bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Event segmentation reveals working memory forgetting rate

Anna Jafarpour [1,*], Elizabeth A Buffalo [1,2], Robert T Knight [3,4,5], Anne GE Collins [3,5]

1- University of Washington, Department of Physiology and Biophysics, Seattle, WA
2- National Primate Center, University of Washington, Seattle, WA
3- University of California, Department of Psychology, Berkeley, CA
4- University of California, Department of Psychology, San Francisco, CA
5- Helen Wills Neuroscience Institute, University of California, Berkeley, CA

* Corresponding author: Anna Jafarpour, Department of Physiology and Biophysics, University of Washington, 1705 NE Pacific St., HSB Box 357290, Seattle, WA 98195; phone: +1 (206) 543-4372, annaja@uw.edu

1. Abstract

We perceive the world as a series of events and fluidly segment them into episodes. Although individuals generally agree on the segmentation at the occurrence of a salient event, the number of determined segments is variable. Working memory plays a key role in tracking and segmenting a sequence of events; however, it is unclear which aspect of working memory is related to individual variability in event segmentation. We used computational modeling to extract the working memory capacity and forgetting rate of healthy adults (n=36) from an association learning task, and we studied a link between individuals' working memory limitations and the subjective number of determined events in three movies with different storylines. We found that memory decay, measured in the learning task, is related to event segmentation: Participants who perceived either a very low (under-segmenters) or a very high (over-segmenters) number of events had a higher forgetting rate. We observed that under-segmenters performed better on a temporal recognition task for the movie with a linear storyline and an overarching story, benefiting from the schema. In contrast, the over-segmenters performed better at free recall than under-segmenters for all the movies. The results provide evidence that variability in forgetting rate is linked to the variability in event perception.

2. Introduction

We experience the world as a series of events and perceive events as segments (Zacks et al., 2006; Zacks and Swallow, 2007; Zacks et al., 2007). Every individual's experience is different in terms of how this flow of events is segmented. Segmentation occurs with perceiving a salient event - e.g., a change in the situation, the characters, or the goal (Jafarpour et al., 2019; Radvansky and Zacks, 2017). According to the current model of event segmentation (Zacks et al., 2007), the working memory system tracks events to help define new event boundaries.; segmentation is determined by attention (Bailey et al., 2017) and expectations (Hymel et al., 2016). However, the link between working memory limitations and event segmentation has not been directly shown.

Working memory is limited in two ways: the amount of information retained, and the duration of maintenance of this information (Baddeley, 2003, 2012; Bays and Husain, 2008; D'Esposito and Postle, 2015). These two limitations are often measured as working memory capacity, which accounts for how many items can be held in memory (Baddeley, 1992; Bays et al., 2009; Cowan,

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2010; Vogel and Machizawa, 2004), and the forgetting rate, that limits how long the information persists in the face of interference (Baddeley, 2012; Collins and Frank, 2012). A challenge in measuring these limitations is that other cognitive systems such as attention can contribute to the accuracy of short term memory (Jafarpour et al., 2017; Rose et al., 2016; Zokaei et al., 2014).

Collins and Frank (2012) designed an association learning task that disentangled the working memory and learning systems and enabled the identification of both limitations of working memory (capacity and decay). In each learning block of the task, the stimulus set size was constant and the set size varied across the learning blocks. This design allowed for investigating the working memory forgetting rate and capacity separately. They showed that increasing the number of stimulus-action associations to be learned (the set size) slowed the learning process, since the working memory system became less successful at contributing to the tracking of large sets of stimulus-action associations (Collins, 2018). In addition, associations learned quickly within a small set size were less durable and decayed more rapidly (Collins, 2018).

Here, we examined the relationship between event segmentation and working memory characteristics. Healthy participants watched three movies with different storylines and performed a source memory test. Following this, they performed the association learning task (Collins and Frank, 2012). Finally, the participants watched the movies again and were asked to segment the movies, determining when new events occurred. We applied a reinforcement learning and working memory model (RLWM) to performance on the association learning task to estimate each participant's working memory capacity and forgetting rate based on Collins et al. (2017). We determined the cross-task relationship between the variation in the number of determined events in the movies with working memory characteristics as estimated from the learning task.

3. Methods and Material

3.1. Participants

36 healthy adults (25 female) were recruited through the online University of Berkeley Psychology Department Research Participation Program. Participants provided informed consent and were compensated ($36 or 3 course-credits). The Office for the Protection of Human Subjects of the University of California, Berkeley approved the study protocol. The mean age was 20.3 (SD = 1.9) and ranged from 18 to 27. All were all right-handed by self-report. We discarded one participant because she identified two standard deviations more events than the group average.

3.2. Experimental design

The experiment ran on a desktop PC and a standard TFT monitor, in a sound-attenuated recording room. It consisted of four parts. First, participants watched three mute animations (each ~3 minutes long) with differing storylines (Figure 1A, see the Supplementary material for the details about the animations). They then performed 35 recognition memory tests per movie. At each test, subjects saw two scenes from a movie - located on the left and right sides of the screen for 2 seconds. Then a prompt appeared asking the participant to indicate which of the two scenes happened 'earlier' or 'later' (Figure 1B). Participants used left or right arrow keys to respond.

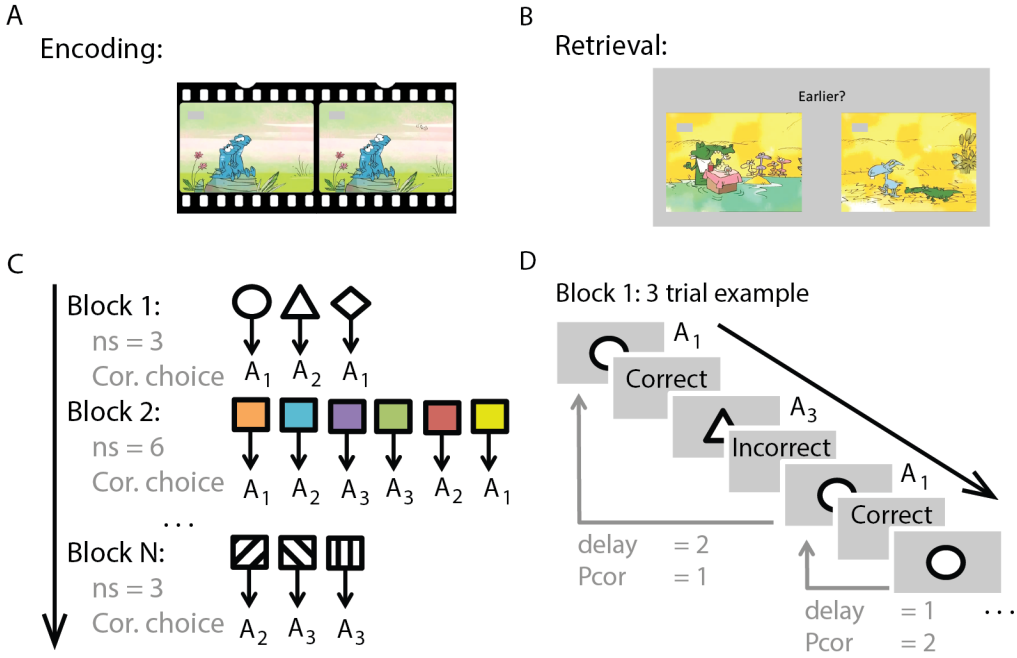
We then ran a version of an association learning task (Collins et al., 2017) to evaluate working memory characteristics, namely working memory capacity and forgetting rate. In this task, participants used their dominant hand to select from three possible actions (J, K, and L from a

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

keyboard) when they saw an image. They used trial and error to learn the correct image-action association. The probability of an action being paired with an image was equal (1/3); thus, an action could pair with more than one stimulus. Participants learned the associations in 12 repetitions of each stimulus; the repetitions were pseudo-randomly interleaved. This procedure repeated in a block-design and included 22 blocks (3 blocks of set sizes 6, 5, and 4; 4 blocks of set size 3, and 6 blocks of set size 2). The stimulus set-size varied in each block to manipulate the requirement for capacity-limited and delay-sensitive working memory (Figure 1C). Participants studied the whole stimuli set at the beginning of each block.

