AVOIDING THE RECENT PAST:
WHICH STIMULUS DIMENSIONS INFLUENCE PROACTIVE INTERFERENCE?

by

Kimberly S. Craig

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2013**

Doctoral Committee:

**Associate Professor Cindy A. Lustig, Chair
Professor John Jonides
Professor John E. Laird
Professor Patricia A. Reuter-Lorenz**

Dedication

For my family.

I would not have made it this far without your inspiration, encouragement, and support.

Acknowledgements

First, I would like to thank my advisor, Dr. Cindy Lustig, for her advice and encouragement at all stages of this project. In many ways this project would not have been possible without her. I would also like to thank Dr. Marc Berman and Dr. John Jonides for their input on Chapter 2 of this work. They provided an additional insight to this project that shaped this work in important ways.

My committee, as well, has provided extremely valuable feedback throughout the dissertation process; Dr. John Laird, Dr. Cindy Lustig, Dr. John Jonides, and Dr. Patricia Reuter-Lorenz have all contributed their perspectives and advice, helping me to focus my ideas and analyses to better answer empirical questions about the nature of proactive interference. Dr. David Meyer has also provided a large amount of support and encouragement throughout the dissertation process.

In large part, this research would not have been possible without the many research assistants who have helped with experimental design, data collection, and data analysis over the years. I would like to thank Laura Argintar, Jennifer Chronis, Halley Feldman, Elyssa Guslits, Yona Isaacs, Emily Kurzawa, Karin Lavie, Xu Li, Lina Lu, Matthew Lugo, Jessica Oakley, Eric Peist, Saqib Rana, Page Sloan, Alexander Tereshchenko, Maria Tocco, Dani Weinberg, and Madison Weisberg for all of their help.

Finally my family, who support me unconditionally and always go above and beyond to encourage me in whatever I choose to pursue. You have given me the confidence and passion for

learning that led me to graduate school and will continue to lead me in a positive direction in years to come.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

Why Do We Forget?

Imagine that you are waiting tables at a busy restaurant. One of your customers orders the mozzarella sticks but, after a few seconds, decides to get the nacho plate instead. When you return to the kitchen you see mozzarella sticks waiting and begin to pick them up, but then realize your error, set them down, and wait for the nacho plate. The temporary confusion you would experience in this situation is an example of proactive interference: information from the past disrupting current memory performance. This type of interference is a major mechanism of forgetting, both in short-term memory (STM) and long-term memory (LTM) (Underwood, 1957).

Interference is a robust, reliable phenomenon in STM tasks (e.g., Berman, Jonides & Lewis, 2009; Mecklinger, Weber, Gunter & Engle, 2003) but also somewhat mysterious. How can we forget information that we have seen only moments ago? Can we ever escape our recent past, or are we always vulnerable to its influence? The following work will focus on proactive interference within STM, investigating the conditions that may be necessary to create this type of interference. Implications for current theories of proactive interference and current models of STM will be discussed.

Theories of Proactive Interference

Two competing theories currently dominate the proactive interference debate. The first suggests that information in STM decays over time (Altmann & Gray, 2002; Brown, 1958). This decay can be represented by activation strength; perceiving an item or retrieving it from LTM gives the item high activation strength, and over time this activation strength decreases. Interference is created when (now-irrelevant) recent items with relatively high activation strengths compete with target items for attention and current information processing. As time passes and activation strength decreases, the item loses its ability to interfere.

The activation-strength idea is simple and intuitive: The more recent an item is, the stronger its residual activation strength, and thus the more likely it is to cause interference. A recent paper by Atkins et al. (2011) appeared to find compelling evidence for this idea: Participants were slower to reject recently-seen information (an index of proactive interference) even when that information belonged to a completely separate category than the currently-relevant information (e.g., being slow to reject ‘England’ when the relevant information only consists of fruit names). From the activation-strength perspective, seeing the recent but now incorrect item (‘England’) automatically re-activates its representation and pulls it back into the focus of attention, creating interference. It is clear that according to activation strength theory, proactive interference should be pervasive, occurring whenever a recently-presented item has the chance to compete for the focus of attention.

A second theory argues that it is a high degree of similarity between recent and current items that creates proactive interference (e.g., Crowder, 1976; Johnson, 1933; Keppel, 1968; McGeogh & McDonald, 1931; Underwood, 1945, 1957). From this perspective, a recent item that is highly dissimilar to the target of current information processing should not interfere. Support for this theory comes from the release-from-proactive-interference effect, where performance declines in a list-learning paradigm over time for similar items (e.g., from the same category), but then is eliminated when a word appears that is highly dissimilar to others on the list (Wickens, Born, & Allen, 1963). Note the key difference between similarity and activation strength theories. Here, an item interferes because it is similar, not because it is recent.

Both the activation-strength and similarity-based competition views are appealing, but both may also be too simple to fully explain interference in STM. However, their strengths and weaknesses are complementary and suggest that both must be taken into account. Returning to our restaurant example, if recent activation was the primary driver of interference, how could a server learn and hold in STM the orders from multiple people at the same table without suffering from catastrophic interference? On the other hand, recency must play some role: The server deals with multiple iterations of a closed set of similar items (i.e., the menu) throughout the day, and it seems unlikely that one patron's lunchtime order of nachos will interfere with the memory for a dinner patron's appetizer. To bring these views together, we consider the structure of STM and the nature of its representations.

The Structure of STM

There are many models of STM's structure, but they tend to fall into two classes that differ in whether they treat STM and LTM as distinct, separate entities or as more of a continuum. The most prominent example of the first class is Baddeley's working memory model (Baddeley& Hitch, 1974; Baddeley, 2000). In the canonical version, STM is comprised of two temporary storage components (the phonological loop and visuospatial sketchpad), and a central executive component that acts as an attentional controller. A more recent modification (Baddeley, 2002) adds an episodic buffer that binds and stores multimodal information. This model has been extremely influential in past decades, both in behavioral (i.e., Burgess & Hitch, 1999) and neuroimaging research (i.e., Paulesu, Frith & Frackowiak, 1993; Wagner, Shannon, Kahn &

Buckner, 2005). However, in the past decade the assumptions of Baddeley's model have been called into question (for a review, see Lustig et al, 2009).

