Published in final edited form as:

J Exp Psychol Learn Mem Cogn. 2017 August ; 43(8): 1183–1202. doi:10.1037/xlm0000367.

Event Segmentation Improves Event Memory Up to One Month Later

Shaney Flores[1], **Heather R. Bailey**[1,2], **Michelle L. Eisenberg**[1], and **Jeffrey M. Zacks**[1]

[1]Washington University in St. Louis

[2]Kansas State University

Abstract

When people observe everyday activity, they spontaneously parse it into discrete meaningful events. Individuals who segment activity in a more normative fashion show better subsequent memory for the events. If segmenting events effectively leads to better memory, does asking people to attend to segmentation improve subsequent memory? To answer this question, participants viewed movies of naturalistic activity with instructions to remember the activity for a later test, and in some conditions additionally pressed a button to segment the movies into meaningful events or performed a control condition that required button-pressing but not attending to segmentation. In five experiments, memory for the movies was assessed at intervals ranging from immediately following viewing to one month later. Performing the event segmentation task led to superior memory at delays ranging from 10 min to one month. Further, individual differences in segmentation ability predicted individual differences in memory performance for up to a month following encoding. This study provides the first evidence that manipulating event segmentation affects memory over long delays and that individual differences in event segmentation are related to differences in memory over long delays. These effects suggest that attending to how an activity breaks down into meaningful events contributes to memory formation. Instructing people to more effectively segment events may serve as a potential intervention to alleviate everyday memory complaints in aging and clinical populations.

Keywords

event segmentation; delay; memory

Organizing information into comprehensible and meaningful structures is important for memory. If one can form a coherent representation of an experience that binds together features of the experience into a structured framework, this reduces the amount of information that needs to be stored while also providing a useful system of indexing for later retrieval, much like a catalog at a library. In the laboratory, imposing organization during encoding of verbal materials helps memory formation and facilitates recall by providing cues to access stored information at retrieval (e.g., Tulving, 1962; Tulving & Pearlstone, 1966). Participants can recall more items from structured word lists than from unstructured

---

Correspondence concerning this article should be addressed to Jeffrey M. Zacks, Department of Psychological & Brain Sciences, Washington University in St. Louis, 1 Brookings Drive, Campus Box 1125, St. Louis, MO 63130. jzacks@wustl.edu.

lists and, when asked to freely recall items from an unstructured list, will tend to cluster their responses by category similarity (Bousfield, 1953; Bousfield & Cohen, 1955; Jenkins & Russell, 1952). Semantic clustering can be reinforced or undermined by the temporal structure of a list, with robust effects on memory (Polyn, Norman, & Kahana, 2009). However, these sorts of materials place an important limit on the study of organizational processes because they are highly organized from the start. Each word or picture comes tightly packaged by the experimenter as an informational chunk.

For more complex information, such as narratives or life events, a wider range of organizational processes can come into play. Event memories are fundamentally structured by organization in time and space, and often exhibit covarying organization on other dimensions including characters, objects, causes, and goals (Radvansky & Zacks, 2014; Rubin & Umanath, 2015). Moreover, memories for particular events are often informed by knowledge acquired from repeated experience with similar events. For example, the structure of one’s memory of going to see a particular film with a friend likely depends not just on information encoded during that episode, but also on information acquired from previous trips to the theatre. Schank and Abelson (1977) proposed that this prior information is stored in scripts, which are generalized knowledge structures that capture the typical patterns in routine activities. Scripts affect memory both by organizing information at encoding and by enabling missing information to be inferred at retrieval. When asked to recall information from stories or movies of routine events, people are more likely to organize their responses to fit their pre-existing scripts, recall more information directly related to the script compared to unrelated information, and distort the remembered order of events to better conform to script norms (Abbott, Black, & Smith, 1985; Bower, Black, & Turner, 1979; Bower & Clark-Meyers, 1980; Brewer & Dupree, 1983; Lichtenstein & Brewer, 1980; Migueles & García-Bajos, 2012).

Script-based studies of memory for events have addressed one important question of how organizational processes operate in memory: how do the units of an activity relate to each other and to knowledge in memory? However, they leave unanswered a complementary question: how are these units formed? Outside the laboratory, our cognitive system is bombarded by a vast amount of continuous and dynamic perceptual information necessary to our functioning and survival. Without an experimenter to provide us with units in the form of discrete stimuli and trials, we are forced to segment continuous real-life experience into organized elements ourselves. This question has been addressed by studies of *event segmentation* (Newtson, 1973, 1976). The segmentation of activity into events is a consequence of the automatic and ongoing perceptual behavior that underlies everyday perception, meaning that people segment ongoing information even when not explicitly asked to do so (Hard, Recchia, & Tversky, 2011; Zacks et al., 2001). Even infants as young as nine to eleven months of age are capable of parsing dynamic action (e.g., Baldwin, Baird, Saylor, & Clark, 2001; Saylor, Baldwin, Baird, & LaBounty, 2007), suggesting that this process may be a fundamental component of human cognition.

One account for how people segment activity into organized elements is *event segmentation theory* (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). According to this theory, a perceiver actively maintains a representation of the current environmental context called an *event model*. The current event model is useful for allowing one to make predictions about upcoming activity and thus for guiding proactive behavior. The perceptual system monitors prediction error, and when prediction error increases, the event model is updated based on current perceptual input, knowledge (including script knowledge), and memory representations of related previous events. This updating process leads to the formation of units in long-term memory. In addition to enabling better predictions, segmentation may have an additional benefit for memory in the form of cognitive economy: Segmentation temporally compresses related perceptual data into more robust long term memory units that can be accessed more quickly.

Studies of memory for filmed events suggest that how an event is segmented is related to how it is subsequently remembered (Boltz, 1992; Hanson & Hirst, 1989; Lassiter, 1988; Lassiter, Stone, & Rogers, 1988; Newtson & Engquist, 1976; Sargent et al., 2013; Schwan, Garsoffky, & Hesse, 2000; Zacks & Tversky, 2001). In addition, several studies have found that individual differences in segmentation predict individual differences in subsequent memory (Bailey et al., 2013; Kurby & Zacks, 2011; Sargent et al., 2013; Zacks, Speers, Vettel, & Jacoby, 2006). For example, a study by Sargent et al. (2013) investigated this relationship across the adult lifespan in a sample of adults aged 20 to 79 years who segmented activity into events and then completed several memory measures for those activities. Each participant's segmentation was compared to that of the group as a whole by calculating a *segmentation agreement* score that measures how well the individual's segmentation corresponds with the typical responses for the group (Kurby & Zacks, 2011). Participants also completed a battery of cognitive tests assessing working memory capacity, processing speed, general knowledge and episodic memory. After controlling for these general cognitive factors, segmentation agreement remained a unique contributor to memory for the activities. This suggests that those who segment activity more normatively form representations that are better organized to support long-term retrieval. However, all studies to date have assessed the relationship between segmentation and memory using short delays, as event memory has always been tested within the same experimental session as the event segmentation task. In the present study, we asked whether these individual differences in segmentation predict memory over longer delays.

If event segmentation predicts memory, is this relationship causal or does it merely reflect some as-yet-unmeasured third variable that affects both segmentation and memory? If the relationship is causal, this opens up the possibility for improving memory for naturalistic activity by facilitating effective segmentation. This is an important potential application, because memory complaints are a prominent feature of healthy aging, age-related diseases such as Alzheimer's disease, and brain injury (Gilewski, Zelinski, & Schaie, 1990; Jorm & Jacomb, 1989; Van Zomeren & Van den Burg, 1985). There are a few hints in the literature that experimentally manipulating stimuli to facilitate or impair segmentation affects subsequent memory. Boltz (1992) showed participants one of two episodes of the TV miniseries *A Perfect Spy*, in which commercial breaks were placed so as to coincide or conflict with event boundaries. Commercials that supported natural event boundaries facilitated subsequent memory, whereas commercials that conflicted with event boundaries impaired memory. A conceptually similar, but more subtle, manipulation was carried out by Schwan and colleagues (2000), who manipulated the presence of film edits rather than commercial breaks in brief movies. Like Boltz (1992), they found that placing interruptions at event boundaries facilitated subsequent memory. Zacks and Tversky (2003) taught participants to perform new everyday tasks, such as building a model, by using computer interfaces that either supported effective segmentation or did not. Learning was better when the interface supported effective segmentation.

Manipulating participants' orientation to event segmentation during event encoding also may affect subsequent memory. Lassiter and colleagues found that instructing participants to segment events at a fine temporal grain led to better memory than instructions to segment at a coarse grain (Lassiter, 1988; Lassiter et al., 1988). However, the generality of this finding has been a matter of debate (see Hanson & Hirst, 1989; Lassiter & Slaw, 1991).

In nearly all previous studies investigating memory and segmentation, memory was assessed within a few minutes following encoding of the event. Sargent et al. (2013), Bailey et al. (2013), Kurby and Zacks (2011), and Lassiter (1988) asked participants to segment movies of everyday actions lasting less than 10 min, and then assessed for memory of the movies directly following each movie. Hanson and Hirst (1989) had participants watch a five-minute distractor video before assessing memory. In Boltz (1992), memory was tested immediately after movie presentation, but because the movies lasted 40–45 min the effective delay was a bit longer, at least for some items. To date, no studies have examined the effect segmentation has on memory over delays longer than at most 45 min—a crucial test if active segmentation is to be relevant to real-world memory.

Certainly, encoding manipulations can affect memory over a long delay. For example, in *survival processing* manipulations (Nairne, Thompson, & Pandeirada, 2007) participants perform an incidental encoding task of rating words on their relevance to surviving in the wilderness. Memory for the words is better following this encoding task as compared to other encoding tasks, such as judging the pleasantness of the words, at delays as long as four days (Abel & Bäuml, 2013; Clark & Bruno, 2015; Raymaekers, Otgaar, & Smeets, 2014). In some cases, the most effective level of an encoding variable depends on the delay between study and test. For example, research on the *spacing effect* shows that increasing the time between restudy sessions can improve memory, and this benefit can persist for up to eight years (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bahrick & Phelphs, 1987; Glenberg & Lehmann, 1980; Pavlik & Anderson, 2005). However, the optimal amount of spacing depends on the delay between the last study episode and the final test: For optimal memory, the lag between study sessions had to increase in parallel with the retention interval (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The effect of an encoding manipulation can even reverse as the retention interval evolves. For example, Roediger and Karpicke (2006) asked participants to read expository passages and then either restudy them or attempt to recall them. Restudying the passages resulted in better memory performance at a five-minute delay; however, at delays of two days or one week the recall group performed better. Thus, it should be of interest whether manipulations of event segmentation facilitate memory over longer delays, and whether the effects change with delay.

One hint that event segmentation effects on memory are durable comes from a recent study of infants' event memory (Sonne, Kingo, & Krøjgaard, 2016). In this experiment, 16-month-old and 20-month-old infants viewed brief cartoons that were edited to occlude scene information either during an event boundary or during an event middle. When they returned to the laboratory two weeks later, they were shown the movies they had seen, each paired with a new movie. The infants showed a familiarity preference, looking more at the previously-seen cartoons than at the new ones, indicating that they remembered information from the movies. This memory effect was stronger for cartoons that had been presented with the middles occluded. Thus, memory after two weeks was better for videos whose event boundaries had been preserved during presentation.

The studies reported here were designed to answer three questions. First, does deliberate attention to event segmentation facilitate memory? Second, do individual differences in segmentation ability predict individual differences in memory? Finally, if these relationships between segmentation and memory are present, are they robust over retention intervals of days to weeks—the sorts of intervals that are relevant for real-world event memory? To answer these questions, we conducted a series of experiments in which participants viewed movies of everyday activities and either segmented them or engaged in a control task, and then attempted to remember them at delays ranging from a few moments to one month.

Experiment 1

In Experiment 1 we examined memory at delay intervals of 10 minutes, one day and one week. During encoding, participants segmented movies into meaningful events or performed one of two control tasks. Memory was tested using a recall test and a recognition test for still pictures from the movie. This allowed us, first, to ask whether segmentation improved memory relative to the control conditions, and second, whether segmentation performance predicted subsequent memory.

We conducted the study using an online pool of human participants from the Amazon Mechanical Turk (AMT) system. Several studies have shown a strong degree of reliability between data produced by AMT participants and participants tested in the laboratory for both perception-based and cognitive experiments (Crump, McDonnell, & Gureckis, 2013; Germine et al., 2012; Goodman, Cryder, & Cheema, 2013; Sprouse, 2011; for a review of AMT, see Mason & Suri, 2012).

Method

Participants—We recruited 482 participants (266 females; age range: 18–66 years; median age: 30 years) from AMT. The study required participants to complete three parts (one at each delay interval). Participants were paid $0.50 for completing session one, $1 for completing session two, and $2.50 for completing session three. All workers were from the United States and had a human intelligence task (HIT) approval rating of at least 95%. Before beginning the study, each participant was presented with a description of the procedure and indicated consent by clicking to continue.