The reward value for a correct response differed across stimuli; an incorrect response yielded no reward. Only one action for a stimulus was correct and each correct stimulus–action association was assigned a probability (p) of yielding a 2-point versus a 1-point reward, and this probability was either high (p = .80), medium (p = .50), or low (p = .20). We counter-balanced the p within participant and blocks to ensure an equal overall value of different set sizes and actions. Participants had 1.4 s to respond. The feedback was displayed for 0.5 s. There was an inter-stimulus interval of 0.5 - 0.8 s in which a fixation cross was shown.

After performing the association learning task, participants watched the movies again; this time they segmented the movies by pressing a space-bar to indicate the start of a new event. We instructed them to press a key "whenever something new happened." We told the participants: "we want to segment this movie into episodes", and we explained that segmentation could occur as often as they liked. The consistency of frequency of event segmentation for each participant was determined across the three movies. Finally, participants performed a surprise free recall test, where they were asked to write a paragraph describing the story of each movie. Participants were instructed not to worry about grammar and wording - simply “write what came to their mind.” The free recall test allowed us to investigate the relationship between individual difference in event segmentation and subsequent memory performance.



bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1. Experimental design: The experiment consisted of four parts; the first two of which are depicted here. Temporal memory test: (A) At encoding, participants watched three mute movies. (B) At retrieval, participants were shown two frames of a movie and determined the temporal order of two movie frames by pressing the left or right key. There were 35 questions about the temporal order of events per movie. Association learning task: (C) Participants performed a block-design association learning task. In each block, participants learned by trial and error the association between a set of images and three possible actions (A_1, A_2, and A_3); feedback was provided. The set size at each block was different (ranging from 2 to 6). (D) A 3-trial example of a learning block. Delay quantifies the number of intervening trials from the last time the stimulus was encountered, and Pcor quantifies the number of trials that the choice was correct.

3.3. Analysis details

3.3.1. Association learning

We analyzed the association learning task in two ways, without modeling and with reinforcement learning and working memory (RLWM) modeling, consistent with previously published studies (Collins, 2018; Collins and Frank, 2018; Collins et al., 2014, 2017). Trials with missed responses or with less than 200 ms response time were discarded. To generate learning curves, we analyzed the proportion of correct choices as a function of the number of iterations (how many times the stimulus was encountered) and set size. Next, we used a logistic regression to evaluate the performance with respect to three parameters - the set size (number of stimuli in a block), delay (number of trials since the last correct choice for the current stimulus), and previously correct answers (number of correct choices made so far for the current stimulus) - and their interactions. We quantified the effect of working memory and learning on a trial-by-trial basis by modeling the probability of a correct choice for each participant as a function of the three parameters: set size, the number of previously correct answers, and delay (see Collins et al., 2017 for details).

3.3.2. Movie segmentation

We evaluated the keypresses during the movies and discarded any key presses that occurred less than 100 ms from the previous key press to remove any multi-clicks. Then we quantified the number of events by counting key presses for each movie. A ranked (Spearman) correlation was utilized to identify the consistency of individual differences across the movies.

3.4. Modeling

We fit three models to the trial-by-trial responses for each subject: two-parameters reinforcement learning (RL2), four-parameters reinforcement learning (RL4), and a modified RLWM model (Collins et al., 2017). We used the Akaike Information Criterion (AIC) to select the best model considering the number of parameters used in each model. RLWM was used as a baseline. We simulated data based on the models for validation.

3.4.1. Two-parameters Reinforcement Learning (RL2)

The basic model was a reinforcement learning model (without the working memory component) with a delta rule learning. For each stimulus, s, and action, a, the expected reward was $Q(s,a)$, and the Q value was updated with observing feedback, r_t, through time. The Q values were updated based on a learning rate, α, and the difference between expected and observed reward

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

at trial t (known as the prediction error: $\delta_t = r_t - Q_t(s,a)$): $Q_{t+1}(s,a) = Q_t(s,a) + \alpha \times \delta_t$. Choosing an action utilized the expected reward value. An action was probabilistically chosen, with a greater likelihood of selecting an action that had a higher Q value, using the SoftMax choice rule:$P(a|s) = e^{\beta Q(s,a)} / \sum_i (e^{\beta Q(s,a_i)})$, where β is an inverse temperature free parameter. This model had two parameters of α and β.

3.4.2. Four-parameters Reinforcement Learning (RL4)

This model in addition to RL2 includes a value for unrewarded correct responses and undirected noise in action selection. In this experiment, a correct response was sometimes rewarded and sometimes not rewarded. We estimated how much a person valued a correct response, irrespective of the reward by estimating the value for correct-but-not-rewarded items, i.e. r_0. The model also considered an undirected noise, $0 < \epsilon < 1$, in the stochastic action selection, to allow for choosing an action that did not have the highest Q value. Accordingly, $P = (1-\epsilon) \times P + \epsilon \times \frac{1}{n_a}$, where $1/n_a$ is a uniform probability of choosing an action.

3.4.3. Reinforcement learning and working memory (RLWM)

We applied RLWM to extract the working memory capacity and forgetting rate of the participants (Collins et al., 2014). Like Collins et al., (2017), this model had 8 parameters and consisted of two components. A working memory component with limited working memory *capacity, C*, and *forgetting rate,* $\emptyset_{WM}$. The Q value was subject to decay with a forgetting rate, $0 < \phi < 1$, so for all the stimuli that are not current, $Q = Q + \phi(Q_0 - Q)$, where $Q_0 = \frac{1}{n_s}$.

The RL component had a *learning rate,* α, value for an unrewarded correct response, r_0, undirected noise, ϵ, and a *forgetting rate,* $\emptyset_{RL}$ (β was set constant at 100). We also allowed for the potential lack of an impact of negative feedback ($\delta < 0$) by estimating a preservation parameter, $pers$. In that case, the learning rate is reduced by $\alpha = (1 - pers) \times \alpha$. Accordingly, $pers$ near 1 indicated lack of an impact of negative feedback (learning rate close to 0; high preservation of Q value), and $pers$ close to 0 indicated equal learning rate for positive and negative feedback.

The WM component was simulated as encoding of stimulus in a Q learning system, like the RL component but the outcome, r_t, was 1 for correct, 0 for incorrect (rather than the observed reward), the learning rate was set to 1 ($\alpha = 1$), and at most C stimuli could be remembered. We formulated the probability of a stimulus being in working memory as:

If $r_t = 1$, $P_{WM}(r_t|s_t,a_t) = \min\left(1, \frac{C}{n_s}\right) \times Q_{wm}(s_t,a_t) + (1 - \min\left(1, \frac{C}{n_s}\right)) \times 1/n_a$,

if the $r_t = 0$, $P_{WM}(r_t|s_t,a_t) = \min\left(1, \frac{C}{n_s}\right) \times (1 - Q_{wm}(s_t,a_t)) + (1 - \min\left(1, \frac{C}{n_s}\right)) \times 1/n_a$

, where n_a is the number of possible actions (= 3). In RL case,

if the $r_t > 0$, $P_{RL}(r_t|s_t,a_t) = Q_{RL}(s_t,a_t)$

if the $r_t = 0$, $P_{RL}(r_t|s_t,a_t) = 1 - Q_{RL}(s_t,a_t)$.

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A *mixture weight*, w_0, formulated how much each of the components was used for action selection. The weight was $w_0 \times \min\left(1, \frac{C}{n_s}\right)$ to represent the confidence in WM efficiency. This initialization reflects that participants are more likely to utilize WM when the stimulus set size is low. The overall policy was:

$P(a|s) = w_t(s) \times P_{WM}(a|s) + (1 - w_t(s)) \times P_{RL}(a|s)$.

