Journal of Experimental Psychology:
Learning, Memory, and Cognition
2014, Vol. 40, No. 1, 86–93

0278-7393/14/$12.00 DOI: 10.1037/a0034250

"A Long Time Ago, in a Context Far, Far Away": Retrospective Time Estimates and Internal Context Change

Lili Sahakyan and James R. Smith
University of North Carolina at Greensboro

This investigation aimed to establish retrospective time judgments as markers of internal context change across 2 memory paradigms, the effects of which have been attributed to internal context change by some researchers. Experiment 1 involved the list-method directed forgetting paradigm and established that the forget group significantly overestimated the duration of the experiment compared with the remember group. Experiment 2 involved the list-before-last paradigm, whereby participants studied 3 lists, and in between encoding of List 2 and List 3, some participants retrieved List 1, whereas the control participants restudied List 1. The results showed that the retrieval group significantly overestimated the duration of the experiment compared with the restudy group. Overall, these results support the context-change interpretation of these paradigms, and they also support the contextual-change hypothesis of retrospective timing (Block & Reed, 1978).

Keywords: mental context change, time estimates, directed forgetting, list-before-last paradigm

Memory retrieval is important in every aspect of daily life, from remembering whether or not you took your medication in the morning to where you parked your car. Although many factors influence how memories are retrieved, critical importance is given to retrieval cues, which are prompts or hints used to trigger retrieval of information because of their association to that information. Our daily experiences unfold in certain temporal-spatial, social-emotional environments, and these background cues become associated with our memories. During retrieval, we rely on them as search cues. Even in the tasks in which the cues are not obvious (like in free recall), people are nonetheless thought to retrieve things from memory by relying on internally generated context cues. Consequently, many computational theories of memory include gradually changing context as an important component of their models (e.g., Anderson & Bower, 1972; Dennis & Humphreys, 2001; Estes, 1955; Howard & Kahana, 2002; Lehman & Malmberg, 2009; Mensink & Raaijmakers, 1988; Polyn, Norman, & Kahana, 2009; Raaijmakers & Shiffrin, 1981; Sederberg, Howard, & Kahana, 2008). Context has been modeled in various ways, such as a common node to which item representations are linked (e.g., Anderson, Bothell, Lebiere, & Matessa, 1998; Anderson & Bower, 1972), a collection of features that randomly change across study trials (e.g., Estes, 1955; Mensink & Raaijmakers, 1988), as a recency weighted average of semantic features of studied items (e.g., Howard & Kahana, 2002), and the like. Regardless of how it is modeled, memory researchers postulate this important theoretical construct in order to explain large quantities of observed data. When the context at test matches that at study, it can serve as a retrieval cue and thereby improve memory.

Despite the critical role of context in memory, it can be difficult to observe in the lab. That is because context features may include external attributes like the physical environment in which we find ourselves, but they may also include internal or mental contextual elements that exist only inside of a person's mind. An important piece of evidence that context affects memory comes from studies that directly manipulated environmental or external context and then observed that matching or mismatching context cues affected memory (see Smith & Vela, 2001, for a review). However, manipulations of mental or internal context are more difficult because of the nature of the construct. *Mental context* refers to our constantly changing blend of thoughts that evolves in response to encoding and retrieval of other events. Because it is difficult to directly manipulate, researchers typically examine memory performance (e.g., accuracy, intrusion rates, response latencies, or response transitions) to infer that a manipulation affected mental context. For example, if memory becomes less accessible as a result of experimental manipulation (e.g., instructing people to forget previously studied information, known as the directed forgetting manipulation), then we might infer that something was changed in the mental environment to produce internal context change, thus lowering memory accessibility. However, this could lead to circular reasoning, and having some other marker of context change besides memory accuracy would be an important tool.

The central goal of this investigation was to find a way of indexing mental context change through some means other than examining memory. To accomplish this goal, we built on research on subjective duration judgments, and we hypothesized that retro-

This article was published Online First August 19, 2013.

Lili Sahakyan and James R. Smith, Department of Psychology, University of North Carolina at Greensboro.

We extend our gratitude to Hannah Hendricks, Andrew Moody, Jackie Palmer, Katherine Henline, Ron Baron, Joshua Perkins, Casey Kelly, Gary Styles, and Beatrice Campbell for their help with data collection and scoring.

Correspondence concerning this article should be addressed to Lili Sahakyan, University of North Carolina at Greensboro, Department of Psychology, 296 Eberhart Building, P.O. Box 26170, Greensboro, NC 27402. E-mail: l_sahaky@uncg.edu

spective time estimates could index whether an experimental manipulation has led to mental context change. Establishing new diagnostic markers of mental context change would have important theoretical implications for memory research because they could be used in conjunction with memory performance to test the context-change interpretations of various phenomena. In addition, having a way of indexing mental context change without having to infer it post hoc from memory performance could allow making a priori predictions about which type of manipulations are likely to cause changes in memory due to mental context change and which ones are not. This could lead to the design of new experiments with specific testable predictions.

