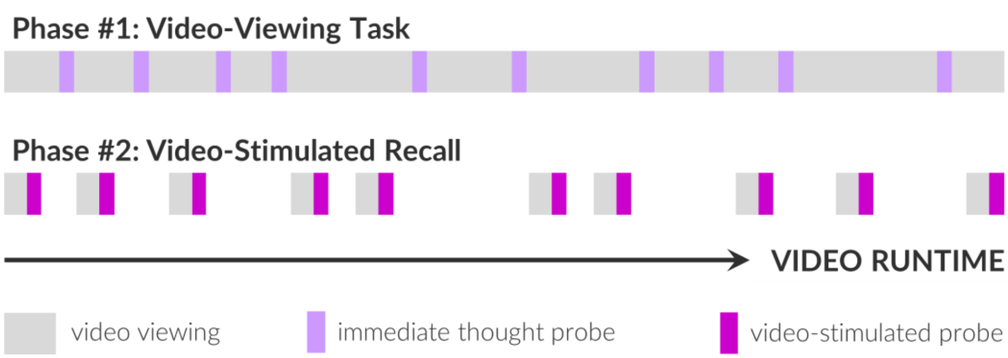

Figure 1. An illustration of the task sequence for the study. In Phase 1, participants viewed a video lecture and were intermittently asked to respond to immediate thought probes. In Phase 2, participants watched short clips from the video lecture and were asked to respond to associated video-stimulated probes. Critically, the video-stimulated probes did not overlap with periods of time that were probed during the initial video-viewing task.

Results

Data Exclusions

Prior to data analysis, we removed data from any participants if they indicated that they had previously viewed the video lecture used for the study, given that prior knowledge of the video content could impact both immediate and video-stimulated ratings of attentional engagement; five participants were excluded per this criterion. We also removed data from any participants if they were non-responsive to 20% of either immediate or video-stimulated probes given the difficulty of tracking attentional engagement accurately in these cases; six participants were excluded for this reason. No other exclusions were implemented. As a result, data from 101 participants (56 women, 42 men, 1 non-binary, 2 non-response; M_{age} = 20.9 years, SD_{age} = 1.9 years) were included in the analysis.

Data Analysis Plan

The aim of the current study was to assess the viability of stimulated recall for capturing attentional fluctuations over time by determining the degree to which video-stimulated reports of attentional engagement converged with immediate reports of attentional engagement. Therefore, we evaluated immediate and video-stimulated probe data using two complementary approaches.

First, as a preliminary check, we evaluated aggregate differences between immediate and video-stimulated probes by assessing overall ratings of attentional engagement using a paired-samples *t*-test as a function of *probe type* (immediate versus video-stimulated). This analysis provided information regarding whether overall engagement ratings differed based on how they were probed.

However, aggregate comparisons of attentional engagement can only provide a *general* assessment of the differences between immediate and video-stimulated ratings, and because timestamps between immediate and video-stimulated probes never overlapped with one another, a *precise* assessment that preserves the temporal relationship between the two is essential. Therefore second, as a more thorough exploration, we evaluated the temporal overlap between immediate and video-stimulated probes. Although prior work using stimulated recall has typically evaluated temporal overlap through qualitative assessments of data (Gibbs, 2007; Willig, 2013), we aimed to take a quantitative approach instead. To do so, we used cross-correlations and linear mixed-effects models to assess the underlying time series of immediate and video-stimulated probes (Brunsdon & Skinner, 1987; Gregson, 2014; Heath, 2000). Full analysis specifications are presented in Supplementary Materials.

Consider the sample data for immediate and video-stimulated probes depicted in Figure 2. Based on past literature (Cheyne et al., 2009; Hylan, 1898; Liddell, 1919), we can assume that discrete responses to thought probes of attentional engagement are merely time points along a continuous underlying time series of attentional engagement. For this reason, we interpolated immediate probe responses to construct an *immediate time series*, and similarly interpolated video-stimulated probe responses to construct a *video-stimulated time series* (Figure 2a). Then, to examine the overlap between the immediate and video-stimulated time series, we used cross-correlations to assess the temporal relationship between the two to determine whether they represented the same underlying fluctuations in attentional engagement. For this, cross-correlation scores between the immediate and video-stimulated time series were calculated when the two directly overlapped, i.e., at a time lag of 0 minutes, and when the time series were shifted forwards or backwards in time from one another, i.e., at various other time lags (Figure 2b). Given that cross-correlation scores can range from +1 (perfectly correlated) to 0 (not correlated) to –1 (negatively correlated), a maximally-positive cross-correlation at a time lag of 0 minutes would represent temporal convergence between immediate and video-stimulated time series, whereas increasing time lags would represent more temporally disparate comparisons.

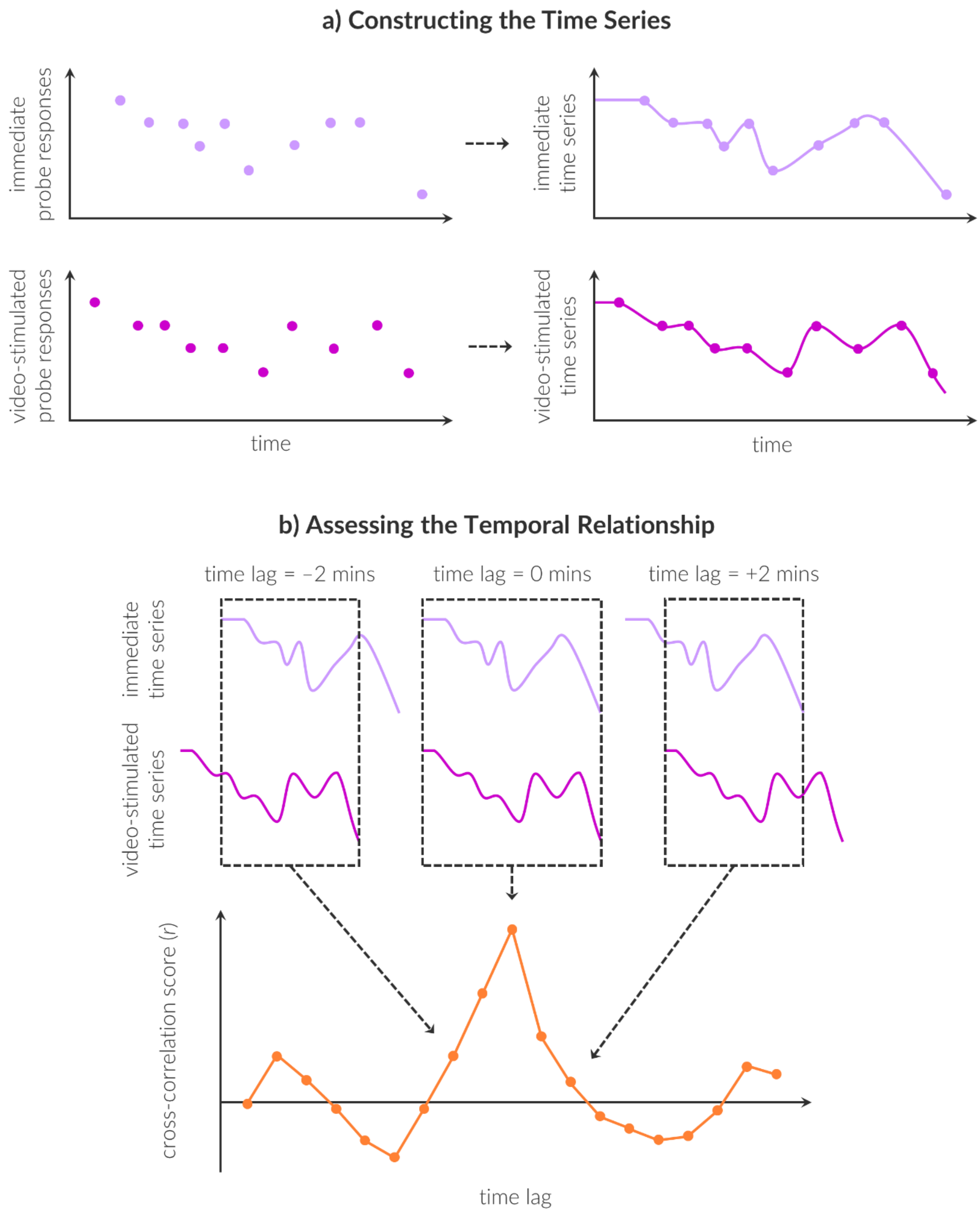

Figure 2. A depiction of the temporal analysis used in the study: (a) We first constructed the underlying immediate and video-stimulated time series by interpolating the immediate and video-stimulated probe responses, respectively; (b) We then assessed the temporal relationship between immediate and video-stimulated time series at a time lag of 0 compared to various other time lags.

Once cross-correlation scores across all time lags were computed, we then used linear mixed-effects models to estimate the cross-correlation score at a time lag of 0 minutes (intercept) and the impact of time lag on cross-correlation scores (slope). This approach was selected because cross-correlations of behavioural and cognitive data are known to have a large range and variability (see Baayen et al., 2008; Paxton & Dale, 2013) and because time lag variables are by nature correlated with one another. All models were run using fully specified random-effects structure across all intercepts and slopes (Barr et al., 2013).

Across aggregate measures, if immediate and video-stimulated probe responses were equivalent such that they both represented similar overall ratings of attentional engagement, we would expect to find similar ratings across both types of probes (Hypothesis 1). Across temporal measures, if immediate and video-stimulated time series were equivalent such that they both represented the same underlying fluctuations in a given individual's attentional engagement, we would expect to find peak positive cross-correlation scores at a time lag of 0 minutes (Hypothesis 2a) and decreases in cross-correlation scores with a shift in time lag away from 0 minutes (Hypothesis 2b).

While finding the foregoing patterns would provide compelling evidence for temporal overlap between the two time series, it is possible that these patterns could arise due to intrinsic properties of the video or random fluctuations that could be expected by chance (Dean & Dunsmuir, 2016), and not by immediate and video-stimulated time series effectively capturing the same underlying fluctuations within each individual. Therefore, to test this possibility, we created a noise distribution from our dataset by obliterating the association between each individual's immediate and video-stimulated time series. To do so, we first randomly shuffled our data such that each participants' immediate time series was paired without replacement with *another* participants' video-stimulated time series rather than their own. Cross-correlation scores were calculated for these shuffled pairs, and linear mixed-effects models were run to estimate the cross-correlation at a time lag of 0 minutes (intercept) and the impact of time lag on cross-correlation scores (slope). This procedure was then repeated 1,000 times to construct noise distributions for both the intercept and the slope, which were then assessed in two ways.

First, across all 1,000 repetitions, we examined the average effects of the noise distributions for all linear mixed-effects models. If stimulated recall were accurately capturing a given individual's attentional engagement across the task and not intrinsic or random fluctuations that could be expected by chance, we would expect these noise distributions to fail Hypotheses 2a and 2b (Hypotheses 2c and 2d). Second, for each of the 1,000 repetitions, we computed the *z*-score of our original data's true effect with respect to the noise distributions to directly compare these results against one another. We did this by first subtracting the intercept and slope of the noise distribution (i.e., the data associated with Hypotheses 2c and 2d) from the true intercept and slope of our original data (i.e., the data associated with Hypotheses 2a and 2b), and dividing each of these differences by the standard deviation of the noise distribution. If stimulated recall were capturing attentional engagement that was significantly different than our noise distribution, we would expect our *z*-scores to exceed the critical *z*-value at a 95% confidence level, i.e., an absolute value of +/–1.96 (Hypothesis 2e).

Aggregate Assessment of Attentional Engagement

A paired-samples *t*-test conducted as a function of *probe type* (immediate versus video-stimulated) found no differences in average attentional engagement across immediate [M = 5.15, SD = 1.59] and video-stimulated responses [M = 5.05, SD = 1.67; $t(100)$ = 1.20, p = .23, d_z =

.12], providing evidence for an overall similarity in measures between the two methods of assessment.

Temporal Assessment of Attentional Engagement

Cross-correlation scores at a time lag of 0 minutes ranged from –.60 to +.90 for the video lecture [*M* = .16, *SD* = .39]. Mean cross-correlations as a function of time lag are presented in Figure 3a (see Supplementary Materials for individual participant data).

