Escaping the recent past: Which stimulus dimensions influence proactive interference?

Kimberly S. Craig · Marc G. Berman · John Jonides · Cindy Lustig

Abstract Proactive interference occurs when information from the past disrupts current processing and is a major source of confusion and errors in short-term memory (STM; Wickens, Born, & Allen, Journal of Verbal Learning and Verbal Behavior, 2:440–445, 1963). The present investigation examines potential boundary conditions for interference, testing the hypothesis that potential competitors must be similar along task-relevant dimensions to influence proactive interference effects. We manipulated both the type of task being completed (Experiments 1, 2, and 3) and dimensions of similarity irrelevant to the current task (Experiments 4 and 5) to determine how the recent presentation of a probe item would affect the speed with which participants could reject that item. Experiments 1, 2, and 3 contrasted STM judgments, which require temporal information, with semantic and perceptual judgments, for which temporal information is irrelevant. In Experiments 4 and 5, task-irrelevant information (perceptual similarity) was manipulated within the recent probes task. We found that interference from past items affected STM task performance but did not affect performance in semantic or perceptual judgment tasks. Conversely, similarity along a nominally irrelevant perceptual dimension did not affect the magnitude of interference in STM tasks. Results are consistent with the view that items in STM are represented by noisy codes consisting of multiple dimensions and that interference occurs when items are similar to each other and, thus, compete along the dimensions relevant to target selection.

Keywords Interference · Inhibition in memory retrieval · Memory · Recognition · Short term memory · Working memory

K. S. Craig (✉) · J. Jonides · C. Lustig (✉)
University of Michigan, 530 Church Street,
Ann Arbor, MI 48109, USA
e-mail: kscraig4@gmail.com
e-mail: clustig@umich.edu

M. G. Berman
University of South Carolina, Columbia, SC 29208, USA

M. G. Berman
Rotman Research Institute, Baycrest, Toronto, M6A 2E1, ON, Canada

Imagine that you are waiting tables at a busy restaurant. One of your customers initially orders the nachos appetizer but, after a few seconds, changes the order to mozzarella sticks. When you return to the kitchen, you see a plate of nachos waiting and begin to pick it up, but then you realize your error, set it down, and wait for the mozzarella sticks. The temporary confusion you experienced in this situation is an example of proactive interference: information from the past disrupting current memory performance. Proactive interference is one of the major mechanisms of forgetting in both short-term memory (STM) and long-term memory (LTM) (Underwood, 1957). In the present article, we test two common explanations for proactive interference: activation strength and similarity-based response competition.

Several models describe the contents of STM in terms of activated representations from LTM and/or perception (e.g., Cowan, 2001; McElree, 2001; Oberauer, 2002; see Jonides et al., 2008, for a review). These models generally consist of a *focus of attention* that contains the target(s) of current processing and a *region of direct access* that contains previously relevant or potentially relevant items that exceed the capacity of the focus of attention but maintain high levels of activation. The items in the region of direct access have the potential to easily move into the focus and compete with the target for access to processing resources. In such models, recently presented items remain in the region of direct access, and so they retain high activation/familiarity, allowing proactive interference to occur. The decay or active suppression of this residual activation may be necessary to prevent or overcome interference effects (e.g., Altmann & Gray, 2002; Anderson & Spellman, 1995).

Other explanations of interference emphasize the critical role that similarity plays in producing competition (Keppel & Underwood, 1962; Watkins & Watkins, 1975; Wickens, 1970; for reviews, see Crowder, 1976; Lustig et al., 2009). From this perspective, response competition may occur only when some form of similarity exists between target and nontarget items (Underwood, 1945). For example, memory for a list of adjectives may be poor if preceding lists also consisted of adjectives (and therefore can interfere). However, if the previous list consisted of unrelated information (e.g., three-digit numbers), memory for the list can be approximately as good as if no prior list had been studied (e.g., Johnson, 1933; McGeoch & McDonald, 1931; for reviews, see Crowder, 1976; Keppel, 1968; Underwood, 1945, 1957). The release-from-proactive-interference procedure (Wickens, Born, & Allen, 1963) provides a classic demonstration of the role of similarity in STM interference: Performance declines within as few as four trials if all trials use materials of the same class (e.g., letters vs. numbers). However, performance is "released" from this detrimental effect if the fourth study item is drawn from the other class of materials (i.e., after seeing three sequences of letters, a sequence of numbers would be remembered better than a fourth sequence of letters). Comparable buildup and release from proactive interference effects can also be seen in more modern working memory tasks, such as operation span (Bunting, 2006), and similarity-based interference is also observed in the short-term version of the false memory task (Atkins & Reuter-Lorenz, 2008; Flegal, Atkins, & Reuter-Lorenz, 2010).

New data from a modified version of the recent probes task suggest that activation strength (due to recent presentation) and similarity-based competition may each contribute to proactive interference (Atkins, Berman, Reuter-Lorenz, Lewis, & Jonides, 2011). On each trial of the standard version of the recent probes task (Jonides, Smith, Marshuetz, Koeppe, & Reuter-Lorenz, 1998; Monsell, 1978), participants study a set of four words displayed for several seconds (see STM trials in Fig. 1). Following a delay, a probe word appears, and participants are asked to indicate whether the word was part of the current trial's memory set. The speed at which participants can reject negative probes (those not part of the current memory set) is influenced by the contents of prior trials. In particular, participants are slower to reject a probe if it was a member of the memory set on the previous trial (a recent negative) than probes that have not been recently seen (nonrecent negatives). These results appear to support activation strength accounts, since the recent probes presumably retain strong residual activation, making it more difficult to reject the "yes" response in favor of the accurate "no" response.

To examine the contributions of recent activation and semantic similarity to interference in STM, Atkins et al.

(2011) manipulated the degree to which the memory set and the probe item were semantically similar. On critical trials, all of the memory set items were drawn from the same semantic category (countries or fruits); the probe item was then drawn from either that category or the complementary one. The *mismatch* trials, on which the probe and memory set items come from separate categories, were obviously negative trials and should have allowed participants to reject the probe immediately. For example, if given the memory set *Canada*, *France*, *Australia*, *Brazil*, participants should have been able to immediately reject the negative probe *orange* because of the category mismatch. Notably, although proactive interference was reduced on these *mismatch* negative trials, as compared with *match* negative trials (where the memory set and probe item came from the same category, but the probe item was not a member of the current memory set), the recent probes effect was not eliminated. Even on mismatch trials, participants were slower to reject recent negatives than nonrecent ones, indicating that temporal recency still produced interference despite a complete semantic mismatch. In addition, recency and semantic similarity had similarly-sized effects on interference: The time needed to reject recent negatives on category mismatch trials was equivalent to the time needed to reject nonrecent negatives on category match trials. These results, along with the very long response times needed to reject items that were both recent and category matches to the memory set, suggest that recent activation and semantic similarity each make separate, possibly equivalent, contributions to interference.

The remaining interference on category mismatch recent-negative trials is remarkable. It appears to provide strong evidence for residual activation strength as a source of interference and suggests that interference resulting from this residual activation may be hard to escape. However, similarity-based explanations of interference offer an alternative account. Rather than explaining recency-based interference as the result of residual activation strength, they note that recently presented probe items are very similar to the current target memory set in terms of when they were presented. (Similarity in trial order may be more important than similarity in time per se; see Berman, Jonides, & Lewis, 2009.) Furthermore, these temporal or trial order characteristics are exactly the ones that are critical to making the decision required by the task; that is, the participant's decision as to whether to accept or reject the probe depends on whether it belongs to the current memory set (accept) or a previous one (reject). Recently presented items may be hard to reject not because they are still highly activated, but rather because they are very similar to the target set in terms of when they were presented.

The present experiments were designed to test how similarity along task-relevant dimensions affects the occurrence
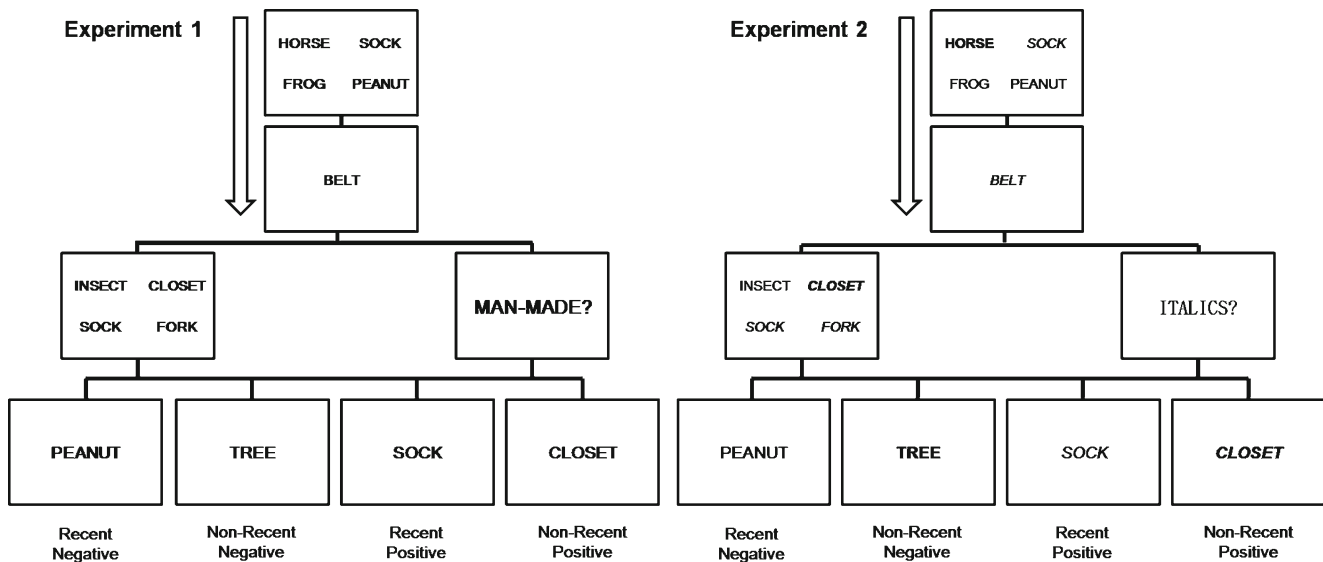
Fig. 1 Sample trial sequences for Experiments 1 and 2; time progresses linearly from top to bottom. The trial type is indicated below each critical probe (bottom row). In both experiments, the first trial of a critical sequence was a short-term memory (STM) trial, during which participants saw a set of four words followed by a probe word and indicated whether or not the probe was part of the current memory set. The second trial in the sequence could be either another STM trial—in which case, the same procedure as that on the initial trial was followed (four memory set words followed by probe word)—or a category judgment trial. For category judgment trials, instead of a memory set,

the category label was displayed, and the participant was to decide whether or not the probe was a member of that category. In Experiment 1, the category judgment relied on semantic information (e.g., is a peanut a man-made object?); in Experiment 2, the category judgment relied on perceptual information (e.g., is the word PEANUT shown in italicized font?). Probe words varied (2 × 2 design) in whether they were positive or negative (members or nonmembers of the current memory or category set) and whether their prior presentation was recent (on the previous trial; e.g., PEANUT, SOCK) or nonrecent (not present on the previous trial; e.g., TREE, CLOSET)

and degree of proactive interference in STM. The first three experiments examined whether changing the judgment to be made on the probe item (either requiring the use of temporal order information or not) would influence the amount of proactive interference observed. The final two experiments kept the requirement to use temporal order information constant and tested whether proactive interference effects varied with manipulations of similarity along other dimensions. Together, the results point to similarity along task-relevant dimensions as a critical factor in producing proactive interference.