We examined retrospective time estimates in two different memory paradigms that have evolved rather independently and have been explained by some researchers in the field in terms of mental context change. Experiment 1 involved the list-method directed forgetting paradigm (e.g., Sahakyan & Kelley, 2002), and Experiment 2 involved the list-before-last paradigm (e.g., Jang & Huber, 2008; Shiffrin, 1970). Obtaining converging evidence with time estimates across these two paradigms would provide further support for the context-change mechanism that was advanced to explain them. In turn, obtaining differences in time estimates in the context-change conditions of the two heterogeneous memory paradigms would provide a test and a conceptual replication of the context-change hypothesis of retrospective timing (Block & Reed, 1978) and would challenge the event-memory hypothesis of retrospective timing (e.g., Ornstein, 1969). We describe these hypotheses later in the current article.

The relationship between mental context and temporal coding has been recognized by many memory scholars (e.g., Bower, 1972; Estes, 1955; Hintzman, 2002; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988). For example, according to Bower (1972), the role of mental context is to help with temporal tagging of the items. In order to infer how long ago an item was encountered, people compare the current state of the context with the state of the context associated with an item, and the similarity between the two states of context serves as a marker of recency of an item (see also Hintzman, 2002; Hitzman, Block, & Summers, 1973). Thus, if mental context changed as a result of specific experimental manipulations, there should be greater dissimilarity between the two states of context. Therefore, we predict a systematic overestimation bias in the context-change conditions compared with conditions that do not experience context change.

Researchers in the duration judgments literature have also noted the importance of context in time estimation processes, especially in reference to judgments that people make about past events and episodes. More specifically, retrospective time estimates are estimates solicited *unexpectedly* upon completion of a task, and hence they must be constructed from information available in memory. However, the nature of information that people rely on to make those judgments has been debated. The early notions emphasized *event memory*, suggesting that the more people remember from the time interval, the longer they will judge it retrospectively (e.g., James, 1890; Ornstein, 1969). For example, an interval containing eight songs is retrospectively judged as having lasted longer than an interval of equivalent duration containing four longer songs (Bailey & Areni, 2006). The event-memory view has been challenged by Block and colleagues (Block, 1990, 1992; Block & Reed, 1978), who proposed that the critical factor in retrospectively assessing duration is not the number of events that we remember from an interval, but rather the extent to which those events constitute a change in cognitive context, known as the *contextual-change* view. Support for the context-change idea came from two experiments, where by manipulating the orienting task performed on the words, Block and Reed (1978) created a "deep processing" interval that led to substantially better memory than the "shallow processing" interval. Contrary to the event-memory view, despite robust differences in memory across the two intervals, the duration estimates for the two intervals were identical. Importantly, in a subsequent experiment, Block and Reed (1978) showed that remembered duration increased when people performed different orienting tasks during a target interval instead of a single task. The authors attributed the results to changes in cognitive or processing context induced by different orienting tasks. In subsequent research, Block (1982) showed that changing the environmental context between the two study lists of equal duration also affects retrospective time estimates by eliminating the bias observed in the "same context" condition to estimate the first list as lasting longer than the second list. Finally, segmented intervals are retrospectively judged as longer than unsegmented intervals (e.g., Poynter, 1983; Zakay, Tsal, Moses, & Shahar, 1994), and event segmentation can be considered as another source of contextual change. Overall, retrospective duration judgments show an overestimation bias as a result of changes in processing context, environmental context, and event segmentation.

Although we explain the directed forgetting paradigm and the list-before-last paradigm in greater detail in corresponding experiments, we note that both paradigms induce some form of forgetting, which has been attributed to changes in mental context by some researchers. According to the context-change hypothesis of retrospective timing, we should obtain significant overestimation bias in the alleged conditions of context change in both experiments compared with the respective control conditions. Obtaining such findings would provide a conceptual replication of the context-change hypothesis of retrospective timing through a different approach than has been used before—by relying on two different memory paradigms that have been explained in terms of context change. Furthermore, obtaining significant overestimation in the predicted conditions would pose a challenge for the event-memory hypothesis of retrospective timing, according to which the more people remember from an interval, the longer they will judge it to have lasted (e.g., Ornstein, 1969). We further elaborate on these predictions in each experiment.

Experiment 1: Directed Forgetting

In Experiment 1, we used the list-method directed forgetting paradigm, which involves instructing participants to forget previously encoded set of items. In recent years, the designs of list-method directed forgetting have evolved from the more frequently used two-list design to multilist designs (for a recent review, see Sahakyan, Delaney, Foster, & Abushanab, 2013). In our study, we used the three-list design variant used by Lehman and Malmberg (2009), because it avoids the surprise element (we further elaborate on this issue below). Participants are told upfront that they would study several lists of items but that they would only be tested on one of the lists, which would be chosen randomly. After the first list (L1), they are reminded to keep that list in mind because they might be tested on it later. After the second list (L2), half the participants receive the forget cue, whereas the remaining half receive the remember cue, which reminds them to keep L2 in mind because they might be tested on it later. The forget cue instructs participants to focus on the upcoming list (L3) because they would only be tested on L3 and specifies that they should forget the previous lists as they would never be tested. After L3 encoding, participants complete a free-recall test of either L2 or L3 (for the forget group participants, the L2 test comes as a surprise). L1 is never tested; it is primarily included in the design to control for various sources of confounds in the two-list design. The three-list procedure produces impaired recall of L2 and enhanced recall of L3 in the forget group compared with the remember group, known as the directed forgetting effect.[1] We entertain predictions of the two major accounts of directed forgetting—the mental context-change account (Sahakyan & Kelley, 2002) and the inhibitory account (e.g., Bjork, 1989; Geiselman, Bjork, & Fishman, 1983)—because depending on what produces directed forgetting, one would expect different consequences for time estimates.