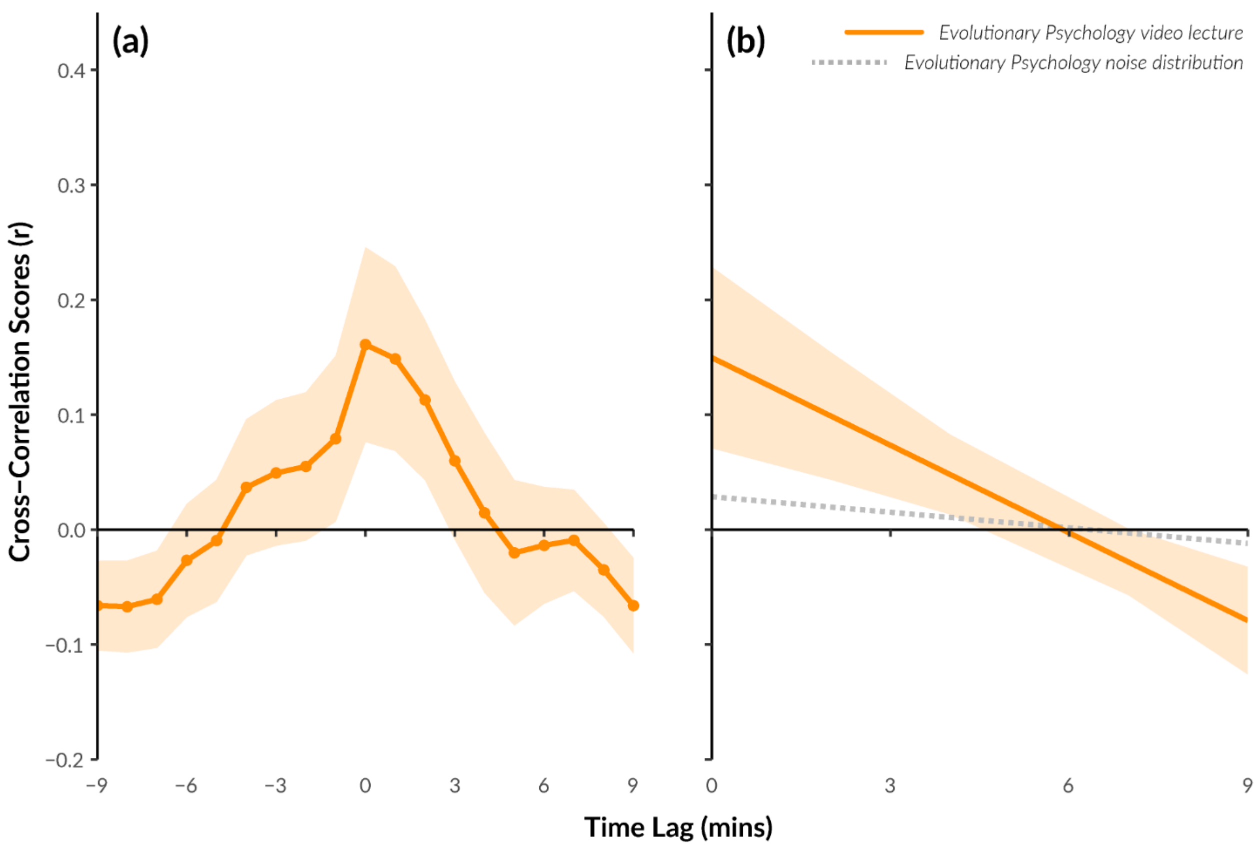

Figure 3. Data and results from Experiment 1. (a) Mean cross-correlation scores as a function of time lag. (b) Results from the linear mixed-effects model testing the effects of positive time lag (0 to +9 minutes) on cross-correlation scores. Model 1 (solid line) found robust temporal overlap at a time lag of 0 minutes, which decreased as time lag increased; however, when examining the average temporal overlap that could be expected by chance across all 1,000 repetition of Model 2 (dashed line), no such effects were found. Shaded bands depict +/– 95% CIs.

For all linear mixed-effects models presented, the coefficients are described using the unstandardized cross-correlation values rather than standardized beta weights. However, for reference on the strength of these effects, Table 1 presents both unstandardized (*b*) and standardized (β) results for all models conducted. In addition, for ease of interpretation, *p*-value calculations are provided using Satterthwaite approximations (Luke, 2017; Satterthwaite, 1941).

Table 1. Results of the linear mixed-effects models for Experiment 1.

Model predicting cross-correlation scores	*b*	β	t
Model 1: Evolutionary Psychology video lecture			
Intercept	.15	.00	3.73***
Time lag (0 to +9 minutes)	–.03	–.24	–3.93***
Model 2: Evolutionary Psychology noise distribution[a]			
Intercept	.03	.00	0.70
Time lag (0 to +9 minutes)	–.01	–.04	–0.63

* $p < .05$, ** $p < .01$, *** $p < .001$
[a] b, β, and t for Model 2 represents average values across all 1,000 repetitions.

Overlap as a function of time lag. Model 1 examined the role of time lag (0 to +9 minutes in 1-minute increments)[2] on cross-correlation scores (see Figure 3b). Linear mixed-effects models demonstrated a significant positive intercept ($b = .15$, $p < .001$), indicating that temporal overlap at a time lag of 0 minutes was significantly different than a null correlation. In

[2] To avoid overparameterization within the model, only positive time lags were used in the analysis; however, analyzing the data using only negative time lags replicated the pattern of effects found. See Supplementary Materials for complete analyses.

addition, cross-correlations were found to decrease as time lag increased (b = –.03, p < .001), illustrating that the temporal overlap between immediate and video-stimulated time series was greater when they were closer in time.

Overlap for a noise distribution. To confirm that cross-correlation scores within our data were not due to intrinsic or random fluctuations, we randomly paired each participant's immediate time series with another participant's video-stimulated time series, calculated cross-correlation scores across all time lags, and used linear mixed-effects models to examine the role of time lag (0 to +9 minutes) on cross-correlation scores. This procedure was then repeated 1,000 times to construct our noise distribution. Across all repetitions (see Figure 3b), no significant differences were found for the average intercept (mean b = .03, mean p = .45) or for the average time lag (mean b = −.01, mean p = .46). In addition, z-scores (intercept = 3.23, time lag = –3.21), representing the extent to which our original data's true effect differed from an intercept and slope that could be expected by chance, both exceeded the critical z-value at a 95% confidence level (i.e., +/–1.96). As such, our noise distributions indicated that the temporal overlap found within our original data is statistically different than patterns occurring by chance.

Discussion

In Experiment 1, we evaluated the utility of a new method for capturing dynamic changes in attentional engagement by examining the statistical convergence between ratings assessed via immediate and video-stimulated thought probes. Across aggregate measures, we found no differences between immediate and video-stimulated ratings (Hypothesis 1), demonstrating that engagement on average did not differ based on the method of assessment. Then across temporal measures, we found a statistically robust overlap between the underlying time series of the two. Namely, cross-correlations were highest at a time lag of 0 minutes (Hypothesis 2a) and decreased with an increase in time lag (Hypothesis 2b), indicating a temporally precise convergence between immediate and video-stimulated time series. Additionally, when examining noise distributions constructed via random pairings, cross-correlation patterns did not reach significance and were statistically different than our original data (Hypotheses 2c and 2d), signifying that the convergence found could not have been expected by chance. Taken together, our data indicate that video-stimulated recall effectively captures a similar pattern of attention engagement to in-the-moment thought probes collected during the task.

Experiment 2

Given that video-stimulated recall captured accurate assessments of attentional engagement in Experiment 1, we next aimed to test the generality of this effect. In Experiment 2, we studied the degree to which these effects held when utilizing video lectures that varied on different engagement metrics. Prior work has used variable stimuli when studying dynamic changes in attention, including short instructional videos to probe attentional selection strategies (Lu et al., 2021; Wibawa & Muhidin, 2021), open course lectures to study the impact of attention in online learning (Risko et al., 2012; Schacter & Szpunar, 2015; Szpunar, Khan, et al., 2013), TED Talks to examine attentional flow (Nadeem, 2021; Xia & Hafner, 2021), and episodic or movie content to investigate free-flowing engagement (Cutting, 2016; Smith, 2013; Song et al., 2021). This range not only reflects the diversity of attentional research for narrative material, but also highlights the various ways that engagement has been studied within the literature, often taking into account one's interest (Schraw et al., 1995), perceived relevance (Astleitner & Wiesner, 2004; Axelson & Flick, 2010), effort (Chen & Wu, 2015), and experiences of flow (Csikszentmihalhi, 2020; Shernoff et al., 2014). To account for the diversity of video stimuli used in past work, and in turn, the characteristics of those stimuli that might drive engagement, it is important for stimulated recall to be effective beyond a single lecture on Evolutionary Psychology (Experiment 1). As such, the utility of stimulated recall hinges on its ability to capture accurate assessments of attentional engagement over time, regardless of potential changes in video content. Therefore, in Experiment 2, we examined whether using videos that varied on several metrics of engagement would affect the statistical overlap found in Experiment 1.

Methods

Participants

One hundred and eighteen participants (89 women, 26 men, 3 non-response; M_{age} = 21.2 years, SD_{age} = 8.3 years) who had not participated in Experiment 1 were recruited through the university's online participant pool.

Materials and Stimuli

Video Lectures. Two 15-minute video lectures were used as the stimuli for Experiment 2: an open-source Yale lecture on art history titled *Lifestyles of the Rich and Famous: Houses and villas at Pompeii* (Kleiner, 2009) and a TEDx Talk on computer science titled *The Super Mario Effect: Tricking your brain into learning more* (Rober, 2018). Similar to Experiment 1, the videos contained both audio and visual elements and were displayed at full-screen resolution, embedded in the browser, with no playback controls visible. Prior to the video-viewing task, participants reported their familiarity with the video topics. Overall, no differences were found between the two [art history, $M = 2.3$, $SD = 1.4$; computer science, $M = 2.7$, $SD = 1.8$; $t(101) = 1.76$, $p = .08$, $d_z = .17$], though familiarity ratings were statistically lower than the video topic used in Experiment 1 [all ps < .001, all ds > .60].

To better capture differences across the video lectures used for the task, both videos were presented to a group of independent raters (30 participants: 16 women, 12 men, 1 non-binary, 1 non-response; M_{age} = 20.9 years, SD_{age} = 2.2 years) who assessed each video on various engagement metrics on a scale of 0 to 10 (see Supplementary Materials for a full listing of descriptors). These metrics were broadly defined to capture different aspects of the viewing experience (e.g., "*to what extent was the video lecture enjoyable?*"), the lecturer (e.g., "*to what extent was the speaker animated?*"), and the overall quality of the video (e.g., "*to what extent was the audio quality clear?*"). Importantly, based on these ratings, the art history video had significantly lower engagement metrics compared to the computer science video [all ps < .013, all d_zs > .49], reflecting that the two did differ across various theoretical dimensions, an important prerequisite for this experiment[3].

Immediate / Video-Stimulated Probes. All material and stimuli were identical to Experiment 1, except that 10 immediate thought probes were utilized for each lecture during the video-viewing task to ensure consistent sampling frequency across the experiments, i.e., that thought probes still appeared every 92 seconds throughout the video lectures. Correspondingly, 10 video-stimulated probes were used for each lecture during video-stimulated recall.

---

[3] To offer comparison to the video lecture used in Experiment 1, another group of independent raters (32 participants: 17 women, 15 men; M_{age} = 20.4 years, SD_{age} = 1.3 years) were asked to assess the evolutionary psychology video on various engagement metrics on a scale of 0 to 10 (see Supplementary Materials). Compared to the video lectures used in Experiment 2, the evolutionary psychology video had significantly higher engagement metrics compared to the art history video [most ps < .036, most ds > .55, except for ratings of flow, appeal, and video quality where ps > .11, ds < .41], but was statistically similar to the computer science video [most ps > .05, most ds < .51, except for ratings of appeal and video quality where ps < .025, ds > .59].

Procedure

The procedure was identical to Experiment 1. To account for the two video lectures used in the experiment, we counterbalanced the order of the video lectures between participants prior to the start of the task. During the study, participants first completed Phase 1 (i.e., the video-viewing task) for both video lectures before proceeding to Phase 2 (i.e., the video-stimulated recall), which presented the video clips in the same counterbalanced order as in Phase 1.

Results

Data Exclusions

The same exclusion criteria were used as in Experiment 1, with 13 participants excluded for prior knowledge of the video lectures and three participants excluded for non-response to 20% of either immediate or video-stimulated probes. As such, data from 102 participants (82 women, 19 men, 1 non-response; M_{age} = 21.5 years, SD_{age} = 8.9 years) were included in the analysis.