Experiment 1

The purpose of Experiment 1 was to test whether recent activation of an item is enough to slow its rejection or whether this slowing can be eliminated by removing the need to consider temporal order information. To this end, we compared the magnitude of the recent-negative effect on STM trials, which require the use of temporal order information, with its magnitude on semantic memory trials, which do not.

Semantic judgments were chosen as the comparison task on the basis of a large body of research suggesting that the speed of semantic judgments can be influenced by recent activation. Repeated semantic category judgments of the

same stimuli result in shorter response times and reduced neural activations, which are thought to represent a reduction in the attentional requirements of searching for and activating the representation of the to-be-judged item (e.g., Buckner et al., 1995). These effects are extremely robust and widely studied under the rubric of repetition priming (see Henson & Rugg, 2003, for a review). Importantly, they are not isolated to the perceptual or response levels; changes in perceptual presentation, response mapping, or the specific category to be judged reduce but do not eliminate the benefits of recent activation (e.g., O'Kane, Insler, & Wagner, 2005). Important for the present experiment, these behavioral priming and neural effects are sensitive to the "lag," or number of intervening items between presentations (e.g., Henson, Rugg, Shallice, & Dolan, 2000; compare with the STM lag results of Berman et al., 2009). Recent presentation of a semantically related item can significantly impact the speed of responses in lexical (or other) decisions about a current word, as shown in semantic priming tasks (Collins & Loftus, 1975; Masson, 1995; Meyer & Schvaneveldt, 1971). In short, there is substantial evidence that recent presentation of an item can facilitate its acceptance in a semantic judgment context; the present experiment tested whether it can also interfere with its rejection.

We hypothesized that if residual activation is what makes participants slow to reject recent-negative items, this effect

should occur equally for STM judgment and semantic judgment trials. On the other hand, if proactive interference occurs only when temporal information is relevant to the task requirements, participants should be equally fast in rejecting recent- and nonrecent-negative probes on semantic judgment trials, for which temporal order information is not relevant.

Method

Participants

Forty individuals (18 female; average age = 18.73 years, SD = 1.09) participated in this experiment. All individuals were recruited through the University of Michigan Subject Pool and received course credit for participation. For this and all subsequent experiments, exclusion criteria included failure to pass screening measures (medication or health conditions that could affect cognition) and/or a score less than 9 (out of a possible 48) on the Extended Range Vocabulary Test (ERVT, Version 3; Educational Testing Service, 1976) or failure to maintain at least 80 % accuracy on both STM and category trials. The ERVT was used to screen for participants with low verbal ability (since the memoranda were words) or who were generally noncompliant and not putting effort into correctly completing the experimental tasks. Our lab generally uses a cutoff score of 9 (out of 48 possible) to screen out such participants in both verbal and nonverbal tasks (see also Demeter, Sarter, & Lustig, 2008; Lustig & Flegal, 2008). In Experiment 1, 6 participants were excluded due to health conditions and/or current medications, 3 participants were excluded due to poor performance on the ERVT, and data from 2 participants were lost for technical reasons. Twenty-nine healthy individuals (13 female) were included in the final analysis. These participants had ERVT scores ranging from 9.75 to 31.00 (M = 18.87, SD = 6.58). They had an average age of 18.62 years (SD = 1.01) and had completed an average of 12.55 years of formal education (SD = 0.74).

Design and materials

All aspects of the research were approved by the Behavioral Sciences Institutional Review Board at the University of Michigan. Stimuli were displayed in 18-point bold MS Sans Serif font using E-Prime 2.0 software (Psychology Software Tools, Inc.).

Trials consisted of two distinct types: STM trials and semantic category trials (Fig. 1). Each STM trial consisted of a black fixation cross appearing for 2,000 ms, followed by a red fixation cross, which appeared for 1,000 ms and was accompanied by an alerting tone. The target set of four words was then presented for 2,000 ms, followed by a 3,000-ms delay before presentation of the probe word in the center of the screen. The probe word appeared for 2,000 ms or until the participant made a keypress response on a standard computer keyboard indicating whether it was (positive probe) or was not (negative probe) a member of the current memory set. A keypress of "1" indicated a positive response, while a keypress of "0" indicated a negative response. Participants were instructed to perform all keypresses with the left and right index fingers.

Semantic category trials proceeded in the same manner as STM trials, with one exception. Following the red fixation cross and warning tone, instead of a set of four words, a semantic category judgment prompt appeared for the same duration. The category prompts ("MAN-MADE?" or "LARGER THAN A COMPUTER SCREEN?") indicated which category dimension was relevant. Two categories were used, rather than one, to reduce the likelihood that contrasts with STM trials were an artifact of the particular category chosen and to discourage participants from covertly making the category judgments on the items when given the memory set. As on the STM trials, when the probe appeared, participants were to make a keypress of "1" to indicate a positive response or a keypress of "0" to indicate a negative response, and participants were instructed to perform all keypresses with the left and right index fingers. In short, semantic category trials were procedurally identical to STM trials, with the exception that the probe was to be judged on category membership rather than memory set membership.

For *recent* trials, the probe was a member of the previous trial's memory set; that is, both category and STM recent trials were always preceded by an STM trial. However, preceding trial type did not allow a participant to predict the current trial's type (category or STM), recency, or correct response: Nonrecent trials could be preceded by either an STM or a category trial, and both recent and nonrecent trials could be either positive or negative. (Experiment 3 directly addresses the concern that the preceding trial type constraint might have led to confounds related to task switching.) Trials were distributed evenly across a 2 (trial type: STM or category) × 2 (recency: recent, nonrecent) × 2 (correct response: positive, negative) design and were presented pseudorandomly, with the constraint that no more than three responses of one type (positive or negative) could occur in a row.

The categories and words used here were drawn from those used by Braver, Reynolds and Donaldson (2003). All words were chosen from this pool and then judged by two independent raters to be unambiguous with regard to category membership. Four groups of words were created: small and man-made items, small and natural items, large and man-made items, and large and natural items. Both STM and semantic category trials could be classified along two dimensions, probe type and recency. Probe type was either

positive or negative depending on whether or not the probe was a member of the currently relevant memory or category set. Recency was defined by membership on previous trials: Recent probes were members of the previous trial's memory set; nonrecent probes had not appeared as memory set members (or as probes) for at least three trials prior to the current trial. All trial types were randomly interspersed throughout each of four blocks. Block order was counterbalanced across participants, using an approximate Latin square design; that is, although participants were initially assigned to block orders according to the Latin square, dropout was slightly uneven across cells. Secondary analyses for this and all other experiments showed that block order did not influence the results (all $ps > .20$).

Procedure

After providing written informed consent, all participants completed practice on the task before beginning the experiment. Practice consisted of four STM trials and two category trials, one of each category type. Fewer category trials were used in practice because of the limited stimulus pool; the main intent of the practice was to ensure that participants understood the task. Participants were able to repeat practice if desired. Following practice, each participant completed four blocks of 64 trials, with 60 s of rest in between blocks.

Results and discussion

For this and all subsequent experiments, response time analyses were limited to correct responses falling within 3 standard deviations of the median response time for that individual and trial type. The total percentage of trials removed as outliers varied by experiment and was between 1.43 % and 2.37 % of the total number of trials completed across participants. Median (rather than mean) response times were analyzed to further reduce the possibility that an individual's results might be unduly influenced by outlying values.

Proactive interference in the recent probes paradigm is indexed by the contrast between recent- and nonrecent-negative probes, and so these were the focus of our analyses. For completeness, means and standard deviations for all trial types are given in Table 1; for statistics on positive trials, see Table 2.

Both response time and accuracy measures were analyzed using a repeated measures design, with two independent variables (recency, recent or nonrecent; trial type, STM or category). Because we used a repeated measures design, effect sizes are reported in generalized η^2 values (abbreviated as η_G^2), rather than partial η^2 values. Effect size heuristics for η_G^2 are as follows: .02 is a small effect, .13 a medium effect, and .26 a large

effect (Bakeman, 2005). To calculate η_G^2, we used the following formula: $\eta_G^2 = SS_{effect} / (SS_{effect} + SS_{subjects})$. Where necessary, Greenhouse–Geisser sphericity corrections were applied to reported p-values; original degrees of freedom are used in the text for easier reading.

Response time

As can be seen in Fig. 2a, the standard recent-negative effect was found for STM trials (cf. Berman et al., 2009; Monsell, 1978), but there was no recent-negative effect on semantic judgment trials [interaction of recency and trial type, $F(1, 28) = 17.45$, $p < .001$, $\eta_G^2 = .02$]. Confirming this impression, post hoc t-tests revealed a significant difference between recent-negative and nonrecent-negative STM trials, $t(28) = 5.56$, $p < .001$, $d = 1.03$, but no difference between recent-negative and nonrecent-negative category trials, $t < 1$.

Although not relevant to our theoretical question, for completeness, we note a statistical main effect of trial type, with negative semantic judgment trials slower, overall, than negative STM trials, $F(1, 28) = 64.57$, $p < .001$, $\eta_G^2 = .34$. There was also a statistical main effect of recency, with recent-negative trials slower than nonrecent-negative trials, $F(1, 28) = 9.87$, $p < .01$, $\eta_G^2 = .01$; as was noted above, this effect was driven by STM trials and did not occur on semantic judgment trials.

Accuracy

As can be seen in Fig. 2b, the accuracy results followed a pattern consistent with the response time data, with a significant interaction between recency and trial type, $F(1, 28) = 9.39$, $p < .01$, $\eta_G^2 = .05$. Participants were less accurate in rejecting recent trials than nonrecent trials in the STM condition, $t(28) = -2.51$, $p = .02$, $d = -0.47$, but showed, if anything, the opposite trend in the semantic judgment condition, $t(28) = 1.96$, $p = .06$, $d = 0.36$.

Just as they were, overall, slower than STM trials, semantic judgment trials were also less accurate [main effect of trial type, $F(1, 28) = 41.77$, $p < .001$, $\eta_G^2 = .24$]. The main effect of recency did not reach statistical significance, due to the interaction effect described above, $F < 1$.

In summary, the response time and accuracy results replicated standard findings of proactive interference effects in STM trials, but there was no evidence of such interference for semantic judgment trials. These data support the hypothesis that the temporal characteristics of a stimulus (i.e., its recency of presentation) are one dimension along which it can be similar to other items and that this similarity will create interference only if that dimension is relevant to the task.