According to the inhibitory viewpoint, the forget cue initiates a process that at the time of retrieval inhibits L2 items. If L2 items are inhibited, then according to the event-memory account of retrospective timing, the forget group should *underestimate* the length of the target interval because the time estimate would be based on fewer items from memory. However, this prediction might be a bit extreme because if L3 items benefit from directed forgetting, then the combined memory of both lists will be equivalent across the forget and remember groups, and hence time estimates would be based on a similar amount of retrievable information from memory. With that said, directed forgetting enhancement is not always observed in the literature. Hence, if we obtain L3 enhancement, there should be no significant differences between the groups in time estimates according to the event-memory hypothesis. If we do not obtain L3 enhancement, then the overall memory would be lower in the forget group, and hence it should underestimate time. To summarize, if directed forgetting arises from inhibition of L2 items, then the event-memory hypothesis of retrospective timing predicts either no differences between the groups or an underestimation bias in the forget group, depending on whether one observes L3 enhancement. Under no circumstances would the event-memory view predict overestimation in the forget group. A completely different prediction is made by the context-change account of retrospective timing if one assumes that directed forgetting arises as a result of mental context change rather than inhibition of items.

Specifically, according to the mental context-change account of directed forgetting (e.g., Sahakyan & Kelley, 2002), during learning, participants encode both the meaning of the items and various contextual features. The forget instruction encourages participants to abandon the contextual cues prevalent during encoding of L2 and to sample new contextual cues for L3 encoding, thereby segregating the precue and postcue lists as different events. During the final test, retrieval context better matches L3 than L2, producing L2 forgetting in the forget group. Thus, according to the context-change account of directed forgetting, the forget cue leads to segmentation of internal experience, whereas the remember cue leads to more blended internal representation of all the lists. Therefore, the context-change account of retrospective timing would predict that the forget group should *overestimate* the duration of the experiment compared with the remember group.

The three-list directed forgetting design was appealing for our experiment because it allowed extending the length of the experiment beyond a 1- to 2-min interval. Most importantly, it avoids the surprise element that often accompanies the forget cue in the two-list designs (e.g., "oops, we showed you the wrong list"). Because participants are warned in advance to expect a forget cue, there are no major surprises or violations of expectations, which could affect the time estimates (e.g., Avni-Babad & Ritov, 2003; Boltz, 1993, 1998).

Because our primary interest centers on time estimates, we explain how they will be analyzed. Time estimates will be evaluated following the established approach in the timing literature. Namely, it is common to compute the ratio of subjective verbal estimates to objective length of the interval as opposed to merely analyzing verbal estimates (e.g., Block & Zakay, 1997; Roy & Christenfeld, 2007, 2008). This is because finding differences in verbal estimates between the two groups does not allow inferring about overestimation or underestimation biases that might be present in those groups. If Group 1 has a higher time estimate than Group 2, then both groups could be underestimating or overestimating the interval, but one group might do so to a greater extent than the other. Thus, to infer who is overestimating and who is not, researchers typically evaluate the duration judgment ratios in addition to reporting verbal estimates. Last but not least, dividing the subjective estimates by objective duration often leads to large positive skew in duration judgment ratios because values larger than 1 indicate overestimation, and they are unbounded by an upper limit, whereas values lower than 1 indicate underestimation, which are bounded by the lower limit (e.g., zero) and cannot be negative. Hence, if there is a positive skew, the recommended approach is to log transform the duration judgment ratios (Roy & Christenfeld, 2007, 2008). Doing so centers the values around zero, which reflects perfect accuracy, and positive and negative values represent overestimation and underestimation, respectively.

Method

Participants. A total of 104 University of North Carolina at Greensboro (UNCG) undergraduates participated in this experiment for course credit. They were randomly divided into 52 participants per forget and remember groups. Half the participants within each group were tested for memory of L2, whereas the remaining half was tested for memory of only L3.

Materials. Thirty-six unrelated English nouns of medium frequency were randomly divided into three lists of 12 words each. The presentation order of the three lists was counterbalanced.

Procedure. To ensure that the objective duration of experiment was controlled, the instructions were prerecorded and presented through speakers. After signing the consent form, participants were informed that they were about to see a red screen and

[1] In the two-list version of the paradigm, the forget cue is presented after L1 (usually as a surprise), and *directed forgetting effect* refers to impaired recall of L1 and enhanced recall of L2 in the Forget group compared with the Remember group. Due to the inclusion of the additional list in the three-list design version, the impairment, respectively, refers to L2, and the enhancement, respectively, refers to L3.