A Bayesian model averaging scheme inferred the relative reliability of WM compared with the RL system over time, *t*:

$$w_{t+1}(s) = \frac{P_{WM}(r_t|s_t, a_t)\, w_t(s)}{P_{WM}(r_t|s_t, a_t)\, w_t(s) + P_{RL}(r_t|s_t, a_t)\,(1 - w_t(s))}$$

, where P_{WM} is the probability that action a is selected for stimulus s according to the WM component at time t and P_{RL} is the probability of action selection according to the RL component. We assumed that although the w_0 is the same for all stimuli, the development of mixture weight over time would be different for each stimulus because the probability of retaining a stimulus in working memory or another retention system is not equal (Jafarpour et al., 2017).

3.5. Cross-task comparison

We determined the correlation between the logistic regression-estimated Beta values (from the trial-by-trial modeling of choices) for the three parameters - delay, number of previously correct answers, and set size - and their interactions and the number of events segmented. This analysis led to 6 correlations. We adjusted the significance level using Bonferroni correction to 0.0083. This analysis was replicated for the three movies. Then, we evaluated a relationship between the number of events and working memory capacity and forgetting rate that were extracted by the WMRL model using a linear and quadratic model fitting. The models were compared using AIC. We report the difference between the linear and quadratic AICs – larger positive difference means that the quadratic model was a better fit, despite the increased number of parameters. We hypothesized that the number of events would reveal the estimated working memory span. We also hypothesized that the undirected noise in the association learning task would predict the participants' general performance, so those with higher noise in RLWM system would demonstrate lower performance in the memory test.

Because we found a U-shaped relationship between the number of segmented events and working memory forgetting rate, we ran a post-hoc analysis on the two extreme groups who had very large (one standard deviation more than the mean) or very low segmentation numbers (one standard deviation less than the mean). We refer to these participants as under- and over-segmenters. We hypothesized that the under-segmenters and over-segmenters used different long-term memory strategies which each led to a fast forgetting rate. We examined the temporal order recognition performance for the movie with a strong overarching cliché story (movie 3) to determine which group relied on the story structure for memory performance, and we counted the number of written words in the free recall task to determine which group recalled more information about the movies.

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4. Results

4.1. Event segmentation task

We first examined individual differences in perception of event boundaries. We tested whether individual variability in event segmentation holds for movies irrespective of the movie storyline by asking participants to segment three different movies (see the Supplementary material for the stories). The consistency in segmenting movies was determined using a ranked correlation. The number of segmented events (NE) between each pair of movies was strongly correlated (Figure 2A, Spearman ranked correlation; movie 1 vs. movie 2: $r = 0.85$, $p < 0.001$; movie 2 vs. movie 3: $r = 0.86$, $p < 0.001$; movie 1 vs. movie 3: $r = 0.71$, $p < 0.001$). One participant segmented the movies more than two standard deviations higher from the average, and we excluded this participant's data from further analysis (included in Figure 2A). However, all results hold when including this participant (the estimated working memory forgetting rate was 0.64). On average participants segmented movie 1 into 13.45 events (SD = 8.86), movie 2 into 9.78 events (SD = 5.8), and movie 3 into 11.27 (SD = 8.01).

Participants accurately recalled the order of movie scenes for all three movies (movie 1: M = 81.30%, SD = 9.09; movie 2: 68.65%, SD = 8.47; movie 3: M = 81.68%, SD = 7.43; Figure 2B; chance was 50%). No one performed below chance, and performance was distributed normally. Paired-sample t-tests showed that participants remembered the order of events for the movies with overarching storylines (1 and 3) better than the movie without an overarching story (movie 2; movie 1 vs 2: $t(34) = 7.85$, $p < 0.001$; movie 2 vs 3: $t(34) = 10.7$, $p < 0.001$; movie 1 vs 3: $t(34) = 0.32$, $p = 0.74$; Figure 2B). We also examined if the observation, that a linear storyline helped with temporal order memory, holds for movie 1 when we split the tests into those with interchangeable order (interchanging the order of the epochs does not hurt the overarching story) and those with an overarching story. Consistent to the comparison between movie 2 and movie 3, the temporal memory for the overarching story of movie 1 (M = 82.18%, SD = 7.29) was better than for interchangeable epochs of movie 1 (M = 73.88%, SD = 17.14; $t(34) = 3.18$, $p = 0.003$).

The response time for movies 1, 2 and 3 did not differ (movie 1 vs 2: $t(34) = 0.73$, $p = 0.46$; movie 1 vs 3: $t(34) = 1.27$, $p = 0.21$; movie 2 vs 3: $t(34) = 0.65$, $p = 0.51$). The tested movie scenes were between 1.6 second to 60 seconds apart for movie 1, 1.6 second to 62 seconds for movie 2, and 1.6 second to 48 seconds for movie 3 (for movie 1: the mean probe distance was 18.84 seconds, SD = 18.31; for movie 2: M = 16.66, SD = 17.27; for movie 3: M = 15.02, SD = 13.22). In general, the number of segmented events did not predict participants' accuracy in temporal source memory (movie 1: linear regression $F(1,34) = 1.1$, $p = 0.3$; movie 2: $F(1,34) = 0.38$, $p = 0.53$; movie 3: $F(1,34) = 0.82$, $p = 0.37$).

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

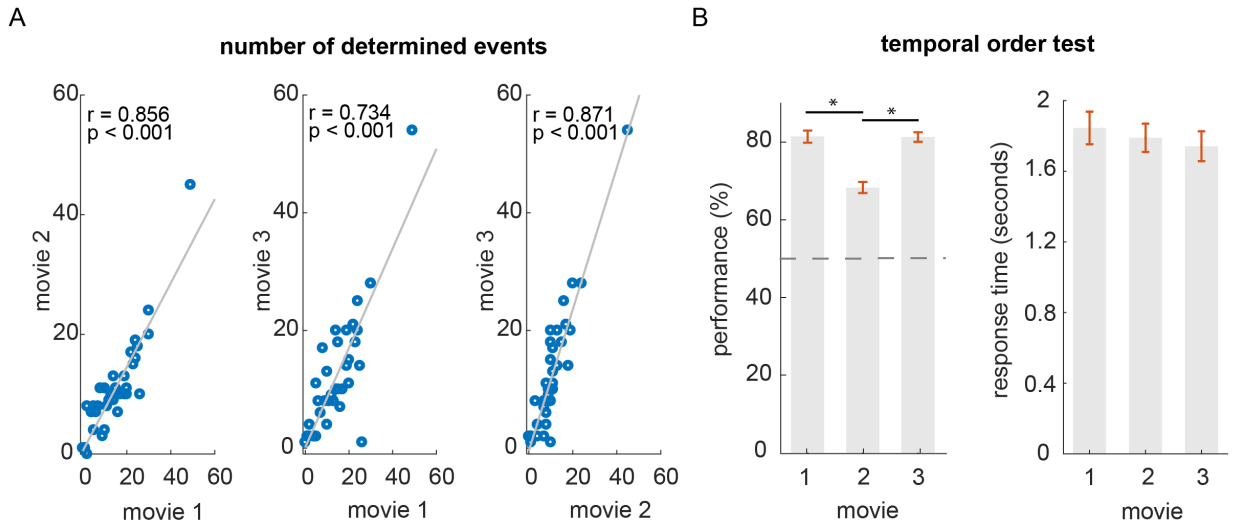


Figure 2. Event segmentation results: (A) The number of events that participants segmented was constant across the movies. Correlation between the number of events for movie 1 and 2 (left), for movie 1 and 3 (middle), and for movie 2 and 3 (right). Each dot represents an individual participant; r values describe the ranked correlation coefficient. (B) Recall performance was above chance for all three movies (left). The performance for movie 2 was less accurate than for movie 1 and 3 (* $p < 0.05$). Response times were not different across movies (right). Error bars represent the standard error of the mean.