More recently, unitary models of memory have gained popularity (e.g., Cowan, 2001; Oberauer, 2002). In these models, STM is conceptualized as an activated portion of LTM, rather than a separate component. Typically, these models consist of a *focus of attention* and a *region of direct access*. Item(s) currently being attended to have the highest activations and can be found in the focus. These items may move into the focus either from the inactive portion of LTM or via a perceptual process, thus allowing new items to enter STM. When an item moves from the focus it enters the *region of direct access*, a privileged portion of LTM that contains previously or potentially relevant items. Items in the *region of direct access* retain a relatively high degree of activation, and are easily moved back into the focus.

To fully understand proactive interference, it is important to consider how interference may be created within these models. For instance, in unitary models, items in the region of direct access may retain the residual activation strength important for creating proactive interference for recently-seen items (according to an activation strength account). As noted above, however, the activation strength account may be too simple to fully explain proactive interference; instead, concepts from both activation strength and similarity-based competition accounts may be important to consider in order to thoroughly understand this phenomenon.

A more detailed consideration of what it means to be an "item" in STM may help resolve the previously-described tension between activation-strength and similarity-based competition views of interference. According to feature-based theories (e.g., Nairne, 2002; Oberauer & Lange, 2008), STM does not hold whole items but instead retains activated cues or feature codes which can be used to recall the associated item. In other words, rather than discrete "items", the representations in STM may be better conceptualized as multi-dimensional bundles of features or pointers to the codes (e.g., perceptual, lexical, phonologic, semantic, temporal) for such features in LTM. When two items have a high degree of similarity, they share many of these features or cues, leading to decreased distinctiveness.

Importantly, an item's temporal context can be considered another featural dimension along which it may be more or less similar to other items (see Howard & Kahana, 2002). At first glance, temporal information is seen as distinct and separate from other stimulus dimensions, such as size, shape, or color. This might occur because temporal information about a stimulus is not a part of the item itself, but is instead about the context in which that item was seen. Hasher and Zacks (1979) argue that although dimensions such as temporal and spatial information are based on context, items are always experienced contextually, and contextual information about an item is processed automatically. This automatic processing may allow contextual stimulus dimensions to be treated in the same way as information about an item's size, shape, or color.

Futher, recent work by Howard and Kahana (2002) suggests that temporal information may be coded neurally based on oscillatory frequencies, allowing for comparisons along the temporal dimension in addition to similarities along other stimulus dimensions. If temporal information is indeed simply another featural dimension, feature-based views could be interpreted by positing that recently-presented items cause interference not because of residual "activation strength" in an energetic sense, but rather because they are similar to (and thus compete with) target items along the temporal dimension of similarity.

The experiments in this dissertation test a refinement of similarity theory: Items will only compete and cause interference if their similarity is along a dimension relevant to the test cue. Lustig and Hasher (2001) proposed that this task-relevance might explain why implicit (indirect) memory tests appeared to be immune to interference affecting explicit (direct) memory tests. Specifically, the dimension of temporal similarity is almost by definition irrelevant to implicit memory tests, but is integral to explicit tests. Within the context of STM, this hypothesis suggests that proactive interference based on recent presentation depends upon whether temporal information is relevant to the task being performed; if not, recent presentation should not create interference.

During the performance of any given task, comparisons of similarity should be made between items only along dimensions of similarity that are task-relevant. When items are similar along these task-relevant dimensions, interference will occur. However, if items are similar along task-irrelevant dimensions, this will not create interference because these dimensions will not be compared.

The present experiments test this hypothesis in part by manipulating the task relevance of several dimensions of similarity and examining the effects of these manipulations on interference in STM.

Overview

The current work examines the extent to which traditional theories of proactive interference can accurately predict the creation of interference. Specifically, it questions the universality implied by activation strength and similarity theories, asking questions such as: What boundary conditions might exist for the creation of interference? Under what circumstances might interference be avoided?

Chapter 2 includes a group of experiments that together suggest a new boundary condition for proactive interference: that an item's similarity must be along a task-relevant dimension if it is to create interference within a given task. This is demonstrated using variants of the Recent Probes task (Berman, Jonides & Lewis, 2009; Monsell, 1978) and a category-matching task (Braver, Reynolds & Donaldson, 2003). Similarity is manipulated along recency and perceptual dimensions; when this similarity is task-relevant, it creates forgetting as measured by decreased accuracy and increased response times to probes (i.e., recency in a STM task), but when this similarity is not task-relevant, forgetting is not observed (i.e., perceptual similarity in a STM task).

Chapter 3 follows up on a secondary question raised by the results presented in Chapter 2: Do task-irrelevant dimensions of similarity serve as a basis for interference if the primary task is very difficult? As will be discussed, in one of the experiments in Chapter 2 a number of participants appeared to have difficulty with the perceptual-judgment categorization task. Chapter 3 describes an experiment which improves performance on this task by training subjects on perceptual judgments before completing the interference experiment. As predicted, after training participants to criterion accuracy (90%) on the difficult task, interference is contained to a task-relevant context. This training addresses individual differences in perceptual-judgment difficulties by improving the abilities of subjects through training. Future studies should investigate the extent of such individual differences and how task-relevance affects proactive interference.

Finally, Chapter 4 presents a double dissociation as further strong evidence for task-relevance as necessary for the creation of proactive interference. In a 1-back task (Smith &Jonides, 1997; Szmalec, Verbruggen, Vandierendonck & Kemps, 2011), similarity was manipulated in number identity and color dimensions, and task-relevance was controlled between-subjects by different instruction sets. Interference was measured in accuracy for repeated 2-back items along task-relevant and task-irrelevant dimensions. Results suggest that task-relevance is important for the creation of interference, and also that the automaticity of information processing may be considered as well.

Together these results show that activation strength or similarity alone cannot fully explain patterns of interference creation. Instead an additional criterion must be added stipulating that similarity must be along a task-relevant dimension in order to create interference. Further, there may be additional criteria related to individual differences or changes in task difficulties. These data suggest that interference is not as pervasive or unavoidable as initially assumed. Rather, it may be possible to avoid the effects of the recent past and protect STM.

References

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

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.

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 Bower, G.H. (Ed.), *The Psychology of Learning and Motivation, Vol. 8.* NY: Academic Press.

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.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12-21.

Burgess, N.& Hitch, G.J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106,* 551-581.