hear a tone, which would mark the beginning of the experiment. The procedures followed that of Lehman and Malmberg (2009). The instructions warned participants that they would study several lists but that they would have to recall only one of those lists, which would be chosen at random later in the experiment. Given that they would not know until later in the experiment which list they would be asked to recall, they were encouraged to pay attention to all the lists. The words were presented at a rate of 4 s followed by 1 s of interstimulus interval. After L1, all participants were reminded to keep that list in mind because they might be tested on it later. Then they studied L2, upon completion of which they received either the forget or the remember cue. Participants in the Forget group were instructed to forget the previous lists and focus on the next list because they would only be tested on the upcoming list. Participants in the Remember group were reminded again to keep the lists in mind because they might be asked later to recall any of those lists later on. After encoding L3, the computer screen changed the color to mark the end of presentation. At that point, all participants were asked to estimate in minutes and seconds how much time they thought had elapsed since the start of the experiment—that is, since the moment they heard the tone and saw the red screen. The objective duration of the estimated interval was 5 min. The time estimates were recorded by the experimenter. Afterwards, half the participants within each group were given 90 s to recall only L2, whereas the remaining half was told to recall only L3 (List 1 was never tested). In the Forget group, the experimenter apologized to participants for asking them to recall L2 despite being told earlier that they would not be tested on it. Finally, all participants were fully debriefed by the experimenter and had an option to withdraw their data given the minor deception (none did).

Results and Discussion

We first analyzed the recall to ensure we obtained the typical directed forgetting effect. The analysis of time estimates is presented next.

Recall findings. The recall findings are summarized in Figure 1. The List (List 2 vs. List 3) × Cue (forget vs. remember) analysis of variance (ANOVA) on proportion of recalled items revealed an overall recency effect, with List 3 being better recalled (M = .47, SE = .03) than List 2 (M = .25, SE = .03), confirmed by the main effect of the list, $F(1, 100) = 35.93$, $p < .001$, $\eta^2 = .26$. Importantly, there was a significant directed forgetting effect implied by the interaction, $F(1, 100) = 10.90$, $MSE = .035$, $p < .01$, $\eta^2 = .10$. Specifically, participants recalled fewer items from List 2 if they were in the Forget group compared with the Remember group, $t(50) = 2.63$, $p < .05$, and they recalled more from List 3 if they were in the Forget group than in the Remember group, $t(50) = 2.05$, $p < .05$. Together, these results confirm the directed forgetting effect. Note that there was no main effect of cue ($F <$ 1). In other words, the overall memory combined across the lists was not different across the two groups (in the Forget group: M = .35, SE = .03; in the Remember group: M = .37, SE = .03).

Time estimation findings. The critical question was whether the Forget group would overestimate the length of the target interval compared with the Remember group. Verbal estimates were significantly longer in the Forget group (M = 6.04, SE = .43) than the Remember group (M = 4.74, SE = .30), $t(102) = 2.51$, $p < .05$, Cohen's d = .49.

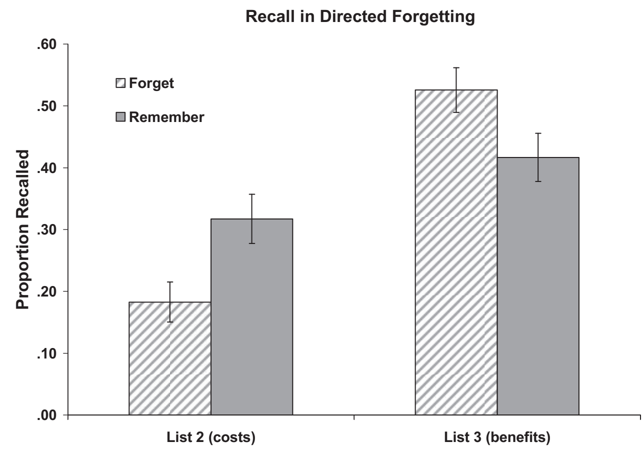

Figure 1. Proportion recalled from List 2 and List 3 in the Forget and Remember groups in Experiment 1. The error bars represent ± standard error of the mean.

To evaluate whether an overestimation or an underestimation bias was present in the Forget group, we computed the ratio of subjective duration to the objective duration of the interval. There was a positive skew in the data (skewness of time estimates was = 2.21), and therefore we log transformed the duration ratio scores. There was a significant overestimation bias present in the Forget group (M = .06, SE = .03, significantly different from zero), $t(51) = 2.31$, $p < .05$, Cohen's d = .45. In contrast, the Remember group had a tendency to underestimate the duration (M = −.04, SE = .02), but the underestimation was not statistically significant, $t(51) = 1.67$, $p = .10$. The difference in duration ratios between the Forget and Remember groups was also significant, $t(102) = 2.85$, $p < .01$, Cohen's d = .56.

Overall, these findings support the context-change account of directed forgetting, and they also support the context-change account of retrospective timing. Participants experiencing internal context change as a result of instructional manipulation experienced subjective lengthening of time, and they overestimated the amount of time that had passed in the experiment. This pattern is problematic for the event-memory hypothesis of retrospective timing, which would not predict differences in time estimates given that the overall memory was equivalent between the Forget and Remember groups. It is also problematic for the inhibitory interpretation of directed forgetting because if items were inhibited as a result of the forget instruction, then one would not expect overestimation in the forget group.

Experiment 2: List-Before-Last Paradigm

Experiment 2 involved the relatively less known memory paradigm, known as the *list-before-last paradigm* (Shiffrin, 1970), whereby participants encode multiple lists, and after each list (except the first), they are told to retrieve not the current list but rather the prior list. Research with this paradigm shows that retrieving prior lists in between the encoding of other lists has both positive and negative consequences for memory (Jang & Huber, 2008; Sahakyan & Hendricks, 2012; Shiffrin, 1970; Unsworth, Spillers, & Brewer, 2012; Ward & Tan, 2004). We focus on the abbreviated version of the list-before-last paradigm that we used in the current experiment as well as in prior research (Sahakyan & Hendricks, 2012). Figure 2 describes the design.