Table 1 Average median response times (RTs, in milliseconds; with standard errors) and accuracy scores (with standard errors) by trial type for both Experiment 1 and Experiment 2

			Recent negative	Nonrecent negative	Recent positive	Nonrecent positive
Experiment 1	Memory	RT	739.78 (14.77)	677.31 (17.25)	686.69 (21.11)	679.81 (19.36)
		Accuracy (%)	96.55 (0.65)	98.28 (0.43)	96.66 (0.60)	96.55 (0.72)
	Category	RT	869.57 (27.12)	881.69 (26.74)	863.33 (27.26)	857.07 (25.32)
		Accuracy (%)	93.86 (0.91)	91.70 (0.96)	89.66 (0.99)	92.03 (0.87)
Experiment 2	Memory	RT	707.15 (19.74)	642.43 (16.10)	669.99 (18.59)	682.66 (18.56)
		Accuracy (%)	96.37 (0.69)	98.40 (0.38)	94.34 (1.05)	95.44 (0.85)
	Category	RT	613.34 (24.49)	607.61 (21.39)	612.76 (24.41)	582.88 (22.35)
		Accuracy (%)	95.19 (0.65)	96.45 (0.65)	92.74 (1.22)	87.58 (1.49)

Experiment 2

Our first experiment indicated that while STM judgments were vulnerable to proactive interference, semantic category judgments were not. This provides evidence that temporal similarity affects trials where temporal information is relevant to the judgment the participant is asked to make (i.e., STM trials) but does not create interference on trials where temporal information is not relevant (i.e., semantic judgment trials). However, perceptual information is often considered more important than semantic information within the context of STM (e.g., Baddeley, 1966, 1986; Baddeley & Hitch, 1974). We therefore asked whether our findings would generalize to judgments based on perceptual, rather than semantic, categorization.

To answer this question, we again interleaved STM trials with category judgment trials. In this experiment, the judgments were based on visual information about the probe item, rather than semantic knowledge. If temporal similarity influences responses regardless of task relevance, we should see the effects of temporal recency on both STM trials and perceptual judgment trials; if, however, proactive interference occurs only when similarity is relevant to the task, temporal recency should affect STM trials but not perceptual judgment trials.

Method

Participants

Fifty-four participants (33 female; average age = 20.37 years, SD = 2.41) participated in this experiment. Two participants were excluded due to medication/health conditions, 5 participants were excluded for failure to reach the criterion ERVT

Table 2 Statistical analyses and effect sizes for Experiment 1 and Experiment 2

	Experiment 1			Experiment 2		
	F value	p value	η_G^2	F value	p value	η_G^2
2 × 2 ANOVA for negative RTs						
Main effect of trial type	64.57	<.001	.34	35.80	<.001	.06
Main effect of recency	9.87	<.01	.01	19.81	<.001	.02
Trial type × recency interaction	17.45	<.001	.02	10.19	<.01	.01
2 × 2 ANOVA for positive RTs						
Main effect of trial type	100.06	<.001	.34	44.59	<.001	.09
Main effect of recency	0.51	.48	<.001	1.41	.24	.001
Trial type × recency interaction	0.00	.97	<.001	10.77	<.01	.007
2 × 2 ANOVA for negative accuracies						
Main effect of trial type	41.77	<.001	.24	6.87	.01	.04
Main effect of recency	0.11	.75	0.00	10.14	<.01	.05
Trial type × recency interaction	9.39	<.01	.05	0.63	.43	.01
2 × 2 ANOVA for positive accuracies						
Main effect of trial type	48.83	<.001	.31	11.33	<.01	.10
Main effect of recency	3.05	.09	.02	8.70	<.01	.02
Trial type × recency interaction	4.09	.05	.02	30.98	<.001	.05

Fig. 2 Average median response times (top panels) and accuracy (bottom panels) for negative trials in Experiments 1 (left) and 2 (right). Error bars on this and subsequent figures represent between-subjects standard errors and should not be used for evaluating the significance of within-subjects comparisons
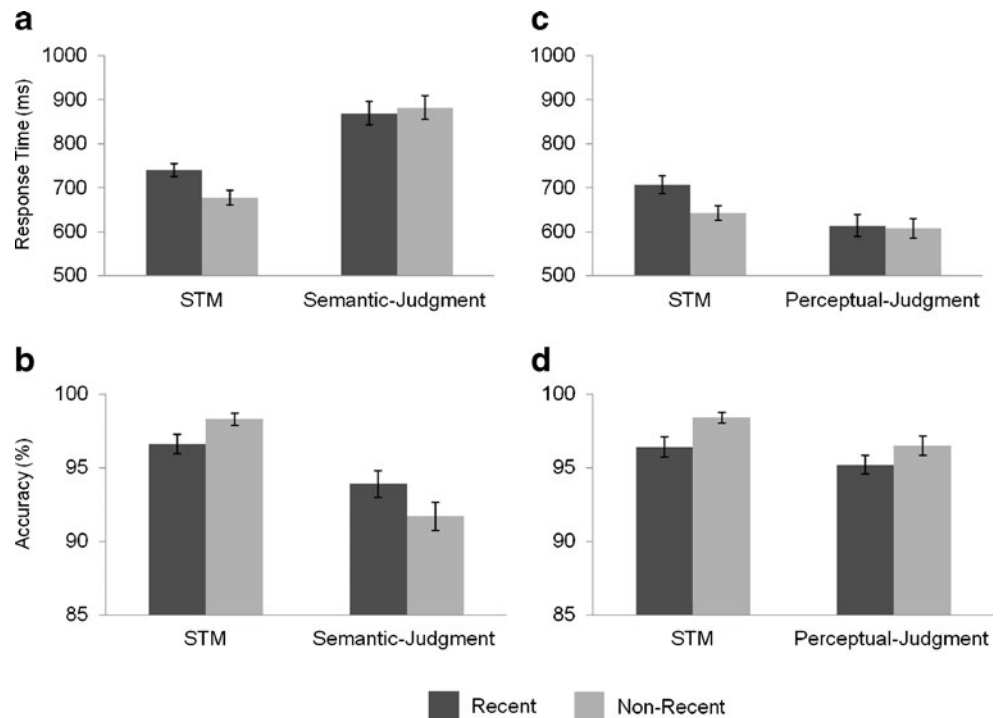
score, and 10 for failing to meet accuracy criteria. While only 2 of these participants had performance below 80 % accuracy for STM trials, all 10 had performance below 80 % for category trials. In particular, participants had a difficult time correctly identifying trials on which they were required to classify the words as italicized or nonitalicized.

Thirty-seven healthy individuals (22 female) were included for analysis. All individuals either received course credit as part of the University of Michigan Subject Pool or were paid for their participation ($15/h). No significant differences were found between paid and unpaid participants in overall response times, $t(35) = 1.00$, $p = .33$, or accuracy, $t < 1$. Participants had a mean age of 20.05 years ($SD = 2.42$) and had completed an average of 13.78 years ($SD = 1.89$) of formal education. ERVT scores for included participants ranged from 9.75 to 39.25 ($M = 18.64$, $SD = 6.42$).

Stimuli

To create consistency in comparing Experiments 1 and 2, all trials from Experiment 1 were repeated exactly, with the exception that the physical characteristics (fonts) of the words were changed and category trials required participants to judge probe items on this basis, rather than semantic category. The words appeared in standard font, italics, bold, or both italics and bold (see Fig. 1). The semantic categories used in Experiment 1 were mapped directly onto the font categories in Experiment 2; that is, a "yes" item for the manmade judgment in Experiment 1 became a "yes" item for the

italics judgment in Experiment 2, while a "yes" item for the larger-than-a-computer-screen judgment in Experiment 1 became a "yes" item for the bold judgment in Experiment 2. Because of this, each item appeared with the same perceptual features each time it appeared. All stimuli were displayed in 16-point font. Nonbold words were displayed in Copperplate Gothic Light font; bold words were displayed in Copperplate Gothic Bold font and also had the bold format option applied. The category cue (ITALICS? BOLD?) was presented in an entirely different font (Courier 18 point) so as not to bias participants toward a particular judgment.

Results and discussion

Response time

As in Experiment 1, proactive interference influenced STM trials, but not category judgment trials, yielding a significant interaction, $F(1, 36) = 10.19$, $p < .01$, $\eta_G^2 = .01$. For STM trials, recent trials took longer than nonrecent trials, $t(36) = 5.88$, $p < .001$, $d = 0.97$; for perceptual judgment trials, the two trial types had similar response times, $t < 1$ (Fig. 2c).

Also replicating Experiment 1, there was a statistically significant main effect of recency, $F(1, 36) = 19.81$, $p < .001$, $\eta_G^2 = .02$, that was driven by the STM trials and did not occur for the perceptual judgment trials. While in Experiment 1 category judgments were slower than STM judgments, here they were faster, $F(1, 36) = 35.80$, $p < .001$, $\eta_G^2 = .06$. The opposite patterns when overall response

times for STM versus category judgments were compared across experiments suggest that their shared finding of interference on STM but not category trials is not easily explained by differences in task difficulty or response time.

Accuracy

Recent trials were less accurate, overall, when compared with nonrecent trials, $F(1, 36) = 10.14$, $p < .01$, $\eta_G^2 = .05$. In addition, STM trials were more accurate than category trials, $F(1, 36) = 6.87$, $p = .01$, $\eta_G^2 = .04$. However, for the accuracy data, trial type did not influence the effect of recency (interaction, $F < 1$). In this experiment, recency impaired the accurate rejection of recent probes for both STM and category trials (Fig. 2d).

This result was surprising in comparison with what we had found in Experiment 1, and so we examined the data more closely. An examination of the perceptual judgments suggested that participants had particular difficulty with the *italics* judgment, being both less accurate, $t(36) = 5.30$, $p < .001$, $d = 0.87$, and slower, $t(36) = -11.43$, $p < .001$, $d = -1.88$, than for judgments about whether it was displayed in bold font. We therefore considered the possibility that recency effects might contaminate the category judgment if that judgment were difficult to make. That is, participants who found the italics dimension difficult to judge may have allowed the nominally irrelevant temporal dimension to influence their response.

To explore this possibility, we split participants into two groups on the basis of their relative accuracy on italics judgments. Specifically, we calculated a difference score for each participant's accuracy on bold judgments versus italics judgments; those participants with difficulty making italics judgments (>5 % accuracy difference between judgments) made up the less accurate group ($n = 19$), as determined by a median split. The more accurate group ($n = 18$) had comparable accuracy scores on both judgment types or performed better in the italics condition. We used relative rather than absolute accuracy on the italics judgment as the basis for group membership to distinguish specific problems with the italics judgment from general low performance (which might be influenced by motivation, fatigue, or other factors). Descriptive statistics for each group are presented in Table 3.

When the analysis was limited to participants in the group with similar accuracies for the two category judgments, the results more closely replicated those seen in Experiment 1. For this subset, there was a difference between recent and nonrecent trials for STM accuracy, $t(17) = -2.41$, $p = .03$, $d = -0.57$, but not for category accuracy, $t < 1$. In contrast, for the group that had difficulty with (low accuracy on) italics judgments, there was no difference between recent and nonrecent trials for either STM accuracy, $t(18) = -1.57$, n.s., $d = -0.36$, or category accuracy, $t(18) = -1.60$, n.s., $d = -0.37$.