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

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

Hasher, L. & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356-388.

Howard, M.W., & Kahana, M.J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269-299.

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

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.

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 Berntson, G.G. & Cacioppo, J.T. (Eds.), *Handbook of neuroscience for the behavioral sciences, Vol. 1.* Hoboken, N.J.: John Wiley & Sons, Inc.

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

Mecklinger, A., Weber, K., Gunter, T.C.,& Engle, R.W. (2003). Dissociable brain mechanisms for inhibitory control: Effects of interference content and working memory capacity. *Cognitive Brain Research, 18,* 26-38.

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

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

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. (2006).Is the focus of attention in working memory expanded through practice? *Journal of Experimental Psychology: Learning, Memory and Cognition, 32*, 197-214.

Oberauer, K. & Lange, E.B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language, 58*, 730-745.

Paulesu, E., Frith, C.D.,&Frackowiak, R.S.J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362,* 342-345.

Peterson, L.R.,& Peterson, M.J. (1959). Short-term retention of individual verbal items.*Journal of Experimental Psychology, 58,* 193-198.

Smith, E.,&Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33,* 5-42.

Szmalec, A., Verbruggen, F., Vandierendonck, A.,& Kemps, E. (2011).Control of interference during working memory updating.*Journal of Experimental Psychology: Human Perception and Performance, 37,* 137-151.

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.

Wagner, A.D., Shannon, B.J., Kahn, I.,& Buckner, R.L. (2005). Parietal lobe contributions to episodic memory retrieval.*Trends in Cognitive Sciences*, *9*, 445-453.

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.

Chapter 2

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 (Wickens, Born, & Allen, 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 short-term memory 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 short-term memory (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.

Introduction

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 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- and long-term memory (Underwood, 1957). In the present chapter, we test two common explanations for proactive interference: activation strength and similarity-based response competition.

Several models describe the contents of short-term memory (STM) in terms of activated representations from long-term memory 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., 3-digit numbers), memory for the list can be approximately as good as if no prior list had been studied (e.g., Johnson, 1933; McGeogh & McDonald, 1931; see reviews by Crowder, 1976; Keppel, 1968, Underwood, 1945, 1957). The release-from-proactive interference procedure (Wickens et al., 1963) provides a classic demonstration of the role of similarity in STM interference: Performance declines within as few as four trials if all trials used materials of the same class (e.g., letters versus 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). In 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 Figure 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 (non-recent negatives). These results appear to support activation-strength accounts, as 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 either from that category or the complementary one. The "mismatch" trials, in 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 compared to "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 non-recent 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 non-recent 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 not (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 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-negatives effect on STM trials, which require the use of temporal-order information, to its magnitude on semantic-memory trials, which do not.

Semantic judgments were chosen as the comparison task based on 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 faster reaction 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 current 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 to 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 tests 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 non-recent negative probes on semantic judgment trials, for which temporal-order information is not relevant.

Methods

Participants

Forty individuals (18 female, average age = 18.73 years, SD = 1.09) participated in this study. 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), a score less than 9 (out of a possible 48) on the Extended Range Vocabulary Test (ERVT, Version 3; Educational Testing Service, 1976), and/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, six participants were excluded due to health conditions and/or current medications, three participants were excluded due to poor performance on the ERVT, and data from two 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 (SD = 1.01) years, 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 (Figure 1). Each STM trial consisted of a black fixation cross appearing for 2000 msec followed by a red fixation cross, which appeared for 1000 msec and was accompanied by an alerting tone. The target set of four words was then presented for 2000 msec, followed by a 3000 msec delay before presentation of the probe word in the center of the screen. The probe word appeared for 2000 msec, 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 and warning tone, instead of a set of four words, a semantic-category judgment prompt appeared for the same duration of time. 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 in 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: non-recent trials could be preceded by either an STM or a category trial, and both recent and non-recent 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) X 2 (recency: recent, non-recent) X 2 (correct response: positive, negative) design and presented pseudo-randomly with the constraint that no more than 3 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 in previous trials: recent probes were members of the previous trial's memory set; non-recent probes had not appeared as memory-set members (or as probes) for at least 3 trials prior to the current trial. All trial types were randomly interspersed throughout each of four blocks, and the order of block was counterbalanced across subjects using an approximate Latin square design.

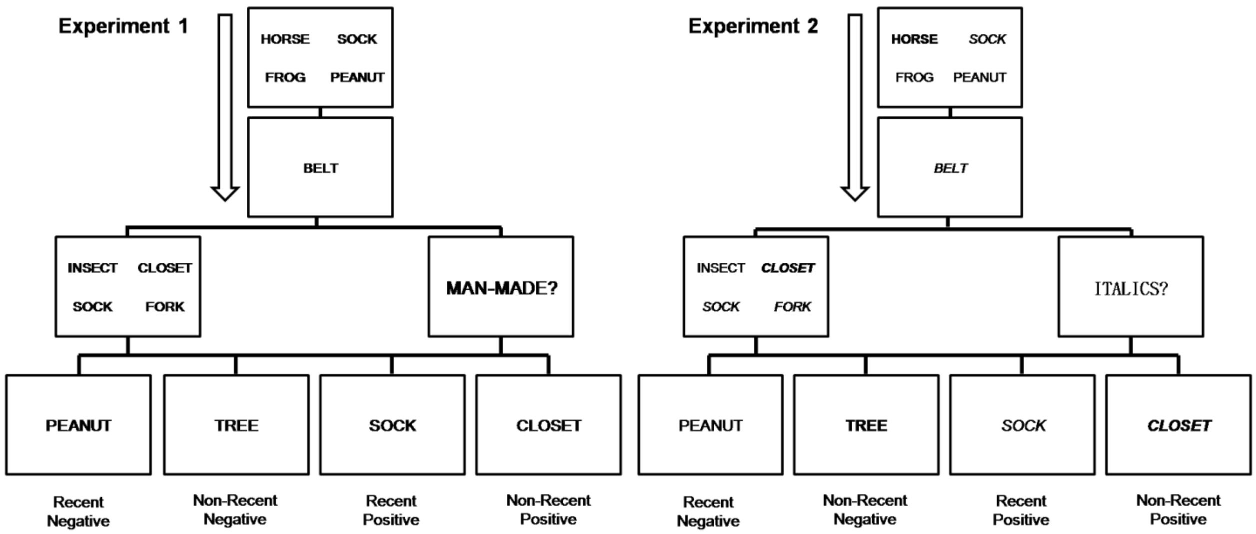

Figure 1. Sample trial sequences for Experiments 1 and 2. Time progresses linearly from top to bottom. The trial type is indicated below each critical probe (bottom row). In both experiments, the first trial of a critical sequence was an STM trial, during which participants saw a set of four words followed 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 either be another STM trial, in which case the same procedure as the initial trial was followed (4 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 x 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 non-recent (not present on the previous trial, e.g., TREE, CLOSET).