4.2. Association learning task

Limitations in working memory affected learning performance as hypothesized. Overall, participants learned the stimulus-action associations. For all set sizes, the accuracy of the last two iterations was on average more than 90% (M= 93.5%, SD = 3%; Figure 3A); however, the accuracy decreased with increasing set size ($r = -0.94$, $p = 0.018$). We analyzed the trial-by-trial performance with respect both to set size and to maintenance of correct associations across intervening trials of a particular stimulus. The result of a logistic regression revealed that performance was reduced with increasing set size ($t(34) = -10.12$, $p < 0.001$) and delay (the number of trials since a correct response to the current stimulus ($t(34) = -8$, $p < 0.001$). By contrast, performance was improved with increasing total number of previously correct responses to the current stimulus ($t(34) = 5.95$, $p < 0.001$). The interactions between set size and delay ($t(34) = -8.29$, $p < 0.001$), set size and previous correct responses ($t(34) = 4.19$, $p < 0.001$), and delay and previous correct responses ($t(34) = 5.18$, $p < 0.001$) also affected the performance (Figure 3B), consistent with previous studies (Collins et al., 2017).

The RLWM model provided the best fit to the data, and model simulations with fit parameters were able to capture the overall learning pattern (Figure 3A). The AICs of model fitting using RL2, RL4, and RLWM models were subtracted and divided by the AIC of RLWM (RL2: M = 7.4, SD = 1.68 and RL4: M = 0.06, SD = 0.03; note that smaller AIC reflects better fit; Figure 3C). Pairwise t-test showed that RL4 was a better model than RL2 ($t(34) = -25.99$, $p < 0.001$), and RLWM provided the best fit (RL2: $t(34) = 25.9$, $p < 0.001$; RL4: ($t(34) = 9.63$, $p = 0.03$).

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

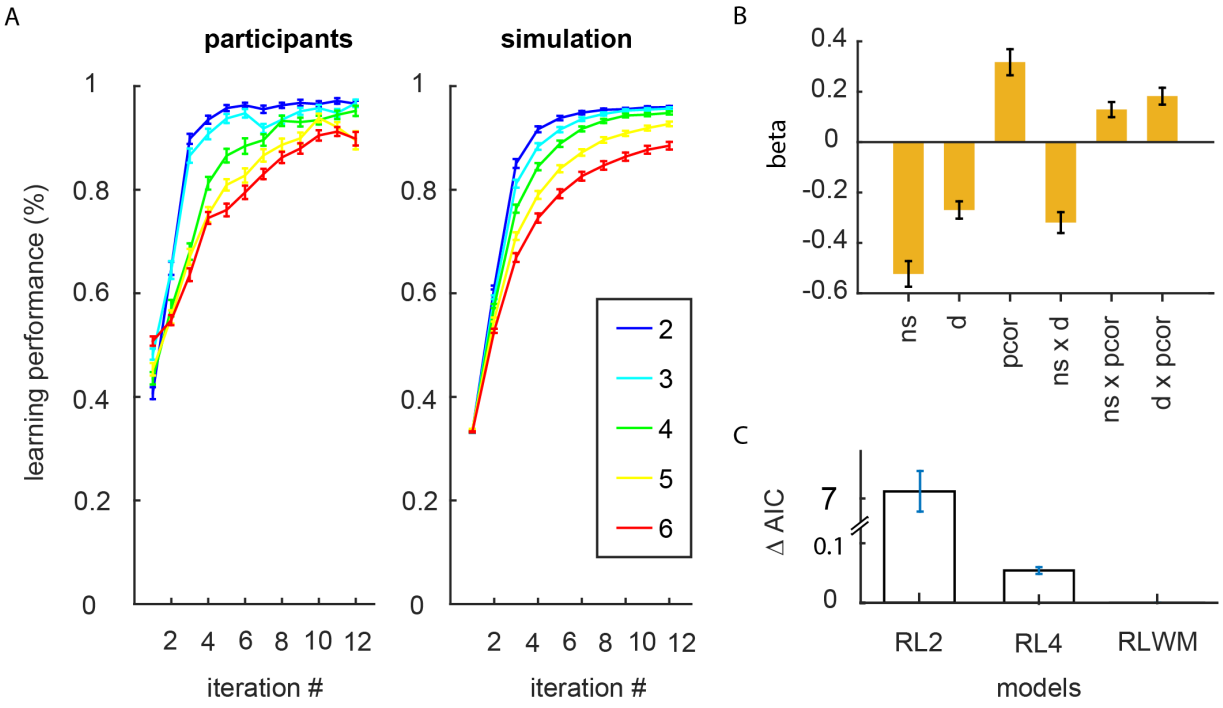


Figure 3. Association learning results: (A) Participants' learning performance (in left) and the RLWM simulation (in right). Learning was slower with increasing set size (ranging from 2 to 6). (B) The stimulus set size (ns), number of trials since the last correct answer (delay; d), number of previously correct responses (pcor), and their paired interactions affected the trial-by-trial performance in the learning task. (C) The AIC relative to the RLWM shows that the RLWM is a better fit than the RL4 and RL2 models. Error bars represent the standard error of the mean.

4.3. Cross-task comparison

We observed a linear correlation between the number of determined events and the beta-value for the interaction between set size and previously correct responses (movie 1: ranked r = 0.31, p = 0.064; movie 2: ranked r = 0.45, p = 0.006; movie 3: ranked r = 0.32, p = 0.057; Figure S1). This interaction indicates that learning is slower with higher set sizes, revealing a working memory limitation on association learning. However, the beta value does not clarify what aspect of working memory is correlated with the number of segmented events. Using the RLWM model we inferred the parameters related to working memory to systematically test for the relationship between working memory and number of segmented events.

We found that the estimated noise in the association learning task correlated with the individual's performance on the movie temporal recognition test. Participants with higher estimated undirected noise performed worse in the temporal recognition tests for all the movies (movie 1: r = -0.31, p = 0.063; movie 2: r = -0.35, p = 0.034; movie 3: r = -0.43, p = 0.009; Figure 4A). On average, the estimated noise was 0.057 (SD = 0.034).

The results also showed a U-shaped relationship between the number of segmented events and WM forgetting rate (Figure 4A). The number of events (NE) in all the movies predicted WM forgetting rate. (Figure 4A; for movie 1, the linear component (NE): p = 0.005, the quadratic

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

component (NE^2): p = 0.005; movie 2, NE: p = 0.003, NE^2: p = 0.004; movie 3: NE: p = 0.014, NE^2 , p = 0.012; see Table 1 for model statistics), but there was no significant linear relationship between NE and forgetting rate without considering the quadratic component (movie 1: p = 0.73; movie 2: p = 0.58; movie 3: p = 0.88; Table 1; Figure 4A).

Table 1. Linear and quadratic model comparisons including all participants. AIC difference represents the difference between the linear and quadratic AICs (n = 35). * p-value < 0.05

All		R^2	p-value	AIC	AIC difference
Movie 1	Linear	-0.026	0.735	-15.07	6.66
	Quadratic	0.173	0.018 *	-21.74	
Movie 2	Linear	-0.02	0.581	-15.27	7.149
	Quadratic	0.189	0.013 *	-22.42	
Movie 3	Linear	0.021	0.88	-14.97	4.92
	Quadratic	0.128	0.041 *	-19.90	

We considered whether this effect was driven by participants who had a WM forgetting rate that was two standard deviations larger than the mean ($\emptyset_{WM}$ > 0.6) and a problematic model fit (See Supplementary Figure 2 for this learning behavior; n = 3): with very high estimated learning rate (α > 0.9, two standard deviations larger than the mean) and low mixture weight (w_0< 0.6), indicating that the working memory module of the model was not functioning in a regime representative of cognitive working memory function. Instead, we found that the U-shaped effect was stronger when excluding these participants (for all components and movies p < 0.001; Table 2; Figure 4B).