These patterns suggest a potential boundary condition on our proposal that interference depends on similarity on the task-relevant dimensions. That is, if a participant has difficulty with making the judgment on task-relevant dimensions, information from other dimensions (in this case, the temporal dimension) may influence or contaminate the judgment. This possibility, while interesting, is post hoc and somewhat tangential to the main thrust of our experiments. We therefore do not discuss it extensively here but, for the interested reader, present further analyses exploring the issue (including response time data) in the Appendix.

In summary, the results of Experiment 2 replicated the important aspects of Experiment 1, especially with regard to response time. These findings provide further support for the hypothesis that temporal similarity creates proactive interference on tasks where temporal information is relevant but does not create proactive interference when temporal information is irrelevant to the task—at least, when participants are performing that task well.

Experiment 3

Experiments 1 and 2 provide support for the idea that similar information must be along a task-relevant dimension in order to create interference. However, because *recent* probes were defined as those that had been in the immediately-prior STM trial's memory set, there was a potential confound in the design: Because all *recent* trials were preceded by an STM trial, recent STM trials were always preceded by the same trial type, whereas recent category trials were always preceded by the other trial type. To test whether this *task-switching* aspect of our design influenced the results, Experiment 3 modified the procedures so that either STM or category trials could serve as a source of recency for the subsequent trial.

Method

Participants

Forty-four individuals (22 female; average age = 20.61 years, $SD = 2.48$) participated in this experiment. Three were excluded due to failure to adhere to instructions on color task mapping, 2 due to failing to meet accuracy criteria, 1 due to technical problems, and 8 for failing to meet the minimum ERVT score.

Thirty healthy individuals (15 female; average age = 20.50 years, $SD = 2.26$) were included for analysis. All individuals received course credit for participation in the study. Participants had completed an average of 14.10 years ($SD = 1.67$) of formal education and scored between 9.75 and 35.25 on the ERVT ($M = 20.06$, $SD = 6.04$).

Table 3 Average median response times (RTs, in milliseconds; with standard errors) and accuracies (with standard errors) for high- and low-performing groups in Experiment 2, as determined by a comparison between italics and bold category judgments

			Recent negative	Nonrecent negative	Recent positive	Nonrecent positive
High-accuracy group	Memory	RT	724.97 (31.90)	653.67 (25.15)	684.50 (29.70)	696.78 (30.43)
		Accuracy (%)	95.66 (1.24)	98.44 (0.58)	93.23 (1.70)	95.49 (1.16)
	Category	RT	617.22 (32.12)	620.86 (35.67)	633.86 (36.81)	611.25 (35.93)
		Accuracy (%)	95.66 (1.02)	96.35 (1.08)	94.97 (1.58)	91.67 (0.98)
Low-Accuracy group	Memory	RT	690.26 (23.97)	631.79 (20.74)	656.24 (23.14)	669.29 (22.21)
		Accuracy (%)	97.04 (0.65)	98.36 (0.50)	95.40 (1.27)	95.40 (1.27)
	Category	RT	609.66 (37.57)	595.05 (25.00)	592.76 (32.58)	556.00 (26.60)
		Accuracy (%)	94.74 (0.83)	96.55 (0.79)	90.63 (1.76)	83.72 (2.46)

Stimuli

The word pool used as verbal stimuli in Experiments 1 and 2 was also used here. As in our prior experiments, participants completed both STM and category judgment tasks. However, the trial structure was altered so that category as well as STM trials could serve as a source of recency (Fig. 3). Due to the constraints of this procedure, only one category judgment ("Manmade?") was used, rather than two.

Each trial, regardless of type, began with a display of four items presented for 2,000 ms. Following this set, a colored square outline appeared for 1,500 ms, with a fixation cross

centered within it. The square appeared in either red or blue, and the color indicated the task (STM or category judgment) the participant should perform. Each color was mapped to the same task throughout the entire experiment, and the mapping was counterbalanced across participants. After the colored square, a probe word appeared, and participants responded with a buttonpress, either "1" or "0," to indicate a "yes" or "no" response to the probe. The mappings of the buttonpresses were also counterbalanced across participants.

As before, the critical comparisons were between recent and nonrecent probes. The altered structure of the category trials now allowed items from those trials to serve as a
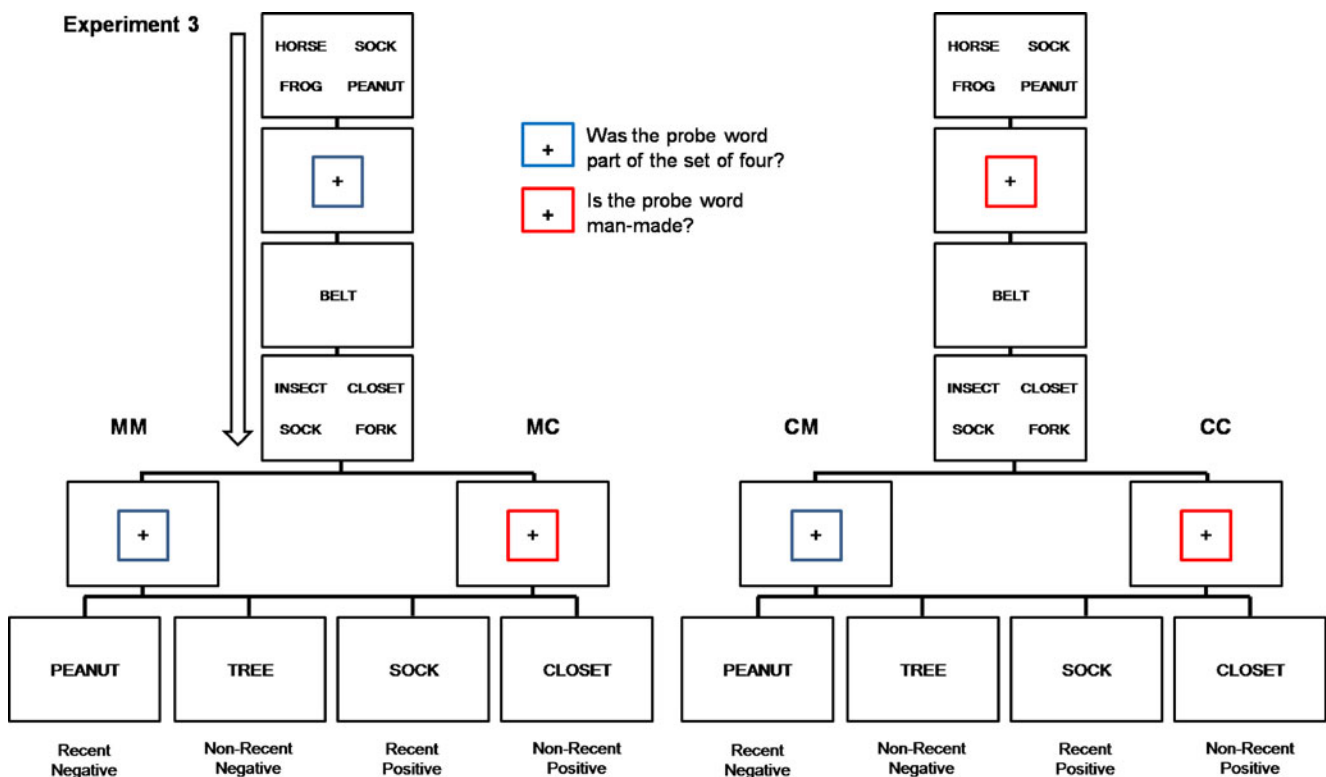

Fig. 3 Sample trial sequences for Experiment 3. After presenting a set of four items, a color cue during fixation indicated which task to perform. Recency was established on the basis of the items presented as part of the previous set of four items and could derive from either prior STM or prior category trials, eliminating the task-switching confound present in Experiments 1 and 2

source of recency. Thus, both recent and nonrecent trial sequences could consist of two consecutive memory trials (MM), a memory trial followed by a category trial (MC), a category trial followed by a memory trial (CM), or two consecutive category trials (CC). As before, a factorial design was used to ensure equal distributions of STM versus category, positive versus negative, and switch versus nonswitch trials, with a pseudorandom order of presentation and the constraint that no more than three negative responses could occur in a row.

Procedure

The overall procedure followed the same format as that in Experiment 1. Practice consisted of 20 trials, evenly distributed among STM and category trials. As in Experiments 1 and 2, participants were able to repeat practice as desired. Following practice, each participant completed six blocks of 64 trials. In between each block, a short (2- to 6-min) nonverbal paper-and-pencil "break" task was completed in order to reduce fatigue and boredom with the computerized task. Each "break" task was drawn from the Kit of Factor Referenced tests (ETS, 1976). These tasks served only as fillers to keep participants engaged in the session, and their results are not discussed further.

Results and discussion

Analyses were again limited to negative trials; positive trial data can be found in Tables 4, 5 and 6.

Response time

The $2 \times 2 \times 2$ design used here allows for a large number of comparisons. We focus our discussion on those most relevant to our theoretical questions (the full ANOVA table is presented in Tables 5 and 6). Means and standard errors are presented in Table 4. The three-way interaction between recency, previous trial type, and current trial type was not significant, $F(1, 29) = 2.72$, $p = .11$. As was shown in planned follow-up analyses, regardless of previous trial type, recency lengthened response times in STM trials [for MM trials, $t(29) = 6.54$, $p < .0005$, $d = 1.19$; for CM trials, $t(29) = 2.66$, $p < .05$, $d = 0.49$] but not in category judgment trials [for MC trials, t < 1; for CC trials, $t(29) = 1.36$, $p = .18$, $d = 0.25$].

Although the results generally fit with our predictions, a close inspection of the means suggested that for STM trials, interference effects might be larger in the *nonswitch* condition and that there were trends for an interference effect (regardless of switch condition) on the category trials. These possibilities were explored using 2×2 ANOVAs (switch × recency) within each current trial type (STM or category).

For current STM trials, the switch × recency interaction was significant, $F(1, 29) = 6.28$, $p < .05$, $\eta_G^2 = .01$, indicating greater interference for MM trials than for CM trials. (See Table 6 for main effects and positive-trial analyses.) For current category trials, the interaction was not significant, $F < 1$. These trials showed a numerical trend toward a main effect of recency, but it did not reach significance, $F(1, 29) = 3.28$, $p = .08$, $\eta_G^2 < .005$. As was noted earlier, planned t-tests indicated that the recency effect was significant for both types of STM trials, both $ps < .05$, $d > 0.45$, but for neither type of category trial, both $ps > .15$, $d < 0.30$.