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. Participants were able to repeat practice if desired. Following practice, each participant completed four blocks of 64 trials, with 60 seconds 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 non-recent 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, non-recent; and trial type: STM, category). Because we used a repeated-measures design, effect sizes are reported in generalized η^2 values (abbreviated as η^2_G), rather than partial η^2 values. Effect-size heuristics for η^2_G are as follows: 0.02 is a small effect, 0.13 a medium effect, and 0.26 a large effect (Bakeman, 2005). To calculate η^2_G, we used the following formula: $\eta^2_G = SS_{effect} / (SS_{effect} + SS_{subjects})$. Where necessary, Greenhouse-Geisser sphericity corrections were applied to reported p-values; original degrees of freedom are used in the text for easier reading.

Table 1. Average median response times (msec) and accuracy scores (%) by trial type for both Experiment 1 and Experiment 2. Standard errors are presented in parentheses.

			Recent Negative	Non-Recent Negative	Recent Positive	Non-Recent 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)

Response Time

As seen in Figure 2a, the standard recent-negatives effect was found for STM trials (cf., Berman et al., 2009; Monsell, 1978) but there was no recent-negatives effect on semantic-judgment trials; interaction of recency and trial type, $F(1,28) = 17.45$, $p < .001$, $\eta^2_G = .02$. Confirming this impression, post-hoc t-tests revealed a significant difference between recent negative and non-recent negative STM trials ($t(28) = 5.56$, $p < .001$, d = 1.03), but no difference between recent negative and non-recent 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, $F(1,28) = 64.57$, $p < .001$, $\eta^2_G = .34$. There was also a statistical main effect of recency, with recent negative trials slower than non-recent negative trials, $F(1,28) = 9.87$, $p < .01$, $\eta^2_G = .01$; as noted above this effect was driven by STM trials and did not occur on semantic-judgment trials.

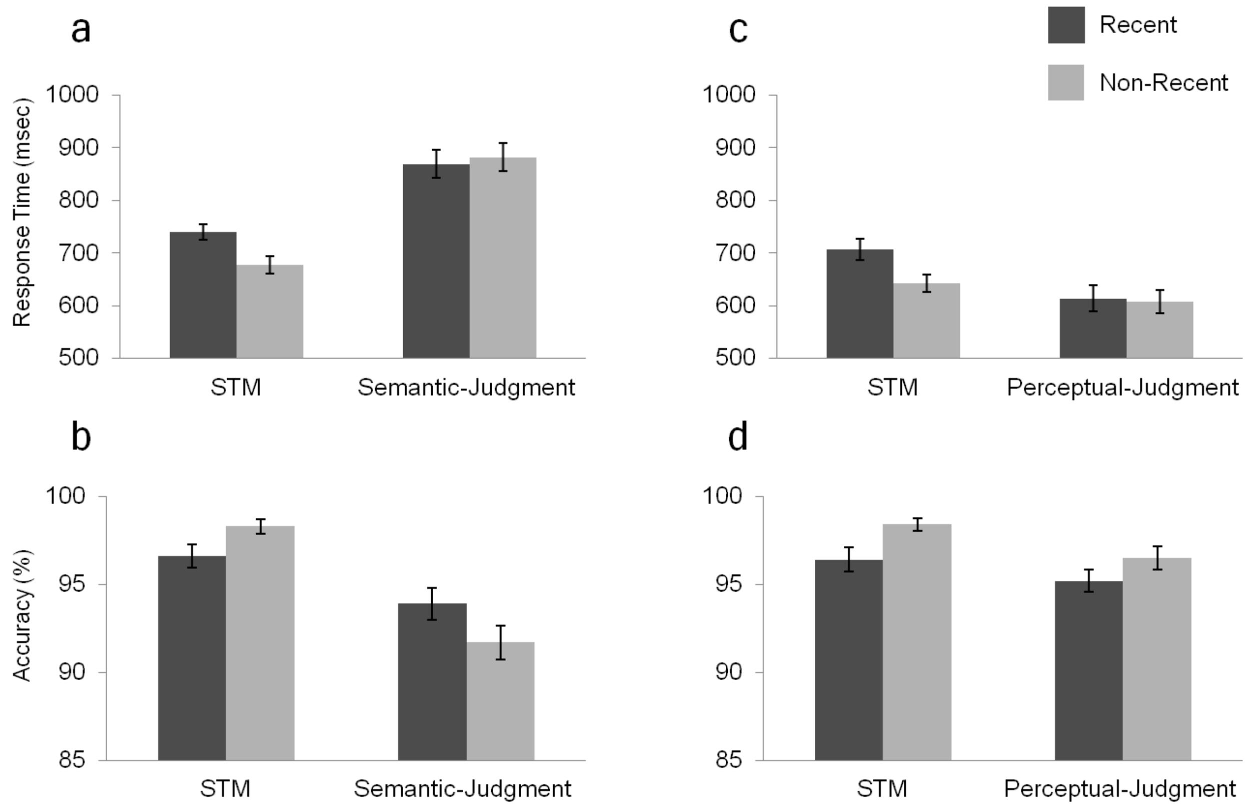

Figure 2. Average median response times (top panels) and accuracy (bottom panels) for negative trials in Experiments 1 (left) and 2 (right). Error bars on this and subsequent figures represent between-subjects standard error and should not be used for evaluating the significance of within-subjects comparisons.