Participants encoded only three lists (termed L1, L2, and L3), and after L2, half of them were told to retrieve L1, whereas the remaining half was shown L1 again. Afterwards, both groups encoded L3 and were then tested for recall of L2. On the basis of prior work, we expected to observe forgetting of L2 in the retrieval group compared with the restudy condition, along with reduced intrusions from L3.

Theoretically, these effects have been attributed to internal context change instigated by L1 retrieval (Jang & Huber, 2008; Sahakyan & Hendricks, 2012). It is assumed that in the absence of retrieval trials, context drifts from one list to the next list in a somewhat gradual fashion. The act of L1 retrieval disrupts the contextual similarity/continuity between the adjacent lists, leading to an internal context change between L2 and L3. During the final test, reinstating L2 context becomes more difficult in the retrieval condition compared with the control condition (i.e., restudy condition), explaining the forgetting of L2. In addition, because retrieval creates an internal context change, it leads to better contextual differentiation between the adjacent lists, and hence L3 does not cause much interference on L2. In prior research, reduction of interference between L2 and L3 was observed either through lower L3 intrusions during the L2 test in the retrieval conditions compared with the control condition (e.g., Sahakyan & Hendricks, 2012) or through reduced effects of list-length manipulation of L3, whereby longer L3 did not cause more retroactive interference on L2 than shorter L3 (Jang & Huber, 2008; Shiffrin, 1970; Unsworth et al., 2012; Ward & Tan, 2004).

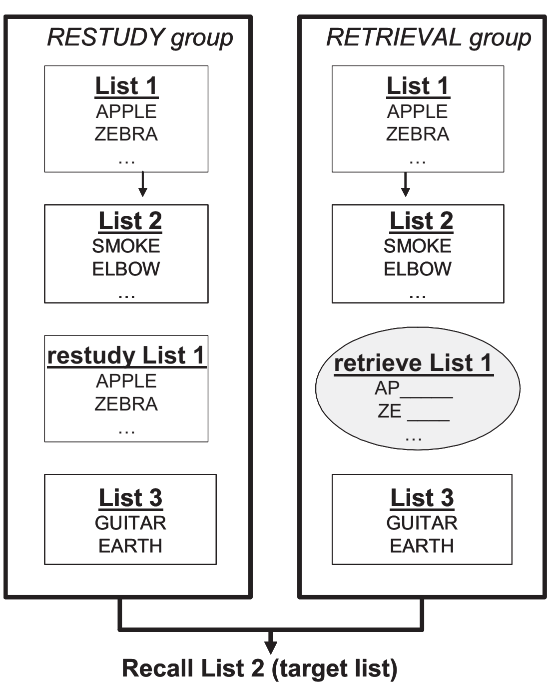

Figure 2. Abbreviated version of the list-before-last paradigm, used in Experiment 2.

Although the list-before-last paradigm was invented quite some time ago (Shiffrin, 1970), there has been relatively little research to date with this paradigm, and we are not aware of theoretical interpretations aside from the internal context change. Alternative interpretations in terms of inhibition or interference of items arising from retrieval of prior lists have been entertained and rejected by earlier research (e.g., Sahakyan & Hendricks, 2012). We nevertheless revisit these interpretations after describing the memory findings. Note that if the effects in the list-before-last paradigm arise from item-level inhibition or interference rather than internal context change, then predictions for time estimates from the event-memory point of view would be similar to those described in the previous experiment. That is, there is no a priori reason to expect overestimation bias in the retrieval condition if time estimates are based on a fewer (or equivalent) number of items retrieved from memory from the overall target interval. However, if retrieval leads to internal context change, then we expected significant overestimation of the duration in the retrieval group compared with the restudy group.

Method

Participants. A total of 60 UNCG undergraduates participated in this experiment for course credit. They were randomly assigned to retrieval and restudy groups, with 30 participants per group. All testing was done individually.

Materials. A set of 36 unrelated English nouns of medium frequency were randomly divided into three lists of 12 words each. The presentation order of the lists was counterbalanced.

Procedure. The instructions specified that several lists of words would be presented and that participants should memorize them for a later test by performing a yes/no pleasantness rating task on all the words. The beginning of presentation was again marked with a red colored screen and a tone. Items were presented at a rate of 4 s, separated by 1 s of interstimulus interval. After the last word of L1 was presented, a blank screen appeared along with prerecorded instructions specifying that L1 had been presented and that L2 is about to be shown. Participants were reminded to perform the pleasantness rating task on the words. After L2 presentation, half the participants were given a sheet of paper that contained the first two letters of L1 words and were asked to recall L1 words. The two letter cues were unique to L1 words. One minute was allotted for retrieval of L1, although participants were unaware of the length of recall task. They were told to begin recall until the experimenter told them to stop. While participants in the retrieval group were told to recall L1, the remaining participants were told that they would see L1 words again, and that they should again rate them for pleasantness. Afterwards, all participants studied L3, the end of which was marked by a screen color change. At that point, participants were asked to write down how much time (in minutes and seconds) had elapsed from the moment they saw the red screen and heard the tone. The objective length of the estimated interval was 4.5 min. After providing the time estimate, participants were given 90 s to free recall L2 on a sheet of paper.