Data Analysis Plan

To account for the addition of video lectures as a factor in Experiment 2, a modified analysis procedure was used to examine potential differences and interactions across various engagement metrics. Namely, across aggregate measures, overall differences were assessed using a repeated-measures ANOVA, with *probe type* (immediate versus video-stimulated) and *video lecture* (art history versus computer science) as factors. Similar to Experiment 1, we expected to find similar ratings of attentional engagement across immediate and video-stimulated probe responses within each video lecture (Hypothesis 1). Across temporal measures, overlap was assessed using cross-correlations and then evaluated as a function of time lag, noise distribution, and video lecture using linear mixed-effects models. Similar to Experiment 1, for each video lecture, we expected to find peak positive cross-correlation scores at a time lag of 0 minutes (Hypothesis 2a), decreases in cross-correlation scores with a shift in time lag away from 0 minutes (Hypothesis 2b), a failure of Hypotheses 2a and 2b when examining noise distributions (Hypotheses 2c and 2d), and *z*-scores that exceed the critical *z*-value at a 95% confidence level (Hypothesis 2e). In addition, we also expected to find similarly strong patterns across the two video lectures (Hypothesis 2f).

Aggregate Assessment of Attentional Engagement

Average attentional engagement across all probe responses was statistically lower for the art history video lecture [$M = 4.19$, $SD = 1.75$] than the computer science video lecture [$M = 5.61$, $SD = 1.52$; $F(1,101) = 74.36$, $p < .001$, $\eta^2_p = .42$]. However, engagement was equivalent between immediate [$M = 4.93$, $SD = 1.42$] and video-stimulated responses [$M = 4.87$, $SD = 1.51$; $F(1,101) = .51$, $p = .48$, $\eta^2_p = .01$]. The interaction between the two variables also did not reach significance [$p = .25$, $\eta^2_p = .01$]. As such, aggregate measures indicated no difference in average attentional engagement for video lectures when probed immediately relative to after-the-fact.

Temporal Assessment of Attentional Engagement

Cross-correlation scores at a time lag of 0 minutes ranged from –.74 to +.97 for the art history video lecture [$M = .34$, $SD = .44$] and –.59 to +1 for the computer science video lecture [$M = .26$, $SD = .40$]. Mean cross-correlations as a function of time lag are presented in Figure 4a (see Supplementary Materials for individual participant data).

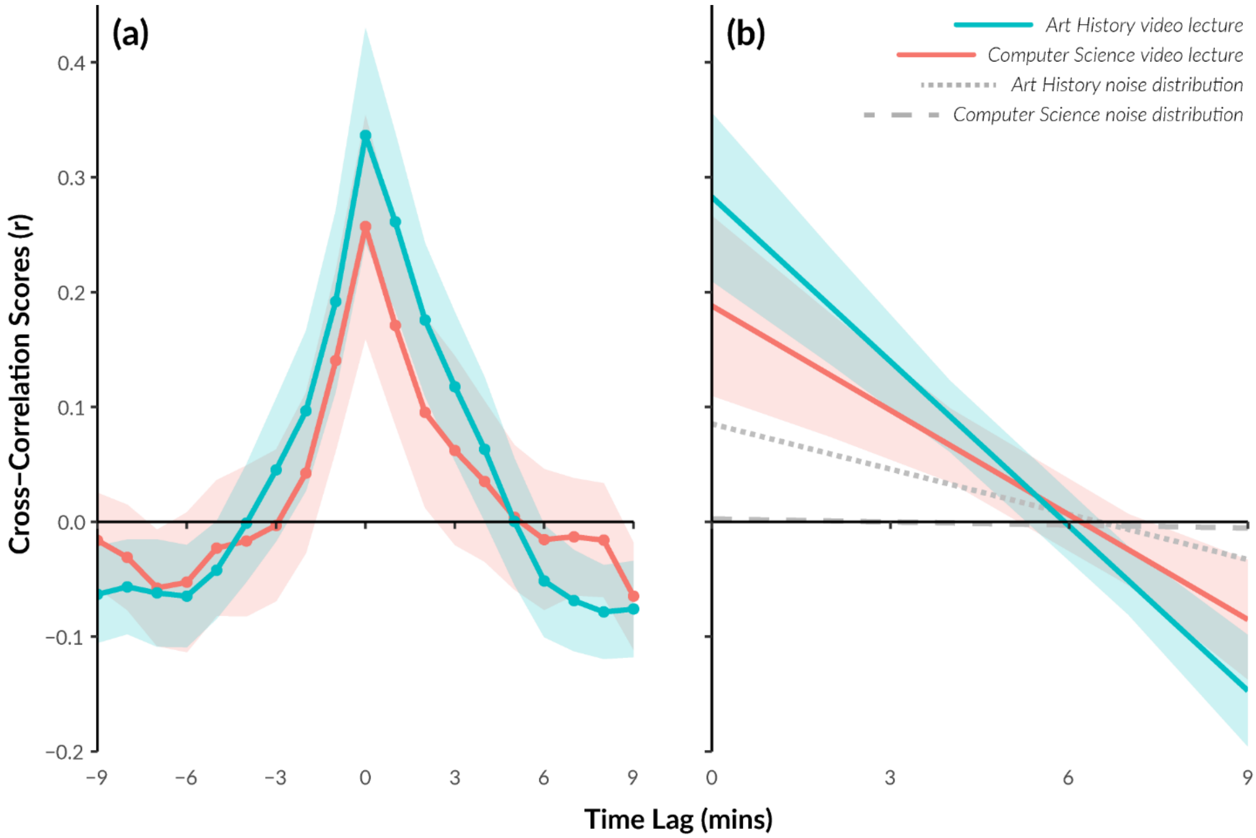

Figure 4. Data and results from Experiment 2. (a) Mean cross-correlation scores as a function of time lag. (b) Results from the linear mixed-effects models testing the effects of positive time lag (0 to +9 minutes) on cross-correlation scores. Models 1 and 2 (solid lines) found robust temporal overlap at a time lag of 0 minutes for both video lectures, which decreased as time lag increased; however, when examining the average temporal overlap that could be expected by chance across all 1,000 repetitions of Models 3 and 4 (dashed lines), no such effects were found. Shaded bands depict +/– 95% CIs.

Table 2 presents the unstandardized (*b*) and standardized (β) results for all linear mixed-effects models.

Table 2. Results of the linear mixed-effects models for Experiment 2.

Model predicting cross-correlation scores	*b*	β	t
Model 1: Art History video lecture			
Intercept	.28	.00	7.58***
Time lag (0 to +9 minutes)	–.05	–.43	–7.58***
Model 2: Computer Science video lecture			
Intercept	.19	.00	4.72***
Time lag (0 to +9 minutes)	–.03	–.29	–4.46***
Model 3: Art History noise distribution[a]			
Intercept	.09	.00	2.25
Time lag (0 to +9 minutes)	–.01	–.13	–2.08
Model 4: Computer Science noise distribution[a]			
Intercept	.00	.00	0.06
Time lag (0 to +9 minutes)	–.00	–.01	–0.12
Model 5: Effect of video lecture			
Time lag (0 to +9 minutes)	–.05	–.37	–7.74***
Video Lecture (Art History = 0)	–.10	–.02	–1.96
Interaction	.02	.08	2.13*

* $p < .05$, ** $p < .01$, *** $p < .001$
[a] b, β, and t for Models 3 and 4 represents average values across all 1,000 repetitions.

Overlap as a function of time lag. Models 1 and 2 examined the role of time lag (0 to +9 minutes)[4] on cross-correlation scores for each video lecture (see Figure 4b). Here, we found a

[4] When analyzing the data using only negative time lags, we replicated the pattern of effects found within positive time lags (see Supplementary Materials).

significant positive intercept for both the art history video lecture (*b* = .28, *p* < .001) and the computer science video lecture (*b* = .19, *p* < .001), demonstrating that temporal overlap at a time lag of 0 minutes was statistically different than a null correlation. In addition, we found a significant effect of time lag for both the art history video lecture (*b* = –.05, *p* < .001) and the computer science video lecture (*b* = –.03, *p* < .001), illustrating that cross-correlations decreased as time lag increased.

Overlap for a noise distribution. Similar to Experiment 1, for each video lecture, we randomly paired each participant's immediate time series with another participant's video-stimulated time series (i.e., video lecture was preserved within each random pair), calculated cross-correlation scores across all time lags, and used linear mixed-effects models to examine the role of time lag (0 to +9 minutes) on cross-correlation scores. We then repeated this procedure 1,000 times to create our noise distribution, and across all repetitions (see Figure 4b), average intercept (art history: mean *b* = .09, mean *p* = .10; computer science: mean *b* = .00, mean *p* = .51) and average time lag (art history: mean *b* = −.01, mean *p* = .13; computer science: mean *b* = −.00, mean *p* = .52) did not reach significance for either video lecture. In addition, *z*-scores for both the art history video lecture (intercept = 5.96, time lag = –6.06) and the computer science video lecture (intercept = 4.77, time lag = –4.32) exceeded the critical *z*-value at a 95% confidence level (i.e., +/–1.96).

Overlap as a function of time lag and video lecture. Model 5 examined the role of time lag (0 to +9 minutes) and video lecture (art history versus computer science) on cross-correlation scores. Similar to Models 1 and 2, we found an overall decrease in cross-correlations with an increase in time lag (*b* = −.05, *p* < .001). Although no overall differences were found as a function of video lecture (*b* = −.10, *p* = .053), the interaction term showed that temporal overlap had a higher peak at a time lag of 0 minutes and a sharper drop-off as time lag increased for the art history versus computer science video lecture (*b* = .02, *p* = .036).

Discussion

In Experiment 2, we evaluated whether video-stimulated recall could generalize and accurately capture dynamic changes in attentional engagement across videos that varied on engagement metrics. Similar to Experiment 1, no aggregate differences were found (Hypothesis 1) and temporal overlap between immediate and video-stimulated time series was statistically robust, with the strongest convergence occurring at a time lag of 0 minutes (Hypothesis 2a), convergence decreasing with an increase in time lag (Hypothesis 2b), no discernable pattern of effects when examining noise distributions that could be expected by chance (Hypotheses 2c and 2d), and statistical differences when comparing our original data against a noise distribution (Hypothesis 2e). Additionally, when assessing for effects across video lectures that differed on engagement metrics, we did not find any overall effect for video content, however, an interaction between time lag and video lecture revealed a higher convergence at a time lag of 0 minutes for the art history versus computer science video lecture (Hypothesis 2f). This finding suggests that temporal overlap may be stronger for videos that have lower engagement metrics. Regardless of this interaction, the data indicated that video-stimulated recall can provide accurate measures of attentional engagement even when assessing videos that differ in their ratings of overall engagement.

Experiment 3

Experiments 1 and 2 found consistent evidence that stimulated recall is an effective approach for assessing prior moment-to-moment attentional states across multiple kinds of videos. However, in both cases, the video-stimulated probe clips maintained the same sequential ordering as when they were initially presented, which may have unknowingly aided individuals in accurately responding. This chronological ordering is consistent with approaches in prior work utilizing video-stimulated recall. That is, past work has typically used chronologically ordered mnemonic stimuli from the original event by either presenting brief clips in the same order as they were experienced in real-time (Conway & Loveday, 2015; Lyle, 2003; Morgan, 2007) or using complete and full recordings of the event as it occurred (MacIntyre & Legatto, 2010; O'Brien, 1993). This chronological ordering was designed to maximize the recall of prior encounters, particularly given that temporal order has been shown to enhance memorial accuracy (Anderson & Conway, 1993). However, it is possible that individuals could have used this ordering to strategically respond to probes based on general or predictable patterns within the data. For example, if individuals reported that their attentional engagement during in-the-moment thought probes was higher at the start and lower at the end of the video lecture, they may have simply responded with similar trends for sequentially presented video clips during stimulated recall, instead of remembering their prior attentional states. It is also plausible that this type of strategic responding may have impacted the art history video lecture more than the computer science video lecture given that the former had lower ratings across all engagement metrics (e.g., interest, memorability), resulting in a higher likelihood of general or predictable patterns within in-the-moment probe data. To rule out these concerns, it is critical to establish whether stimulated recall remains an effective metric of prior attentional states, even when the probe clips are presented out of order, disrupting sequential and chronological ordering. Therefore, in Experiment 3, we examined whether varying the sequential ordering of the video-stimulated probe clips would affect the statistical overlap found in Experiments 1 and 2.

Methods

Participants

One hundred and twenty-two participants (74 women, 47 men, 1 non-response; M_{age} = 20.3 years, SD_{age} = 2.3 years) who had not participated in prior experiments were recruited through the university's online participant pool.

Materials and Stimuli

Video Lectures. All material and stimuli were identical to Experiment 2. As per prior experiments, participants did not report any differences in familiarity between the two video topics [art history, $M = 2.1$, $SD = 1.1$; computer science, $M = 2.3$, $SD = 1.7$; $t(99) = 1.38$, $p = .17$, $d_z = .14$]. As before, familiarity for the video topics was statistically lower than Experiment 1 [all ps < .001, all ds > .82].