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

	Experiment 1			**Experiment 2**		
2x2 ANOVA for negative RTs	*F* value	*p* value	η^2_G	*F* value	*p* value	η^2_G
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 x recency interaction	17.45	< .001	.02	10.19	< .01	.01
2x2 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 x recency interaction	0.00	.97	< .001	10.77	< .01	.007
2x2 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 x recency interaction	9.39	< .01	.05	0.63	.43	.01
2x2 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 x recency interaction	4.09	.05	.02	30.98	< .001	.05

Accuracy

As seen in Figure 2b, the accuracy results followed a pattern consistent with the response time data, with a significant interaction between recency and trial type, $F(1,28) = 9.39$, $p < .01$, $\eta^2_G = .05$. Participants were less accurate in rejecting recent trials than non-recent 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, $F(1,28) = 41.77$, $p < .001$, $\eta^2_G = .24$). 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 only create interference if that dimension is relevant to the task.

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 subject 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, then 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, then temporal recency should affect STM trials but not perceptual-judgment trials.

Methods

Participants

Fifty four participants (33 female, average age = 20.37 years, SD = 2.41) participated in this study. Two participants were excluded due to medication/health conditions, five participants were excluded for failure to reach the criterion ERVT score, and 10 for failing to meet accuracy criteria. While only two of these participants had performance below 80% accuracy for STM trials, all ten had performance below 80% for category trials. In particular, subjects had a difficult time correctly identifying trials in which they were required to classify the words as italicized or non-italicized.

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 ($15/hour). No significant differences were found between paid and unpaid subjects in overall response times ($t(35) = 1.00$, $p = .33$) or accuracy ($t < 1$). Participants had a mean age of 20.05 (SD = 2.42) years and had completed an average of 13.78 (SD = 1.89) years 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 Figure 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 man-made 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. Non-bold 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 pt) so as not to bias participants towards 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, $F(1,36) = 10.19$, $p <.01$, $\eta^2_G = .01$). For STM trials, recent trials took longer than non-recent trials, $t(36) = 5.88$, $p < .001$, d = 0.97; for perceptual-judgment trials, the two trial types had similar response times, $t < 1$ (Figure 2c).

Also replicating Experiment 1, there was a statistically significant main effect of recency, $F(1,36) = 19.81$, $p < .001$, $\eta^2_G = .02$) 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, ($F(1,36) = 35.80$, $p < 0.001$, $\eta^2_G = .06$). The opposite patterns when comparing overall response times for STM vs. category judgments across experiments suggests 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 to non-recent trials, $F(1,36) = 10.14$, $p < .01$, $\eta^2_G = .05$. In addition, STM trials were more accurate than category trials, $F(1,36) = 6.87$, $p = .01$, $\eta^2_G = .04$. 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 (Figure 2d).

This result was surprising in comparison to 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 based on their relative accuracy on italics judgments. Specifically, we calculated a difference score for each participant between the accuracy on bold judgments and italics judgments; those 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 non-recent 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 non-recent trials for either STM accuracy ($t(18) = -1.57$, *n.s.*, d = -0.36) or category accuracy ($t(18) = -1.60$, *n.s.*, d = -.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 Appendix A.

Table 3. Average median response times (standard error) and accuracies (standard error) for high and low performing groups in Experiment 2, as determined by a comparison between italics and bold category judgments.

			Recent Negative	Non-Recent Negative	Recent Positive	Non-Recent 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)

In summary, the results of Experiment 2 replicated the important aspects of Experiment 1, especially with regard to reaction time. These findings provide further support for the hypothesis that temporal similarity creates proactive interference on tasks where temporal information is relevant, but it does not create proactive interference when temporal information is irrelevant to the task – at least when subjects 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 study. Three were excluded due to failure to adhere to instructions on color-task mapping, two due to failing to meet accuracy criteria, one due to technical problems, and eight 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 (SD = 1.67) years of formal education, and scored between 9.75 and 35.25 on the ERVT, M = 20.06, SD = 6.04.

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 (Figure 3). Due 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 2000 msec. Following this set, a colored square outline appeared for 1500 msec with a fixation cross centered within it. The square appeared either in 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 button press, either '1' or '0', to indicate a "yes" or "no" response to the probe. The mappings of the button presses were also counterbalanced across participants.

As before, the critical comparisons were between recent and non-recent probes. The altered structure of the category trials now allowed items from those trials to serve as a source of recency. Thus, both recent and non-recent 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 vs. category, positive vs. negative, and switch vs.

nonswitch trials, with a pseudorandom order of presentation and the constraint that no more than three negative responses could occur in a row.