Materials and Tasks—Participants watched three movies of actors engaging in everyday activities (see Figure 1), and completed two memory tests for each movie. The movies included a female actor making breakfast (329 s), a male actor decorating a room for a party

(376 s), and a male actor gardening (354 s). A short practice movie of a man building a boat out of Duplo blocks (155 s) also was presented. During the initial encoding, each participant performed one of three tasks, *event segmentation, intentional encoding,* or *timing,* with all three movies. Participants in the event segmentation group were instructed to press the space bar whenever, in their judgment, one meaningful unit of activity ended and another began (e.g., Newtson, 1976). Participants were instructed to identify the smallest units that were meaningful to them, while trying to remember as much as possible. Those in the intentional encoding group were instructed simply to watch the movie and try to remember as much as possible. Those in the timing group were instructed to press the keyboard space bar every 15 s during the movie, again while trying to remember as much as possible. No external prompt was given; they were asked to try to estimate the 15-s intervals. The timing task controlled for potential effects of the motor task on event memory. However, it should be noted that attending to counting may have artificially interfered with viewers' event encoding. All three task groups were informed that their memory for the movies would be tested.

After watching all three movies one time each while performing their assigned task, participants completed two memory tests for each film: recall and recognition memory. For the recall memory measure, participants were instructed to type what had occurred in the movie in as much detail as possible. They were given five minutes to respond. Testing procedures for recognition memory were similar to those described in Zacks et al. (2006). On each trial, participants were shown two still pictures: a target picture from the movie viewed and a lure from footage of the same actor performing similar activities in the same setting. (For example, for the breakfast video, foil pictures depicted the actor preparing different ingredients or assembling them in a different order.) Participants were instructed to select the picture that had come from the movie they had viewed. Twenty trials were presented for each movie, and recognition memory was scored as the proportion of trials answered correctly for each movie.

Finally, to assess the potential contribution of knowledge about everyday events to event memory, participants completed a measure of their script knowledge (Rosen, Caplan, Sheesley, Rodriguez, & Grafman, 2003). In this task, participants were given three minutes to type the sequence of events, line-by-line, associated with three everyday activities (getting ready for work, shopping for groceries, going out to dinner). An example response of a different activity (putting a child to bed) was provided to participants to help them structure their responses.

Design and Procedure—Testing occurred in three sessions over a period of one week. In session one, participants viewed the three movies while performing either the intentional encoding, timing, or event segmentation tasks. Each movie was only viewed one time, and task was manipulated between participants. Demographic information by group for all studies is presented in Table 1. The timing and event segmentation groups were given the opportunity to practice their assigned task on a separate practice movie before proceeding to the main experiment. Movie presentation order was counterbalanced. Following the presentation of the third movie, participants completed the recall and recognition tests for the first movie viewed. The delay period between the presentation of the first movie and its memory test was approximately 10 minutes.

Session two occurred 24 – 48 hours later. During this session, participants completed memory tests for the second movie viewed during session one and completed the script knowledge task. Session three occurred seven to nine days following session one. During this session, participants completed the recall and recognition tests for the last movie viewed during session one.

In short, the study design included the between-participants variable of task (segmentation, intentional encoding, or timing) and the within-participants variable of delay (10 min, 1 day, 1 week).

Recall and Script Elicitation Scoring—The primary dependent measures were recall and recognition performance, and segmentation performance for the event segmentation group. Here, we describe how the recall protocols and segmentation agreement were scored, and also how the independent variable of script knowledge was scored.

Due to the large number of responses, hand scoring of the recall protocols was not practical. Therefore, we validated a simple alternative measure, starting with scored recall data from a previous study (Sargent et al., 2013). In that study, recall protocols from 208 participants were scored using an adaptation of the action coding system developed by Schwartz, Reed, Montgomery, Palmer, and Mayer (1991), which divides the actor's stream of behavior into fine-grained action units; for example, “walks into kitchen”, “puts soap on hand”, “turns on water.” In that study, inter-rater reliability was found to be 0.84, $p < .001$, 95% CI [0.79, 0.90]. To create a recall measure that could be calculated automatically, we simply counted the number of words in each protocol and divided each count by the number of fine-grained actions in the scoring rubric to give a *normalized word count*. For the Sargent et al. (2013) study, the normalized word count measure correlated well with the hand-scored recall measure, $r = 0.77$, $p < .001$, 95% CI [0.71, 0.82]. To ensure the normalized word count measure captured the previously-observed relationship between event segmentation and memory, we computed linear mixed effects models with hand-scored recall or normalized word count as the dependent measures, a fixed effect of segmentation agreement, and random effects of movie and participant. For both analyses, there was a strong effect of segmentation agreement: For the hand-scoring, $F(1, 579.55) = 30.2$, $p < .001$; for normalized word count, $F(1, 568.33) = 9.96$, $p = .002$.

Therefore, for the present studies, we used the normalized word count as the criterial measure of recall memory. We adopted the same approach to score the script elicitation task: Responses were scored as the mean number of lines for the three everyday activities assessed.

Segmentation Agreement—To assess how well each participant in the event boundary condition segmented each movie, we calculated *segmentation agreement* scores for each viewing (Kurby & Zacks, 2011). Each movie was divided into 1 s bins. For each participant, each bin was coded as “1” if the participant segmented during that bin and “0” otherwise. These distributions were summed across participants to create a normative probability-of-segmentation time series. Segmentation agreement was calculated by computing the point-biserial correlation between an individual's segmentation and the normative time series for the group, and then scaling each correlation based on the number of boundaries the person identified (see Kurby & Zacks, 2011). This scaling accounts for the fact that the minimum and maximum possible correlation for a particular viewing depends on the number of boundaries identified by the viewer, and on the particular values of the normative time series. Scaled values range from zero to one, with zero being the worst possible agreement and one being the best.

Data Preparation—Of the participants enrolled, 157 failed to complete all sessions within specified time windows and thus were excluded from analyses. An additional 29 participants were dropped because valid data were not recorded for one or more memory measures, and 89 participants from both the segmentation and timing task were dropped because valid button pressing data were not recorded for at least one of the experimental movies. For the remaining segmentation and timing task groups who provided valid data, we calculated the mean number of button presses per minute for each participant, and excluded outlying participants whose measure was more than 2.5 SDs from the group mean; this resulted in the loss of 11 participants. For recall memory, participants whose word counts were fewer than 10 words for one or more of the recall responses were identified as outliers and excluded. Response times (RTs) for the recognition memory task were trimmed by excluding RTs from incorrect trials and RTs greater than 2.5 SDs from the within subject mean. Participants whose mean trimmed RT was greater than 2.5 SDs from the mean across participants or were faster than 500ms for at least one movie's recognition test were excluded from further analyses. We did not exclude participants on the basis of recognition accuracy. In total, 14 participants were excluded for being outliers on the memory tasks. Our final sample consisted of $n = 177$ (71 segmentation, 52 intentional encoding, and 54 timing).

Results and Discussion

Analyses were conducted in the R statistical environment (version 3.0.2; R Core Team, 2013) using the *lme4* package for linear mixed effect modeling (Bates et al., 2015). We used linear mixed effects models because they can simultaneously account for the influence of multiple random effects, in this case the effects of participants and of movies (for a review, see Baayen, Davidson, & Bates, 2008). For the current experiment, all linear mixed models treated both participant and movie as random effects, and treated encoding task and delay as fixed effects. In the analyses assessing the effects of segmentation agreement or script knowledge on memory these variables were also treated as fixed effects. The *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2014) was used to test the fixed main effects and interactions by comparing the nested model, using the Satterthwaite approximation for degrees of freedom. To compare the different encoding conditions within each delay condition, we fit separate models for each delay and then used the *lsmeans* (Lenth, 2016) and *multcomp* (Hothorn, Bretz, & Westfall, 2008) packages to compute Tukey tests comparing the different encoding conditions within delay.

Recall—We first fit a linear mixed model to predict recall performance from the fixed effects of task group and delay along with their interaction.[1] As can be seen in Figure 2, more information was recalled by the segmentation group ($M = 1.00$, $SD = 0.49$) than either the intentional encoding ($M = 0.91$, $SD = 0.46$) or timing ($M = 0.81$, $SD = 0.46$) groups, resulting in a main effect of task, $F(2,173.99) = 8.58$, $p < .001$. Mean recall performance decreased across the delay intervals (10 min: $M = 1.00$, $SD = 0.49$; 1 day: $M = 0.91$, $SD = 0.46$; 1 week: $M = 0.82$, $SD = 0.46$), resulting in a main effect of delay, $F(2,345.98) = 18.97$, $p < .001$. The interaction of task and delay was not significant ($F(4,345.93) = 0.57$, $p = .68$), indicating that task group difference was relatively stable across the delay intervals. As is shown in Table 3, the segmentation group out-performed the intentional encoding group significantly at all three delays. This provides the first evidence that explicitly segmenting everyday activity into meaningful events can incidentally produce a memory benefit when memory for those events is accessed much later in time. The segmentation group also was significantly better than the timing group at each delay. The intentional encoding and timing groups did not differ significantly at any delay.

Recognition—Separate analyses were conducted for recognition accuracy and response time. As can be seen in Figure 3, the segmentation group had higher recognition accuracy ($M = 0.69$, $SD = 0.13$) than either intentional encoding ($M = 0.64$, $SD = 0.15$) or timing ($M = 0.64$, $SD = 0.14$), resulting in a significant main effect of task, $F(2,173.96) = 9.23$, $p < .001$. The linear mixed model also detected a main effect of delay, $F(2,345.82) = 21.81$, $p < .001$, such that the recognition performance dropped significantly between 10 min ($M = 0.71$, $SD = 0.12$) and 1 day ($M = 0.64$, $SD = 0.14$), t(346) = 4.64, $p < .001$. Performance on recognition remained relatively stable between 1 day and 1 week (M = 0.63, SD = 0.15), t(346) = 1.72, $p = .20$. The interaction of task and delay was not significant, $F(4,345.76) = 2.06$, $p = .08$. As is indicated in Table 3, participants in the segmentation condition performed significantly better than the intentional encoding condition for the 10-minute and 1-day delays. Further, at a 1-week delay the segmentation group performed significantly better than the timing group and marginally better than the intentional encoding group. The intentional encoding and timing groups did not differ significantly at any delay.

The segmentation group responded slower ($M = 4225$ ms, $SD = 1858$ ms) to recognition probes than either the intentional encoding ($M = 3480$ ms, $SD = 1833$ ms) or timing ($M = 3930$ ms, $SD = 1747$) groups (see Figure 4), resulting in a main effect of group on recognition response time, $F(2,173.99) = 3.92$, $p = .02$. On average, participants responded faster to recognition memory trials at the 1 week delay ($M = 3590$ ms, $SD = 1810$ ms) than either the 10 min ($M = 4113$ ms, $SD = 1748$ ms) or 1 day ($M = 4046$ ms, $SD = 1924$ ms) delays, yielding a main effect of delay, $F(2,346.04) = 10.28$, $p < .001$. The task-by-delay interaction was found to be non-significant for response time, $F(4,346.02) = 0.55$, $p = .70$, suggesting that group differences in response times were relatively stable across the different delay intervals. In only one case was a pairwise difference across conditions significant: The segmentation group was significantly slower than the intentional encoding group at the 1-day delay (see Table 3).

---

[1] The distribution of mean normalized word count was positively skewed, so a log transformation was applied and the data were reanalyzed; the results were equivalent. Similar skew was observed in Experiments 2A, 2B, and 3, but in each case, transforming the data did not change the results of the statistical tests. In Experiment 4, the distribution of mean normalized word count was already normal. A log transformation resulted in a negatively skewed distribution; therefore we did not conduct further analyses on the log-transformed data.

In sum, participants who segmented the ongoing flow of perceptual activity into meaningful events remembered those events better than participants who simply watched with the intention to remember or who pressed a button to mark 15 s intervals. This benefit was seen in both recall and recognition tests, and was stable for up to a week following encoding.

For recognition memory, participants who had encoded by segmenting responded more slowly than the other groups, which may indicate that their superior accuracy was in part due to a speed-accuracy trade-off. However, given that the test conditions were identical for all three encoding groups, this seems unlikely. A more plausible account is that when participants were tested on movies they had segmented, they retrieved more information and thus took longer to make their recognition judgments.

Segmentation Agreement—Descriptive statistics for the segmentation task in the boundary segmentation group are given in Table 2, with data for the timing task for comparison. To test the hypothesis that segmentation agreement predicted memory, we fit linear mixed models predicting recall word count from segmentation agreement. All measures were z-scored to control for differences between movies and delays; segmentation agreement was z-scored within movie, and memory measures were z-scored within movie and delay. The full model included fixed effects of segmentation agreement and delay (dummy coded as 3 binary indicator variables), plus their interaction (coded as 3 additional variables). Because we were interested in the relationship between segmentation and memory, we constrained the intercepts in the model to be zero. The random effect of participant was modeled as an effect on memory performance at the 10 min delay; in other words, the 10 min delay was treated as the reference level characterizing differences between people. (Models using the other two delays as the reference level produced equivalent results.) For recall word count, there was a strong effect of segmentation, $F(1, 202.2) = 12.8$, $p < .001$, as can be seen in Figure 5. To test whether the relationship varied with delay, we compared the full model to a reduced model with no interaction terms. The comparison was not significant, meaning there was not evidence that the strength of the relationship varied with delay, $\chi^2(2) = 0.50$, $p = .78$. For recognition accuracy, neither the effect of segmentation nor the interaction with delay was significant, $F(1, 209.0) = 0.52$, $p = .47$; $\chi^2(2) = 1.52$, $p = .47$ (See Figure 6).