Immediate / Video-Stimulated Probes. All material and stimuli were identical to Experiment 2, except that the video-stimulated probe clips were presented in *non-sequential* runtime order, i.e., in a randomized order that was different to the chronological order from the original video lecture.

Procedure

The procedure was identical to Experiment 2.

Results

Data Exclusions

The same exclusion criteria were used as in prior experiments, with 18 participants excluded for prior knowledge of the video lectures and four participants excluded for non-response to 20% of either immediate or video-stimulated probes. As such, data from 100 participants (65 women, 34 men, 1 non-response; M_{age} = 20.3 years, SD_{age} = 2.4 years) were included in the analysis.

Data Analysis Plan

The same analysis procedures were used as in Experiment 2. To summarize, for aggregate measures, overall differences were assessed using a repeated-measures ANOVA as a function of *probe type* (immediate versus video-stimulated) and *video lecture* (art history versus computer science), and we expected to find similar ratings of attentional engagement across probe type within each video lecture (Hypothesis 1). For temporal measures, overlap was explored by assessing immediate and video-stimulated time series using cross-correlations and evaluating them as a function of time lag, noise distributions, and video lecture using linear mixed-effects models. Similar to Experiment 2, for each video lecture, we expected peak positive cross-correlation scores at a time lag of 0 minutes (Hypothesis 2a), decreases in cross-correlation scores with an increase in time lag (Hypothesis 2b), a failure of Hypotheses 2a and 2b when examining noise distributions (Hypotheses 2c and 2d), *z*-scores that exceed the critical *z*-value at a 95% confidence level (Hypothesis 2e), and similarly strong patterns when comparing between video lectures (Hypothesis 2f).

Aggregate Assessment of Attentional Engagement

Average attentional engagement across probe responses was statistically lower for the art history video lecture [M = 4.07, SD = 1.65] versus computer science video lecture [M = 5.65, SD = 1.34; $F(1,99) = 95.18$, $p < .001$, $\eta^2_p = .49$]. However, similar to Experiment 2, engagement was equivalent across immediate [M = 4.89, SD = 1.28] and video-stimulated responses [M = 4.83,

SD = 1.33; $F(1,99) = 1.00$, $p = .32$, $\eta^2_p = .01$], with no interaction between the two variables [$p = .69$, $\eta^2_p < .01$].

Temporal Assessment of Attentional Engagement

Cross-correlation scores at a time lag of 0 minutes ranged from –.82 to +.94 for the art history video lecture [$M = .18$, $SD = .40$] and –.68 to +.94 for the computer science video lecture [$M = .14$, $SD = .37$]. Mean cross-correlations as a function of time lag are presented in Figure 5a (see Supplementary Materials for individual participant data).

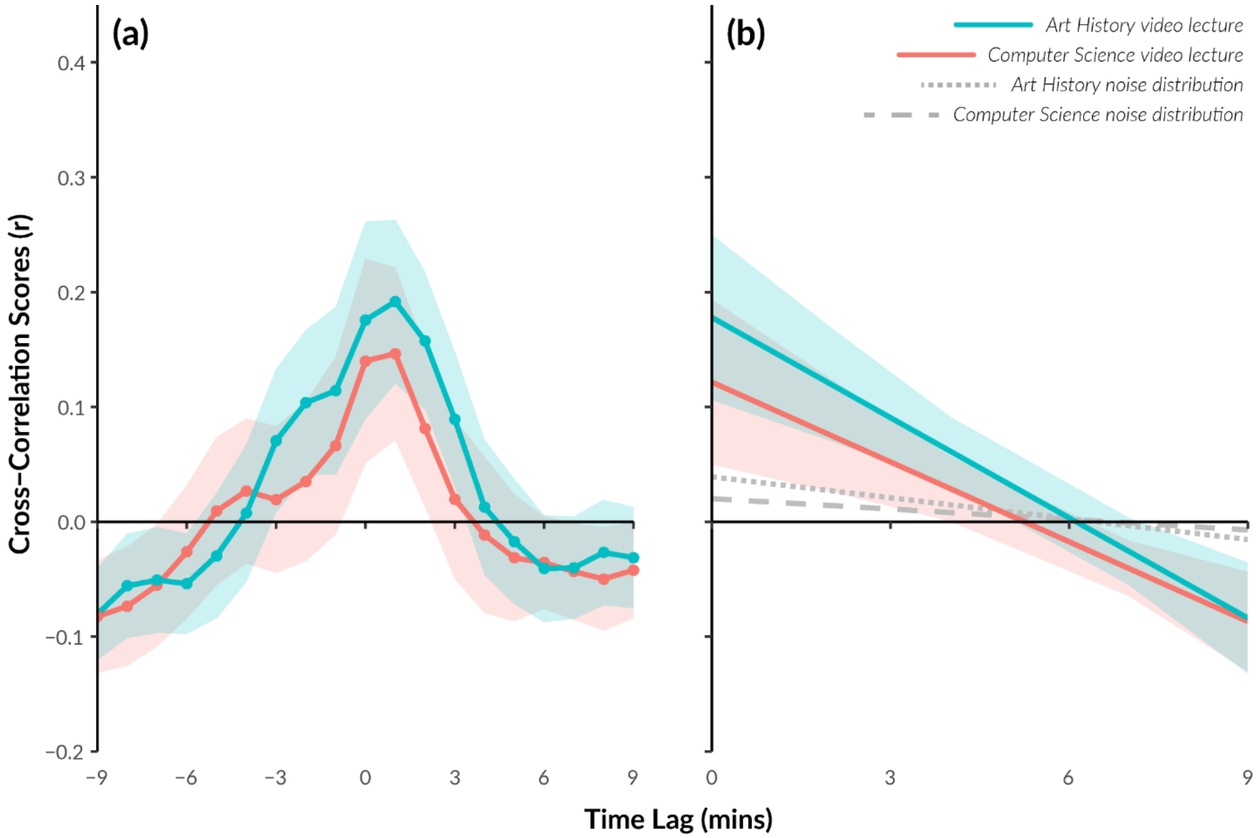

Figure 5. Data and results from Experiment 3. (a) Mean cross-correlation scores as a function of time lag. (b) Results from the linear mixed-effects models testing the effects of positive time lag (0 to +9 minutes) on cross-correlation scores. Models 1 and 2 (solid lines) found robust temporal overlap at a time lag of 0 minutes for both video lectures, which decreased as time lag increased; however, when examining the average temporal overlap that could be expected by chance across all 1,000 repetitions of Models 3 and 4 (dashed lines), no such effects were found. Shaded bands depict +/– 95% CIs.

Table 3 presents the unstandardized (*b*) and standardized (β) results for all linear mixed-effects models.

Table 3. Results of the linear mixed-effects models for Experiment 3.

Model predicting cross-correlation scores	*b*	β	t
Model 1: Art History video lecture			
Intercept	.18	.00	4.86***
Time lag (0 to +9 minutes)	–.03	–.29	–4.65***
Model 2: Computer Science video lecture			
Intercept	.12	.00	3.33**
Time lag (0 to +9 minutes)	–.02	–.25	–3.79***
Model 3: Art History noise distribution[a]			
Intercept	.04	.00	1.05
Time lag (0 to +9 minutes)	–.01	–.06	–0.93
Model 4: Computer Science noise distribution[a]			
Intercept	.02	.00	0.60
Time lag (0 to +9 minutes)	–.00	–.04	–0.51
Model 5: Effect of video lecture			
Time lag (0 to +9 minutes)	–.03	–.27	–4.75***
Video Lecture (Art History = 0)	–.06	–.06	–1.42
Interaction	.01	.04	0.99

* $p < .05$, ** $p < .01$, *** $p < .001$
[a] b, β, and t for Models 3 and 4 represents average values across all 1,000 repetitions.

Overlap as a function of time lag. Models 1 and 2 examined the role of time lag (0 to +9 minutes)[5] on cross-correlation scores for each video lecture (see Figure 5b). Similar to Experiment 2, we found a significant positive intercept for both the art history video lecture (*b* =

[5] Negative time lags replicated the pattern of effects found within positive time lags (see Supplementary Materials).

.18, $p < .001$) and computer science video lecture ($b = .12$, $p = .001$), with cross-correlations decreasing as time lag increased for both the art history video lecture ($b = –.03$, $p < .001$) and computer science video lecture ($b = –.02$, $p < .001$).

Overlap for a noise distribution. When examining the effect of time lag (0 to +9 minutes) on cross-correlation scores for all 1,000 repetitions (see Figure 5b), average intercept (art history: mean $b = .04$, mean $p = .37$; computer science: mean $b = .02$, mean $p = .45$) and average time lag (art history: mean $b = −.01$, mean $p = .39$; computer science: mean $b = −.00$, mean $p = .47$) did not reach significance. *Z*-scores for both the art history video lecture (intercept = 4.13, time lag = –3.81) and computer science video lecture (intercept = 3.02, time lag = –3.43) both exceeded the critical *z*-value at a 95% confidence level.

Overlap as a function of time lag and video lecture. Model 5 examined the role of time lag (0 to +9 minutes) and video lecture (art history versus computer science) on cross-correlation scores. Similar to Models 1 and 2, and similar to Experiment 2, we found an overall decrease in cross-correlations with an increase in time lag ($b = −.03$, $p < .001$), and no overall differences as a function of video lecture ($b = −.06$, $p = .16$). However, unlike Experiment 2, there was no significant interaction between time lag and video lecture, indicating no differences in temporal overlap between the art history and computer science video lectures ($b = .01$, $p = .33$).

Discussion

In Experiment 3, we evaluated the accuracy of capturing dynamic changes in attentional engagement via video-stimulated recall when probes were assessed in non-sequential runtime order. Similar to Experiment 2, we found no aggregate differences between immediate and video-stimulated ratings (Hypothesis 1). In addition, we found a robust temporal overlap between immediate and video-stimulated time series across both video lectures, with the strongest convergence at a time lag of 0 minutes (Hypothesis 2a) that decreased as time lag increased (Hypothesis 2b), no patterns when examining a noise distribution that could have been expected by chance (Hypotheses 2c and 2d), and statistical differences when comparing our original data against the noise distribution (Hypothesis 2e). Unlike Experiment 2 though, there was no interaction found between time lag and video lecture, indicating consistent convergence regardless of the characteristics of the video (Hypothesis 2f). In summary, our data indicate that probing video-stimulated recall outside of sequential order results in coinciding measures of attentional engagement.

General Discussion

In the present paper, we tested whether stimulated recall, a method wherein individuals are asked to retrospectively assess specific subjective experiences with the aid of a corresponding memorial stimulus (Bloom, 1953; Gass & Mackey, 2016; Lyle, 2003), could prove beneficial within the attentional domain to capture dynamic fluctuations in engagement over time. Across all experiments, we found that overall engagement levels did not differ between in-the-moment and video-stimulated reports, and more importantly, that there was strong convergence between the time series of attentional engagement for in-the-moment and video-stimulated responses. We confirmed the generality of this convergence by replicating the finding across video lectures that differed in topic (i.e., evolutionary psychology, art history, and computer science), in independent metrics of engagement (e.g., interest, memorability, cognitive effort), in length (i.e., 30 and 15 minutes), and in instances where recall stimuli were presented in the same versus random chronological order. Supporting these findings, we found statistically significant differences between our original data and a noise distribution generated by randomly pairing a given participant's in-the-moment time series with a *different* participants' video-stimulated time series. Together, the foregoing results support the notion that stimulated recall produces temporally accurate and statistically similar ratings of attentional engagement compared to in-the-moment thought probes. Given these findings, stimulated recall becomes a new and practical methodological approach to probing dynamic states of attention.

Stimulated recall of attentional states: Benefits, limitations, and future possibilities

The present implementation of stimulated recall has several methodological benefits. First, this method has the potential of assessing attentional states at much higher levels of temporal frequency than what has been possible in the past with in-the-moment thought probes and aggregate retrospective reports. Conceivably, individuals could be given recall cues (e.g., video clips) that sample the original task (e.g., video lecture) at intervals in the order of seconds rather than minutes to capture detailed moment-to-moment fluctuations in attentional states. Second, this method provides the flexibility to assess a wide range of subjective experiences, including various states of attention (e.g., attentiveness, mind wandering) and specific dimensions of attention (e.g., depth and intentionality of mind wandering) using a variety of response options (e.g., dichotomous, rating, semantic differences, free response). Third, this method can be flexibly applied across numerous task paradigms (e.g., video watching, live lecture settings, social interactions) since stimulated recall has already been successfully used with these activities across communication (Gregersen et al., 2014; MacIntyre & Legatto, 2010), education (Meade & McMeniman, 1992; Morgan, 2007; O'Brien, 1993), and clinical domains (Paskins et al., 2014; Sinnott et al., 2017).