Accuracy

For accuracy, the three-way interaction between previous trial type, recency, and current trial type was not significant, $F < 1$. However, the 2×2 interaction between recency and current trial type was significant, $F(1, 29) = 6.88$, $p < .05$, $\eta_G^2 = .02$, once again indicating a larger interference effect for STM trials than for category trials (Fig. 4b). Planned t-tests confirmed that recency-based interference reduced

Table 4 Average median response times (RTs, in milliseconds; with standard errors) and accuracies (with standard errors) broken down by trial type for Experiment 3

			Recent negative	Nonrecent negative	Recent positive	Nonrecent positive
Current memory trials	MM	RT	716.33 (27.45)	614.82 (18.85)	612.23 (25.37)	601.62 (20.49)
		Accuracy (%)	94.72 (0.91)	98.89 (0.49)	98.19 (0.38)	97.50 (0.59)
	CM	RT	686.97 (25.97)	641.88 (23.08)	618.35 (25.26)	627.12 (26.03)
		Accuracy (%)	94.03 (0.89)	96.25 (0.64)	98.19 (0.43)	93.33 (0.71)
Current category trials	MC	RT	859.30 (38.91)	840.97 (34.47)	899.23 (40.89)	916.45 (37.36)
		Accuracy (%)	93.47 (1.21)	94.03 (1.05)	88.33 (1.30)	92.78 (0.98)
	CC	RT	838.12 (34.83)	814.40 (33.55)	883.57 (34.11)	855.45 (32.62)
		Accuracy (%)	92.36 (1.08)	92.64 (1.19)	91.39 (1.18)	90.97 (1.15)

Note. M, memory; C, category

Table 5 Statistics for all 2 × 2 × 2 ANOVAs in Experiment 3

		Negative trials			Positive trials		
		F value	p value	η^2_G	F value	p value	η^2_G
Response Time (ms)	Main effect of recency	24.31	<.001	.02	0.18	.67	<.001
	Main effect of previous trial type	5.63	.03	<.01	2.13	.16	<.01
	Main effect of current trial type	46.93	<.001	.22	132.62	<.001	.40
	Recency × previous trial type interaction	1.81	.19	<.01	0.43	.52	<.001
	Recency × current trial type interaction	15.31	<.01	.01	0.08	.79	<.001
	Previous × current trial type interaction	3.42	.08	<.01	10.03	<.01	.01
	Three-way interaction	2.72	.11	<.01	4.43	.04	<.01
Accuracy (% correct)	Main effect of recency	11.68	<.01	.03	0.82	.37	<.01
	Main effect of previous trial type	6.48	.02	.02	2.13	.16	.01
	Main effect of current trial type	10.32	<.01	.07	64.20	<.001	.27
	Recency × previous trial type interaction	0.69	.42	.03	15.58	<.001	.05
	Recency × current trial type interaction	6.88	.01	.02	23.45	<.001	.06
	Previous × current trial type interaction	0.21	.65	<.001	6.44	.02	.02
	Three-way interaction	0.62	.44	<.01	0.14	.72	<.001

accuracy for STM trials, with a significant effect in the MM condition, $t(29) = -4.26$, $p < .0005$, $d = -0.78$, and a marginal one in the CM condition, $t(29) = -1.94$, $p = .06$, $d = -0.35$. However, there was no interference effect for category trials, both ts < 1.

In summary, the results of this experiment generally replicated the patterns seen in Experiments 1 and 2 and did not support the hypothesis that trial type sequence or switching was responsible for those patterns. One caveat to this conclusion is that in this experiment, there was a

nonsignificant numerical trend for recency effects in the response time data for category trials that was not seen in the prior experiments. It is possible that the intermixing of STM and category trials and the arbitrary cue (red or blue box) used to indicate trial type led to some difficulties maintaining task set, which, in turn, allowed contamination from irrelevant task dimensions. We mention this caveat for completeness and as a possible avenue for further investigation, but since is it is a post hoc explanation of a nonsignificant effect, we do not consider it further here. Overall,

Table 6 Statistics for all 2 × 2 ANOVAs in Experiment 3

|			Negative trials			Positive trials		
			F value	p value	η^2_G	F value	p value	η^2_G
Current memory trials	Response time (ms)	Main effect of recency	39.13	<.001	.07	0.01	.92	<.001
		Main effect of previous trial type	0.02	.89	<.001	3.57	.07	<.01
		Recency × previous trial type interaction	6.28	.02	.01	1.55	.22	<.01
	Accuracy (% correct)	Main effect of recency	19.03	<.001	.14	33.14	<.001	.18
		Main effect of previous trial type	5.87	.02	.04	20.71	<.001	.11
		Recency × previous trial type interaction	1.59	.22	.01	20.71	<.001	.11
Current Category Trials	Response time (ms)	Main effect of recency	3.28	.08	<.01	0.18	.68	<.001
		Main effect of previous trial type	8.40	.01	<.01	7.52	.01	.01
		Recency × previous trial type interaction	0.03	.86	<.001	2.03	.17	<.01
	Accuracy (% correct)	Main effect of recency	0.30	.59	<.01	6.61	.02	.03
		Main effect of previous trial type	2.63	.12	.01	0.45	.51	<.01
		Recency × previous trial type interaction	.02	.88	<.001	6.65	.02	.04
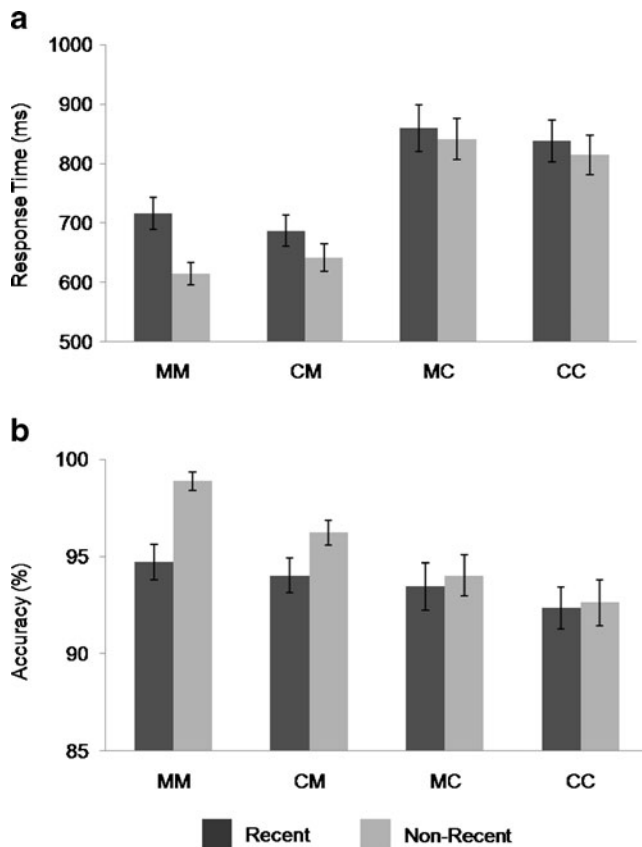
Fig. 4 Average median response times (top panel) and accuracy (bottom panel) for critical trials in Experiment 3. Note that M stands for short-term memory (STM) trials, and C for category judgment trials. The combination of abbreviations indicates trial order (e.g., MM indicates an STM trial preceded by an STM trial; CM indicates an STM trial preceded by a category trial)

the results indicate that regardless of trial sequence or switching, recency led to interference on STM trials but not on category trials.

Experiment 4

In Experiments 1, 2, and 3, we manipulated whether the temporal dimension was relevant to the judgment being made about probe items. In the following experiments, we keep the relevance of the temporal dimension constant and examine whether manipulating similarity along other dimensions influences the magnitude of the proactive interference effect.

Experiment 4 conceptually replicated the design of Atkins et al. (2011) but manipulated the perceptual match (rather than the category match) between the memory set and the probe items. All trials were STM trials, in which participants were asked to judge whether the probe item was a member of the current memory set. The color of the memory set and probe items varied (red or blue); if the probe item was a different color than the memory set, it was always a negative item and should be rejected. However, participants were not told to use color to make their decisions about the probes; from their perspective, color was irrelevant.

If similarity between target and nontarget items generally produces interference regardless of temporal recency, it should take more time to reject *color-match* negative probes, which are similar to (match) the memory set items along the nominally irrelevant color dimension, than to reject *color-mismatch* negative probes, which do not share this similarity with the memory set. In addition, if temporal and perceptual similarity each contribute to interference, the recent-negative effect should be larger for color-match than for color-mismatch trials. On the other hand, if similarity between the memory set and the probe item contributes to interference only when that similarity is along task-relevant dimensions, color mis/match, which from the participants' perspective is not relevant, should not influence either overall response times or the size of the recent-negative effect.[1]

Method

Participants

Thirty-four participants (24 female; average age = 18.47 years, $SD = 0.51$) participated in this experiment for course credit via the Introductory Psychology Subject Pool at the University of Michigan. Two participants were excluded for failing to meet the ERVT criterion score; no participants failed to meet the accuracy criterion.

The final sample consisted of 32 participants (23 female). Participants had a mean age of 18.47 years ($SD = 0.51$) and had completed an average of 12.06 years ($SD = 0.25$) of formal education. ERVT scores for included participants ranged from 9.00 to 29.50 ($M = 18.21$, $SD = 5.67$).

Stimuli

The same pool of word stimuli as that used for Experiments 1, 2, and 3 was used here; the trial structure was the same as that for the STM trials in those experiments. The major added manipulation was the color of the memory set and probe items. All items within a memory set were the same color (red or blue); probe items were presented in either the

---

[1] We ran a small additional experiment to ensure that the color dimension could have an effect if participants knew that it was relevant. This was indeed the case. With these instructions, the recency effect on response time was halved for mismatch trials (from 59 ms on match trials to 26 ms on mismatch trials) [interaction test, $F(1, 12) = 7.24$, $p < .05$] and eliminated it for accuracy data (from 5 % in the match condition to a reversed difference of less than 1 % in the mismatch condition) [interaction test, $F(1, 12) = 4.97$, $p < .05$].

same color (match) or the complementary color (mismatch) as the memory set (Fig. 5). Thus, color-match trials were similar to (the same as) the memory set items along the color dimension, but color-mismatch trials were not.

Color-match trials could be either positive or negative. For color-mismatch trials, the probe was never a member of the current memory set, and thus the correct response for color-mismatch trials was always negative. As was noted earlier, this allowed us to make competing predictions regarding the influence of the color dimension. If similarity along the color dimension influences performance, color-match trials should take longer to reject than color-mismatch trials, since the latter were never members of the current memory set and could hypothetically be rejected on the basis of the color mismatch alone. Furthermore, if similarity along the color and temporal dimensions interacts, interference as indexed by the recent-negative effect ought to be larger for color-match than for color-mismatch trials. The competing (and preferred) hypothesis was that because the color dimension was not relevant from the participants' perspective, it would not influence performance. That is, if the color dimension is irrelevant from the participant's perspective and only task-relevant dimensions influence interference, both overall response time and the recent-negative effect should be equivalent for color-match and color-mismatch trials.