In sum, we replicated previous findings that more normative segmentation predicts better subsequent memory (Bailey et al., 2013; Sargent et al., 2013; Zacks et al., 2006). Importantly, for the first time, we showed that this effect is stable for up to one week.

Script Knowledge—To investigate whether script knowledge could account for the relationship between segmentation and memory, we constructed two linear mixed models. The first model was constructed in a manner similar to the model used to test recall and recognition performance (the ‘null’ model); it included the fixed effects of task and delay, and their interaction, but not the effect of script knowledge. A second model included the effect of script knowledge but no interaction with either task group or delay (the ‘no interaction’ model). We first analyzed the influence of script knowledge as a fixed effect by comparing both the null and no interaction models. For recall, the comparison was significant, meaning that pre-existing script knowledge was a significant predictor of recall performance, $\chi^2(1) = 25.41$, $p < .001$. In this model, the effects of both task group and delay were still significant [task: $F(2,172.98) = 8.44$, $p < .001$; delay: $F(2,345.98) = 26.85$, $p < .001$]. To test whether the effect of script knowledge varied with either delay or task, we constructed a third model that included the two-way interactions of script knowledge with task and delay, and the three-way interaction of task, delay and script knowledge (the 'full' model) and compared this to the no interaction model. The addition of the three-way interaction did not significantly improve model fit, $\chi^2(8) = 4.76$, $p = .78$. For recognition accuracy, the no interaction model did not provide any additional explanatory power beyond the null model,, $\chi^2(1) = 1.85$, $p = .17$. These findings replicate those by Sargent et al. (2013), suggesting that script knowledge facilitates event memory, but does so through a mechanism that is independent of segmentation.

Together, the group differences and individual differences provide the first evidence that attending to everyday event structure produces an event memory benefit up to a week after encoding. This is important for theories of long term memory, because it suggests that memory representations that undergo consolidation and support stable long-term memory reflect the event structure of the experiences that were encoded (Rubin & Umanath, 2015). These results are also of potential practical significance because they suggest that variables affecting event structure encoding can determine memory performance long past the encoding episode, allowing for the development of potential interventions to alleviate memory complaints for everyday activities.

As noted previously, the timing control has both advantages and disadvantages. We included it so as to have a comparison condition that, like the segmentation task, involves making motor decisions and executing actions during movie viewing. However, it could well have been that the requirement to count off time would interfere with attending to the events in the movie, imposing an artificial dual-task cost. In fact, the timing group performed similarly to the intentional encoding group; therefore in studies 2A, 2B, and 3 it was eliminated. (We brought this condition back in Experiment 4, for reasons we will discuss later.)

Experiments 2A and 2B

The primary aim of Experiment 2 was to ask whether the effect of segmentation on memory was retained at a delay of one month. In addition, we asked whether the relationship between individual differences in segmentation agreement and memory extended to this longer delay. Finally, in Experiment 2 we addressed a limitation of the design of Experiment 1. In that study, movies were always tested in the order in which they were presented. Thus, the movie tested in the encoding session was always the first movie presented, and the movie tested at the one week delay was always the last movie presented, confounding presentation order with delay. This leaves open the possibility that differences between the delays could be due to proactive interference rather than delay. In Experiment 2, the relationship between presentation order and delay was counterbalanced to control for this confound. We were interested in further exploring the possibility that the advantage of the segmentation group in recognition accuracy was due to a speed-accuracy trade-off, so we again measured recognition memory accuracy and response time.

When conducting Experiment 2, we found that the one month delay substantially increased our attrition rates. We therefore collected a second sample; the two samples are reported separately as Experiments 2A and 2B.

Method

Participants—For Experiment 2A, we recruited 456 participants (295 female, median age: 32 years, age range: 18–77 years) from AMT. For Experiment 2B, we recruited 575 participants (369 female, median age: 31 years, age range: 18–77 years). All workers were from the United States and had a HIT approval rating of 95% or above. The study required participants to complete two experimental sessions. Participants received $1 for completing session one and $1.50 for completing session two. Informed consent was obtained as for Studies 1A and 1B.

Materials and Tasks—Movies and memory tests were the same as in Experiment 1. An additional experimental movie of a female actor checking out a book at the library (249 s) was included, bringing the total number of test movies to four. Participants were assigned to the intentional encoding or event segmentation tasks; the timing task was dropped. To encourage participants to segment at a finer and more consistent grain, we provided feedback on the number of boundaries identified by participants in the segmentation group for the practice movie (as in Bailey et al., 2013; Magliano & Zacks, 2011; Sargent et al., 2013; Zacks, Speer, Swallow, & Maley, 2010). Participants who identified at least six boundaries were told to continue identifying boundaries in the same manner for the experimental movies. Those who identified fewer than six boundaries were told to attempt to identify more boundaries in the experimental movies.

Design and Procedure—Testing occurred in two sessions over a period of one month. In the first session, participants viewed four experimental movies while performing either the intentional encoding or the event segmentation task. Task was manipulated between participants and the event segmentation group was given the opportunity to practice their task using the practice movie. Movie presentation order was counterbalanced within each task group. Following the presentation of the fourth experimental movie, participants completed recall and recognition tests for two of the movies. The order in which memory for the movies was tested was counterbalanced in sets of two. For example, participants could have received memory tests for either the first and second movies viewed or for the third and fourth movies viewed during this session. In every case, at least one additional movie was presented before a given movie was tested. Session two occurred approximately 30 – 44 days after session one. In session two, participants completed recall and recognition tests for the set of movies not tested in session one. They also completed the script knowledge task.

The study design included the between-participants variables of task and of the order in which the movies were tested, and the within-participants variable of delay. As in Experiment 1, the primary dependent measures were recall and recognition performance, and segmentation performance for the event segmentation group.

Data Preparation—Of the 2A sample, 232 participants (50.8% of the sample) were lost to attrition between the two experimental sessions; for 2B, the figure was 335 (58.2% of the sample). Additional participants were excluded due to: completing the study outside of the allowable time frame (50 participants in Study 2A, one in study 2B), missing valid data on one or more of the memory measures (17 in 2A, 23 in 2B), failing to record valid button presses (16 in 2A, six in 2B), having a button presses per minute rate over 2.5 SDs from the mean of the group (six in 2A, eight in 2B), reporting that they were not attending to movies during movie viewing (one in each sample), writing fewer than 10 words for any recall response (seven in 2A, nine in 2B), and being an outlier on recognition memory (five in 2A, eight in 2B). The final sample sizes consisted of $n = 122$ in Experiment 2A (56 segmentation, 66 intentional encoding) and $n = 184$ in Experiment 2B (88 segmentation, 96 intentional encoding).

Results and Discussion

Analysis procedures were the same as in Experiment 1. We treated participant and movie as random effects when constructing the linear mixed models.

Recall—As can be seen in Figure 2, the segmentation group recalled more information overall, but this difference was not significant for either sample. [Experiment 2A: segmentation $M = 1.01$, $SD = 0.59$; intentional encoding $M = 0.89$, $SD = 0.58$; $F(1,117.99) = 2.62$, $p = .10$. Experiment 2B: segmentation $M = 0.93$, $SD = 0.59$; intentional encoding $M = 0.89$, $SD = 0.59$; $F(1,179.97) = 0.35$, $p = .55$.] Consistent with this, pairwise comparisons found no significant group differences at either delay (Table 3). In neither sample was there an effect of presentation order [Experiment 2A: $F(1,117.99) = 0.00$, $p = .98$; Experiment 2B: $F(1,179.97) = 0.49$, $p = .48$]. Nor was there a significant interaction between presentation order and delay [Experiment 2A: $F(1,117.99) = 0.44$, $p = .51$; Experiment 2B: $F(1,179.97) = .01$, $p = .92$], which suggests that the patterns observed in Experiment 1 were not strongly affected by the confounding of presentation order and delay. Participants recalled significantly more information at the 10 min delay than at the one month delay, leading to significant main effects of delay in both samples. [Experiment 2A: 10 min $M = 1.18$, $SD = 0.59$; one month $M = 0.71$, $SD = 0.47$; $F(1,359) = 215.15$, $p < .001$. Experiment 2B: 10 min $M = 1.13$, $SD = 0.61$; one month $M = 0.68$, $SD = 0.47$; $F(1,544.88) = 217.26$, $p < .001$.] Task performance did not appear to vary across the delays, as there were no significant interactions between task and delay [Experiment 2A: $F(1,358.93) = 0.50$, $p = .47$; Experiment 2B: $F(1,544.91) = 0.22$, $p = .63$.]. In Experiment 2B but not 2A, there was a significant three-way interaction between task, delay, and order group, such that intentional encoding viewers who were tested on the third and fourth movies recalled more information at the one month delay compared to the other conditions [Experiment 2A: $F(1,358.93) = 0.12$, $p = .72$; Experiment 2B: $F(1,544.87) = 20.30$, $p < .001$.]. This interaction was not readily interpretable. No other main effects or interactions reached statistical significance.

Recognition—We again analyzed recognition accuracy and RT. As can be seen in Figure 3, the segmentation group had higher recognition accuracy than the intentional encoding group in both samples. This resulted in main effects of task that were marginally significant in Experiment 2A and significant in Experiment 2B [Experiment 2A: segmentation $M =$

0.71, *SD* = 0.14; intentional encoding *M* = 0.68, *SD* = 0.15; *F*(1,117.97) = 2.93, *p* = .09. Experiment 2B: *M* = 0.70, *SD* = 0.14; intentional encoding *M* = 0.67, *SD* = 0.15; *F*(1,179.96) = 6.52, *p* = .01]. Pairwise comparisons found a significant difference at the 10 minute delay and a marginally significant difference at the 1-month delay in Experiment 2B (see Table 3). Again, participants had significantly higher recognition accuracy at 10 min than at one month [Experiment 2A: 10 min *M* = 0.76, *SD* = 0.13; one month *M* = 0.63, *SD* = 0.13; *F*(1,359.10) = 26.44, *p* < .001. Experiment 2B: 10 min *M* = 0.74, *SD* = 0.14; one month *M* = 0.62, *SD* = 0.13; *F*(1,544.59) = 152.39, *p* < .001.]. In neither sample was there a significant interaction between group and delay [Experiment 2A: *F*(1, 358.79) = 0.09, *p* = .76; Experiment 2B: *F*(1, 545.03) = 0.00, *p* = .98.]. As with recall, no effect of testing order on task performance was observed [Experiment 2A: *F*(1,117.97) = 1.63, *p* = .20; Experiment 2B: *F*(1, 179.96) = 1.59, *p* = .21].

Given the suggestive and consistent differences between the groups in recognition accuracy, we combined the data from Experiments 2A and 2B at each delay and performed pairwise comparisons on the combined sample. These indicated a significant difference at both the 10-minute delay [*t*(302) = 2.31, *p* = .02] and the 1-month delay [*t*(301) = 2.38, *p* = .018]. However, these results should be interpreted with some caution because the two samples were collected sequentially and thus the observations are not fully independent.

For recognition RT, no differences in response speed were detected between the segmentation and intentional encoding groups across either sample [See Figure 4 and Table 3. Experiment 2A: segmentation *M* = 3837 ms, *SD* = 1808 ms; intentional encoding *M* = 3946 ms, *SD* = 2086 ms; *F*(1,117.99) = 0.15, *p* = .69. Experiment 2B: segmentation *M* = 3627 ms, *SD* = 1633 m; intentional encoding *M* = 3767 ms, *SD* = 1673 ms; *F*(1,179.98) = 0.61, *p* = .43.]. Participants responded slower at the 10 min delay than at the one month delay in both samples, leading to significant main effects of delay [Experiment 2A: 10 min *M* = 4195 ms, *SD* = 2024 ms; one month *M* = 3598 ms, *SD* = 1854 ms; *F*(1,359.10) = 26.44, *p* < .001. Experiment 2B: 10 min *M* = 3951 ms, *SD* = 1765 ms; one month *M* = 3438 ms, *SD* = 1490 ms; *F*(1,545.01) = 41.03, *p* < .001.]. No other main effects or interactions were significant. In particular, there were no main effects or interactions involving presentation order [Experiment 2A: largest *F* = 2.58; Experiment 2B: largest *F* = 1.81].