Table 2. Linear and quadratic model comparison including 32 participants. ** p-value < 0.001

Subset		R^2	p-value	AIC	AIC difference
Movie 1	Linear	0.015	0.23	-68.86	21.59
	Quadratic	0.5237	<0.001 **	-90.46	
Movie 2	Linear	-0.015	0.46	-67.91	16.94
	Quadratic	0.429	<0.001 **	-84.86	
Movie 3	Linear	-0.033	0.87	-67.36	32.76
	Quadratic	0.651	<0.001 **	-100.12	

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

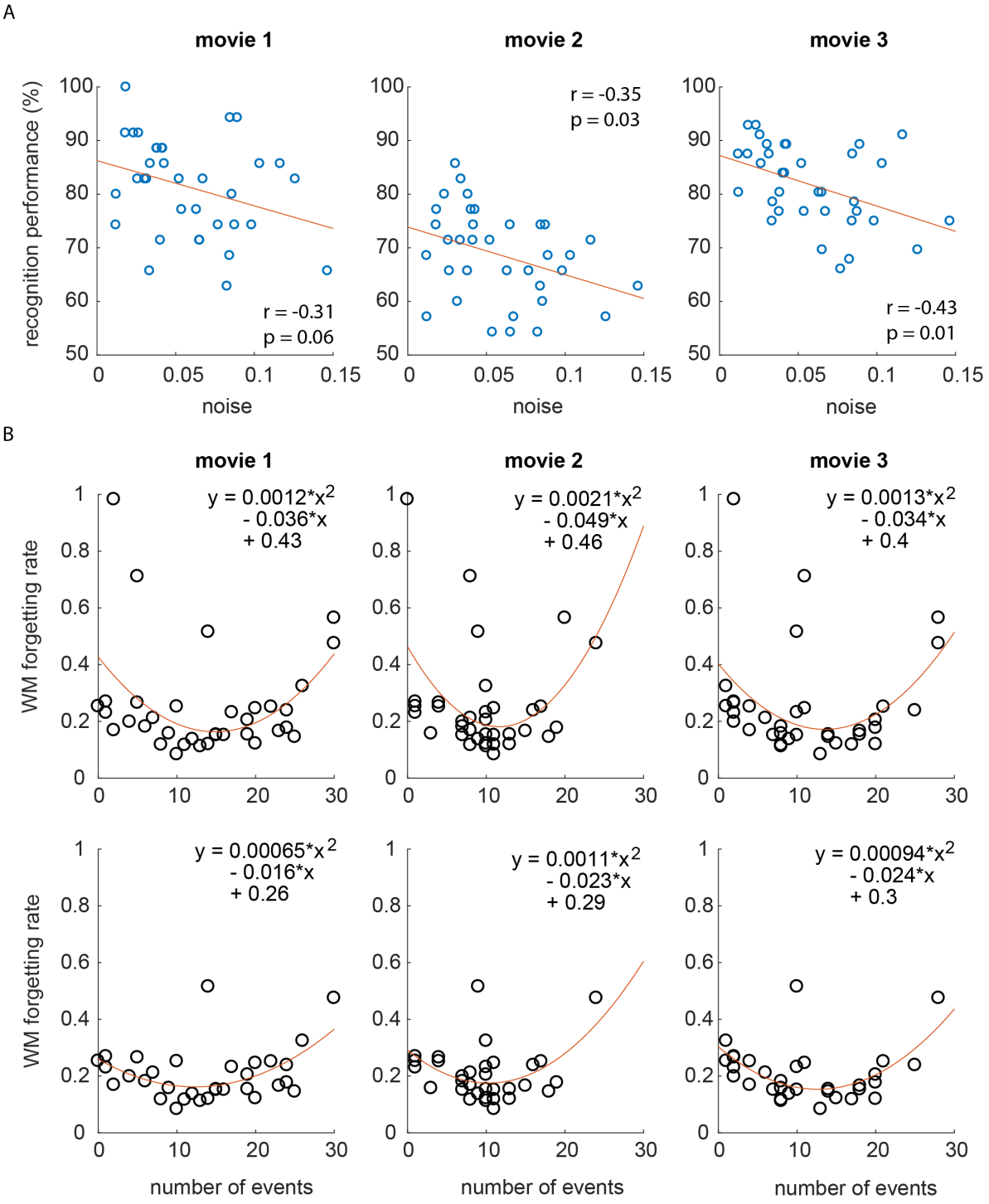


Figure 4. Cross-task results: (A) For all three movies, recognition performance in the movie task decreased with an increase in estimated noise in the association learning task. (B) Working memory forgetting rate had a U-shaped relationship with the number of determined events (top row) for all the movies. This relationship is significant even after discarding 3 participants with a very high estimated learning rate (more than 0.9) and a low estimated mixture weight (less than

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

0.6; bottom row). The assumption was that these participants relied more on their RL components and their WM forgetting rates may be inaccurate. The excluded participants had a high estimated WM forgetting rate. Each dot represents a participant.

Post-hoc analyses were employed to study the recognition and recall performance of participants with a high forgetting rate including (1) those who were more than one standard deviation higher than the mean number of events (i.e., over-segmenters) and (2) those who were more than one standard deviation lower than the mean (i.e., under-segmenters); there were 7 participants in each of these two groups. We observed that temporal recognition performance for movie 3 (which had a linear storyline) was better for under-segmenters than over-segmenters (two sample t-test $t(12) = 3.26$, $p = 0.0068$; Figure 5A). However, the over-segmenters wrote more words during the free recall task (sum of all movies $M = 204.6$, $SD = 54.8$) than the under-segmenters ($M = 107.1$, $SD = 70.57$; two-sample t-test $t(10) = -2.67$, $p = 0.023$; Figure 5B). The quality of stories was the same across the movies (see Supplementary material for examples) and this difference was observed for all the movies (table S1; movie 1: $t(10) = -2.84$, $p = 0.017$; movie 2: $t(10) = -2.66$, $p = 0.023$; movie 3: $t(10) = -1.90$, $p = 0.085$).

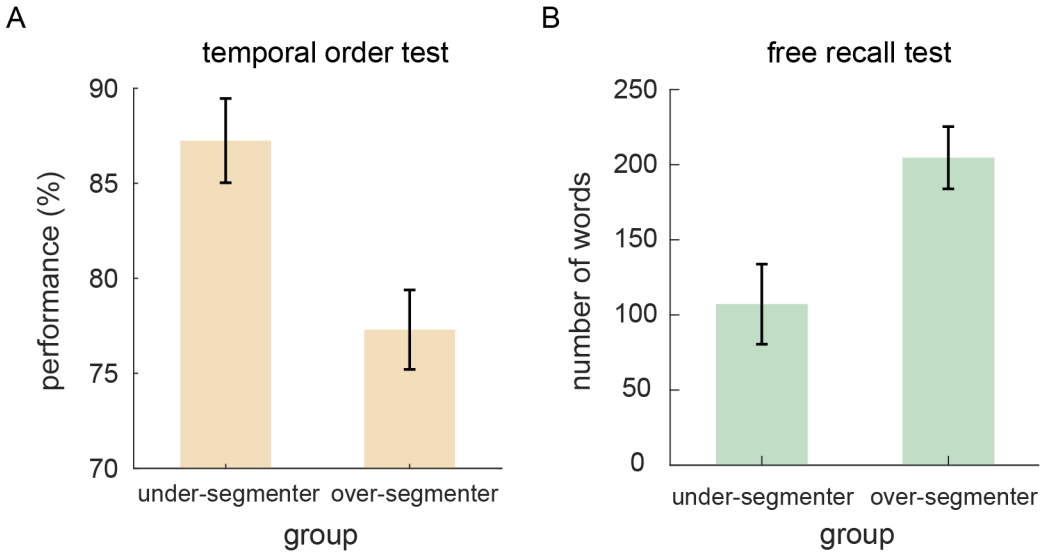


Figure 5. Under-segmenters and over-segmenters performed differently on temporal recognition and free recall tasks. (A) Under-segmenters performed better at the temporal recognition test for movie 3, which had a linear storyline ($p < 0.01$). (B) Over-segmenters performed better at the free recall; they wrote more words about the movies at the recall than under-segmenters ($p < 0.05$; see Supplementary material for examples). Error bars represent the standard error of the mean.