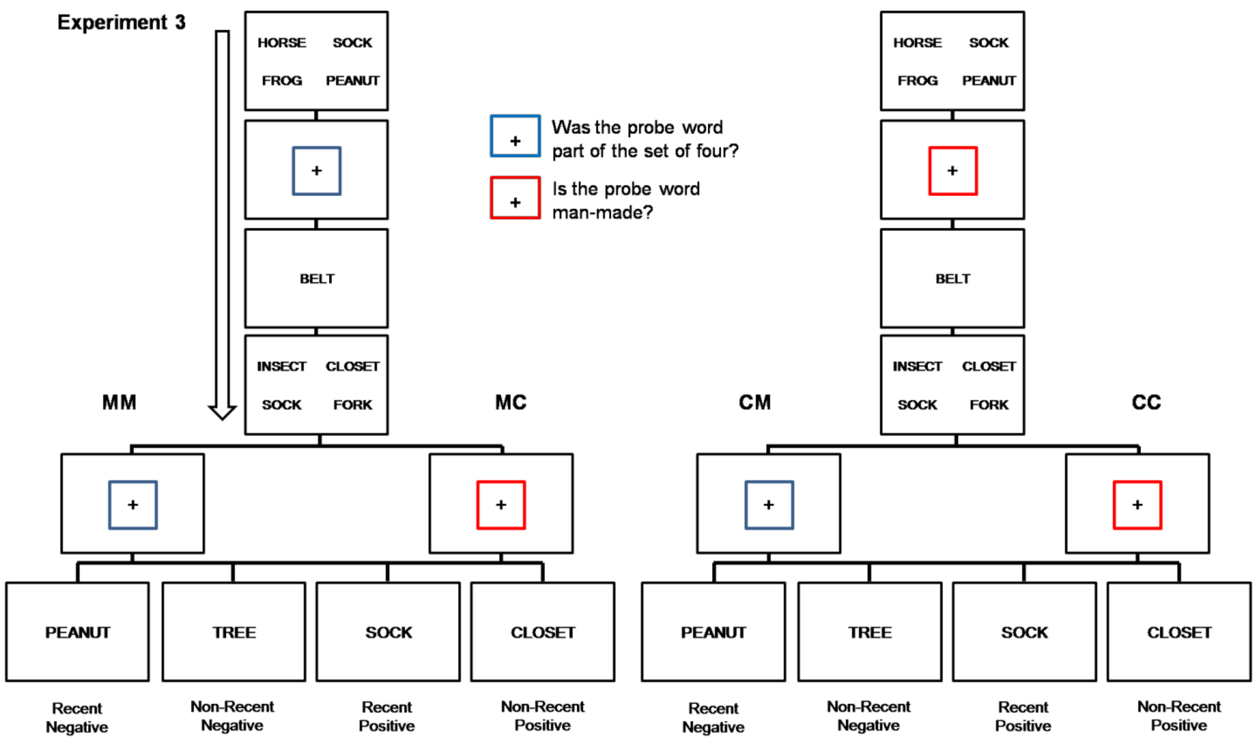

Figure 3. Sample trial sequences for Experiment 3. After presenting a set of 4 items, a color cue during fixation indicated which task to perform. Recency was established based on the items presented as part of the previous set of 4 items and could derive from either prior-STM or prior-category trials, eliminating the task-switching confound present in Experiments 1 and 2.

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 6 blocks of 64 trials. In between each block, a short (2-6 minute) 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 only served as fillers to keep subjects 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 2x2x2 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 SEs are presented in Table 4. The three-way interaction between recency, previous trial type, and current trial type was not significant, $F(1,29) = 2.72$, $p = .11$, indicating that prior trial type did not influence the size of the difference in interference effects between STM and category trials (Figure 4). Planned follow-ups indicated that while recency slowed response times in the MM ($t(29) = 6.54$, $p < .0005$, d = 1.19) and CM ($t(29) = 2.66$, $p < .05$, d = 0.49) trials, recency did not affect response times for MC (t < 1) or CC ($t(29) = 1.36$, $p = .18$, d = .25) trials.

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 X 2 ANOVAs (switch X recency) within each current trial type (STM or category).

Table 4. Average median response times (standard error) and accuracies (standard error) broken down by trial type for Experiment 3.

			Recent Negative	Non-Recent Negative	Recent Positive	Non-Recent 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)

Table 5. Statistics for all 2x2x2 ANOVA analyses in Experiment 3.

		Negative Trials			**Positive Trials**		
		F value	*p* value	η^2_G	*F* value	*p* value	η^2_G
Response Time (msec)	*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 x previous trial type interaction	1.81	.19	<.01	0.43	.52	<.001
	Recency x current trial type interaction	15.31	<.01	.01	0.08	.79	<.001
	Previous x current trial type interaction	3.42	.08	<.01	10.03	<.01	.01
	3-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 x previous trial type interaction	0.69	.42	.03	15.58	<.001	.05
	Recency x current trial type interaction	6.88	.01	.02	23.45	<.001	.06
	Previous x current trial type interaction	0.21	.65	<.001	6.44	.02	.02
	3-way interaction	0.62	.44	<.01	0.14	.72	<.001

Table 6. Statistics for all 2x2 ANOVA analyses 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 (msec)	*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 x 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 x previous trial type interaction	1.59	.22	.01	20.71	<.001	.11
Current Category Trials	Response Time (msec)	*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 x 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 x previous trial type interaction	.02	.88	<.001	6.65	.02	.04

For current STM trials, the switch X recency interaction was significant, $F(1,29) = 6.28$, $p < .05$, $\eta^2_G = .01$, indicating greater interference for MM trials than 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 towards a main effect of recency, but it did not reach significance, ($F(1,29) = 3.28$, $p = .08$, $\eta^2_G < .005$). As noted earlier, planned t-tests indicated that the recency effect was significant for both types of STM trials, both $p < .05$, $d > .45$, but neither type of category trial, both $p > .15$, $d < .30$.