Segmentation Agreement—Descriptive statistics for the segmentation task in the segmentation group are given in Table 2. As in Experiment 1, all measures were z-scored to control for differences between movies and delays: segmentation agreement was z-scored within movie, and memory measures were z-scored within movie and delay. The full model included fixed effects of segmentation agreement and delay (dummy coded as 2 binary indicator variables), plus their interaction (coded as 2 additional variables). Because we were interested in the relationship between segmentation and memory, we constrained the intercepts in the model to be zero. As in Experiment 1, the random effect of participant was modeled as an effect on memory performance at the 10 minute delay. (Models using the one month delay produced equivalent results.) For recall word count, there were strong effects of segmentation in both samples such that participants who segmented more normatively displayed better recall performance [Experiment 2A: *F*(1, 218.99) = 26.7, *p* < .001; Experiment 2B: *F*(1, 336.69) = 12.25, *p* < .001.]. To test whether the relationship varied with delay, we compared the full model to a reduced model with no interaction terms. The comparison was not significant, meaning there was not evidence that the strength of the relation varied with delay [Experiment 2A: $\chi^2(1) = .35$, $p = .55$; Experiment 2B: $\chi^2(1) = .16$, $p = .68$]. For recognition accuracy, we found an effect of segmentation on recognition accuracy in Experiment 2B but not 2A [Experiment 2A: $F(1, 219.98) = 2.91$, $p = .09$; Experiment 2B: $F(1, 347.96) = 7.05$, $p < .01$]. For neither sample did this effect interact with delay [Experiment 2A: $\chi^2(1) = 0.26$, $p = .61$; Experiment 2B: $\chi^2(1) = 0.68$, $p = .40$].

Script Knowledge—As for Experiment 1, we compared three models: a null model, a no-interaction model, and a full model including the two-way interactions of script knowledge with task and delay, plus the three-way interaction. For recall, the no-interaction model fit significantly better than the null model, suggesting that script knowledge predicted recall performance [Experiment 2A: $\chi^2(1) = 25.769$, $p < .001$; Experiment 2B: $\chi^2(1) = 42.936$, $p < .001$]. For recognition accuracy, adding script knowledge to the model did not improve fit in Experiment 2A but did in Experiment 2B, [Experiment 2A: $\chi^2(1) = 2.9375$, $p = .09$; Experiment 2B: $\chi^2(1) = 8.3432$, $p = .003$]. Adding script knowledge to the models did not change the effects of task on memory: The effect of task on recall in the no interaction model was not significant for either experiment [Experiment 2A: $F(1,116.99) = 1.612$, $p = .20$; Experiment 2B: $F(1,178.97) = 0.00$, $p = .97$]. The effect of task on recognition was marginally significant for Experiment 2A, $F(1,116.97) = 2.873$, $p = .09$, and significant for Experiment 2B, $F(1,178.95) = 5.418$, $p = .02$. We next compared the no interaction model with another model that included an interaction among script knowledge, task, delay, and the order in which movies were tested to determine whether there were underlying differences in schematic knowledge for everyday events among the groups. For both recall and recognition accuracy, the comparison was not significant in either Experiment 2A or 2B, suggesting that the effect of script knowledge on memory did not vary with task, delay, or the order the movies were tested in [Recall: Experiment 2A $\chi^2(7) = 12.234$, $p = .09$; Experiment 2B $\chi^2(7) = 12.096$, $p = .10$; Recognition: Experiment 2A $\chi^2(7) = 3.6373$, $p = .82$; Experiment 2B $\chi^2(7) = 3.589$, $p = .82$].

In summary, Experiments 2A and 2B generally replicated the results of Experiment 1 and extended them to a one month delay. Viewers who segmented movies during encoding remembered those movies better than those who simply watched and encoded the movies for subsequent memory, though this difference was significant only in Experiment 2B and only for the recognition memory measure. Of those who segmented, higher segmentation agreement was associated with better subsequent recall and recognition. Neither the group differences nor the within-group relations between segmentation agreement and memory changed substantially with delay, indicating that these effects are consistent up to a month after encoding.

These experiments also cleared up two ambiguities raised by Experiment 1. First, because the relationship between delay and encoding order was counterbalanced, they provide evidence that the delay effects were due to delay rather than to encoding order. Second, they showed that the segmentation group's advantage in recognition memory in Experiment 1 was not likely due to a speed-accuracy trade-off, because in these experiments the segmentation groups performed the recognition trials slightly and non-significantly faster than the intentional encoding groups.

Experiment 3

Experiments 1, 2A and 2B established that segmenting a movie while viewing it improved subsequent memory. They also found that participants with higher segmentation agreement scores consistently had better memory than those with lower segmentation agreement scores over time. These effects were present when memory was tested a few minutes after presentation, and were stable over delays of at least a week and possibly a month. But, is the effect of active segmentation at encoding present *immediately* after encoding, or does it require time to develop? As noted in the Introduction, the effect of encoding strategies on memory can depend on the retention interval, and in particular what is optimal encoding for immediate retrieval can differ from what is optimal for retrieval after a delay (e.g., Roediger & Karpicke, 2006). Is this the case for encoding everyday events?

To address this question, we conducted an experiment in which we tested memory immediately after each movie was viewed. Somewhat to our surprise, segmentation did *not* improve memory when tested immediately.

Method

Participants—We recruited 450 participants (259 females; ages range: 18–67; median age: 29 years) from AMT. All workers were from the United States and had a HIT approval rate of at least 95%. Participants received $2 as compensation for time and effort. Informed consent was obtained as for Studies 1 and 2.

Materials and Tasks—Experiment 3 used the same movies, tasks, and memory tests as Experiments 2A and 2B. Unlike in Experiments 2A and 2B, participants in the event segmentation task did not receive feedback regarding the number of units identified. In addition, participants completed a test of *order memory*. Participants viewed 12 still pictures arranged in a random order in a 4x3 grid on the screen, and were instructed to order the still frames in the correct temporal order (Zacks et al., 2006). They did so by clicking on arrows next to the pictures to move them. Order memory was scored as an error measure, which was the mean absolute deviation from the correct position (i.e., lower scores meant better performance). We replaced the practice movie of a man building a boat out of Duplo blocks with a shorter movie of a woman making a sandwich (146 s), because participants in our Mechanical Turk studies had reported that they preferred shorter procedures.

Design and Procedure—Participants performed either the intentional encoding or event segmentation task while viewing a single movie. Then, in contrast to the previous experiments, memory was tested immediately. The recall test was given first, followed by the recognition test, then the order memory test. The design included the between-participants variable of task. The primary dependent measures were recall, recognition and order memory performance.

Data Preparation—One hundred and twenty-nine participants were excluded from analyses because valid data were not recorded for one or more of the memory measures. Preparation procedures for recall and recognition were the same as in Experiment 1. As a result, seven participants were excluded for being outliers on memory measures. Thus, the final sample consisted of $n = 314$ participants, 120 in the segmentation condition and 194 in the intentional encoding condition.

Results and Discussion

As in Experiment 1, linear mixed models were constructed to predict the dependent measures. (Further pairwise comparisons were not required because the design included only a single delay.) Participant was treated as the only random effect because each participant viewed only one movie. Because the memory tests were given directly following movie presentation, all models had a single fixed effect: task.

Recall—As can be seen in Figure 2, the intentional encoding group recalled a higher proportion of information from the movies than the segmentation group: segmentation $M = 1.28$, $SD = 0.50$; intentional encoding $M = 1.34$, $SD = 0.50$. This difference was not statistically significant, $F(1,310.49) = 0.08$, $p = .78$.

Recognition—Separate analyses were conducted on accuracy and response time. For accuracy, see Figure 3. The intentional encoding group had slightly higher accuracy($M = 0.74$, $SD = 0.14$) than the segmentation group ($M = 0.76$, $SD = 0.11$). This difference was not significant, $F(1,310.59) = 0.13$, $p = .72$.

For response time, see Figure 4. No differences were detected between the segmentation ($M = 5470$ ms, $SD = 2200$ ms)and intentional encoding groups ($M = 5525$ ms, $SD = 2322$ ms), $F(1,310.44) = 2.04$, $p = .15$. The fact that the intentional encoding group performed as well as or better than the segmentation group on an immediate recognition test, which highlights the importance of a delay interval is necessary for effective event memory encoding.

Order Memory—The segmentation group made slightly fewer errors ($M = 1.12$, $SD = 1.32$) than the intentional encoding group ($M = 1.18$), but this difference was not statistically significant, $F(1,310.25) = 1.41$, $p = .23$.

Segmentation Agreement—Descriptive statistics for the segmentation task are given in Table 2. Given that memory tests were administered immediately following encoding, we treated segmentation agreement as the only fixed factor in our model. As in the previous studies, segmentation agreement predicted both recall performance and recognition accuracy [recall: $F(1,113.97) = 13.16$, $p < .001$; recognition accuracy: $F(1,113.91) = 4.30$, $p = .04$; see Figure 6.].

In sum, Experiment 3 found that with an immediate test, memory was *not* better in the segmentation condition. This result stands in contrast to Experiments 1, 2A, and 2B, in which a benefit or a trend towards a benefit was found when a delay was present. However, these results are consistent with the results of Roediger and Karpicke (2006), who found that a normatively less effective strategy (i.e., restudying) produced better memory at an immediate test, whereas a normatively effective strategy (i.e., testing) produced better memory after a delay.[2] In addition, we replicated the finding of the previous experiments that participants with better segmentation agreement also had better memory; this also replicates previous results with immediate tests (Bailey et al., 2013; Kurby & Zacks, 2012; Sargent et al., 2013; Zacks et al., 2006).

Experiment 4

In Experiments 1, 2A, and 2B, participants completed memory tests after delays varying from 10 min to one month from the viewing of a movie. Across all of these delayed test conditions, we consistently found that (a) the act of segmenting during encoding improved memory, and (b) those who segmented more effectively remembered more. (Of these effects, the least robust is the benefit of performing the segmentation task at the one month; however, even here there was a strong relationship between effective segmentation and memory.) In contrast, when participants were tested without a delay in Experiment 3 there was no benefit associated with performing the segmentation task or with segmenting more effectively. This pattern invites a hypothesis that when people recall and recognize an activity *immediately*, they rely on memory systems that are less affected by event segmentation, but once a brief delay has elapsed they depend on memory systems whose effectiveness depends on segmentation during encoding. To directly test this hypothesis, we compared an immediate delay condition with a brief 10 min delay condition in Experiment 4.

Method

Participants—We recruited 776 participants (481 females; age range: 18–70; median age: 32 years) from AMT. All workers were from the United States and had a HIT approval rate of at least 95%. Participants received $1.50 as compensation for time and effort. Informed consent was obtained as for the previous studies.

Materials and Tasks—Movies, tasks and memory tests were the same as in Experiment 1. We brought back the timing control task for this study, because we thought it was possible that the requirements to make motor decisions and execute actions in the segmentation task might be particularly detrimental to immediate event memory. Participants received feedback during the event segmentation practice as in Experiments 2A and 2B.

Design and Procedure—Participants performed either the *intentional encoding*, *event segmentation* or *timing* task while watching the three experimental movies. Memory tests were administered either directly following each movie (*immediate*) or after all movies had been viewed (*10 min delay*). Task and delay conditions were manipulated between participants. (Here, delay was manipulated between participants, in contrast to the within-participants manipulations in Experiments 1, 2A, and 2B. This was done to provide as pure as possible of delayed memory, given our interest in directly comparing immediate and delayed memory.) Both the event segmentation and timing groups were given the

---

[2] In another experiment conducted in our laboratory with 75 older adults and using similar methods, intentional encoding led to significantly *better* recall and recognition accuracy when tested immediately after presentation.

opportunity to practice their task using the practice movie. Movie presentation order was counterbalanced. As in the previous studies, the primary dependent measures were recall and recognition performance, and segmentation for the event segmentation group.

Data Preparation—Data preparation procedures were the same as in Experiment 1. From the sample, 145 participants were excluded for not having valid data on one or more of the memory measures. In addition, 56 participants were identified as being outliers on at least one of the memory measures and were excluded from analyses. The final sample for the experiment consisted of 575 participants (192 segmentation, 194 intentional encoding, and 189 timing).

Results and Discussion

Analysis procedures were the same as in Experiments 1, 2A and 2B. We treated participant and movie as random effects in constructing the linear mixed models.

Recall—Averaging across delay conditions, the segmentation group recalled the most information from the movies (*M* = 0.98, *SD* = 0.47), followed by the intentional encoding group (*M* = 0.95, *SD* = 0.49) and finally the timing group (*M* = 0.92, *SD* = 0.43); however, this effect was not significant, *F*(2,568.96) = 1.25, *p* = .28. Participants in the 10 min delay condition recalled significantly less information (*M* = 0.91, SD = 0.46) than participants who received memory tests immediately after viewing a movie (*M* = 0.98, *SD* = 0.46), resulting in a main effect of delay, *F*(1, 568.96) = 5.37, *p* = .02. Our main question of interest this analysis was whether we would observe an interaction between task and delay, replicating the findings from Experiment 3 in the immediate condition and the findings from Experiments 1, 2A, and 2B in the 10 min delay condition. The data replicated our previous results, resulting in a significant Task x Delay interaction, *F*(2, 568.96) = 3.41, *p* = .03. Pairwise comparisons indicated no significant group differences in the immediate condition (see Table 3). This replicated the finding from Experiment 3 that no benefit of segmentation on memory is present when memory is tested immediately after encoding. Pairwise comparisons for the 10 min delay condition indicated the segmentation group recalled significantly more information than the intentional encoding group. This pattern of results replicates the general pattern observed in Experiments 1, 2A and 2B.