It is important to note, however, that while stimulated recall promises to be a useful tool across many research situations, it is possible that not all paradigms will benefit equally from this method. Specifically, stimulated recall may be highly beneficial when attentional processes are measured during naturalistic, free-flowing activities, like watching a video lecture (Wammes & Smilek, 2017), having a conversation (Maran et al., 2020), operating a driving simulator (Albert et al., 2018), or monitoring an air-traffic signal (Endsley & Rogers, 1996), as these activities are comparable to prior use cases of stimulated recall. For example, short video clips from a recorded session in a driving simulator could be used to gauge moment-to-moment changes in attentional states after-the-fact. However, this method may be less useful when attentional process are measured across more controlled, trial-based paradigms, such as the *n*-back (Kirchner, 1958), go/no-go (Donders, 1969), and sustained attention to response tasks (Robertson et al., 1997), given that these paradigms likely contain insufficient variability within the audio and/or visual stimuli to produce distinct momentary memories of attention. For example, discrete trial-specific memories are not likely to be formed during the metronome response task, which requires a synchronous response to an auditory tone presented every 1,300 milliseconds for 19.5 minutes (Seli, Cheyne, et al., 2013). Although thoughtful and practical consideration is needed when utilizing this approach, the ability to capture temporal detail, investigative range, and task flexibility should prove fruitful within the field of attention.

The present findings also demonstrate how quantitative analytics can be used to enhance research conducted using stimulated recall. Prior work using stimulated recall has primarily relied on qualitative assessments of video-stimulated responses, either through data-driven or concept-driven coding schemes (Gibbs, 2007; Willig, 2013). Supplementing these assessments with nonlinear techniques such as cross-correlations or modelling techniques such as linear mixed-effects models can be useful for assessing dynamic patterns of data in an individualized manner while allowing for higher degrees of complexity to be built within datasets (e.g., assessing joint convergence between sets of cross-correlations or nesting hierarchical data at more than one level). Though attention should be paid to spurious relationships that can occur within behavioural time series (see Dean & Dunsmuir, 2016), the practical and functional utility of these analytical techniques should prove advantageous for the domains that currently utilize stimulated recall and for those that hope to do so in the future.

In addition to the aforementioned benefits of stimulated recall, one future use of this method can be to combine it with existing techniques to produce a more fine-grained understanding of attentional states over time. For example, studies could probe both in-the-moment and video-stimulated ratings of attentional states and combine these non-overlapping assessments to offer unique insights into the temporal dynamics of attention (i.e., the depth and structure of attentional patterns within individuals). Alternatively, studies could probe both retrospective reports and video-stimulated ratings of attentional states to assess how individuals arrive at overall aggregate metrics (i.e., whether this measure is based on a cumulative assessment or on a smaller combination of time points). These novel and unconventional approaches could prove valuable in asking more complex questions about how and why attentional states change over time and across embedded contexts.

Stimulated recall of attentional states: Theoretical implications

A broader theoretical implication of our finding that individuals can provide accurate retrospective reports of their attentional engagement is that it indicates that individuals have persistent memories of prior attentional experiences, i.e., *remembered momentary attentional states* (RMAS). To date, little is known about the nature of memories for prior attentional states. The recall of one's specific past attentional experiences would certainly be useful for guiding future attentional strategies, particularly in light of the notion that individuals have control over their attentional experiences (e.g., the capability to intentionally mind wander; Seli, Risko, & Smilek, 2016; Seli, Risko, Smilek, et al., 2016). For example, students reviewing a lecture might recall that they were not paying attention during a certain section of the lecture, and thus allocate more time to studying that specific content. Studies aimed at understanding the nature of remembered momentary attentional states could shed light on the parameters and boundaries of these memorial events and how they may be used in future episodes of attending.

One place to start would be to further explore our finding of subtle differences in the strength of convergence across video lectures (i.e., the temporal overlap between in-the-moment and retrospective probes). Specifically, convergence was stronger for the video lecture with lower engagement metrics, but only when video-stimulated ratings were probed sequentially in Experiment 2 and not when they were probed in random order in Experiment 3. Because stimulated recall uses memorial aids as a reference point to accurately recall prior subjective states (Bloom, 1953; Gass & Mackey, 2016; Lyle, 2003), one likely explanation is that order effects led to higher and more temporally precise convergence in cases where videos might have predictable attentional trends. For instance, videos that are considered less enjoyable or less interesting may lead to predictable engagement ratings that gradually decrease over time. By that same principle, more enjoyable or more interesting videos might have less predictable attentional trends due to ratings that are more variable and less tightly linked over time. If this were true, utilizing stimulated recall with video clips presented in a random order would likely remove temporal order cues, thus reducing predictability and dampening differences in remembered momentary attentional states across videos that varying in levels of engagement. Indeed, this may explain why a difference in convergence was not observed in Experiment 3. The present conjectures are consistent with stimulated recall work speculating that individuals may impose order when recollecting prior subjective states (Lyle, 2003), and with memory work showing enhanced recall when events occur in temporal sequence (Franklin et al., 2020; Rubin & Umanath, 2015).

Another concrete step would be to determine the specific characteristics of remembered momentary attentional states. One question that is fascinating from an attentional perspective is whether memories of ‘mind wandering’ are as detailed and accurate as those of ‘engagement’. Prior work has demonstrated that mind wandering is only one form of attentional *disengagement* that can occur (Danckert, 2018), with others including external distractions (Stawarczyk et al., 2011) and mind-blanking (Ward & Wegner, 2013). For this reason, it is possible that memories for mind wandering states might be less accurate than memories of engagement because the former requires the recollection of more specific thought content. Another question that is intriguing from a memory perspective is to what extent remembered momentary attentional states overlap with known concepts from the memory literature. Prior work suggests that event and episodic memories (i.e., memories of specific narrative or personal experiences) are temporarily stored as short, ordered visual excerpts that are later consolidated with general knowledge (Conway, 2009; Rubin & Umanath, 2015). This conceptualization is consistent with our current findings and suggest that attentional states might be stored both discretely and holistically, such that chronology may be an informative but not necessary requirement for its recall. Although more work is needed to uncover the characteristics of remembered momentary attentional states, our current findings present new lines of research for how subjective experiences can be leveraged to better predict and control individual attentional allocation (Breed & Moore, 2016; Fernandez-Duque et al., 2000; Roebers, 2017).

Conclusion

The present work provides converging evidence that stimulated recall can be highly beneficial for use within the field of attention. Stimulated recall was found to be similar to in-the-moment thought probes, with strong convergence between the two sets of responses when their time series were directly overlaid, decreased convergence with an increase in time lag, and no convergence when comparing our original time series with randomly shuffled and paired responses. Together, these results demonstrate that stimulated recall is a viable and practical alternative for unobtrusively capturing various qualities of attentional states in a variety of settings. Importantly, the method allows for the measurement of dynamic changes in attentional processes over time with a high degree of temporal precision.

Declarations

Funding

This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Banting and Postdoctoral Fellowships awarded to EJP, the NSERC Post-graduate Scholarship awarded to SAG, and NSERC Discovery Grants awarded to JDW and DS.

Conflicts of interest

All authors declare no conflict of interest.

Ethics approval

All studies were conducted in accordance with the Declaration of Helsinki, and all protocol and procedures were approved by the University of Waterloo's research ethics board (#42793).

Consent to participate

Informed written consent was obtained from all individuals who participated in the studies in this paper.

Consent for publication

Informed written consent was obtained from all individuals regarding the publishing of their anonymized data.

Availability of data and materials

Anonymized data for all studies in this paper are publicly available on the Open Science Framework at https://osf.io/r8ub7/. The experiments were not preregistered.

Code availability

Analysis scripts for this paper are publicly available on the Open Science Framework at https://osf.io/r8ub7/.

Open Practices Statement

Anonymized data as well as the analysis scripts for the studies contained within this paper are publicly available on the Open Science Framework at https://osf.io/r8ub7/. The experiments were not preregistered.

References

Albert, D. A., Ouimet, M. C., Jarret, J., Cloutier, M. S., Paquette, M., Badeau, N., & Brown, T. G. (2018). Linking mind wandering tendency to risky driving in young male drivers. *Accid Anal Prev*, *111*, 125-132. https://doi.org/10.1016/j.aap.2017.11.019

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367-10371. https://doi.org/10.1073/pnas.1104047108

Anderson, S. J., & Conway, M. A. (1993). Investigating the structure of autobiographical memories. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *19*(5), 1178-1116.

Astleitner, H., & Wiesner, C. (2004). An Integrated Model of Multimedia Learning and Motivation. *Journal of Educational Multimedia and Hypermedia*, *13*(1), 3-21.

Axelson, R. D., & Flick, A. (2010). Defining Student Engagement. *Change: The Magazine of Higher Learning*, *43*(1), 38-43. https://doi.org/10.1080/00091383.2011.533096

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. https://doi.org/https://doi.org/10.1016/j.jml.2007.12.005

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

Barker, P. M., & McDougall, T. J. (2020). Two Interpolation Methods Using Multiply-Rotated Piecewise Cubic Hermite Interpolating Polynomials. *Journal of Atmospheric and Oceanic Technology*, *37*(4), 605-619. https://doi.org/10.1175/jtech-d-19-0211.1

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bastian, M., & Sackur, J. (2013). Mind wandering at the fingertips: automatic parsing of subjective states based on response time variability. *Frontiers in Psychology*, *4*, 573. https://doi.org/10.3389/fpsyg.2013.00573

Beers, P. J., Boshuizen, H. P. A., Kirschner, P. A., Gijselaers, W., & Westendorp, J. (2006). Cognitive load measurements and stimulated recall interviews for studying the effects of information and communications technology. *Educational Technology Research and Development*, *56*(3), 309-328. https://doi.org/10.1007/s11423-006-9020-7

Bloom, B. S. (1953). Thought-processes in lectures and discussions. *The Journal of General Education*, *7*(3), 160-169.

Breed, M. D., & Moore, J. (2016). Cognition. In M. D. Breed & J. Moore (Eds.), *Animal Behavior (2nd edition)* (pp. 175-209). Academic Press.

Brunsdon, T. M., & Skinner, C. J. (1987). The analysis of dependencies between time series in psychological experiments. *British Journal of Mathematical and Statistical Psychology*, *40*(2), 125-139. https://doi.org/https://doi.org/10.1111/j.2044-8317.1987.tb00873.x

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences*, *107*(37), 16048-16053.

Calderhead, J. (1981). Stimulated recall: A method for research on teaching. *British Journal of Educational Psychology*, *51*(2), 211-217.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484-1525. https://doi.org/10.1016/j.visres.2011.04.012

Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, *80*, 108-121.

Cheyne, J. A., Solman, G. J., Carriere, J. S., & Smilek, D. (2009). Anatomy of an error: a bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, *111*(1), 98-113. https://doi.org/10.1016/j.cognition.2008.12.009

Christoff, K., Irving, Z. C., Fox, K. C. R., 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

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/https://doi.org/10.4324/9780203771587

Conway, M. A. (2009). Episodic memories. *Neuropsychologia*, *47*(11), 2305-2313. https://doi.org/10.1016/j.neuropsychologia.2009.02.003

Conway, M. A., & Loveday, C. (2015). Remembering, imagining, false memories & personal meanings. *Consciousness and Cognition: An International Journal*, *33*, 574-581. https://doi.org/10.1016/j.concog.2014.12.002

Csikszentmihalhi, M. (2020). *Finding flow: The psychology of engagement with everyday life*. Hachette UK.

Cutting, J. E. (2016). Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin & Review*, *23*(6), 1713-1743. https://doi.org/10.3758/s13423-016-1051-4

Dan, E. L., Dînşoreanu, M., & Mureşan, R. C. (2020, 21-23 May 2020). Accuracy of Six Interpolation Methods Applied on Pupil Diameter Data. 2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR),

Danckert, J. (2018). Special topic introduction: understanding engagement: mind-wandering, boredom and attention. *Exp Brain Res*, *236*(9), 2447-2449. https://doi.org/10.1007/s00221-017-4914-7

Dean, R. T., & Dunsmuir, W. T. (2016). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior Research Methods*, *48*(2), 783-802. https://doi.org/10.3758/s13428-015-0611-2

Delorme, A., & Brandmeyer, T. (2019). When the meditating mind wanders. *Current Opinion in Psychology*, *28*, 133-137. https://doi.org/https://doi.org/10.1016/j.copsyc.2018.12.006

Dempsey, N. P. (2010). Stimulated Recall Interviews in Ethnography. *Qualitative Sociology*, *33*(3), 349-367. https://doi.org/10.1007/s11133-010-9157-x

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *J Abnorm Psychol*, *111*(2), 225-236. https://doi.org/10.1037//0021-843x.111.2.225

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193-222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412-431.