5. Discussion

Individuals fluently segment a flow of events into episodes (Zacks and Swallow, 2007; Zacks et al., 2007), a process closely linked to working memory (Kurby and Zacks, 2008; Radvansky, 2017; Richmond et al., 2017). Here, we examined what aspect of working memory is linked to event-segmentation, instantiated as the number of determined events. Although participants can

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

adaptively modulate event segmentation to a fine- or course-grain (Speer et al., 2007), across a population, individuals vary in terms of the number of determined events (Ben-Yakov and Henson, 2018; Jafarpour et al., 2018; Zacks et al., 2006). We hypothesized that working memory ability would differentiate the individual according to the number of determined events. Two separate tasks were administered to a group of healthy participants, a movie segmentation task and an association learning task, followed by a surprise free recall memory test about the movies. From the movie segmentation task, we assessed the subjective number of determined events (Figure 2), and from the association learning task we inferred working memory capacity and forgetting rate (Figure 3). The cross-task results revealed that the number of determined events had a U-shaped relationship with working memory forgetting rate, so that participants with a high forgetting rate either segment too many or too few events (Figure 4).

All participants performed well in the association learning task although their working memory limitation affected performance (Figure 3). A logistic regression was used to estimate the effect of set size (number of stimuli in a block), delay (number of trials since the last correct choice for the current stimulus), and previously correct answers (number of correct choices made so far for the current stimulus) and their interactions on the trial-by-trial responses in the association learning task. The results showed that all these factors, which are relevant to working memory limitations, played a role in the responses (Figure 3). Response accuracy decreased with increasing set size and delay. By contrast, an increase in the number of previously correct answers improved response accuracy. Comparing the beta estimates of these parameters to the number of determined events in the movie segmentation task, we observed a correlation between the number of determined events and the effect of working memory limitation on learning (Figure S1). This effect was further investigated by isolating the working memory component of learning using a reinforcement learning and working memory model (RLWM) which successfully simulated the learning behavior (Figure 3). The RLWM model provided an estimate of working memory capacity and forgetting rate along with six other parameters including the undirected noise.

The cross-task comparison revealed that the undirected noise correlated with temporal recognition accuracy for all the movies (Figure 4A). Participants with higher estimated undirected noise in the association learning task had a lower accuracy in the movie recognition memory test. Such correlation may reflect the level of attention to the tasks in general. This finding, along with the simulation performance and the AIC comparison (Figure 4), validates the RLWM's performance in modeling the learning behavior. Given that attention helps with retaining a similar representation of context (Aly and Turk-Browne, 2016), participants with more estimated undirected noise may have less consistent representation of context, resulting in less accurate memory of temporal order of events (Horner et al., 2016).

The main cross-task observation, a U-shaped relationship between working memory forgetting rate and the number of determined events, was replicated in three novel movies with different storylines (Figure 4). One of the movies did not have an overarching story, instead contained flows of discrete events that involved similar characters across the events (like a Tom & Jerry cartoon; movie 2). Another movie had a linear and overarching story of a pig's life (movie 3). The

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

third movie was a combination of temporally interchangeable stories and an overarching story of a lion falling in love and fighting over a lioness (movie 1). Participants' rank in the number of determined events was consistent across the movies, and they remembered the order of events better than the chance for all the movies (Figure 2).

The U-shaped result indicates that participants with a higher forgetting rate than the norm used different strategies for event-segmentation; some participants were under-segmenters and others were over-segmenters. Temporal order memory for the under-segmenters was better than for over-segmenters when the movie had a linear overarching story, suggesting that these participants benefited more from the schema of the movie than the over-segmenters (Figure 5A). By contrast, the over-segmenters update the context, e.g., "walk through doorways" (Radvansky & Copeland, 2006), more often than under-segmenters. This behavior resulted in a worse performance in the temporal recognition task (Hanson and Hirst, 1989). Over-segmenters, however, had a better recall performance (Figure 5B), writing more words at free recall compared to under-segmenters. In general, participants remember event-boundaries better than other events (Newtson and Engquist, 1976) and a fine-grained event segmentation benefits source memory but not recognition memory (Hanson and Hirst, 1989, 1991). As also shown by Sargent et al. (2013), participants who segmented a movie in fine grain had an accurate recall (also see Bailey et al., 2013).

An outstanding question is the causality of the relationship between working memory and event-segmentation. One possibility is that working memory is a primary cognitive mechanism and event-segmentation is determined by limits in working memory. Accordingly, participants with a high forgetting rate determine a high number of events. An alternative possibility is that event-segmentation is a primary cognitive mechanism (Radvansky, 2017; Richmond et al., 2017). In this case, participants who segment more often either lose access to the information from the previous events (Ezzyat and Davachi, 2014; Horner et al., 2016) leading to a high forgetting rate, or they rely on schematic storylines to keep a track of what happened. A third possibility is that both working memory and event segmentation rely on a common cognitive mechanism, such as utilizing a schema or script. Utilizing a schema facilitates memory and segmentation (Farag et al., 2010; Gobet et al., 2015; Sargent et al., 2013; Zacks et al., 2010). For example, utilizing a phone number's schema enables effective segmentation and memory for a 10-digit number (Miller, 1956). Future research is necessary to clarify the neural mechanisms of these alternatives.

In conclusion, we observed that the working memory forgetting rate reflects individual differences in event-segmentation. A U-shaped relationship between the number of determined events on one task and the forgetting rate on another task showed that participants with a high forgetting rate used two different strategies for event perception. Memory during conditions of under-segmentation was benefited more from a linear storyline, while free recall performance benefitted from over-segmentation. Taken together, these data suggest that an individual's forgetting rate can be inferred by that individual's rate of event segmentation.

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6. Acknowledgments

We thank Cassandra Lei, Crystal H. Shi, and Megan Schneider for data collection.

7. References

Aly, M., and Turk-Browne, N.B. (2016). Attention Stabilizes Representations in the Human Hippocampus. Cereb Cortex *26*, 783–796.

Baddeley, A. (1992). Working memory. Science *255*, 556–559.

Baddeley, A. (2003). Working memory: looking back and looking forward. Nature Reviews Neuroscience *4*, 829–839.

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. Annual Review of Psychology *63*, 1–29.

Bailey, H.R., Kurby, C.A., Giovannetti, T., and Zacks, J.M. (2013). Action perception predicts action performance. Neuropsychologia *51*, 2294–2304.

Bailey, H.R., Kurby, C.A., Sargent, J.Q., and Zacks, J.M. (2017). Attentional focus affects how events are segmented and updated in narrative reading. Mem Cogn *45*, 940–955.

Bays, P.M., and Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. Science *321*, 851–854.

Bays, P.M., Catalao, R.F.G., and Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. J Vis *9*.

Ben-Yakov, A., and Henson, R. (2018). The hippocampal film-editor: sensitivity and specificity to event boundaries in continuous experience. J. Neurosci. 0524–18.

Collins, A.G.E. (2018). The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory. Journal of Cognitive Neuroscience *30*, 1422–1432.

Collins, A.G.E., and Frank, M.J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. European Journal of Neuroscience *35*, 1024–1035.

Collins, A.G.E., and Frank, M.J. (2018). Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. PNAS 201720963.

Collins, A.G.E., Brown, J.K., Gold, J.M., Waltz, J.A., and Frank, M.J. (2014). Working Memory Contributions to Reinforcement Learning Impairments in Schizophrenia. J. Neurosci. *34*, 13747–13756.

Collins, A.G.E., Albrecht, M.A., Waltz, J.A., Gold, J.M., and Frank, M.J. (2017). Interactions Among Working Memory, Reinforcement Learning, and Effort in Value-Based Choice: A New Paradigm and Selective Deficits in Schizophrenia. Biological Psychiatry *82*, 431–439.

Cowan, N. (2010). The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? Curr Dir Psychol Sci *19*, 51–57.