Recognition—For recognition accuracy, averaging over delay, the timing group performed slightly worse (*M* = 0.68, *SD* = 0.15) than either the segmentation (*M* = 0.70, *SD* = 0.16) or intentional encoding (*M* = 0.70, *SD* = 0.15) groups, resulting in a marginal main effect of task, *F*(2,568.9) = 2.91, *p* = .06. Participants in the 10 min delay condition also performed slightly worse (*M* = 0.68, *SD* = 0.15) than those in the immediate delay condition (*M* = 0.71, *SD* = 0.15), yielding a main effect of delay, *F*(1, 568.9) = 10.79, *p* = .001. Performance amongst the task groups in both delay conditions remained relatively consistent, resulting in no Task x Delay interaction, *F*(2, 568.9) = 1.50, *p* = .22. Numerically, the pattern was similar to that seen for recall memory: the intentional encoding group performed better on the immediate test but worse on the delayed test. The only significant pairwise comparison was a finding that the timing group performed better than the passive group in the immediate condition (see Table 3).

For response time, the timing group responded significantly faster (M = 4164 ms, SD = 1881) than either the intentional encoding (M = 4530 ms, SD = 2189 ms) or segmentation (M = 4690 ms, SD = 2218 ms) groups, resulting in a main effect of task, F(2,568.96) = 4.81, p = .008. Given that the timing group had worse recognition accuracy, this again seems consistent with the notion that participants responded more quickly when they were able to retrieve less information. The only significant pairwise comparison was between the segmentation and timing groups at the immediate delay (see Table 3). Although the 10 min delay group responded faster (M = 4357 ms, SD = 2053 ms) than the immediate delay group (M = 4566 ms, SD = 2167 ms), the difference was not significant, F(1, 568.96) = 2.33, p = .13. Furthermore, response times amongst the task groups were relatively consistent between the two delays, resulting in no Task x Delay interaction, F(2, 568.96) = 0.34, p = .71.

Segmentation Agreement—Descriptive statistics for the segmentation task are given in Table 2, with data for the timing task for comparison. To test the hypothesis that segmentation agreement predicted memory, we again fit linear mixed models predicting our memory measures from segmentation agreement. Recall word count, recognition accuracy, and segmentation agreement were z-scored within movie and delay to control for differences between movies and delays. The full model included fixed effects of segmentation agreement and delay. As delay was a between-subject variable, we included another fixed effect that modeled the interaction of segmentation agreement and delay (computed as the product of segmentation agreement and a binary indicator of delay). As in the segmentation analysis in Experiment 1, we constrained the intercepts in the model to be zero. The random effect of subject was also included in the model. For recall word count, there was a strong effect of segmentation, F(1, 187.99) = 15.3, p < .001. To test whether the relationship varied with delay, we compared the full model to a reduced model with no separate fixed effect modeling the interaction. The comparison was not significant, meaning there was not evidence that the strength of the relation varied with delay, $\chi^2(1) = 1.37$, p = .24. For recognition accuracy, there was a strong effect of segmentation, F(1,187.96) = 8.44, p = .004. We again compared the full model to a reduced model with no separate fixed effect modeling the interaction. The comparison was not significant, meaning there was not evidence that the strength of the relation varied with delay for recognition accuracy, $\chi^2(1) = 1.79$, p = .18. Again, more normative segmentation was predictive of later recall performance in our segmentation group at both the immediate and 10 minute delays.

In sum, Experiment 4 replicated the finding from Experiments 1, 2A, and 2B that active segmentation produces more robust and coherent memory representations than intentional encoding the movies at a 10 min delay, and also replicated the findings from Experiments 3 that segmentation was not superior when memory was assessed immediately following encoding. One possibility is that when participants are tested immediately after encoding, they are able to rely on perceptual details that are not substantially affected by segmentation. However, such cues may be highly subject to interference or decay, such that by 10 min later they are less able to support good memory performance. The requirement to perform a motor task in the segmentation condition (and the timing condition) may interfere with the encoding of these perceptual features. Numerically, the timing task performed worst on both the recall and recognition tests at the immediate delay. This is consistent with the idea that dual-task interference with the encoding of detailed perceptual features suppresses performance of the segmentation group on the initial test.

We also found that individual differences in segmentation predicted memory in both the immediate and 10 min delay conditions. This replicates what was observed in the previous experiments and suggests that even though perceptual cues may contribute to immediate memory, processes associated with segmentation also are important, such that participants who segment more normatively remember more.

General Discussion

The current study sought to answer three questions. First, does attention to segmentation contribute to improved memory performance? Second, do individual differences in segmentation ability predict memory? Finally, are these effects sustained over delay intervals of moments to weeks? A series of experiments demonstrated that as viewers watched movies with the intent to remember them later, attending to segmentation improved their memory. However, this effect required a brief retention interval to emerge; when memory was assessed immediately following encoding, the memory benefit was absent. Importantly, event memory for segmented activity was robust to degradation for up to a month following encoding, suggesting that segmentation is a significant contributor to long-term memory. To our knowledge, this is the first evidence that attending to event segmentation produces durable improvements in memory.

Additionally, this is the first study to demonstrate that individual differences in segmentation ability predict memory at long delays. Individuals whose segmentation aligns better with group norms recall more information about an event even one month after encoding. Together, the results from the experimental manipulation and from the individual difference analyses converge to suggest that segmenting ongoing activity into effective units contributes to durable long-term memory. This result is consistent with a recent study showing that segmentation agreement accounts for unique variance in event memory above and beyond laboratory measures of processing speed, working memory capacity, verbal episodic memory, and crystallized knowledge (Sargent et al., 2013).

Event segmentation theory (EST; Zacks et al., 2007) provides a potential mechanism for the effects of segmentation on subsequent memory: During comprehension one constructs a model of the current situation and uses this event model to predict what is likely to occur in the near future. The theory proposes that when prediction error spikes above a threshold, the current event model is updated to better represent the current situation. One consequence of this updating process is that the timing of event model updating determines the contents of the representations that are available to be consolidated into long-term memory. When event boundaries are identified in more normative parts of the event, features of the ongoing activity are chunked together adaptively, resulting in memory representations that are more likely to be durable and retrievable. This converges with evidence from other studies showing that information contained within an event is more likely to be recalled than information that spans multiple events (DuBrow & Davachi, 2013; Ezzyat & Davachi, 2011, 2014; Swallow, Zacks, & Abrams, 2009; Zwaan, 1996; Zwaan, Langston, & Graesser, 1995). Thus, EST provides a straightforward explanation of why people with higher segmentation agreement subsequently remember more: they have formed more adaptive event representations.

The finding that performing the segmentation task leads to better memory also is consistent with EST, though it is not so clearly entailed by the theory. According to EST, the process of event segmentation is not typically part of conscious awareness. This proposal is supported by data showing that, with little training, viewers are able to segment activity with good agreement across observers (Bailey et al., 2013; Sargent et al., 2013; Zacks et al., 2006) and within observers across time (Speer, Swallow, & Zacks, 2003), and by data showing that some parts of the cortex increase in activity during event boundaries when observers are simply watching movies or reading stories without the intention of segmenting them (Speer, Zacks, & Reynolds, 2007; Whitney et al., 2009; Zacks et al., 2001, 2010). Deliberate segmentation may improve memory by at least two possible pathways. First, attending to the segmentation mechanism may increase its fidelity by increasing resources devoted to prediction error monitoring or to event model updating. Second, the segmentation task may increase attention to features of the activity that are better suited for forming adaptive long-term memories. For example, the segmentation instructions may encourage viewers to focus on features of the actor and her actions and decrease distraction by extraneous features such as objects that are not used in the activity.

Why, then, did performing the segmentation task no facilitate event memory when tested immediately after viewing? One likely possibility is that memory over short delays depends more heavily on representations other than event models. In memory for discourse, researchers routinely distinguish between the representation of the “surface form” of a story—the details of word and grammar of exactly what was written—and the event models that represent what is happening in the story (e.g., Bransford, Barclay, & Franks, 1972; Kintsch & Van Dijk, 1978). Memory after shorter delays tends to reflect the surface structure, whereas memory after longer delays is almost entirely dependent on the event model representations. One possibility is that performing the segmentation task competes with deliberate encoding of the surface structure of the activity—the details of exactly which low-level actions are performed on precisely which objects. However, memory for this surface structure rapidly decays, and with delay the superior encoding of the event models afforded by the segmentation task dominates.

In Experiments 1, 2A, and 2B, we found that event knowledge also predicted event memory at all delays. This result replicates the findings of Sargent et al. (2013), who found that event knowledge predicted event memory independently of segmentation agreement and other psychometric measures of cognitive ability. In terms of EST, a natural mechanism for this effect is that when viewers update their event models in response to spikes in prediction error, they draw on event knowledge in addition to immediately available perceptual cues. This replication strongly supports the proposal that a rich knowledge base facilitates memory for events, as it does for other sorts of materials (e.g., Chase & Simon, 1973; Voss, Vesonder, & Spilich, 1980).

Theoretically motivated manipulations to improve memory encoding have a rich history in cognitive psychology, including manipulations to encourage imagery (Paivio, 1970), elaboration (Craik & Lockhart, 1972), or survival processing (Nairne et al., 2007)—not to mention older techniques that may lack theoretical bona fides but are highly effective nonetheless, such as the method of loci (Ross & Lawrence, 1968) and the peg word method (Wood & Pratt, 1987). However, event segmentation is novel in that whereas most other mnemonic interventions are effective for remembering materials that are already highly structured, such as lists of words, the event segmentation intervention can be applied easily to any everyday experience, movie, or narrative text.

The ongoing segmentation of everyday activities can be affected by healthy aging and also by clinical disorders such as Alzheimer's disease (Bailey et al., 2013; Zacks et al., 2006), schizophrenia (Zalla, Verlut, Franck, Puzenat, & Sirigu, 2004), Parkinson's disease (Zalla et al., 1998; 2000), and post-traumatic stress disorder (Eisenberg, Sargent, & Zacks, in press). In the case of Alzheimer's disease, lower event segmentation performance has been shown to be associated with greater genetic risk (Bailey et al., 2015) and greater anatomic evidence of neuropathology (Bailey et al., 2013), and in both cases accounts for some of the associated memory deficits. This population may have degraded semantic knowledge about events, resulting in segmentation that produces less coherent memory representations. It is possible that interventions to improve segmentation in these groups could help remediate their memory deficits.

In other studies of memory and aging, encoding manipulations that help structure information adaptively have been found to be beneficial to older adults (Naveh-Benjamin, Craik, & Ben-Shaul, 2002). Previous studies in event segmentation have found that in healthy young adult populations, edits in videos that reinforce the natural segmentation of events can improve memory (Boltz, 1992; Schwan et al., 2000). Recently, Gold and Zacks (2014) tested whether providing external cues could help memory in both a younger and older adult sample. They edited movies to add cues to influence viewers' segmentation: an alarm sound, slower presentation speed for the video, and an arrow pointing at an object relevant to the event. The cues were placed either at event boundaries or at the midpoints between events, and these time points were reinforced in a post-viewing summary presentation of frames from the movies. After watching each movie and summary, participants completed recall and recognition memory tests for the movie. They found cuing at event boundaries selectively increased recall of boundary information for both groups. These results suggest that providing an external event structure can help facilitate memory for the event.

These considerations suggest a two-pronged strategy for memory scaffolding. First, people can be provided with materials that provide external structural cues to aid in promoting more normative segmentation. Second, they can be instructed to attend to how they segment in order to impose a more effective self-generated structure onto events. Such a clinical intervention would have several substantial advantages. Namely, clinical interventions utilizing these methods could be adapted to a variety of material, from remembering the order in which to take prescription medication to how to properly brush one's teeth or dress one's self. Such interventions would also be easy to implement, as the event segmentation task itself does not require much in terms of material beyond the stimulus video. The task is also portable and could be conducted in either a hospital or residential setting depending on the current living situation of the client. Future studies should investigate the efficacy of adapting the event segmentation task as a clinical intervention for use in dementia populations.

Of course, there is no free lunch. It is likely that the instruction to attend to event segmentation imposes cognitive costs which, depending on the task, may compromise ongoing task performance. For example, an individual attending to the order in which to take prescription medication may not be aware of exactly which medications they are taking. An important topic for future research is to assess the relative costs and benefits of attending to event segmentation, so as to maximize benefits, minimize costs, and identify situations in which the costs are slim.

As cognitive creatures, we bring forth order from the chaos of sensory signals. One fundamental aspect of this order is the structuring of events in time. By adaptively chunking ongoing experience, we form the basic units that allow us to look back and remember days and weeks later.

Acknowledgments

This project was funded by NIH grant R01 AG031150, PI Jeffrey M. Zacks, and DARPA grant DA13AP00009, PI Jeffrey M. Zacks. We would like to thank Mike Strube for his help with constructing the mixed linear models for segmentation agreement.

References

Abbott V, Black JB, Smith EE. The representation of scripts in memory. Journal of Memory and Language. 1985; 24(2):179–199.