Ellison, W. D., Trahan, A. C., Pinzon, J. C., Gillespie, M. E., Simmons, L. M., & King, K. Y. (2020). For whom, and for what, is experience sampling more accurate than retrospective report? *Personality and Individual Differences*, *163*, 110071. https://doi.org/https://doi.org/10.1016/j.paid.2020.110071

Endsley, M. R., & Rogers, M. D. (1996). Attention distribution and situational awareness in air traffic control. Human Factors and Ergonomics Society 40th Annual Meeting, Los Angeles.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Feng, S., D’Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, *20*(3), 586-592. https://doi.org/10.3758/s13423-012-0367-y

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Conscious Cogn*, *9*(2 Pt 1), 288-307. https://doi.org/10.1006/ccog.2000.0447

Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M., & Gershman, S. J. (2020). Structured Event Memory: A neuro-symbolic model of event cognition. *Psychological Review*, *127*(3), 327-361. https://doi.org/10.1037/rev0000177

Fritsch, F. N., & Carlson, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, *17*, 238-246. https://doi.org/10.1137/0717021

Gable, S. L., Hopper, E. A., & Schooler, J. W. (2019). When the Muses Strike: Creative Ideas of Physicists and Writers Routinely Occur During Mind Wandering. *Psychological Science*, *30*(3), 396-404. https://doi.org/10.1177/0956797618820626

Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. Routledge.

Giambra, L. M. (1995). A Laboratory Method for Investigating Influences on Switching Attention to Task-Unrelated Imagery and Thought. *Consciousness and Cognition*, *4*, 1-21.

Gibbs, G. (2007). *Analyzing Qualitative Data* https://doi.org/10.4135/9781849208574

Gregersen, T., Macintyre, P. D., & Meza, M. D. (2014). The Motion of Emotion: Idiodynamic Case Studies of Learners' Foreign Language Anxiety. *The Modern Language Journal*, *98*(2), 574-588. https://doi.org/https://doi.org/10.1111/modl.12084

Gregson, R. A. M. (2014). *Time series in psychology*. Psychology Press.

Greve, M., & Was, C. A. (2022). Mind wandering probes as a source of mind wandering depends on attention control demands. *Consciousness and Cognition*, *103*, 103355. https://doi.org/https://doi.org/10.1016/j.concog.2022.103355

Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E., & Barsalou, L. W. (2012). Mind wandering and attention during focused meditation: A fine-grained temporal analysis of fluctuating cognitive states. *Neuroimage*, *59*(1), 750-760. https://doi.org/https://doi.org/10.1016/j.neuroimage.2011.07.008

Heath, R. A. (2000). *Nonlinear Dynamics: Techniques and Applications in Psychology* (1st ed.). Psychology Press. https://doi.org/https://doi.org/10.4324/9781410605702

Hylan, J. P. (1898). *Fluctuation of attention* (Vol. 2). Macmillan.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Neuroscience Review*, *2*(3), 194-203. https://doi.org/10.1038/35058500

Kahaner, D., Moler, C., & Nash, S. (1988). *Numerical Methods and Software*. Prentice Hall.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: the day reconstruction method. *Science*, *306*(5702), 1776-1780. https://doi.org/10.1126/science.1103572

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For Whom the Mind Wanders, and When. *Psychological Science*, *18*(7), 614-621.

Karapanos, E., Martens, J.-B., & Hassenzahl, M. (2010). *On the retrospective assessment of users' experiences over time: memory or actuality?* CHI '10 Extended Abstracts on

Human Factors in Computing Systems, Atlanta, Georgia, USA. https://doi.org/10.1145/1753846.1754105

Killingsworth, M. A., & Gilbert, D. T. (2010). A Wandering Mind Is an Unhappy Mind. *Science*, *330*, 932. https://doi.org/10.1126/science.1192439

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology: General*, *55*(4), 352-358.

Kleiner, D. E. E. (2009). *Lifestyles of the rich and famous: Houses and villas at Pompeii*, Open Yale Courses. https://oyc.yale.edu/history-art/hsar-252/lecture-5

Kopp, K., Mills, C., & D'Mello, S. (2016). Mind wandering during film comprehension: The role of prior knowledge and situational interest. *Psychonomic Bulletin & Review*, *23*(3), 842-848. https://doi.org/10.3758/s13423-015-0936-y

Larson, R., & Csikszentmihalyi, M. (1978). Experiential correlates of time alone in adolescence. *Journal of Personality*, *46*(4), 677-693.

Leroy, S., Schmidt, A. M., & Madjar, N. (2020). Interruptions and Task Transitions: Understanding Their Characteristics, Processes, and Consequences. *Academy of Management Annals*, *14*(2), 661-694. https://doi.org/10.5465/annals.2017.0146

Liddell, H. S. (1919). Eye-Movement during Fluctuation of Attention. *The American Journal of Psychology*, *30*(3), 241-252. https://doi.org/10.2307/1413874

Lu, X., Li, Q., & Wang, X. (2021, 5-7 May 2021). Research on the Impacts of Feedback in Instructional Videos on College Students' Attention and Learning Effects. 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Luck, S. J., & Vecera, S. P. (2002). Attention. In *Handbook of experimental psychology: Sensation and perception* (3 ed., Vol. 1, pp. 235-286). John Wiley & Sons Inc.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494-1502. https://doi.org/10.3758/s13428-016-0809-y

Lyle, J. (2003). Stimulated Recall: A Report on Its Use in Naturalistic Research. *British Educational Research Journal*, *29*(6), 861-878. http://www.jstor.org/stable/1502138

MacIntyre, P. D., & Legatto, J. J. (2010). A Dynamic System Approach to Willingness to Communicate: Developing an Idiodynamic Method to Capture Rapidly Changing Affect. *Applied Linguistics*, *32*(2), 149-171. https://doi.org/10.1093/applin/amq037

Maran, T., Furtner, M., Liegl, S., Ravet-Brown, T., Haraped, L., & Sachse, P. (2020). Visual Attention in Real-World Conversation: Gaze Patterns Are Modulated by Communication and Group Size. *Applied Psychology*, *70*(4), 1602-1627. https://doi.org/10.1111/apps.12291

Marty-Dugas, J., Howes, L., & Smilek, D. (2021). Sustained attention and the experience of flow. *Psychological Research*, *85*(7), 2682-2696. https://doi.org/10.1007/s00426-020-01433-x

Matthews, G., Joyner, L., Gilliland, K., S.Campbell, Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state big three. *Personality Psychology in Europe*, *7*, 335-350.

McVay, J. C., & Kane, M. J. (2012). Why Does Working Memory Capacity Predict Variation in Reading Comprehension? On the Influence of Mind Wandering and Executive Attention.

Journal of Experimental Psychology: General, *141*(2), 302-320. https://doi.org/10.1037/a0025250

Meade, P., & McMeniman, M. (1992). Stimulated recall—An effective methodology for examining successful teaching in science. *The Australian Educational Researcher*, *19*(3), 1-18.

Morgan, A. (2007). Using video-stimulated recall to understand young children's perceptions of learning in classroom settings. *European Early Childhood Education Research Journal*, *15*(2), 213-226. https://doi.org/10.1080/13502930701320933

Nadeem, N. (2021). “Stories that are worth spreading”: A communicative model of TED talk narratives. *Narrative Inquiry*, *31*(2), 434-457. https://doi.org/https://doi.org/10.1075/ni.19037.nad

O'Brien, J. (1993). Action research through stimulated recall. *Research in Science Education*, *23*(1), 214-221.

Pashler, H. E. (1998). *The psychology of attention*. The MIT Press.

Paskins, Z., McHugh, G., & Hassell, A. B. (2014). Getting under the skin of the primary care consultation using video stimulated recall: a systematic review. *BMC Medical Research Methodology*, *14*(1), 101. https://doi.org/10.1186/1471-2288-14-101

Paxton, A., & Dale, R. (2013). Argument disrupts interpersonal synchrony. *Quarterly Journal of Experimental Psychology*, *66*(11), 2092-2102. https://doi.org/10.1080/17470218.2013.853089

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023-1031.

Posner, M. I., & Petersen, S. E. (1990). The Attention System of the Human Brain. *Annual Review of Neuroscience*, *13*(1), 25-42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday Attention: Variation in Mind Wandering and Memory in a Lecture. *Applied Cognitive Psychology*, *26*(2), 234-242. https://doi.org/10.1002/acp.1814

Risko, E. F., Buchanan, D., Medimorec, S., & Kingstone, A. (2013). Everyday attention: Mind wandering and computer use during lectures. *Computers & Education*, *68*, 275-283. https://doi.org/https://doi.org/10.1016/j.compedu.2013.05.001

Rober, M. (2018). *The Super Mario Effect: Tricking your brain into learning more*, TEDxPenn. https://www.ted.com/talks/mark_rober_the_super_mario_effect_tricking_your_brain_into_learning_more?language=en

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747-758.

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, *45*, 31-51. https://doi.org/10.1016/j.dr.2017.04.001

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, *122*(1), 1-23. https://doi.org/10.1037/a0037907

Sapolsky, R. (2009). *The uniqueness of humans* Stanford University, TED. https://www.ted.com/talks/robert_sapolsky_the_uniqueness_of_humans

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, *6*(5), 309-316. https://doi.org/10.1007/BF02288586

Schacter, D. L., & Szpunar, K. K. (2015). Enhancing attention and memory during video-recorded lectures. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 60-71. https://doi.org/10.1037/stl0000011

Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your mind wanders weakly, your mind wanders deeply: objective measures reveal mindless reading at different levels. *Cognition*, *125*(2), 179-194. https://doi.org/10.1016/j.cognition.2012.07.004

Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of Situational Interest. *Journal of Reading Behavior*, *27*(1), 1-17. https://doi.org/10.1080/10862969509547866

Schubert, A. L., Frischkorn, G. T., & Rummel, J. (2020). The validity of the online thought-probing procedure of mind wandering is not threatened by variations of probe rate and probe framing. *Psychological Research*, *84*(7), 1846-1856. https://doi.org/10.1007/s00426-019-01194-2

Seli, P., Carriere, J. S. A., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology*, *4*, 430-430. https://doi.org/10.3389/fpsyg.2013.00430

Seli, P., Cheyne, J. A., & Smilek, D. (2013). Wandering minds and wavering rhythms: linking mind wandering and behavioral variability. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 1-5. https://doi.org/10.1037/a0030954

Seli, P., Risko, E. F., & Smilek, D. (2016). On the Necessity of Distinguishing Between Unintentional and Intentional Mind Wandering. *Psychol Sci*, *27*(5), 685-691. https://doi.org/10.1177/0956797616634068

Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-Wandering With and Without Intention. *Trends Cogn Sci*, *20*(8), 605-617. https://doi.org/10.1016/j.tics.2016.05.010

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2014). Student engagement in high school classrooms from the perspective of flow theory. In *Applications of flow in human development and education* (pp. 475-494). Springer.

Sinnott, C., Kelly, M. A., & Bradley, C. P. (2017). A scoping review of the potential for chart stimulated recall as a clinical research method. *BMC Health Serv Res*, *17*(1), 583. https://doi.org/10.1186/s12913-017-2539-y

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, *13*(4), 657-690. https://doi.org/https://doi.org/10.1016/j.concog.2004.06.003

Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review*, *14*(2), 230-236.

Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: the curious incident of the wandering mind. *Memory & Cognition*, *36*(6), 1144-1150. https://doi.org/10.3758/MC.36.6.1144

Smallwood, J., Mrazek, M. D., & Schooler, J. W. (2011). Medicine for the wandering mind: mind wandering in medical practice. *Medical Education*, *45*(11), 1072-1080. https://doi.org/https://doi.org/10.1111/j.1365-2923.2011.04074.x

Smith, T. J. (2013). Watching you watch movies: Using eye tracking to inform film theory.