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

D'Esposito, M., and Postle, B.R. (2015). The Cognitive Neuroscience of Working Memory. Annual Review of Psychology *66*, 115–142.

Ezzyat, Y., and Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. Neuron *81*, 1179–1189.

Farag, C., Troiani, V., Bonner, M., Powers, C., Avants, B., Gee, J., and Grossman, M. (2010). Hierarchical Organization of Scripts: Converging Evidence from fMRI and Frontotemporal Degeneration. Cereb Cortex *20*, 2453–2463.

Gobet, F., Lane, P.C.R., and Lloyd-Kelly, M. (2015). Chunks, Schemata, and Retrieval Structures: Past and Current Computational Models. Front. Psychol. *6*.

Hanson, C., and Hirst, W. (1989). On the representation of events: a study of orientation, recall, and recognition. J Exp Psychol Gen *118*, 136–147.

Hanson, C., and Hirst, W. (1991). Recognizing differences in recognition tasks: A reply to Lassiter and Slaw. Journal of Experimental Psychology: General *120*, 211–212.

Horner, A.J., Bisby, J.A., Wang, A., Bogus, K., and Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. Cognition *154*, 151–164.

Hymel, A., Levin, D.T., and Baker, L.J. (2016). Default processing of event sequences. Journal of Experimental Psychology: Human Perception and Performance *42*, 235–246.

Jafarpour, A., Penny, W., Barnes, G., Knight, R.T., and Duzel, E. (2017). Working-Memory Replay Prioritizes Weakly Attended Events. ENeuro ENEURO.0171-17.2017.

Jafarpour, A., Griffin, S., Lin, J.J., and Knight, R.T. (2018). Orbitofrontal cortex, dorsolateral prefrontal cortex, and hippocampus represent the magnitude of event saliency. BioRxiv 285718.

Kurby, C.A., and Zacks, J.M. (2008). Segmentation in the perception and memory of events. Trends Cogn. Sci. (Regul. Ed.) *12*, 72–79.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review *63*, 81–97.

Newtson, D., and Engquist, G. (1976). The perceptual organization of ongoing behavior. Journal of Experimental Social Psychology *12*, 436–450.

Radvansky, G.A. (2017). Event Segmentation as a Working Memory Process. Journal of Applied Research in Memory and Cognition *6*, 121–123.

Radvansky, G.A., and Zacks, J.M. (2017). Event boundaries in memory and cognition. Current Opinion in Behavioral Sciences *17*, 133–140.

Richmond, L.L., Gold, D.A., and Zacks, J.M. (2017). Event Perception: Translations and Applications. Journal of Applied Research in Memory and Cognition *6*, 111–120.

Rose, N.S., LaRocque, J.J., Riggall, A.C., Gosseries, O., Starrett, M.J., Meyering, E.E., and Postle, B.R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. Science *354*, 1136–1139.

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sargent, J.Q., Zacks, J.M., Hambrick, D.Z., Zacks, R.T., Kurby, C.A., Bailey, H.R., Eisenberg, M.L., and Beck, T.M. (2013). Event segmentation ability uniquely predicts event memory. Cognition *129*, 241–255.

Speer, N.K., Zacks, J.M., and Reynolds, J.R. (2007). Human Brain Activity Time-Locked to Narrative Event Boundaries. Psychological Science *18*, 449–455.

Vogel, E.K., and Machizawa, M.G. (2004). Neural activity predicts individual differences in visual working memory capacity. Nature *428*, 748–751.

Zacks, J.M., and Swallow, K.M. (2007). Event Segmentation. Current Directions in Psychological Science *16*, 80–84.

Zacks, J.M., Swallow, K.M., Vettel, J.M., and McAvoy, M.P. (2006). Visual motion and the neural correlates of event perception. Brain Res. *1076*, 150–162.

Zacks, J.M., Speer, N.K., Swallow, K.M., Braver, T.S., and Reynolds, J.R. (2007). Event perception: A mind-brain perspective. Psychological Bulletin *133*, 273–293.

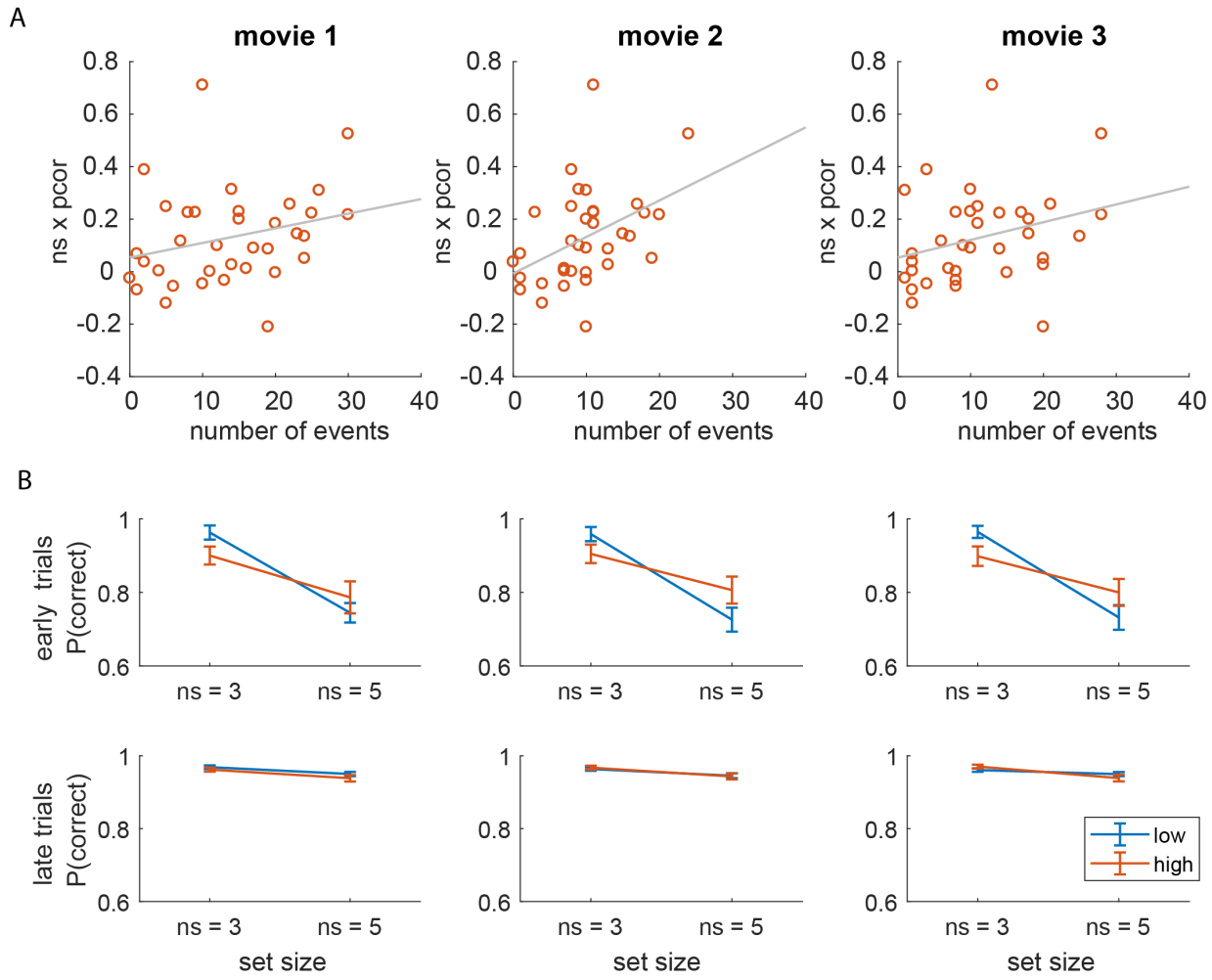
Zacks, J.M., Speer, N.K., Swallow, K.M., and Maley, C.J. (2010). The Brain's Cutting-Room Floor: Segmentation of Narrative Cinema. Front. Hum. Neurosci. *4*.