Abel M, Bäuml KHT. Adaptive memory: The influence of sleep and wake delay on the survival-processing effect. Journal of Cognitive Psychology. 2013; 25(8):917–924. http://doi.org/10.1080/20445911.2013.825621.

Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59(4):390–412. http://doi.org/10.1016/j.jml.2007.12.005.

Bahrick HP, Bahrick LE, Bahrick AS, Bahrick PE. Maintenance of foreign language vocabulary and the spacing effect. Psychological Science. 1993; 4(5):316–321.

Bahrick HP, Phelphs E. Retention of Spanish vocabulary over 8 years. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1987; 13(2):344–349.

Bailey HR, Sargent JQ, Flores S, Nowotny P, Goate A, Zacks JM. *APOE* $\varepsilon 4$ genotype predicts memory for everyday activities. Aging, Neuropsychology, and Cognition. 2015; 22(6):639–666. http://doi.org/10.1080/13825585.2015.1020916.

Bailey HR, Zacks JM, Hambrick DZ, Zacks RT, Head D, Kurby CA, Sargent JQ. Medial temporal lobe volume predicts elders' everyday memory. Psychological Science. 2013; 24(7):1113–1122. http://doi.org/10.1177/0956797612466676. [PubMed: 23630222]

Baldwin DA, Baird JA, Saylor MM, Clark MA. Infants parse dynamic action. Child Development. 2001; 72(3):708–717. [PubMed: 11405577]

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, RHB., Singmann, H., … Grothendieck, G. Package "lme4". 2015. Retrieved from https://cran.r-project.org/web/packages/lme4/lme4.pdf

Boltz M. Temporal accent structure and the remembering of filmed narratives. Journal of Experimental Psychology: Human Perception and Performance. 1992; 18(1):90–105. [PubMed: 1532194]

Bousfield WA. The occurrence of clustering in the recall of randomly arranged associates. The Journal of General Psychology. 1953; 49(2):229–240.

Bousfield WA, Cohen BH. The occurrence of clustering in the recall of randomly arranged words of different frequencies-of-usage. The Journal of General Psychology. 1955; 52(1):83–95.

Bower GH, Black JB, Turner TJ. Scripts in memory for text. Cognitive Psychology. 1979; 11(2):177–220.

Bower GH, Clark-Meyers G. Memory for scripts with organized vs randomized presentation. British Journal of Psychology. 1980; 71(3):369–377.

Bransford JD, Barclay JR, Franks JJ. Sentence memory: A constructive versus interpretive approach. Cognitive Psychology. 1972; 3(2):193–209.

Brewer WF, Dupree DA. Use of plan schemata in the recall and recognition of goal-directed actions. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1983; 9(1):117–129.

Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: A review and quantitative synthesis. Psychological Bulletin. 2006; 132(3):354–380. http://doi.org/10.1037/0033-2909.132.3.354. [PubMed: 16719566]

Chase WG, Simon HA. Perception in chess. Cognitive Psychology. 1973; 4:55–81.

Clark, DPA., Bruno, D. Fit to last: Exploring the longevity of the survival processing effect; The Quarterly Journal of Experimental Psychology. 2015. p. 1-15.http://doi.org/10.1080/17470218.2015.1076864

Craik FIM, Lockhart RS. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior. 1972; 11(6):671– 684.

Crump MJC, McDonnell JV, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS ONE. 2013; 8(3):e57410. http://doi.org/10.1371/journal.pone.0057410. [PubMed: 23516406]

DuBrow S, Davachi L. The influence of context boundaries on memory for the sequential order of events. Journal of Experimental Psychology: General. 2013; 142(4):1277–1286. http://doi.org/10.1037/a0034024. [PubMed: 23957281]

Eisenberg ML, Sargent JQ, Zacks JM. Post-traumatic stress disorder is associated with deficits in event comprehension. Collabra. in press

Ezzyat Y, Davachi L. What constitutes an episode in episodic memory? Psychological Science. 2011; 22(2):243–252. http://doi.org/10.1177/0956797610393742. [PubMed: 21178116]

Ezzyat Y, Davachi L. Similarity Breeds Proximity: Pattern Similarity within and across Contexts Is Related to Later Mnemonic Judgments of Temporal Proximity. Neuron. 2014; 81(5):1179–1189. http://doi.org/10.1016/j.neuron.2014.01.042. [PubMed: 24607235]

Germine L, Nakayama K, Duchaine BC, Chabris CF, Chatterjee G, Wilmer JB. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. Psychonomic Bulletin & Review. 2012; 19(5):847–857. http://doi.org/10.3758/s13423-012-0296-9. [PubMed: 22829343]

Gilewski MJ, Zelinski EM, Schaie KW. The Memory Functioning Questionnaire for assessment of memory complaints in adulthood and old age. Psychology and Aging. 1990; 5(4):482–490. [PubMed: 2278670]

Glenberg AM, Lehmann TS. Spacing repetitions over 1 week. Memory & Cognition. 1980; 8(6):528–538. [PubMed: 7219173]

Gold, DA., Zacks, JM. Facilitating effective event segmentation and memory in younger and older adults. Poster presented at the 26th Annual APS Convention; San Francisco, CA. 2014.

Goodman JK, Cryder CE, Cheema A. Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. Journal of Behavioral Decision Making. 2013; 26(3):213–224. http://doi.org/10.1002/bdm.1753.

Hanson C, Hirst W. On the representation of events: a study of orientation, recall, and recognition. Journal of Experimental Psychology: General. 1989; 118(2):136–147. [PubMed: 2525593]

Hard BM, Recchia G, Tversky B. The shape of action. Journal of Experimental Psychology: General. 2011; 140(4):586–604. http://doi.org/10.1037/a0024310. [PubMed: 21806308]

Hothorn T, Bretz F, Westfall P. Simultaneous Inference in General Parametric Models. Biometrical Journal. 2008; 50(3):346–363. http://doi.org/10.1002/bimj.200810425. [PubMed: 18481363]

Jenkins JJ, Russell WA. Associative clustering during recall. Journal of Abnormal Social Psychology. 1952; 47(4):818– 821.

Jorm AF, Jacomb PA. The informant questionnaire on cognitive decline in the elderly (IQCODE): Sociodemographic correlates, reliability, validity and some norms. Psychological Medicine. 1989; 19(4):1015–1022. [PubMed: 2594878]

Kintsch W, Van Dijk TA. Toward a model of text comprehension and production. Psychological Review. 1978; 85(5):363.

Kurby CA, Zacks JM. Age differences in the perception of hierarchical structure in events. Memory & Cognition. 2011; 39(1):75–91. http://doi.org/10.3758/s13421-010-0027-2. [PubMed: 21264613]

Kurby CA, Zacks JM. Starting from scratch and building brick by brick in comprehension. Memory & Cognition. 2012; 40(5):812–826. http://doi.org/10.3758/s13421-011-0179-8. [PubMed: 22282158]

Kuznetsova, A., Brockhoff, PB., Christensen, RHB. Package "lmerTest". 2014. Retrieved from http://cran.uib.no/web/packages/lmerTest/lmerTest.pdf

Lassiter GD. Behavior perception, affect, and memory. Social Cognition. 1988; 6(2):150–176.

Lassiter GD, Slaw RD. The unitization and memory of events. Journal of Experimental Psychology: General. 1991; 120(1):80–82.

Lassiter GD, Stone JI, Rogers SL. Memorial consequences of variation in behavior perception. Journal of Experimental Social Psychology. 1988; 24(3):222–239.

Lenth RV. Least-Squares Means: The *R* Package **lsmeans**. Journal of Statistical Software. 2016; 69(1) http://doi.org/10.18637/jss.v069.i01.

Lichtenstein EH, Brewer WF. Memory for goal-directed events. Cognitive Psychology. 1980; 12(3): 412–445.

Magliano JP, Zacks JM. The Impact of Continuity Editing in Narrative Film on Event Segmentation: Cognitive Science. Cognitive Science. 2011; 35(8):1489–1517. http://doi.org/10.1111/j.1551-6709.2011.01202.x. [PubMed: 21972849]

Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behavior Research Methods. 2012; 44(1):1–23. http://doi.org/10.3758/s13428-011-0124-6. [PubMed: 21717266]

Migueles M, García-Bajos E. The power of script knowledge and selective retrieval in the recall of daily activities. The Journal of General Psychology. 2012; 139(2):100–113. [PubMed: 24836912]

Nairne JS, Thompson SR, Pandeirada JNS. Adaptive memory: Survival processing enhances retention. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2007; 33(2):263–273. http://doi.org/10.1037/0278-7393.33.2.263.

Naveh-Benjamin M, Craik FIM, Ben-Shaul L. Age-related differences in cued recall: Effects of support at encoding and retrieval. Aging, Neuropsychology, and Cognition (Neuropsychology, Development and Cognition: Section B). 2002; 9(4):276–287. http://doi.org/10.1076/anec.9.4.276.8773.

Newtson D. Attribution and the unit of perception of ongoing behavior. Journal of Personality and Social Psychology. 1973; 28(1):28–38.

Newtson, D. Foundations of attribution: The perception of ongoing behavior. In: Harvey, JH.Ickes, W., Kidd, RF., editors. New directions in attribution research. Vol. 1. Hilldale, N.J: Erlbaum; 1976. p. 223-247.

Newtson D, Engquist G. The perceptual organization of ongoing behavioral. Journal of Experimental Social Psychology. 1976; 12(5):436–450.

Paivio A. On the functional significance of imagery. Psychological Bulletin. 1970; 73(6):385–392.

Pavlik PI, Anderson JR. Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. Cognitive Science. 2005; 29(4):559–586. [PubMed: 21702785]

Polyn SM, Norman KA, Kahana MJ. A context maintenance and retrieval model of organizational processes in free recall. Psychological Review. 2009; 116(1):129–156. http://doi.org/http://doi.org/10.1037/a0014420. [PubMed: 19159151]

Radvansky, GA., Zacks, JM. Event Cognition. New York: Oxford University Press; 2014.

Raymaekers LHC, Otgaar H, Smeets T. The longevity of adaptive memory: Evidence for mnemonic advantages of survival processing 24 and 48 hours later. Memory. 2014; 22(1):19–25. http://doi.org/10.1080/09658211.2013.791321. [PubMed: 23638617]

R Core Team. R: A language and environment for statistical computing (Version 3.0.2). Vienna, Austria: R Foundation for Statistical Computing; 2013. Retrieved from http://www.R-project.org/

Roediger HL, Karpicke JD. Test-enhanced learning taking memory tests improves long-term retention. Psychological Science. 2006; 17(3):249–255. [PubMed: 16507066]

Rosen VM, Caplan L, Sheesley L, Rodriguez R, Grafman J. An examination of daily activities and their scripts across the adult lifespan. Behavior Research Methods, Instruments, & Computers. 2003; 35(1):32–48.

Ross J, Lawrence KA. Some observations on memory artifice. Psychonomic Science. 1968; 13(2): 107–108.

Rubin DC, Umanath S. Event memory: A theory of memory for laboratory, autobiographical, and fictional events. Psychological Review. 2015; 122(1):1–23. http://doi.org/10.1037/a0037907. [PubMed: 25330330]

Sargent JQ, Zacks JM, Hambrick DZ, Zacks RT, Kurby CA, Bailey HR, … Beck TM. Event segmentation ability uniquely predicts event memory. Cognition. 2013; 129(2):241–255. http://doi.org/10.1016/j.cognition.2013.07.002. [PubMed: 23942350]

Saylor MM, Baldwin DA, Baird JA, LaBounty J. Infants' on-line segmentation of dynamic human action. Journal of Cognition and Development. 2007; 8(1):113–128.

Schank, RC., Abelson, R. Scripts, plans, goals, and understanding. Hilldale, N.J: Lawrence Erlbaum Associates; 1977.

Schwan S, Garsoffky B, Hesse FW. Do film cuts facilitate the perceptual and cognitive organization of activitiy sequences? Memory & Cognition. 2000; 28(2):214–223. [PubMed: 10790977]

Schwartz MF, Reed ES, Montgomery M, Palmer C, Mayer NH. The quantitative description of action disorganization after brain damage: A case study. Cognitive Neuropsychology. 1991; 8(5):381–414.

Sonne T, Kingo OS, Krøjgaard P. Occlusions at event boundaries during encoding have a negative effect on infant memory. Consciousness and Cognition. 2016; 41:72–82. http://doi.org/10.1016/j.concog.2016.02.006. [PubMed: 26891192]

Speer NK, Swallow KM, Zacks JM. Activation of human motion processing areas during event perception. Cognitive, Affective, & Behavioral Neuroscience. 2003; 3(4):335–345.

Speer NK, Zacks JM, Reynolds JR. Human brain activity time-locked to narrative event boundaries. Psychological Science. 2007; 18(5):449–455. [PubMed: 17576286]

Sprouse J. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. Behavior Research Methods. 2011; 43(1):155–167. http://doi.org/10.3758/s13428-010-0039-7. [PubMed: 21287108]

Swallow KM, Zacks JM, Abrams RA. Event boundaries in perception affect memory encoding and updating. Journal of Experimental Psychology: General. 2009; 138(2):236–257. http://doi.org/10.1037/a0015631. [PubMed: 19397382]

Tulving E. Subjective organization in free recall of “unrelated” words. Psychological Review. 1962; 69(4):344–354. [PubMed: 13923056]

Tulving E, Pearlstone Z. Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior. 1966; 5(4):381–391.