Trials were split evenly between negative and positive trials. All positive trials within this experiment were non-recent; this was done in order to keep the overall experiment time reasonable and to prevent fatigue effects. Negative trials were split evenly between color-mismatch recent trials, color-mismatch nonrecent trials, color-match recent trials, and color-match nonrecent trials.

To further increase the chance that perceptual information might contribute to proactive interference effects, recent probes (regardless of whether their colors were matched or mismatched with the current memory set) were always presented in the same color on the critical trial as they had been on the immediately prior trial's memory set. This correspondence with the prior trial was implemented to maximize the possibility that shared color could increase the familiarity associated with the probe. All trial types were pseudorandomly interspersed throughout each of four blocks of 64 trials, within the constraints needed for appropriate proportions of match/mismatch and recent/nonrecent trials, and the order of block was counterbalanced across participants using an approximate Latin square design.

Results and discussion

As in the previous experiments, analyses are limited to the negative trials that are of theoretical interest; information on positive trials is presented in Table 7. Response times and
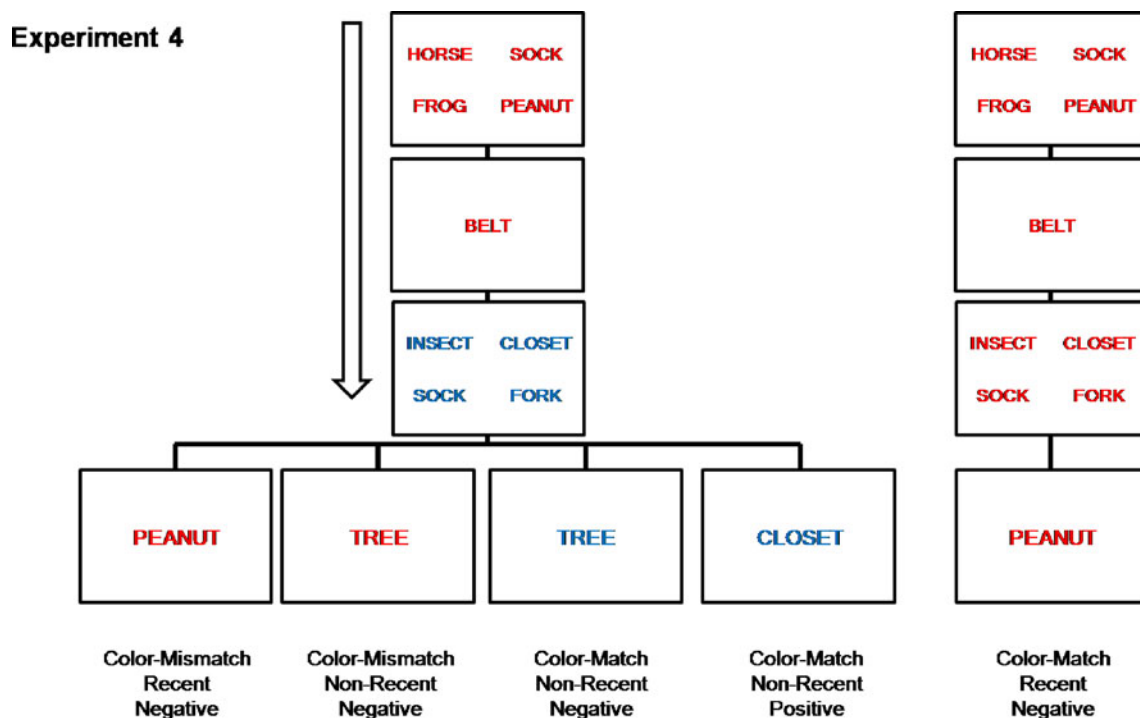

Fig. 5 Sample trial sequences for Experiment 4. The trial type is indicated below each critical probe. Perceptual similarity was manipulated between the current memory set and the probe word such that color-mismatch trials had low perceptual similarity with the memory set, whereas color-match trials had high perceptual similarity with (matched) the memory set along the color dimension

Table 7 Average median response times (RTs, in milliseconds; with standard errors) and accuracy scores (with standard errors) by trial type for Experiment 4

			Recent negative	Nonrecent negative	Positive
Experiment 4	Color mismatch	RT	656.47 (21.00)	593.22 (17.17)	–
		Accuracy (%)	96.00 (0.99)	98.54 (0.47)	–
	Color match	RT	658.81 (21.18)	586.23 (14.85)	589.69 (14.34)
		Accuracy (%)	93.37 (1.19)	99.02 (0.41)	95.24 (0.59)

Note. Because all positive trials were classified as color-match trials, there are no values for color-mismatch positive trials

accuracy were analyzed using 2 × 2 ANOVAs with the factors recency and color match.

Response time

As can be seen in Fig. 6a, this experiment replicated the standard recent-negative effect, $F(1, 31) = 87.81$, $p < .001$, $\eta_G^2 = .10$, and that effect was not influenced by color match, $F < 1$. There was also no main effect of color match, $F < 1$ (see Table 8).

Accuracy

The accuracy data showed a small effect of color match on the recency data, $F(1, 31) = 4.71$, $p = .04$, $\eta_G^2 = .03$. Follow-up t-tests indicated that interference affected both match, $t(31) = -5.13$, $p < .001$, $d = -0.91$, and mismatch, $t(31) = -3.08$, $p < .005$, $d = -0.55$, trials. As in Experiment 2, additional follow-up analyses indicated that only low-accuracy participants showed an effect of color match, suggesting that the nonrelevant dimension may begin to have an influence when

participants have difficulty making a judgment on the relevant dimension (see the Appendix). Replicating standard results, accuracy was lower for recent probes than for nonrecent probes, $F(1, 31) = 38.80$, $p < .001$, $\eta_G^2 = .16$.

Experiment 5

Experiment 4 manipulated the perceptual similarity between the memory set and the probe item and held perceptual similarity between the prior and current presentations of the probe item constant. In the present experiment, we manipulated similarity across subsequent presentations of the probe items, so that when the critical recent-negative probe items appeared, they were either identical in appearance (color, font, and bold/italics) to their presentation in the memory set of the previous trial or very different in appearance. If general familiarity and activation strength affect the degree of proactive interference caused by the probe item's presence on the immediately prior trial, probe items that look exactly the same on the current trial as they did on

Fig. 6 Average median response times (top panels) as well as accuracies (bottom panels) for negative responses in Experiment 4 (left; **a** and **b**) and Experiment 5 (right; **c** and **d**)
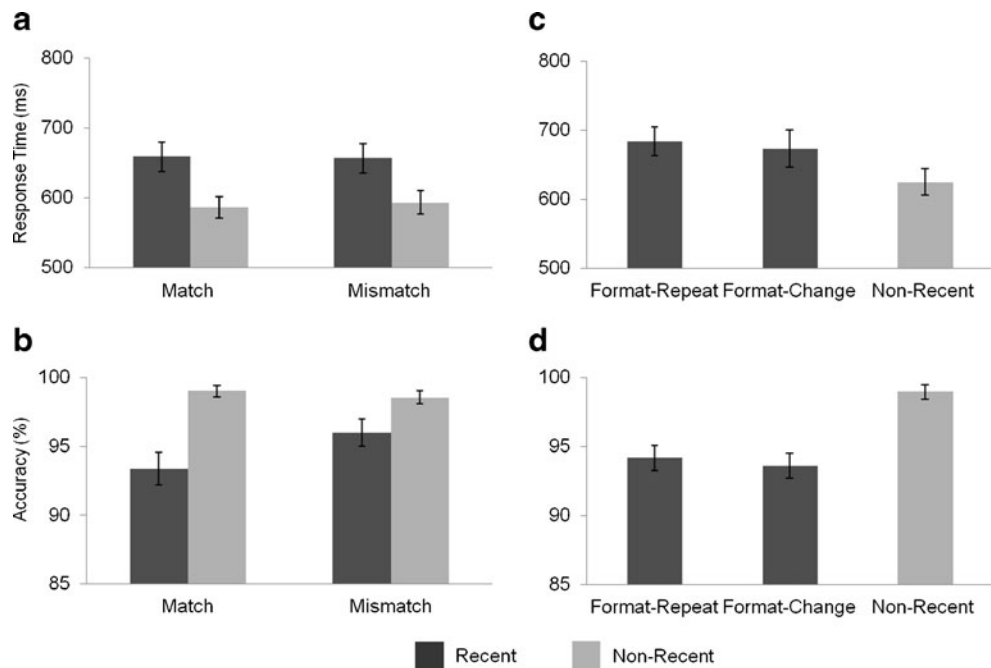
Table 8 Statistical analyses and effect sizes for Experiment 4

	Experiment 4		
	F value	*p* value	η_G^2
2 × 2 ANOVA for negative RTs			
Main effect of trial type	0.16	.69	<.001
Main effect of recency	87.81	<.001	.10
Trial type × recency interaction	0.80	.38	<.001
2 × 2 ANOVA for negative accuracies			
Main effect of trial type	3.13	.09	.01
Main effect of recency	38.80	<.001	.16
Trial type × recency interaction	4.71	.04	.03

the prior trial should be more familiar and, thus, more difficult to reject than probe items that have extensively changed in appearance since their prior presentation.

Method

Participants

Twenty-five participants (16 female; average age = 18.32 years, SD = 0.75) participated in this experiment. One participant was excluded due to medication/health conditions, 1 due to missing data, and 5 for failure to reach the criterion ERVT score; no participants were excluded due to the accuracy criterion.

Eighteen participants (11 female) were included for analysis. Participants had a mean age of 18.39 years (SD = 0.85) and had completed an average of 12.28 years (SD = 0.75) of formal education, with ERVT scores between 10.75 and 29.00 (M = 16.25, SD = 4.72).

Stimuli

All trials followed the same organization as the STM trials in the previous experiments and used the same pool of words. The perceptual attributes of color (orange or blue), font (Arial or Times New Roman), bold (bolded or not), and italics (italicized or not) varied among the words presented on each trial (Fig. 7). The critical manipulation was for recent probes: *format-repeat* recent probe items were presented with exactly the same perceptual attributes (color, bold/not-bold, italicized/not-italicized, Arial/Times New Roman) when shown as current-trial probes as they had been when they were presented as part of the immediately prior trial's memory set. In contrast, *format-change* recent trials were presented with the opposite set of perceptual attributes when shown as probe items on the current trial, as compared with their format in the previous trial's memory set. For example, for a format-change trial, a word that had appeared in orange, bolded, nonitalicized Arial font on the

immediately previous trial would appear in blue, nonbold, italicized Times New Roman font on the current trial. All trial types were randomly interspersed throughout each of four blocks of 64 trials, and the block order was again counterbalanced across participants using an approximate Latin square design.

Trials were evenly balanced between negative and positive, as well as recent and nonrecent, trials. Recent trials were of two types, format change and format repeat, and these two trial types also occurred with equal frequency.

Results and discussion

Both median response times and accuracies for negative trials were analyzed using an ANOVA with trial type (nonrecent negative, format-repeat recent negative, format-change recent negative) as a repeated factor, followed by planned *t*-tests comparing format-repeat versus format-change trials. See Tables 9 and 10 for all trial values and analyses.