Results and Discussion

Although the primary goal of the study was to examine time estimates across the two groups, we first evaluated accurate L2 recall along with L3 intrusions (summarized in Figure 3), followed by time estimation findings.

Recall findings. The retrieval group was asked to retrieve L1 in between encoding L2 and L3, and they recalled $M = .69$, $SD = .21$ of L1 words. The intervening retrieval manipulation led to forgetting of L2 during the final test, as evidenced by a significantly lower proportion of L2 items recalled in the retrieval group ($M = .15$, $SE = .02$) compared with the restudy group ($M = .30$, $SE = .04$), $t(58) = 3.52$, $p < .01$, Cohen's $d = .91$. In previous research, we found that retrieval manipulation lowers intrusions from L3 during L2 recall on the final test compared with the control condition (Sahakyan & Hendricks, 2012). Despite a different control group in the current design (restudy group as opposed to a math group used in previous research), a similar pattern was observed also in the current study, albeit the results fell short of conventional significance. There were more L3 intrusions during L2 recall in the restudy condition ($M = .06$, $SE = .03$) than in the retrieval condition ($M = .01$, $SE = .001$), $t(58) = 1.89$, $p = .064$. Overall, the recall data are consistent with prior research with this paradigm.

Although forgetting of L2 items along with reduced intrusions from L3 have been overall explained in terms of internal context change arising from retrieval of L1 between the remaining two lists, one might also wonder whether these effects could arise from L2 item inhibition. For example, one could also make a case that in order to successfully retrieve L1 items, participants have to inhibit L2 items, and therefore reduced recall of L2 items during the final test reflects their state of inhibition. If this is the case, then one could expect a negative correlation between L1 recall and L2 recall, because the more L2 items are inhibited, the more successful L1 recall should be. However, correlation between L1 and L2 recall was not significant, and it was in the positive rather than the negative direction, $r(30) = .15$, $p = .44$. In addition, prior research directly manipulated L1 accessibility through variety of approaches, and it did not obtain correlation between the magnitude of L2 recall and L1 recall despite obtaining different levels of L1 recall (e.g., Sahakyan & Hendricks, 2012). Finally, it is really hard to explain reduced L3 intrusions from an inhibition point of view.

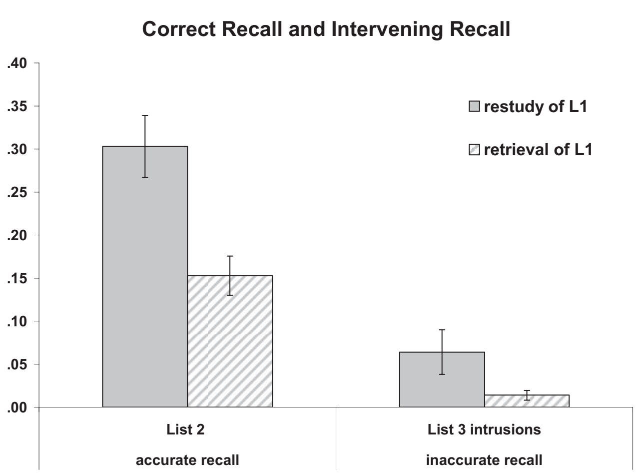

Figure 3. Proportion recalled from List 2 (accurate recall) and List 3 (inaccurate recall) in the retrieval and restudy groups in Experiment 2. The error bars represent ± standard error of the mean. L1 = first list.

If successful retrieval in general requires inhibition of competing items, then the restudy group should also be inhibiting L3 items in order to recall L2 during the final test. However, L3 items are more likely to intrude in the restudy group than in the retrieval group, making the argument for inhibition problematic.

Another alternative interpretation of L2 forgetting might be due to interference from L1 items. For example, one might make a case that had been a final memory test for L1, participants might have recalled more L1 items in the retrieval group than the restudy group (i.e., a testing effect), and the benefit in L1 memory may have been balanced out by the drop in L2 memory. Although we never tested L1 during the final test, it is reasonable to expect that participants with higher L1 memory on the intermediate L1 test would also be likely to recall more L1 items during the final test. Thus, by using an intermediate L1 recall as a proxy, an interference account would predict a negative correlation between L1 and L2 recall, because the more L1 items are available in memory, the more likely they would be to block access to L2 items, thus lowering L2 recall. However, as mentioned above, the correlation between L1 and L2 recall was not significant, and if anything, it was in the positive rather than negative direction. Therefore, an interference account is unlikely to explain L2 forgetting. The overall pattern of recall is more consistent with context-change interpretation than item inhibition or interference interpretations.

Time estimation findings. The critical question was whether the retrieval group would show a significant overestimation bias compared with the control group. In terms of verbal estimates, the retrieval group gave significantly longer time estimates ($M = 5.83$, $SE = .43$) than the restudy group did ($M = 4.37$, $SE = .36$), $t(58) = 2.60$, $p < .05$, Cohen's $d = .67$.