Song, H., Finn, E. S., & Rosenberg, M. D. (2021). Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proceedings of the National Academy of Sciences*, *118*(33), e2021905118. https://doi.org/doi:10.1073/pnas.2021905118

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 Psychol (Amst)*, *136*(3), 370-381. https://doi.org/10.1016/j.actpsy.2011.01.002

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110*(16), 6313-6317. https://doi.org/10.1073/pnas.1221764110

Szpunar, K. K., Moulton, S., & Schacter, D. L. (2013). Mind wandering and education: from the classroom to online learning [Review]. *Frontiers in Psychology*, *4*, 495. https://doi.org/10.3389/fpsyg.2013.00495

Terhune, D. B., Croucher, M., Marcusson-Clavertz, D., & Macdonald, J. S. P. (2017). Time contracts and temporal precision declines when the mind wanders. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(11), 1864–1871. http://research.gold.ac.uk/20499/

Tusche, A., Smallwood, J., Bernhardt, B. C., & Singer, T. (2014). Classifying the wandering mind: Revealing the affective content of thoughts during task-free rest periods. *Neuroimage*, *97*, 107-116. https://doi.org/https://doi.org/10.1016/j.neuroimage.2014.03.076

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 832-842. https://doi.org/10.1037/a0029669

Wammes, J. D., Seli, P., Cheyne, J. A., Boucher, P. O., & Smilek, D. (2016). Mind wandering during lectures II: Relation to academic performance. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 33-48. https://doi.org/10.1037/stl0000055

Wammes, J. D., & Smilek, D. (2017). Examining the Influence of Lecture Format on Degree of Mind Wandering. *Journal of Applied Research in Memory and Cognition*, *6*(2), 174-184. https://doi.org/10.1037/h0101808

Ward, A. F., & Wegner, D. M. (2013). Mind-blanking: when the mind goes away. *Front Psychol*, *4*, 650. https://doi.org/10.3389/fpsyg.2013.00650

Welhaf, M. S., Meier, M. E., Smeekens, B. A., Silvia, P. J., Kwapil, T. R., & Kane, M. J. (2022). A “Goldilocks zone” for mind-wandering reports? A secondary data analysis of how few thought probes are enough for reliable and valid measurement. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01766-4

Wibawa, B., & Muhidin, A. (2021). The effect of instructional videos on learning performance. *AIP Conference Proceedings*, *2331*(1), 060011. https://doi.org/10.1063/5.0041759

Willig, C. (2013). Introducing qualitative research in psychology. http://site.ebrary.com/id/10716899

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495-501. https://doi.org/10.1038/nrn1411

Xia, S. A., & Hafner, C. A. (2021). Engaging the online audience in the digital era: A multimodal analysis of engagement strategies in TED talk videos. *Ibérica*(42), 33-58. https://doi.org/10.17398/2340-2784.42.33

Zeinali, M., Shahmorad, S., & Mirnia, K. (2014). Hermite and piecewise cubic Hermite interpolation of fuzzy data. *Journal of Intelligent & Fuzzy Systems*, *26*, 2889-2898. https://doi.org/10.3233/IFS-130955

Supplementary Materials

for

Attention in hindsight: Using stimulated recall to capture dynamic fluctuations in attentional engagement

Effie J. Pereira*[1], Samantha Ayers-Glassey*[1], Jeffrey D. Wammes[2], & Daniel Smilek[1]

[1] *Department of Psychology, University of Waterloo*

[2] *Department of Psychology, Queen's University*

* equal contribution to the manuscript

Correspondence to:

Effie J. Pereira
Department of Psychology, University of Waterloo
PAS Building, 200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1-519-888-4567 ext. 42813
Email: effie.pereira@uwaterloo.ca

Experiment 1

Results

Data Analysis

Constructing the Time Series. Immediate and video-stimulated probe responses were converted into immediate and video-stimulated time series, respectively, by interpolating the probe responses at 1-minute intervals using a piecewise polynomial function (i.e., PCHIP, Piecewise Cubic Hermite Interpolating Polynomial; Fritsch & Carlson, 1980; Kahaner et al., 1988). This function was selected because of its ability to estimate a time series within the bounds of the data points in order to best preserve and approximate dynamic fluctuations within a dataset (Barker & McDougall, 2020; Dan et al., 2020; Zeinali et al., 2014). Data interpolation resulted in a 16-point time series (i.e., 1 point per minute) for every immediate and video-stimulated time series for each participant.

Assessing the Temporal Relationship. Cross-correlations between each participant's immediate and video-stimulated time series were calculated using a step size of 1-minute to reflect the data interpolation intervals employed in the prior stage. This resulted in cross-correlation scores ranging from time lags of –9 to +9 minutes for each participant. Figure S1 illustrates these cross-correlation scores and demonstrates the range and variability seen within the data.

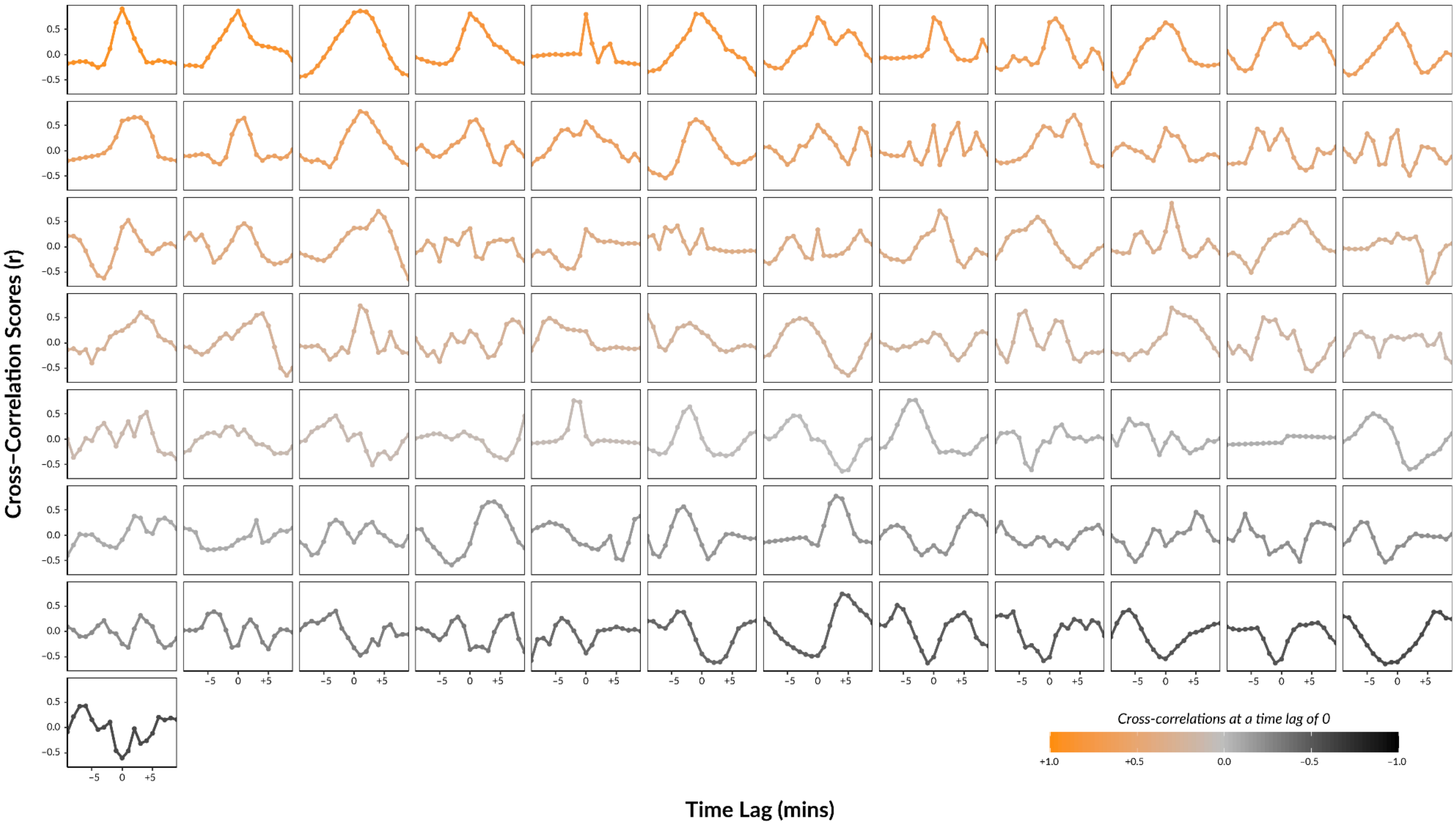

Figure S1. A matrix of cross-correlation plots for each participant in Experiment 1, arranged and coloured in descending order of maximum cross-correlation at a time lag of 0 minutes for the video lecture. Perfect cross-correlations (i.e., when no fluctuations existed between a participant's immediate or video-stimulated time series such that the cross-correlation scores could not be calculated across any time lags) are not depicted within the matrix.

Temporal Assessment of Attentional Engagement

Overlap as a function of time lag. Model 1 examined the role of time lag (0 to –9 minutes in 1-minute decrements) on cross-correlation scores (see Table S1). Similar to positive time lags, we found a significant positive intercept (b = .12, p < .001) and a significant decrease in cross-correlations with an increase in time lag (b = –.02, p < .001).

Overlap for a noise distribution. When examining the effect of time lag (0 to –9 minutes) on cross-correlation scores for all 1,000 repetitions in our noise distribution, no differences were found in average intercept (mean b = .01, mean p = .50) and average time lag (mean b = –.01, mean p = .47). Z-scores (intercept = 2.87, time lag = –3.02), representing the extent to which our original data's true effect differed from an intercept and slope that could be expected by chance, both exceeded the critical z-value at a 95% confidence level (i.e., +/–1.96).

Table S1. Results of the linear mixed-effects models with negative time lags for Experiment 1.

Model predicting cross-correlation scores	*b*	β	t
Model 1: Evolutionary Psychology video lecture			
Intercept	.12	.00	3.43***
Time lag (0 to –9 minutes)	–.02	–.24	–3.72***
Model 2: Evolutionary Psychology noise distribution[a]			
Intercept	.01	.00	0.33
Time lag (0 to –9 minutes)	–.01	–.03	–0.50

* p < .05, ** p < .01, *** p < .001
[a] b, β, and t for Model 2 represents average values across all 1,000 repetitions.

Experiment 2

Methods

Materials and Stimuli

Video Lectures. Table S2 details the full listing of descriptors and responses for engagement metrics that participants assessed for each video lecture.

Table S2. List of descriptors and responses assessed for each video lecture.

#	Descriptive Measure (assessed on a scale of 0 to 10)	Art History		Computer Science		Evolutionary Psychology	
		M	SD	M	SD	M	SD
1.	To what extent was the video lecture enjoyable?	4.2	2.6	6.9	2.3	6.6	2.8
2.	To what extent was the video lecture interesting?	4.4	2.6	6.5	2.4	7.0	2.6
3.	To what extent was the video lecture memorable?	3.6	2.6	6.8	2.4	5.9	2.9
4.	To what extent was the video lecture engaging?	3.2	2.5	6.9	2.3	6.0	3.1
5.	To what extent were you able to completely focus your thoughts on the video lecture without straining to pay attention?	3.6	2.8	6.3	2.5	4.9	3.0
6.	To what extent did you have to exert effort to focus your thoughts on the video lecture?	6.7	2.4	4.1	2.5	5.2	3.0
7.	To what extent was the speaker personable?	4.5	2.4	7.1	2.7	7.4	2.6
8.	To what extent was the speaker animated?	4.1	2.6	7.3	2.8	6.3	2.8
9.	To what extent was the speaker effective at communicating their topic?	6.3	2.5	8.2	1.7	8.0	2.3
10.	To what extent were the graphics / videos user by the lecturer visually appealing?	4.9	2.6	7.6	1.9	3.8	2.7
11.	To what extent was the video quality clear?	5.9	2.6	7.4	2.0	5.9	3.0
12.	To what extent was the audio quality clear?	6.8	2.5	7.8	1.7	8.1	2.2

Results

Data Analysis

Assessing the Temporal Relationship. Figure S2 illustrates the range and variability of cross-correlation scores found within Experiment 2.