Van Zomeren AH, Van den Burg W. Residual complaints of patients two years after severe head injury. Journal of Neurology, Neurosurgery & Psychiatry. 1985; 48(1):21–28.

Voss JF, Vesonder GT, Spilich GJ. Text generation and recall by high-knowledge and low-knowledge individuals. Journal of Verbal Learning and Verbal Behavior. 1980; 19(6):651–667.

Waldorff FB, Siersma V, Waldemar G. Association between subjective memory complaints and health care utilisation: a three-year follow up. BMC Geriatrics. 2009; 9(1) http://doi.org/10.1186/1471-2318-9-43.

Whitney C, Huber W, Klann J, Weis S, Krach S, Kircher T. Neural correlates of narrative shifts during auditory story comprehension. NeuroImage. 2009; 47(1):360–366. http://doi.org/10.1016/j.neuroimage.2009.04.037. [PubMed: 19376237]

Wood LE, Pratt JD. Pegword mnemonic as an aid to memory in the elderly: A comparison of four age groups. Educational Gerontology. 1987; 13(4):325–339.

Zacks JM, Braver TS, Sheridan MA, Donaldson DI, Snyder AZ, Ollinger JM, … Raichle ME. Human brain activity time-locked to perceptual event boundaries. Nature Neuroscience. 2001; 4(6):651–655. [PubMed: 11369948]

Zacks JM, Speer NK, Swallow KM, Braver TS, Reynolds JR. Event perception: A mind-brain perspective. Psychological Bulletin. 2007; 133(2):273–293. http://doi.org/10.1037/0033-2909.133.2.273. [PubMed: 17338600]

Zacks JM, Speer NK, Swallow KM, Maley CJ. The brain's cutting-room floor: segmentation of narrative cinema. Frontiers in Human Neuroscience. 2010; 4 http://doi.org/10.3389/fnhum.2010.00168.

Zacks JM, Speer NK, Vettel JM, Jacoby LL. Event understanding and memory in healthy aging and dementia of the Alzheimer type. Psychology and Aging. 2006; 21(3):466–482. http://doi.org/10.1037/0882-7974.21.3.466. [PubMed: 16953710]

Zacks JM, Tversky B. Event structure in perception and conception. Psychological Bulletin. 2001; 127(1):3– 21. [PubMed: 11271755]

Zacks JM, Tversky B. Structuring information interfaces for procedural learning. Journal of Experimental Psychology: Applied. 2003; 9(2):88–100. http://doi.org/10.1037/1076-898X.9.2.88. [PubMed: 12877269]

Zalla T, Sirigu A, Pillon B, Dubois B, Agid Y, Grafman J. How patients with Parkinson's Disease retrieve and manage cognitive event knowledge. Cortex. 2000; 36(2):163–179. [PubMed: 10815704]

Zalla T, Sirigu A, Pillon B, Dubois B, Grafman J, Agid Y. Deficit in evaluating pre-determinated sequences of script events in patients with Parkinson's Disease. Cortex. 1998; 34(4):621–627. [PubMed: 9800095]

Zalla T, Verlut I, Franck N, Puzenat D, Sirigu A. Perception of dynamic action in patients with schizophrenia. Psychiatry Research. 2004; 128(1):39–51. http://doi.org/10.1016/j.psychres.2003.12.026. [PubMed: 15450913]

Zwaan RA. Processing narrative time shifts. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1996; 22(5):1196–1207.

Zwaan RA, Langston MC, Graesser AC. The construction of situation models in narrative comprehension: An event-indexing model. Psychological Science. 1995; 6(5):292–297.

Figure 1.

Still frames taken from three of the experimental movies: making breakfast, decorating for a party, and gardening.

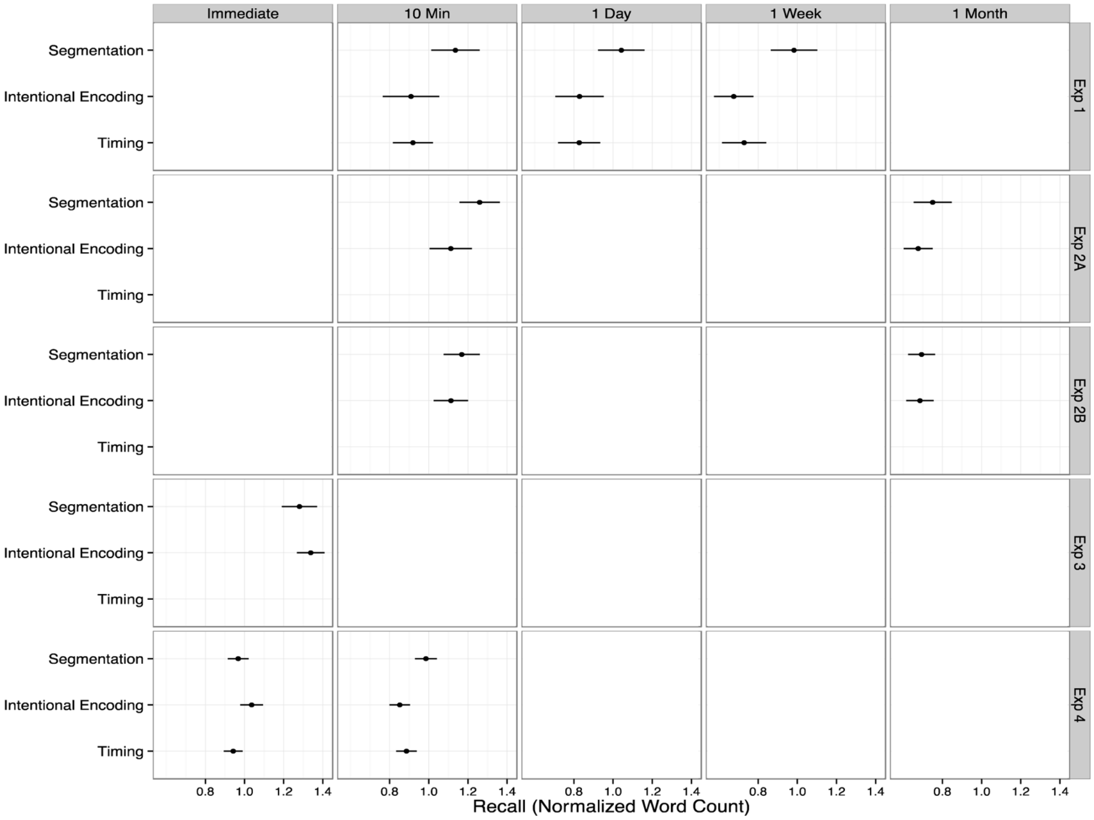

Figure 2.

Mean recall memory performance for task groups at each delay for all six experiments. Segmentation during encoding resulted in better recall performance in the 10 min, 1 day, 1 week and 1 month delay conditions. In the presence of an immediate delay, the intentional encoding group recalled more or an equivalent amount of information than the segmentation group. Error bars are 95% confidence intervals.

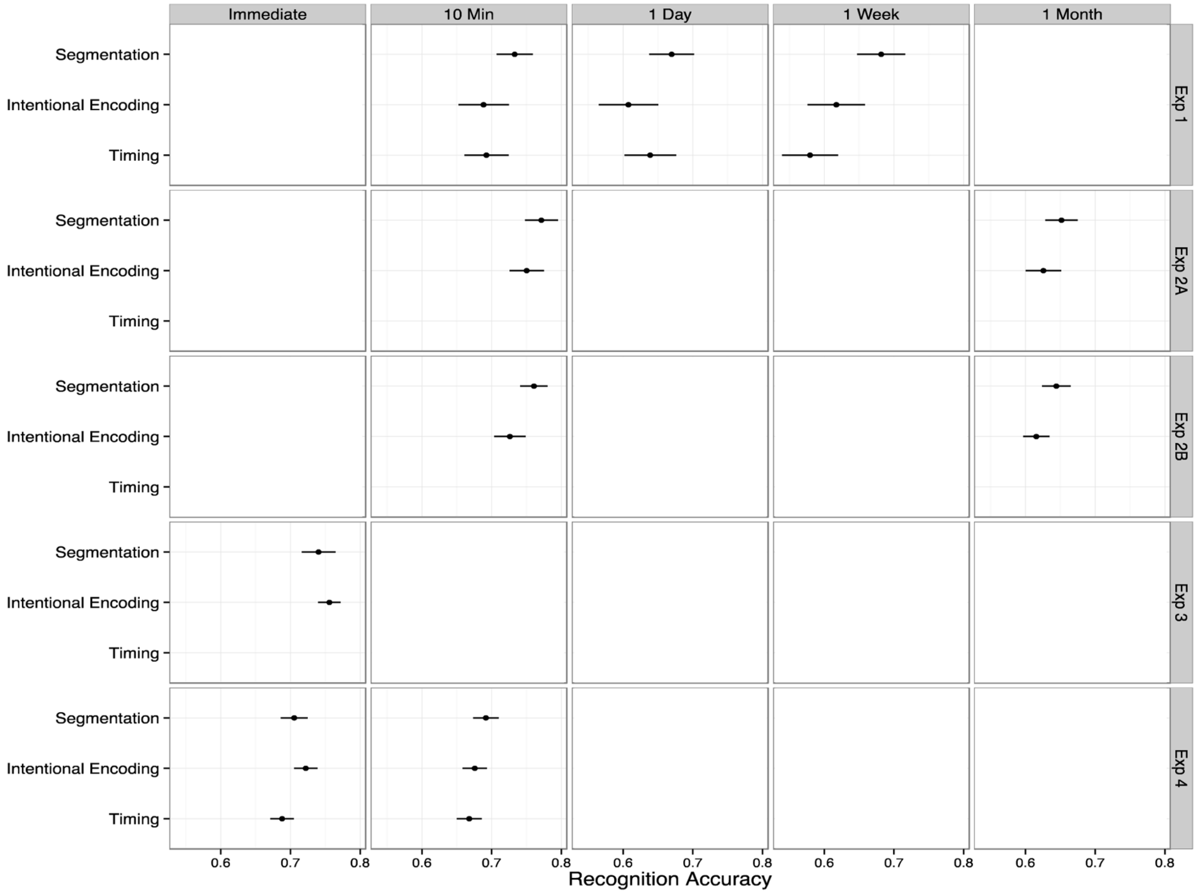

Figure 3.

Mean recognition accuracy for task groups at each delay for all six experiments. Segmentation during encoding resulted in better recognition accuracy in the 10 min, 1 day, 1 week and 1 month delay conditions. In the immediate delay condition, the intentional encoding group had higher recognition accuracy only. Error bars are 95% confidence intervals.

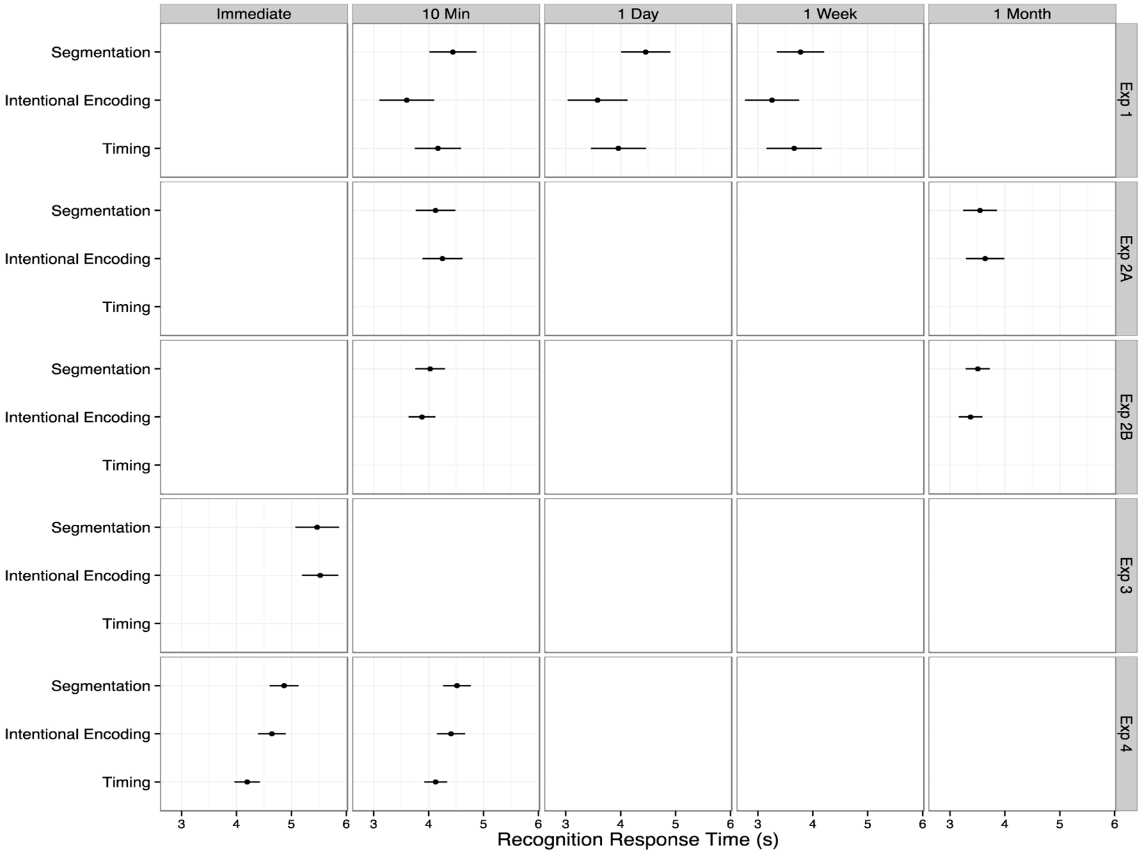

Figure 4.