Similar to the previous experiment, the duration judgment ratios of subjective to objective duration were positively skewed (average skewness = 2.52), and therefore we log transformed the scores. The retrieval group significantly overestimated the length of the interval ($M = .08$, $SE = .03$, significantly different from zero), $t(29) = 2.89$, $p < .01$, Cohen's $d = .75$. In contrast, there was a tendency for underestimation in the restudy group ($M = -.05$, $SE = .04$), although the effect was not significant, $t(29) = 1.54$, $p = .14$. Comparison of duration ratios across the two groups revealed significant differences, $t(58) = 3.02$, $p < .01$, Cohen's $d = .78$. Overall, the findings support the hypothesis that internal context change leads to significant overestimation of time, and they are problematic for the inhibitory interpretation of list-before-last paradigm.

General Discussion

Using two different memory paradigms, we demonstrated that the conditions experiencing internal context change retrospectively overestimated the length of time that had passed in the experiment. In Experiment 1, the Forget group overestimated the duration compared with the Remember group. In Experiment 2, the retrieval group estimated that the experiment lasted longer than the restudy group. Although context was not directly manipulated in these paradigms, the conditions attributed to internal context change experienced significant time distortion and overestimated the magnitude of elapsed time. These findings suggest that retrospective time estimates could be used as a marker of internal context change, and they could indicate whether an overall change in internal or mental context has taken place during the experiment.

One might be tempted to attribute longer time estimates obtained in our studies to the task difficulty rather than internal context change. For example, an argument might be made that some of our tasks are more cognitively demanding than others (e.g., engaging in retrieval vs. restudying). However, the difficulty argument is hard to make for both experiments. Experiment 1 involved instructions to forget, and it is not clear a priori why letting go of information should be more demanding than retaining that information in the control group. Yet the Forget group gave longer time estimates than the Remember group. Most importantly, the meta-analysis of time estimation literature indicates that task difficulty does not affect retrospective duration judgments, although it does affect prospective duration judgments, which are the time judgments people expect to make at the completion of the task (Block, Hancock, & Zakay, 2010; Block & Zakay, 1997). Thus, our results are unlikely to be driven by the task difficulty. They are consistent with empirical findings from the time estimation literature, in which explicit manipulations of environmental context (e.g., Block, 1982), process context (e.g., Block & Reed, 1978), or task segmentation (Poynter, 1983; Zakay et al., 1994) led to subjective lengthening of time.

Although our results are inconsistent with the event-memory hypothesis of retrospective timing at the overall group level, we nevertheless assessed the correlation between the proportion of items recalled and duration estimates. In Experiment 1, we tested recall of both L2 and L3. However, because participants recalled only one of the lists, assessing correlations using only L2 recall or L3 recall was problematic due to few data points in each group. Hence, to maximize the number of observations, we computed the correlation between time estimates and recall collapsed across the lists in the Forget group and in the Remember group. There was no significant relationship in either group (in the Forget group, Pearson's $r = -.09$, $p = .52$; in the Remember group, Pearson's $r = .21$, $p = .13$). In Experiment 2, we only tested recall of L2, and the correlation between time estimates and memory is based only on L2 proportion recalled. There was also no significant relationship either in the retrieval group (Pearson's $r = .12$, $p = .53$) or in the restudy group (Pearson's $r = .07$, $p = .73$). Overall, these findings indicate that memory for the events of the interval were not related to the time estimates. To some extent, these findings are not surprising, given that time assessment were made *before* memory test, and thus memory was unlikely to influence results. Nevertheless, participants may have engaged in some cursory retrieval of information in order to come up with time judgments, and thus various memory heuristics (e.g., number of items immediately retrievable from memory, or the speed with which it comes to mind) could have played a role in time judgments. Our data do not speak to these issues, and additional research is needed. Although we did not directly assess response latencies in either of our experiments, previous research examined response latencies both in the directed forgetting paradigm (Spillers & Unsworth, 2011) and list-before-last paradigm (Unsworth et al., 2012), and revealed longer response latency both in the Forget group and the retrieval group of the respective paradigms. However, in both cases, longer response latencies were driven by intrusion errors in recall. Specifically, as participants initiated recall, they intruded many items from the intervening lists, indicating that they were searching through a larger search set during retrieval (e.g., Rohrer & Wixted, 1994). Thus, longer response latencies in both paradigms were attributed to outputting erroneous items during the recall of a designated list as opposed to difficulty initiating retrieval. Nevertheless, future research is needed to examine the role of memory heuristics in the time estimation process.

Although Block and colleagues proposed the importance of contextual information in retrospective estimates, they also recognized that the difficulty with the contextual-change model of retrospective timing is that its explanation tends to be circular because it does not propose a way of measuring the amount of change in cognitive context independent of time estimates (e.g., Block & Reed, 1978, p. 665). Our results suggest a new approach toward escaping the circularity by building on memory research (rather than time research) and the paradigms that are explained by memory researchers in terms of mental context change. Obtaining converging evidence through examination of memory performance and retrospective time estimates enhances the theoretical understanding of mental context change across both memory field and time perception field.

Using time estimates as a marker of internal context change can be a useful tool for designing targeted memory experiments. For example, during the course of investigating directed forgetting, our lab invented what has become known as the mental context-change paradigm (or the diversionary thought paradigm), whereby between the two lists, participants are asked to imagine various things unrelated to the experiment, such as being invisible (e.g., Sahakyan & Kelley, 2002), or daydreaming about vacations (e.g., Delaney, Sahakyan, Kelley, & Zimmerman, 2010). Such manipulations typically produce effects mimicking directed forgetting (for a review, see Sahakyan et al., 2013). Our lab is currently using a time estimation approach to predict a priori which types of distracting thoughts between the two lists are likely to lead to changes in memory.