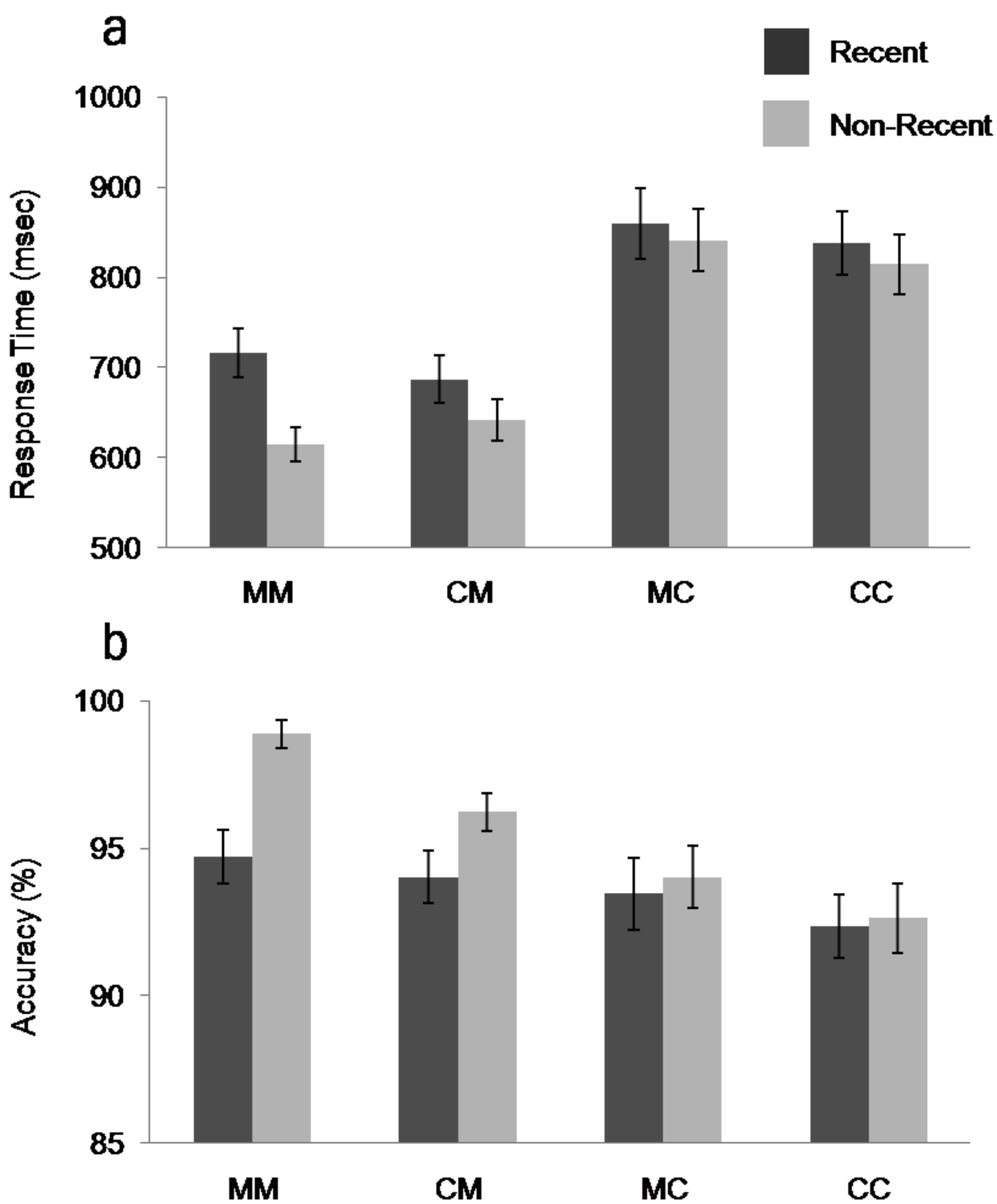

Figure 4. Average median response times (top panel) and accuracy (bottom panel) for critical trials in Experiment 3. Note that M stands for STM trials, and C for category-judgment trials. The combination of abbreviations indicates trial order (e.g., MM indicates a STM trial preceded by an STM trial; CM indicates a STM trial preceded by a category trial).

Accuracy

For accuracy, the 3-way interaction between previous trial type, recency, and current trial type was not significant, $F < 1$. However, the 2 x 2 interaction between recency and current trial type was significant, $F(1,29) = 6.88$, $p < .05$, $\eta^2_G = .02$, once again indicating a larger interference effect for STM trials than category trials (Figure 4). Planned t-tests confirmed that recency-based interference reduced accuracy for STM trials, (MM condition ($t(29) = -4.26$, $p < .0005$, d = -0.78)), and was marginally significant for STM trials which were preceded by a category trial, (CM condition ($t(29) = -1.94$, $p = .06$, d = -0.35)) but recency-based interference did not affect the accuracy of category trials, both t < 1.

In summary, the results of this experiment generally replicated the patterns seen in Experiments 1 & 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 nonsignficant 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 as is it is a post-hoc explanation of a nonsignificant effect we do not consider it further here. Overall the results indicate that regardless of trial sequence or switching, recency led to interference on STM trials but not 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 replicates the design of Atkins et al. (2011), but manipulates the perceptual match (rather than the category match) between the memory set and the probe items. All trials are STM trials, which ask participants 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-negatives 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 only contributes to interference 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-negatives effect.[1]

[1]We ran a small additional experiment to ensure that the color dimension could have an effect if participants knew 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 $F(1, 12) = 4.97$, $p < .05$).

Methods

Participants

Thirty four participants (24 female, average age = 18.47 years, SD = 0.51) participated in this study 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 thirty-two participants (23 female). Participants had a mean age of 18.47 (SD = 0.51) years, and had completed an average of 12.06 (SD = 0.25) years of formal education. ERVT scores for included participants ranged between 9.00 and 29.50 (M = 18.21, SD = 5.67).

Stimuli

The same pool of word stimuli used for Experiments 1, 2, and 3 was used here; the trial structure was the same as 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 same color (match) or the complementary color (mismatch) as the memory set (Figure 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 noted earlier, this allowed us to make competing predictions regarding the influence of the color dimension. If similarity along the color dimension influences performance, then 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, then interference as indexed by the recent-negatives effect ought to be larger for color-match than 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, then both overall response time and the recent-negatives 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 non-recent trials, color-match recent trials, and color-match non-recent 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 randomly interspersed throughout each of four blocks of 64 trials, and the order of block was counterbalanced across subjects using an approximate Latin square design.

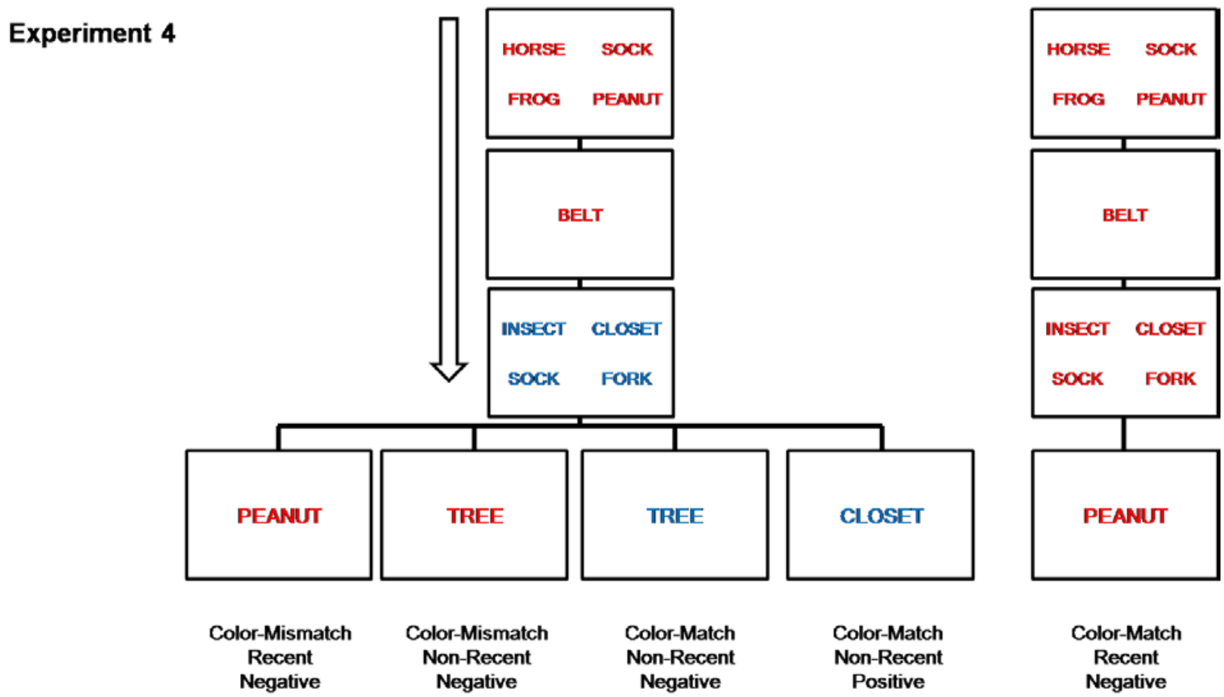

Figure 5. Sample trial sequences for Experiment 4. The trial type is indicated below each critical probe. Perceptual similarity was manipulated between the current memory set and the probe word such that color-mismatch trials had low perceptual similarity with the memory set whereas color-match trials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