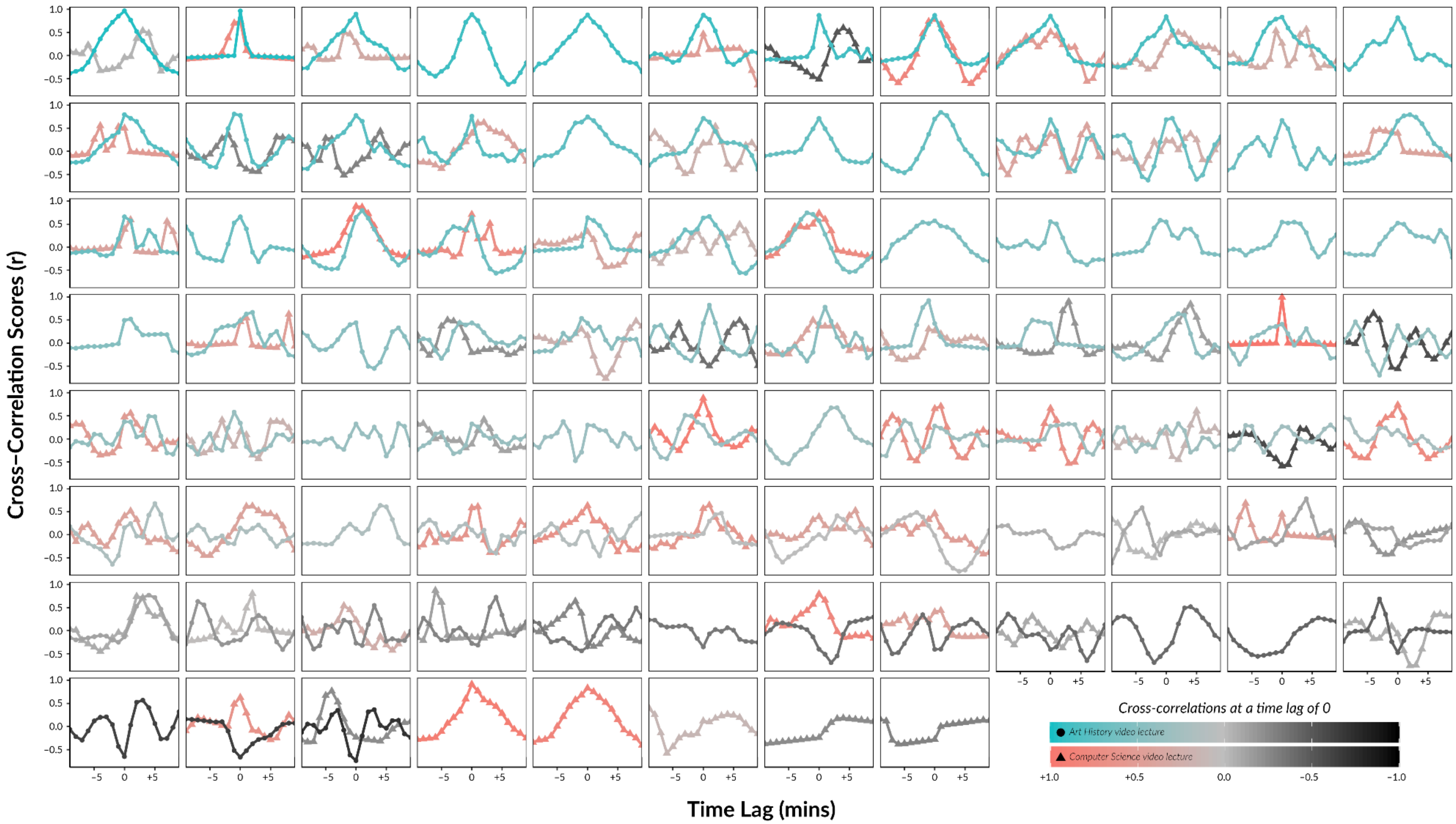

Figure S2. A matrix of cross-correlation plots for each participant for the art history (circles) and computer science (triangle) video lectures in Experiment 2. Matrices are arranged and coloured in descending order of maximum cross-correlation at a time lag of 0 minutes for the art history video lecture. Perfect cross-correlations (i.e., when there were no fluctuations between a participant's immediate or video-stimulated time series such that cross-correlation scores could not be calculated) are not depicted within the matrix.

Temporal Assessment of Attentional Engagement

Overlap as a function of time lag. Models 1 and 2 examined the role of time lag (0 to –9 minutes) on cross-correlation scores for each video lecture (see Table S3). Similar to positive time lags, we found a significant positive intercept for both the art history video lecture (b = .22, p < .001) and computer science video lecture (b = .14, p < .001) and a decrease in cross-correlations as time lag increased for both the art history video lecture (b = –.04, p < .001) and computer science video lecture (b = –.03, p < .001).

Overlap for a noise distribution. When examining the effect of time lag (0 to –9 minutes) on cross-correlation scores for all 1,000 repetitions in our noise distribution, no differences were found in the average intercept (art history: mean b = .07, mean p = .16; computer science: mean b = .01, mean p = .49) or average time lag (art history: mean b = −.01, mean p = .22; computer science: mean b = −.00, mean p = .50). In addition, z-scores for both the art history video lecture (intercept = 4.73, time lag = –5.19) and computer science video lecture (intercept = 3.13, time lag = –3.31) also exceeded the critical z-value at a 95% confidence level (i.e., +/–1.96).

Overlap as a function of time lag and video lecture. Model 5 examined the role of time lag (0 to –9 minutes) and video lecture (art history versus computer science) on cross-correlation scores. Similar to positive time lags, we found an overall decrease in cross-correlations with an increase in time lag (b = −.04, p < .001) and no overall differences as a function of video lecture (b = −.08, p = .14). Here however, unlike positive time lags, we found no differences for the interaction between time lag and video lecture (b = .01, p = .11).

Table S3. Results of the linear mixed-effects models with negative time lags for Experiment 2.

Model predicting cross-correlation scores	*b*	β	t
Model 1: Art History video lecture			
Intercept	.22	.00	5.60***
Time lag (0 to –9 minutes)	–.04	–.37	–6.17***
Model 2: Computer Science video lecture			
Intercept	.14	.00	3.69***
Time lag (0 to –9 minutes)	–.03	–.26	–4.07***
Model 3: Art History noise distribution[a]			
Intercept	.07	.00	1.80
Time lag (0 to –9 minutes)	–.01	–.11	–1.56
Model 4: Computer Science noise distribution[a]			
Intercept	.01	.00	0.26
Time lag (0 to –9 minutes)	–.00	–.02	–0.26
Model 5: Effect of video lecture			
Time lag (0 to –9 minutes)	–.04	–.33	–6.26***
Video Lecture (Art History = 0)	–.08	–.03	–1.49
Interaction	.01	.07	1.61

* $p < .05$, ** $p < .01$, *** $p < .001$
[a] b, β, and t for Models 3 and 4 represents average values across all 1,000 repetitions.

Experiment 3

Results

Data Analysis

Assessing the Temporal Relationship. Figure S3 illustrates the range and variability of cross-correlation scores found within Experiment 3.

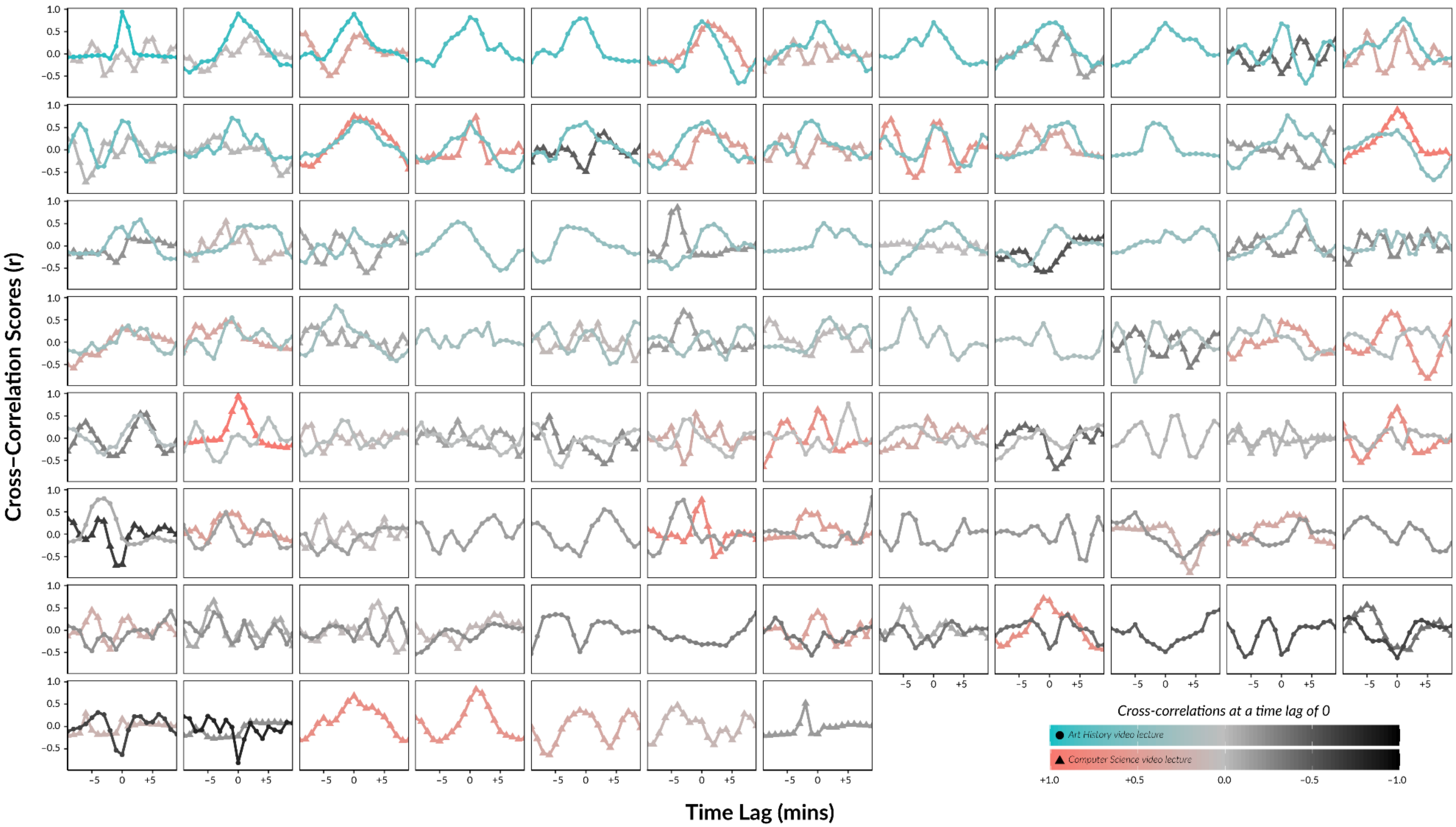

Figure S3. A matrix of cross-correlation plots for each participant for the art history (circles) and computer science (triangle) video lectures in Experiment 3. Matrices are arranged and coloured in descending order of maximum cross-correlation at a time lag of 0 minutes for the art history video lecture. Perfect cross-correlations i.e., when there were no fluctuations between a participant's immediate or video-stimulated time series such that cross-correlation scores could not be calculated) are not depicted within the matrix.

Temporal Assessment of Attentional Engagement

Overlap as a function of time lag. Models 1 and 2 examined the role of time lag (0 to –9 minutes) on cross-correlation scores for each video lecture (see Table S4), and similar to positive time lags, we found a significant positive intercept for both the art history video lecture ($b = .15$, $p < .001$) and computer science video lecture ($b = .10$, $p = .002$), and a decrease in cross-correlations as time lag increased for both the art history video lecture ($b = –.03$, $p < .001$) and computer science video lecture ($b = –.02$, $p < .001$).

Overlap for a noise distribution. When examining the effect of time lag (0 to –9 minutes) on cross-correlation scores for all 1,000 repetitions, no differences were found for average intercept (art history: mean $b = .04$, mean $p = .37$; computer science: mean $b = .01$, mean $p = .50$) and average time lag (art history: mean $b = −.01$, mean $p = .39$; computer science: mean $b = −.00$, mean $p = .50$). Z-scores for the art history video lecture (intercept = 2.95, time lag = –3.60) and computer science video lecture (intercept = 2.50, time lag = –2.97) also exceeded the critical z-value at a 95% confidence level (i.e., +/–1.96).

Overlap as a function of time lag and video lecture. Model 5 examined the role of time lag (0 to –9 minutes) and video lecture (art history versus computer science) on cross-correlation scores. Similar to positive time lags, we found an overall decrease in cross-correlations with an increase in time lag ($b = −.03$, $p < .001$) and no overall differences as a function of video lecture ($b = −.04$, $p = .38$) or for the interaction between time lag and video lecture ($b = .01$, $p = .44$).

Table S4. Results of the linear mixed-effects models with negative time lags for Experiment 3.

Model predicting cross-correlation scores	*b*	β	t
Model 1: Art History video lecture			
Intercept	.15	.00	4.13***
Time lag (0 to –9 minutes)	–.03	–.28	–4.68***
Model 2: Computer Science video lecture			
Intercept	.10	.00	3.15**
Time lag (0 to –9 minutes)	–.02	–.23	–3.68***
Model 3: Art History noise distribution[a]			
Intercept	.04	.00	1.03
Time lag (0 to –9 minutes)	–.01	–.06	–0.92
Model 4: Computer Science noise distribution[a]			
Intercept	.01	.00	0.31
Time lag (0 to –9 minutes)	–.00	–.02	–0.28
Model 5: Effect of video lecture			
Time lag (0 to –9 minutes)	–.03	–.26	–4.71***
Video Lecture (Art History = 0)	–.04	–.02	–0.88
Interaction	.01	.03	0.78

* $p < .05$, ** $p < .01$, *** $p < .001$
[a] b, β, and t for Models 3 and 4 represents average values across all 1,000 repetitions.