References

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38,* 341–380. doi:10.1006/jmla.1997.2553

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123. doi:10.1037/h0033773

Avni-Babad, D., & Ritov, I. (2003). Routine and the perception of time. *Journal of Experimental Psychology: General, 132,* 543–550. doi:10.1037/0096-3445.132.4.543

Bailey, N., & Areni, C. S. (2006). Background music as a quasi clock in retrospective duration judgments. *Perceptual and Motor Skills, 102,* 435–444. doi:10.2466/pms.102.2.435-444

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Block, R. A. (1982). Temporal judgments and contextual change. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 530–544. doi:10.1037/0278-7393.8.6.530

Block, R. A. (1990). Models of psychological time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 1–35). Hillsdale, NJ: Erlbaum.

Block, R. A. (1992). Prospective and retrospective duration judgment: The role of information processing and memory. In F. Macar, V. Pouthas, & W. I. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 141–152). Dordrecht, the Netherlands: Kluwer.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica, 134,* 330–343. doi:10.1016/j.actpsy.2010.03.006

Block, R. A., & Reed, M. A. (1978). Remembered duration: Evidence for a contextual-change hypothesis. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 656–665. doi:10.1037/0278-7393.4.6.656

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review, 4,* 184–197. doi:10.3758/BF03209393

Boltz, M. (1993). Time estimation and expectancies. *Memory & Cognition, 21,* 853–863. doi:10.3758/BF03202753

Boltz, M. G. (1998). Task predictability and remembered duration. *Perception & Psychophysics, 60,* 768–784. doi:10.3758/BF03206062

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (p. 85–121). New York, NY: Wiley.

Delaney, P. F., Sahakyan, L., Kelley, C. M., & Zimmerman, C. (2010). Remembering to forget: The amnesic effect of daydreaming. *Psychological Science, 21,* 1036–1042.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478. doi:10.1037/0033-295X.108.2.452

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62,* 145–154. doi:10.1037/h0048509

Geiselman, R. E., Bjork, R. A., & Fishman, D. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General, 112,* 58–72. doi:10.1037/0096-3445.112.1.58

Hintzman, D. L. (2002). Context matching and judgments of recency. *Psychonomic Bulletin & Review, 9,* 368–374. doi:10.3758/BF03196295

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Contextual associations and memory for serial position. *Journal of Experimental Psychology, 97,* 220–229.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299. doi:10.1006/jmps.2001.1388

James, W. (1890). *The principles of psychology* (Vol. 1). New York, NY: Holt. doi:10.1037/11059-000

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 112–127. doi:10.1037/0278-7393.34.1.112

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 970–988. doi:10.1037/a0015728

Mensink, G., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434–455. doi:10.1037/0033-295X.95.4.434

Ornstein, R. E. (1969). *On the experience of time*. Harmondsworth, United Kingdom: Penguin.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129–156. doi:10.1037/a0014420

Poynter, W. D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition, 11,* 77–82. doi:10.3758/BF03197664

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. doi:10.1037/0033-295X.88.2.93

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition, 22,* 511–524. doi:10.3758/BF03198390

Roy, M. M., & Christenfeld, N. J. S. (2007). Bias in memory predicts bias in future task duration. *Memory & Cognition, 35,* 557–564. doi:10.3758/BF03193294

Roy, M. M., & Christenfeld, N. J. S. (2008). Effect of task length on remembered and predicted duration. *Psychonomic Bulletin & Review, 15,* 202–207. doi:10.3758/PBR.15.1.202

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2013). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 131–190). Amsterdam, the Netherlands: Elsevier.

Sahakyan, L., & Hendricks, H. E. (2012). Context change and retrieval difficulty in the list-before-the-last paradigm. *Memory & Cognition, 40,* 844–860. doi:10.3758/s13421-012-0198-0

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1064–1072. doi:10.1037/0278-7393.28.6.1064

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115,* 893–912. doi:10.1037/a0013396

Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science, 168,* 1601–1603. doi:10.1126/science.168.3939.1601

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8,* 203–220. doi:10.3758/BF03196157

Spillers, G. J., & Unsworth, N. (2011). Are the costs of directed forgetting due to failures of sampling or recovery? Exploring the dynamics of recall in list-method directed forgetting. *Memory & Cognition, 39,* 403–411. doi:10.3758/s13421-010-0038-z

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Evidence of noisy contextual search: Examining the dynamics of list-before-last recall. *Memory, 20,* 1–13. doi:10.1080/09658211.2011.626430

Ward, G., & Tan, L. (2004). The effect of the length of to-be-remembered lists and intervening lists on free recall: A re-examination using overt rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1196–1210. doi:10.1037/0278-7393.30.6.1196

Zakay, D., Tsal, Y., Moses, M., & Shahar, I. (1994). The role of segmentation in prospective and retrospective time estimation processes. *Memory & Cognition, 22,* 344–351. doi:10.3758/BF03200861

Received December 23, 2012
Revision received June 24, 2013
Accepted July 17, 2013 ■