Response time

As can be seen in Fig. 6c, the amount of interference caused by an item was not influenced by changing its perceptual qualities. Participants were slower to reject both types of recent negatives (format repeat or format change) than they were to reject nonrecent negatives, $F(2, 34) = 15.26$, $p < .001$, $\eta_G^2 = .07$, but there was no difference between format-repeat and format-change trials, $t < 1$.

Accuracy

The accuracy results followed the same pattern as the response time data. Participants were more accurate at rejecting non-recent negatives than they were at rejecting either type of recent negative, $F(2, 34) = 15.74$, $p < .001$, $\eta_G^2 = .34$, and correct rejection rates for format-repeat versus format-change recent-negative trials were equivalent, $t < 1$ (Fig. 6d).

In summary, changing the perceptual qualities of the recent probe from its first presentation did not alter the recent-negative effect. Task-irrelevant stimulus dimensions, such as perceptual information within this STM task where temporal/trial-order information was the relevant dimension, failed to modify the size of proactive interference effects.

General discussion

The results of these experiments suggest that neither activation strength (due to familiarity from recent presentation) nor similarity per se is sufficient to cause interference. Recent presentation of an item led to interference when the

Fig. 7 Sample trial sequences for Experiment 5. Perceptual similarity was manipulated between the first presentation of an item and its second presentation as a recent probe word, so that probes on format-change trials had very little perceptual similarity across presentations, while probes on format-repeat trials had exactly the same format across presentations
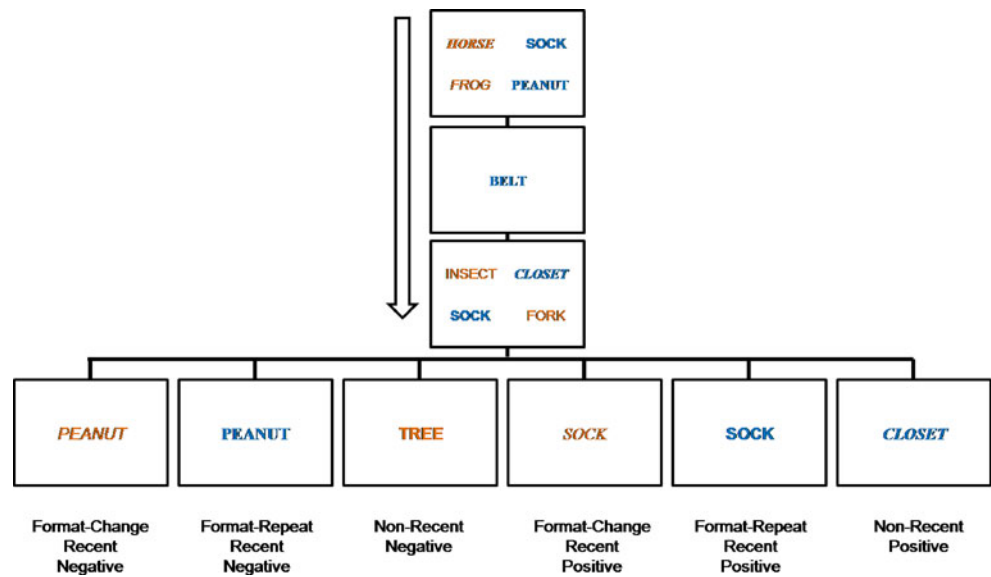
temporal characteristics of that item were important for responding to the test cue, as on STM trials. Conversely, changing nontemporal dimensions of the stimuli did not affect interference on STM trials. Taken together, these results suggest that although previous research has shown that temporal/trial-order and semantic and perceptual characteristics can all influence interference effects, none of these dimensions has a special status in determining interference. Instead, the critical question appears to be whether similarity along a particular dimension allows a nontarget item to be confused with target items in a manner that is relevant for responding to the test cue.

These results are pertinent to recent questions regarding the sources of interference in STM and how they may interact (e.g., Atkins et al., 2011; Jonides & Nee, 2006; Oztekin, Curtis, & McElree, 2009). Rather than describing interference from recent presentation as a result of biased competition or activation strength, temporal or trial order characteristics may be "just another" dimension along which nontarget stimuli can be similar to and compete with target stimuli as possible responses to the task cue. Recent items are hard to reject in STM tasks because they are hard to discriminate from the current memory set along the temporal/trial order dimension. This conceptualization of interference and the contribution of

temporal information to proactive interference has much in common with several models of STM that describe items in terms of collections of features and forgetting as a result of competition among such features or a loss of discriminability among them (e.g., Lewandowsky, Oberauer, & Brown, 2009; Nairne, 2002; Oberauer & Kliegl, 2006).

We have generally confined our theoretical discussion to the STM domain because that is where debates about activation and decay versus similarity-based competition are most prevalent and because the Sternberg task upon which our tasks were based (Sternberg, 1966) is considered a classic STM task. It is important to note that we have tested only proactive interference resulting from recent presentation, which may be of particular relevance for STM tasks, and manipulated only some stimulus dimensions (temporal, perceptual, semantic). It is possible that different patterns would occur when testing PI from other sources (e.g., long-standing habits such as dominant vs. nondominant meanings of homonyms) or when manipulating other dimensions.

However, the principles discussed here are also thought to govern interference in LTM. Indeed, in many cases, our predictions derive from the classic work on interference theory done using LTM paradigms (see reviews by Crowder, 1976; Lustig et al., 2009). The core idea tested here—that

Table 9 Average median response times (RTs, in milliseconds; with standard errors) and accuracy scores (with standard errors) by trial type for Experiment 5

			"No" response	"Yes" response
Experiment 5	Nonrecent	RT	624.89 (19.06)	644.31 (21.35)
		Accuracy (%)	98.96 (0.52)	95.23 (0.85)
	Format Change	RT	673.31 (26.83)	628.94 (23.77)
		Accuracy (%)	93.59 (0.91)	93.65 (0.97)
	Format Repeat	RT	683.61 (21.15)	628.25 (21.91)
		Accuracy (%)	94.17 (0.91)	92.92 (0.85)

Table 10 Statistical analyses and effect sizes for Experiment 5

	Experiment 5		
	F value	p value	η_G^2
3 × 1 ANOVA for negative RTs	15.26	<.001	.07
3 × 1 ANOVA for positive RTs	2.13	.14	.01
3 × 1 ANOVA for negative accuracies	15.74	<.001	.34
3 × 1 ANOVA for positive accuracies	2.07	.15	.06

interference depends critically on whether items are similar on dimensions important for responding to the test cue—has also been used to explain interference in tests of long-term implicit memory (Lustig & Hasher, 2001a, 2001b). These results can thus be seen as supporting the idea that STM and LTM may be better thought of in terms of phenomenology and task parameters (e.g., Anderson et al., 2004; Cowan, 2001; Craik & Lockhart, 1972; Jonides et al., 2008; McElree, 2001; Oberauer, 2002) than as separate systems or stores (e.g., Baddeley, 2000; Baddeley & Hitch, 1974; Goldman-Rakic, 1999).

Another interesting perspective on these issues is offered by signal detection theory, which can be combined with the idea that items in STM are represented by noisy codes consisting of multiple dimensions or features (font, case, formatting, orthography, semantic meaning, trial order, etc.) to explain the size and presence of interference effects on a variety of STM (and possibly LTM) tasks (see Atkins et al., 2011; Lustig, Matell, & Meck, 2005; Nairne, 1990, 2002). When the probe item is presented, it initiates a search *along task-relevant dimensions*. Recent-negative items have reduced signal-to-noise ratios when compared with nonrecent items, due to their high levels of temporal similarity with the current memoranda, and so the discrimination process becomes more difficult.

The phrase "along task-relevant dimensions" is critical. On STM trials, temporal information is relevant by definition and, thus, is included in the search and decision process. Items within the same memory set are presumably the most similar along this dimension but would share high degrees of temporal similarity with items from the previous trial. The more intervening trials between the current memory set and the set to which the probe belonged, the less similar the probe item is to items in the current memory set, and thus the easier (faster) it becomes to discriminate between the two. In contrast, on category judgment trials, the temporal dimension is not relevant and so may not factor into the search and decision process. Conversely, as Experiments 4 and 5 showed, if participants do not perceive perceptual dimensions as relevant to the task, such similarity will not influence interference effects.

However, if task-irrelevant dimensions of similarity do not affect interference, why did Atkins et al. (2011) find that interference was reduced but not eliminated when the probe item (e.g., *orange*) did not match the category of the studied items (e.g., *Canada*, *France*, *Australia*, *Brazil*)? These results initially appear to be at odds with those of Experiment 4, which used the same design but manipulated similarity on the perceptual dimension instead (e.g., red probes vs. blue memory set), and found that interference was just as large as when the negative probe matched the memory set in color. The key difference here is the degree to which the manipulated dimension of similarity was integral to evaluating the probe item. To evaluate whether a word (e.g., *orange*) is a member of the current memory set, as in the recent probes task, one must process the meaning of that word. If the semantic category of the word is clearly different from the target memory set, as in Atkins et al. (2011), that information may be used to speed the response. (It is theoretically possible to construct a situation where participants would not process the word to the level of meaning, but it is highly unlikely that they would adopt this strategy on their own.) In contrast, it is quite easy to decide whether a probe word was a member of a memory set without considering its ink color; indeed, that is what participants had to do on every trial where there was an ink color match. In other words, in both experiments, the recent negative probes were similar to the target set on the temporal dimension. However, only in the Atkins et al. experiments was the dimension along which similarity was manipulated (word meaning) integral to evaluating the probe's match with the words in the target memory set. Participants in those studies could use the meaning mismatch to facilitate rejection of the probe. In contrast, participants in the present study did not use the font color dimension for evaluating the probe, and therefore, its difference from the target set along the color dimension did not influence their efficiency in rejecting it.

Summary

How items are represented in STM and what factors lead to forgetting and interference are issues of long standing (see the discussion by Jonides et al., 2008). Our results and those of Atkins et al. (2011) suggest that on STM tasks, the recent presentation of an item makes it similar to items on the current trial along the temporal dimension and this results in interference because STM tasks require discrimination between target and nontarget items along that dimension. In contrast, when items are similar along dimensions irrelevant to the current task, interference does not result.

These principles of similarity-based interference have also been proposed to govern LTM, including implicit memory. Carefully designed experiments using parallel STM and LTM procedures will be needed to determine whether the same mechanisms in fact govern interference across these domains (see Flegal et al., 2010, for one example). It will also be important to determine the boundary conditions for these ideas; there is some suggestion in our results that nominally task-irrelevant dimensions may begin to influence performance if judgments on the task-relevant dimension are difficult (Experiments 2 and 4). Overall, however, our results suggest that to escape the past, make it irrelevant.