Mean trimmed recognition response times for task groups at each delay for all six experiments. In over half of the experiments, no significant differences in recognition response time were found among the task groups. In Experiment 1, the segmentation group took significantly longer to respond while the timing group was significantly faster to respond in Experiment 4. Error bars are 95% confidence intervals.

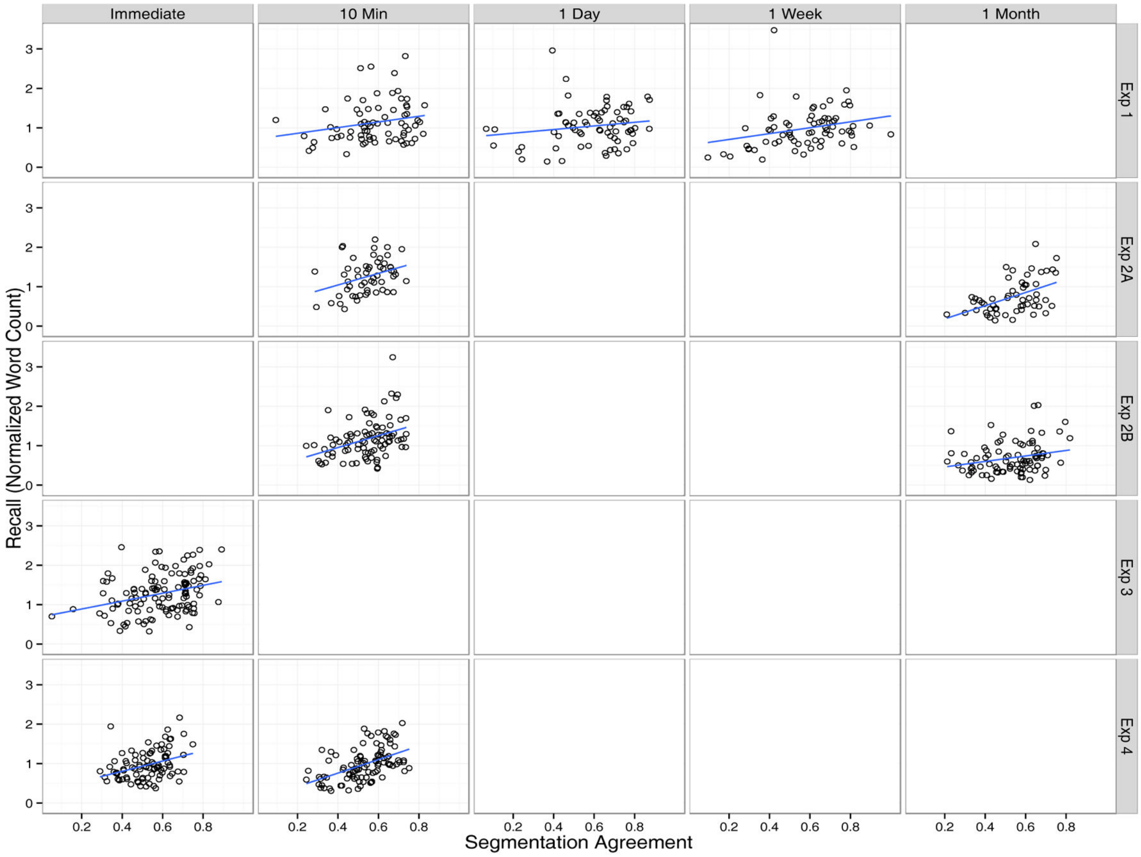

Figure 5.

Relationship between segmentation agreement and recall memory at each delay for all six experiments. In all experiments segmentation agreement predicted recall performance. All plots reflect mean normalized word count for a participant at each delay with a 95% confidence interval for the regression line.

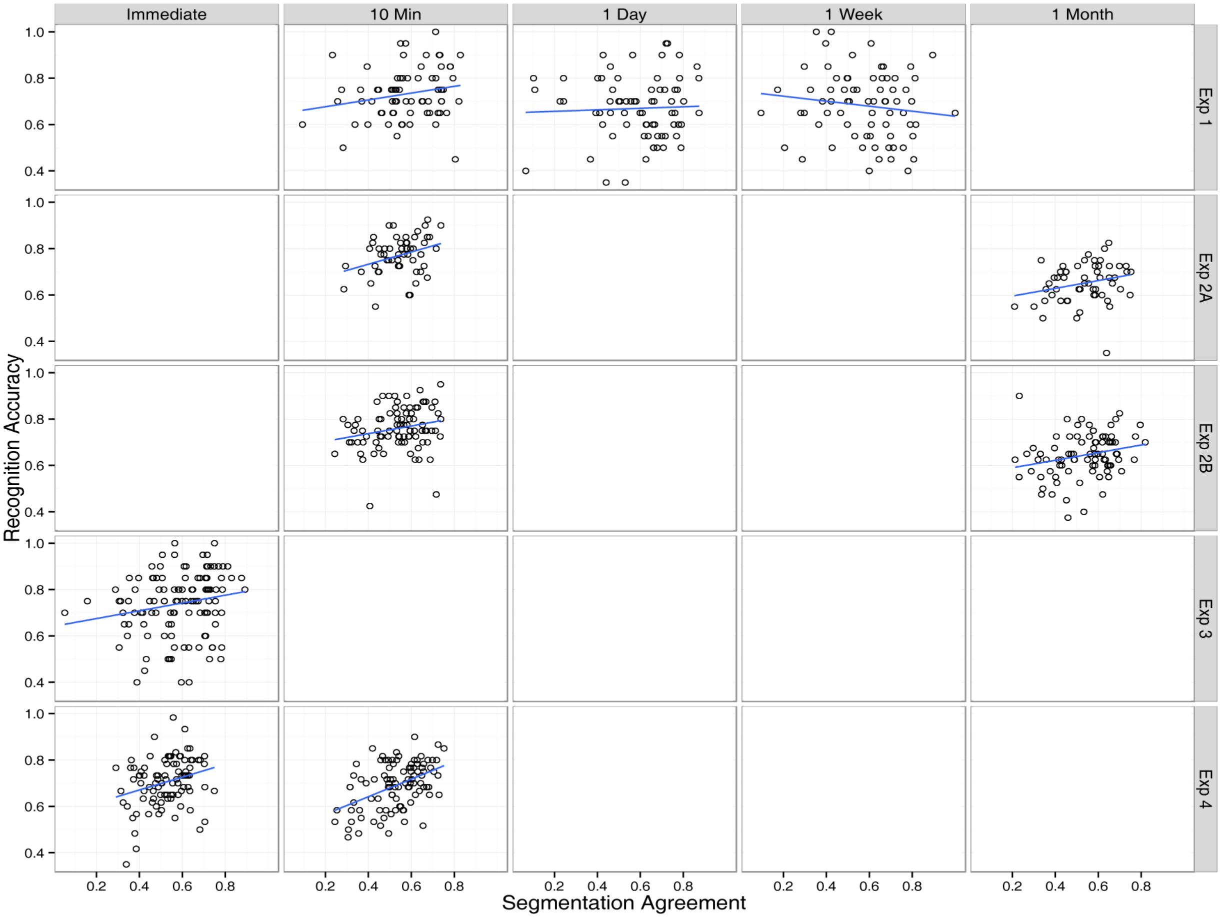

Figure 6.

Relationship between segmentation agreement and recognition accuracy at each delay for all six experiments. The relationship between segmentation agreement and recognition accuracy is mixed. Experiments 2B and 4 showed a significant relationship between segmentation agreement and recognition performance while Experiments 1, 2A and 3 did not. All plots reflect mean recognition accuracy for a participant at each delay with a 95% confidence interval for the regression line.

Table 1

Demographic information for each participant group.

	Condition	N	Number Female	Median Age (yrs)	Age Range (yrs)
Experiment 1	Segmentation	243	140	31	18 – 66
	Intentional Encoding	84	43	29	18 – 61
	Timing	155	83	31	20 – 63
Experiment 2a	Segmentation	260	172	31.5	18 – 65
	Intentional Encoding	196	123	32	19 – 77
Experiment 2b	Segmentation	287	191	31	18 – 75
	Intentional Encoding	288	178	30	18 – 77
Experiment 3	Segmentation	140	90	30	18 – 65
	Intentional Encoding	210	124	29	18 – 67
Experiment 4	Segmentation/Immediate	135	70	32	18 – 70
	Intentional Encoding/Immediate	106	63	30	18 – 68
	Timing/Immediate	136	83	31.5	18 – 69
	Segmentation/10 Minute Delay	142	93	32	20 – 66
	Intentional Encoding/10 Minute Delay	119	76	35	18 – 65
	Timing/10 Minute Delay	138	96	30	18 – 66

Table 2

Descriptive Statistics for Number of Button Presses Identified during Encoding Task by Group

	Segmentation	Timing
Exp. 1		
Range	1 – 107	13 – 37
Mean Number of Boundaries	31.35	22.05
Median Number of Boundaries	24	22
Exp. 2A		
Range	3 – 67	
Mean Number of Boundaries	16.55	
Median Number of Boundaries	14	
Exp. 2B		
Range	3 – 76	
Mean Number of Boundaries	17.73	
Median Number of Boundaries	14	
Exp. 3		
Range	6 – 275	
Mean Number of Boundaries	32.53	
Median Number of Boundaries	20	
Exp. 4		
Range	1 – 117	7 – 49
Mean Number of Boundaries	23.36	23.11
Median Number of Boundaries	16	23

Table 3

Pairwise Comparisons Between Tasks at Each Delay.

			Recall			Recognition Accuracy			Recognition Response Time		
	Delay	**Conditions compared**	**df**	**t**	***p***	**df**	**t**	***p***	**df**	**t**	***p***
Experiment 1	10 Minute	Segmentation vs. Intentional	172	3.154	0.00525	172	2.69	0.0213	172	2.225	0.0697
		Segmentation vs. Timing	172	2.857	0.01333	172	2.129	0.0869	172	0.684	0.7729
		Timing vs. Passive	172	−0.322	0.94451	172	−0.561	0.8408	172	−1.464	0.3105
	1 Day	Segmentation vs. Intentional	172	2.43	0.0424	172	2.524	0.0333	172	2.742	0.0183
		Segmentation vs. Timing	172	2.717	0.0197	172	1.266	0.4162	172	1.527	0.2805
		Timing vs. Passive	172	0.236	0.9697	172	−1.198	0.4557	172	−1.161	0.478
	1 Week	Segmentation vs. Intentional	172	3.938	< 0.001	172	2.216	0.071263	172	2.131	0.0866
		Segmentation vs. Timing	172	3.166	0.00519	172	3.881	0.000417	172	0.511	0.8657
		Timing vs. Passive	172	−0.763	0.72583	172	1.518	0.284829	172	−1.528	0.2802
Experiment 2a	10 Minute	Segmentation vs. Intentional	118	1.495	0.138	117	1.172	0.244	118	−0.433	0.666
	1 Month	Segmentation vs. Intentional	118	1.22	0.225	117	1.517	0.132	118	−0.317	0.752
Experiment 2b	10 Minute	Segmentation vs. Intentional	180	0.587	0.558	180	2.046	0.0422	180	0.664	0.508
	1 Month	Segmentation vs. Intentional	180	0.413	0.68	180	1.898	0.0594	180	0.717	0.475
Experiment 3	Immediate	Segmentation vs. Intentional	300	−1.16	0.247	300	−1.083	0.28	300	−0.094	0.925
Experiment 4	Immediate	Segmentation vs. Intentional	288	−1.23	0.436	288	−1.15	0.4843	288	0.884	0.6507
		Segmentation vs. Timing	288	0.455	0.892	288	1.205	0.4509	288	2.66	0.0224
		Timing vs. Passive	288	1.704	0.205	288	2.383	0.0467	288	1.801	0.171
	10 Minute	Segmentation vs. Intentional	281	2.426	0.0419	281	1.17	0.472	281	0.456	0.892
		Segmentation vs. Timing	281	1.782	0.1776	281	1.742	0.191	281	1.609	0.244
		Timing vs. Passive	281	−0.613	0.813	281	0.58	0.831	281	1.149	0.485