Zokaei, N., Manohar, S., Husain, M., and Feredoes, E. (2014). Causal Evidence for a Privileged Working Memory State in Early Visual Cortex. J. Neurosci. *34*, 158–162.

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

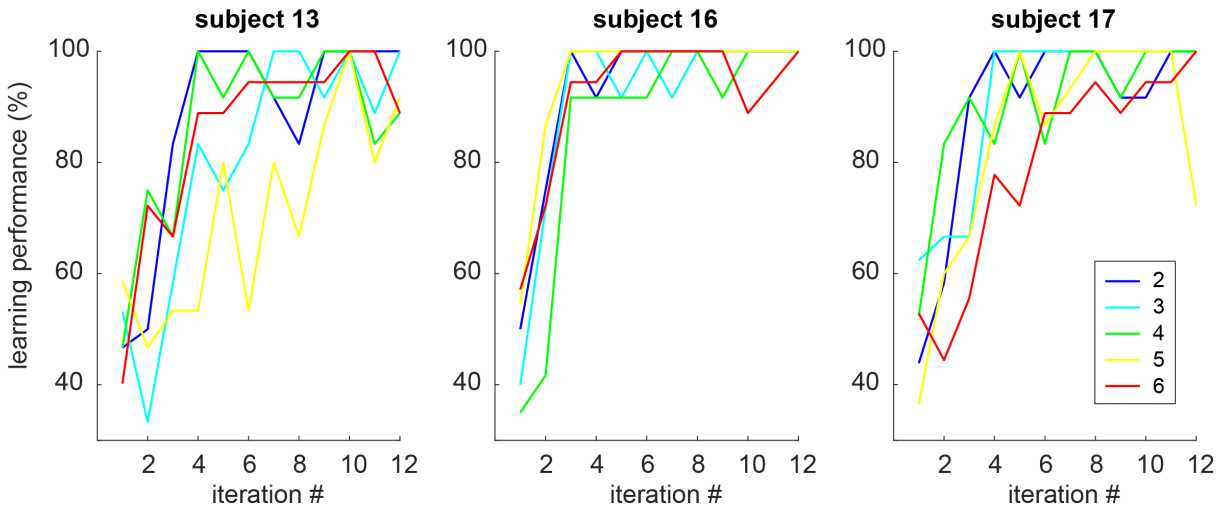
8. Supplementary material

8.1. Figures



Supplementary Figure 1. The number of determined events in the movie task correlated with working memory span during the association learning task. (A) The estimated beta value for the interaction between the number of stimuli and number of previously correct trials correlated with the number of determined events in each movie (left: movie 1, ranked r = 0.32, p = 0.064; middle: movie 2, ranked r = 0.45, p = 0.005; right: movie 3, r = 0.32, p = 0.056). (B) The probability of previously correct trials (pcor) for a small set size (3) and a large set size (5) for the 50% of participants with low number of segments (in blue) and with high number of segments (in red) in early trials (first 2 trials; top row) and in the later trials (averaged across 3 and more trials; bottom row).

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplementary Figure 2. Learning behavior of three participants who were excluded from the analysis of the relationship between the number of determined events and working memory forgetting rate because they had a high WM forgetting rate, more than two standard deviations larger than the mean ($\emptyset_{WM}$ > 0.6) and problematic model fits.

8.2. Tables

Table S1. The mean and standard deviation (in parentheses) of the number of written words by under-and over-segmenters for each of the movies:

	Movie 1	Movie 2	Movie 3	Total
Under-segmenter	32.66 (23.14)	26.5 (18.32)	48.0 (32.39)	107.16 (70.57)
Over-segmenter	64.66 (14.94	57.83 (22.21)	82.16 (29.59)	204.66 (54.80)

8.3. Movie stories

Movie 1:

The first animation depicts an overarching cliché love triangle story, along with some temporally interchangeable stories. It starts by showing a few animal couples going back and forth in a park. Then, there is a small lion that looks heartbroken. The lion sees a lioness, but there is a bigger lion that wants to meet the lioness too. The two lions fight for the lioness's attention through a series of matches that have periods with an anticipated flow. For example, after each match, the score is shown on board. However, the score is not immediately shown after an eating contest. After the bigger lion wins the eating contest, it eats the small lion's food too. Then the score is shown, and the matches continue. The small lion loses the competition and moves on to another part of the world. The end of the movie shows that the small lion meets a lioness.

Movie 2:

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The second animation is a sequence of independent events that involve crocodiles. It starts by showing two zebras listening to music next to a swamp. A crocodile suddenly eats one of them. Then, the crocodiles swim in a swamp just below the surface with only the eye visible. One of the crocodiles is wearing glasses. This crocodile stands up for a moment to clean its glasses and then it continues swimming below the surface. Next, a crocodile attacks a cow that is drinking water. The cow is too big for the crocodile so it cannot bite it. The cow, however, beats the crocodile in one attempt. Next, a crocodile is eating at a table in the swamp that has birds next to it. It puts catchup on the birds and eats them one by one by a fork. A bigger crocodile takes a cow into the swamp, but the cow defeats the crocodile and comes out. Next, a goat is swimming away from a crocodile, clearly scared. The swamp suddenly dries out. The crocodile cannot walk fast, but the goat happily leaves the swamp. Then, a crocodile attacks a cow that is by the swamp, but the cow skins the crocodile and takes it for tanning. Next, a crocodile with dental braces is shown drinking with a straw. A baby zebra plays by the swamp and bothers the crocodile. After that, a crocodile is shown participating in a non-violence resistance group with other animals, holding a peace sign. The movie ends with a scene of a very long crocodile on which a bird is happily picnicking.

Movie 3:

The third movie depicts a linear life story of a pig. It starts by showing a caterpillar on a leaf. Then a big sow appears and gives birth to seven piglets. The piglets follow the sow in a line going around woods and crossing roads. Two of the pigs suddenly disappear; they were killed on the road. The rest of the piglets also disappear one by one, except for one. Then the caterpillar is netting a cocoon – showing the passage of time. The piglet grows up to be an ugly boar, and the sow is old. The sow dies. The pig meets three gilts. They reject him (depicted as a computer error message box) because he does not have money, he is ugly, and one of the gilts is already married. The cocoon is now complete, and the boar is still sad and alone. It bumps into a lion that was hunting for zebras. The lion gets happy for the catch, but the boar is too smelly. The lion puts the boar in a washing machine. It comes out as a red boar which is not desirable to the lion. The lion dumps the boar. The boar passes by the gilts again. This time, the one that was interested in a good-looking boar is interested and follows him.

8.4. Example of recalls

A representative example of an over-segmenter:

Movie 1:

giraffe dragging aligator along the ground. the painting was 2D but looked like the smaller tiger was flattened into it. smaller tiger takes a long time to eat. lots of couples at the beginning. bigger tiger cheats on the test. not sure what that map with the arrow meant but some cannon? but somehow smaller tiger crused bigger tiger yay

Movie 2:

glasses aligator was funny. big bison/ox was hero of the story. wonder why that first zebra or animal just dove into the lake. aligator really wants to eat that bison. not sure what that wrench looking tool was, but aligator got skinned. part where

bioRxiv preprint doi: https://doi.org/10.1101/571380; this version posted March 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the aligator puts sauce on the birds just standing there was unexpected but very funny

Movie 3:

Time passing was kind of conveyed by the worm/caterpillar, which was really cool. lots of crisscrossing across the screen. what did the arrow signify? funny how the arrow lost the color in the wash and dyed him red. error boxes popping up was funny because it was almost like a computer game. not sure what happened to the other little babies along the way....

A representative example of an under-segmenter:

Movie 1:

two different levels of lions competing with each other
the lower level lion only won the writing test

Movie 2:

aligators wants to eat other animals
some animals sold the aligator skin

Movie 3:

the washing machine switched the color of the pig and the color of the arrow
the female pig fall in love with the male pig after he changed his color