Appendix

Experiment 2 analyses

The accuracy results in Experiment 2 suggested that the temporal dimension might have some influence on category judgments for participants who had difficulty with those judgments. To explore this issue, we divided participants into high- and low-accuracy groups on the basis of their relative performance on the more difficult *italics* judgment. Analyses of the accuracy data are presented in the main text. Here, we present the response time data for readers interested in this issue. As in the main text, analyses focus on the negative trials that are of theoretical interest. Like the accuracy data, the response time results suggest that participants who were skilled at making the category judgment showed interference effects on STM but not on category judgment trials, whereas those participants who had difficulty with the category judgment showed interference effects on both trial types.

High-accuracy group

The high-accuracy group replicated the pattern seen in Experiment 1. An ANOVA on median response times to negative probes revealed that STM, but not perceptual judgment, trials were influenced by recency [interaction, $F(1, 17) = 8.86$, $p < .01$, $\eta_{G}^2 = .02$]. For STM trials, t-tests revealed a significant difference between recent and nonrecent probes, $t(17) = 3.82$,

$p = .001$, $d = 0.90$, but no difference on category trials, $t(17) = -0.30$, $p = .77$, $d = -0.07$. Main effects of recency, $F(1, 17) = 12.45$, $p < .01$, $\eta_{G}^2 = .02$, and category, $F(1, 17) = 33.22$, $p < .001$, $\eta_{G}^2 = .07$, were also present.

Low-accuracy group

The low-accuracy group showed effects of recency for both judgment types. An ANOVA on median response times to negative probes did not reveal any interaction, $F(1, 18) = 2.61$, $p = .12$, but main effects of both recency, $F(1, 18) = 8.27$, $p = .01$, $\eta_{G}^2 = .02$, and category, $F(1, 18) = 10.98$, $p < .01$, $\eta_{G}^2 = .06$, were present.

Experiment 4 analyses

The accuracy results in Experiment 4 also suggested that the irrelevant dimension (in this case, color mis/match) might influence recency judgments for low-accuracy participants. These results complement those of Experiment 2; that is, in Experiment 2, low-accuracy participants showed an effect of the nominally irrelevant temporal dimension on the perceptual category judgment; in Experiment 4, low-accuracy participants showed an effect of the nominally irrelevant perceptual category on the temporal (recency) judgment. Since there was no category judgment in this experiment, median splits on accuracy were based on overall accuracy on negative trials (both recent and nonrecent).

High-accuracy group

The high-accuracy group produced the expected results: Color mis/match did not affect the size of the recency effect for response time, $F(1, 15) = 1.90$, $p = .19$, $\eta_{G}^2 = .06$, or accuracy, $F < 1$. Both response time and accuracy measures showed main effects of recency (both $ps < .05$, $\eta_{G}^2 > .10$) and no effects of color mis/match (both $Fs < 1$).

Low-accuracy group

The response time data for the low-accuracy group followed a similar pattern as those for the high-accuracy group: no interaction between color mis/match and response time, $F < 1$, a main effect of recency, $F(1, 15) = 42.28$, $p < .0001$, $\eta_{G}^2 = .09$, and no main effect of color mis/match, $F < 1$. In contrast, the accuracy data did show an interaction between color mis/match and recency, $F(1, 15) = 6.85$, $p < .05$, $\eta_{G}^2 = .09$. There was also a significant main effect of recency on the accuracy data, $F(1, 15) = 8.76$, $p < .05$, $\eta_{G}^2 = .14$, and a marginal effect of color mis/match, $F(1, 15) = 3.68$, $p = .07$, $\eta_{G}^2 = .04$.

Overall

For completeness, we report the response times and accuracies for Experiments 1, 3, 4, and 5, broken down by high- and low-accuracy groups (Table 11; Experiment 2 results are presented in main text Table 3). In Experiments 1 and 2, the median split is accomplished by comparing the "easier" and "more difficult" category judgments and separating those individuals who had greater disparity between accuracy on those judgments from individuals who succeeded similarly in both, on the basis of accuracy. For Experiments 3–5, the median split was created in a more standard way, by separating the individuals along the median level of overall performance, as measured by accuracy. The numerical pattern of greater influence of the irrelevant dimension for lower-accuracy participants holds for some but not all comparisons; overall high accuracy and low variability in accuracy may limit the ability to test for such effects in these data. We therefore note this pattern as a potential important caveat to our claim that similarity must be along task-relevant dimensions to have an effect but caution that further experimentation is needed to fully test this possibility.

Table 11 Response times (RTs, in milliseconds; with standard errors) and accuracies (with standard errors) for negative trials for high- and low-accuracy groups in Experiments 1, 3, 4, and 5

			High performers		Low performers	
			Recent negative	Nonrecent negative	Recent negative	Nonrecent negative
Experiment 1	Memory	RT	729.83 (73.48)	667.03 (93.08)	750.43 (87.06)	688.32 (94.90)
		Accuracy (%)	96.46 (3.10)	98.75 (2.30)	96.65 (3.97)	97.77 (2.27)
	Category	RT	842.40 (132.22)	841.57 (136.47)	898.68 (159.20)	924.68 (144.09)
		Accuracy (%)	94.17 (4.71)	94.17 (3.71)	93.53 (5.27)	89.06 (5.31)
Experiment 3	CC	RT	776.73 (135.30)	724.83 (126.24)	899.50 (221.42)	903.97 (191.91)
		Accuracy (%)	95.83 (2.73)	95.83 (4.73)	88.89 (6.23)	89.44 (6.66)
	CM	RT	651.03 (112.73)	623.00 (109.76)	722.90 (162.62)	660.77 (142.47)
		Accuracy (%)	95.83 (4.45)	96.39 (3.47)	92.22 (4.69)	96.11 (3.68)
	MC	RT	783.67 (177.58)	756.43 (125.43)	934.93 (224.29)	925.50 (206.88)
		Accuracy (%)	96.39 (3.81)	96.94 (3.33)	90.56 (7.63)	91.11 (6.27)
	MM	RT	686.53 (117.81)	586.67 (66.08)	746.13 (176.17)	642.97 (126.58)
		Accuracy (%)	95.83 (4.17)	99.72 (1.08)	93.61 (5.65)	98.06 (3.47)
Experiment 4	Color match	RT	666.09 (115.46)	590.63 (84.73)	651.53 (127.36)	581.84 (85.79)
		Accuracy (%)	98.05 (2.25)	99.61 (1.07)	88.70 (6.45)	98.42 (3.04)
	Color mismatch	RT	654.50 (103.20)	598.16 (87.81)	658.44 (136.03)	588.28 (108.29)
		Accuracy (%)	98.05 (3.40)	99.80 (0.78)	93.95 (6.70)	97.27 (3.20)
Experiment 5	Format repeat	RT	677.56 (68.37)	620.94 (63.40)	689.67 (111.11)	628.83 (99.20)
		Accuracy (%)	95.10 (4.65)	99.65 (0.69)	93.24 (2.88)	98.26 (2.97)
	Format change	RT	670.28 (85.02)	620.94 (63.40)	676.33 (142.46)	628.83 (99.20)
		Accuracy (%)	95.67 (4.23)	99.65 (0.69)	91.50 (2.03)	98.26 (2.97)

Note. For Experiment 5, nonrecent values are equivalent. This is due to the experimental design, where nonrecent values are not classified according to format-repeat or format-change distinctions. Values are repeated in this table for ease of comparison with recent values.

References

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13,* 27–33.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. L. (2004). An integrated theory of the mind. *Psychological Review, 111,* 1036–1060.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68–100.

Atkins, A. S., Berman, M. G., Reuter-Lorenz, P. A., Lewis, R. L., & Jonides, J. (2011). Resolving semantic and proactive interference in memory over the short-term. *Memory and Cognition, 39,* 806–817.

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory and Cognition, 36,* 74–81.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology, 18,* 362–365.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). NY: Academic Press.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37,* 379–384.

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 317–333.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron, 39,* 713–726.

Buckner, R. L., Petersen, S. E., Ojemann, J. G., Miezin, F. M., Squire, L. R., & Raichle, M. E. (1995). Functional anatomical studies of explicit and implicit memory tasks. *Journal of Neuroscience, 15,* 12–29.

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 183–196.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing – framework for memory research. *Journal of Verbal Learning and Behavior, 11,* 671–684.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Demeter, E., Sarter, M., & Lustig, C. (2008). Rats and humans paying attention: Cross-species task development for translational research. *Neuropsychology, 22,* 787–799.

Educational Testing Service. (1976). *Kit of factor-referenced tests.* Princeton, NJ: Author.

Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36,* 1331–1338.

Goldman-Rakic, P. S. (1999). The physiological approach: Functional architecture of working memory and disordered cognition in schizophrenia. *Biological Psychiatry, 46,* 650–661.

Henson, R. N. A., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia, 41,* 263–270.

Henson, R. N. A., Rugg, M. D., Shallice, T., & Dolan, R. J. (2000). Confidence in recognition memory for words: Dissociating right prefrontal roles in episodic retrieval. *Journal of Cognitive Neuroscience, 12,* 913–923.

Johnson, L. M. (1933). Similarity of meaning as a factor in retroactive inhibition. *Journal of General Psychology, 9,* 377–389.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C., Berman, M. G., & Sledge Moore, K. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59,* 193–224.

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience, 139,* 181–193.

Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., & Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Sciences in the United States of America, 95,* 8410–8413.

Keppel, G. (1968). Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior therapy* (pp. 172–213). Edgewood Cliffs, NJ: Prentice-Hall.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153–161.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences, 13,* 120–126.

Lustig, C., Berman, M. G., Nee, D. E., Lewis, R. L., Sledge Moore, K., & Jonides, J. (2009). Psychological and neural mechanisms of short-term memory. In G. G. Berntson & J. T. Cacioppo (Eds.), *Handbook of neuroscience for the behavioral sciences* (Vol. 1). Hoboken, N.J.: Wiley.

Lustig, C., & Flegal, K. (2008). Targeting latent function: Encouraging effective encoding for successful memory training and transfer. *Psychology and Aging, 23,* 754–764.

Lustig, C., & Hasher, L. (2001a). Implicit memory is not immune to interference. *Psychological Bulletin, 127,* 618–628.

Lustig, C., & Hasher, L. (2001b). Implicit memory is vulnerable to proactive interference. *Psychological Science, 12,* 408–412.

Lustig, C., Matell, M. S., & Meck, W. H. (2005). Not "just" a coincidence: Frontal-striatal interactions in working memory and interval timing. *Memory, 13,* 441–448.

Masson, M. E. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 3–23.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 817–835.

McGeogh, J. A., & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *American Journal of Psychology, 43,* 579–588.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology, 10,* 465–501.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53–81.

O'Kane, G., Insler, R. Z., & Wagner, A. D. (2005). Conceptual and perceptual novelty effects in human medial temporal cortex. *Hippocampus, 15,* 326–332.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 411–421.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55,* 601–626.

Oztekin, I., Curtis, C. E., & McElree, B. (2009). The medial temporal lobe and the left inferior prefrontal cortex jointly support interference in verbal working memory. *Journal of Cognitive Neuroscience, 21,* 1967–1979.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.

Underwood, B. J. (1945). The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs, 59,* 1–33.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64,* 49–60.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 442–452.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15.

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 2,* 440–445.