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

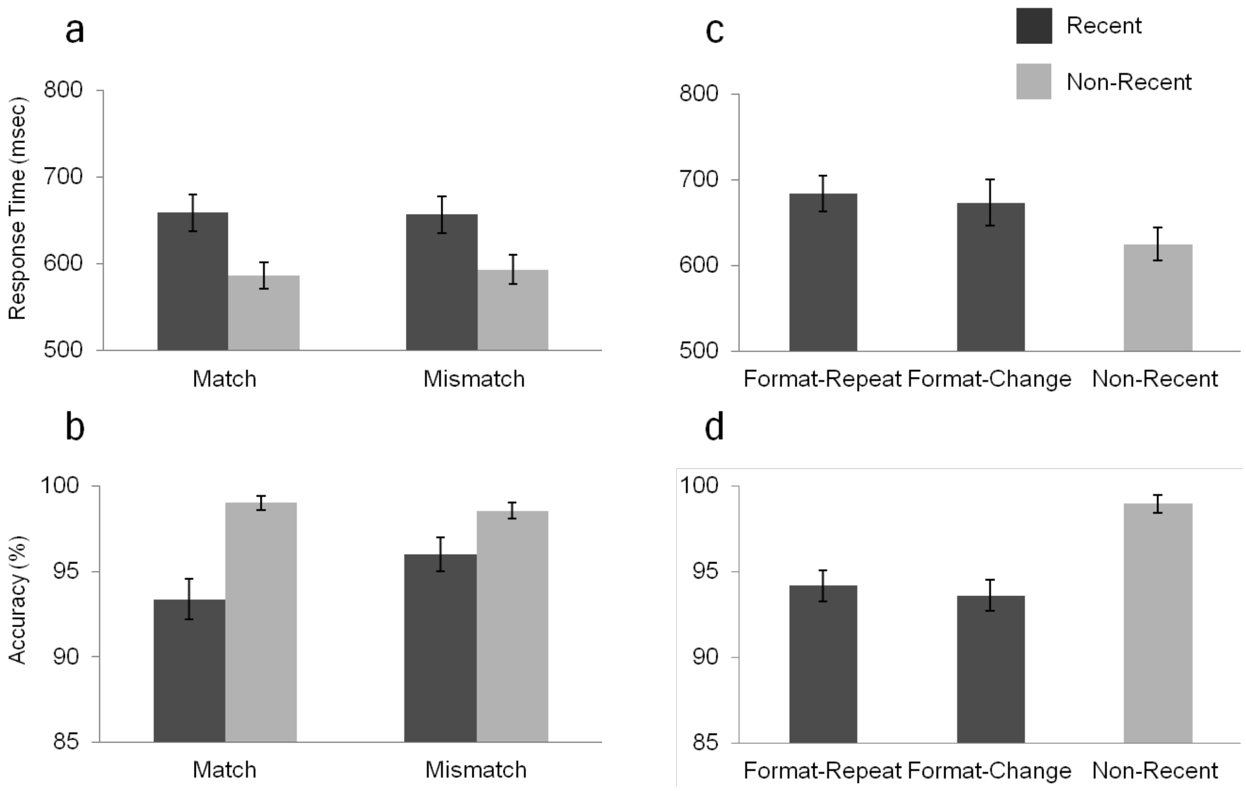

Figure 6. Average median response times (top panels) as well as accuracies (bottom panels) for negative responses in Experiment 4(left; panels a and b) and Experiment 5 (right; panels c and d).

Response Time

As seen in Figure 6a, this experiment replicated the standard recent-negatives effect, $F(1, 31) = 87.81$, $p < .001$, $\eta^2_G = .10$, 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 effect, $F(1,31) = 4.71$, $p = .04$, $\eta^2_G = .03$. 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 subjects showed an effect of color match, suggesting that the non-relevant dimension may begin to have an influence when participants have difficulty making the judgment on the relevant dimension (See Appendix A). Replicating standard results, accuracy was lower for recent probes than for non-recent probes, $F(1,31) = 38.80$, $p < .001$, $\eta^2_G = .16$.

Table 7. Average median response times (standard error) in milliseconds, and accuracy scores (standard error) by trial type for Experiment 4. Note that because all positive trials were classified as color-match trials, there are no values for color-mismatch positive trials.

			Recent Negative	Non-Recent 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)

Table 8. Statistical analyses and effect sizes for Experiment 4.

	Experiment 4		
2x2 ANOVA for negative RTs	*F* value	*p* value	η^2_G
Main effect of trial type	0.16	.69	< .001
Main effect of recency	87.81	< .001	.10
Trial type x recency interaction	0.80	.38	< .001
2x2 ANOVA for negative accuracies			
Main effect of trial type	3.13	.09	.01
Main effect of recency	38.80	< .001	.16
Trial type x recency interaction	4.71	.04	.03

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 presentation 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 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.

Methods

Participants

Twenty five participants (16 female, average age = 18.32 years, SD = 0.75) participated in this study. One participant was excluded due to medication/health conditions, one due to missing data, and five 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 (SD = 0.85) years, and had completed an average of 12.28 (SD = 0.75) years 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 (Figure 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 in the current trial compared to their format in the previous trial's memory set. For example, for a format-change trial, a word that had appeared in orange, bolded, non-italicized Arial font on the immediately-previous trial would appear in blue, non-bold, 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 subjects using an approximate Latin square design.

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

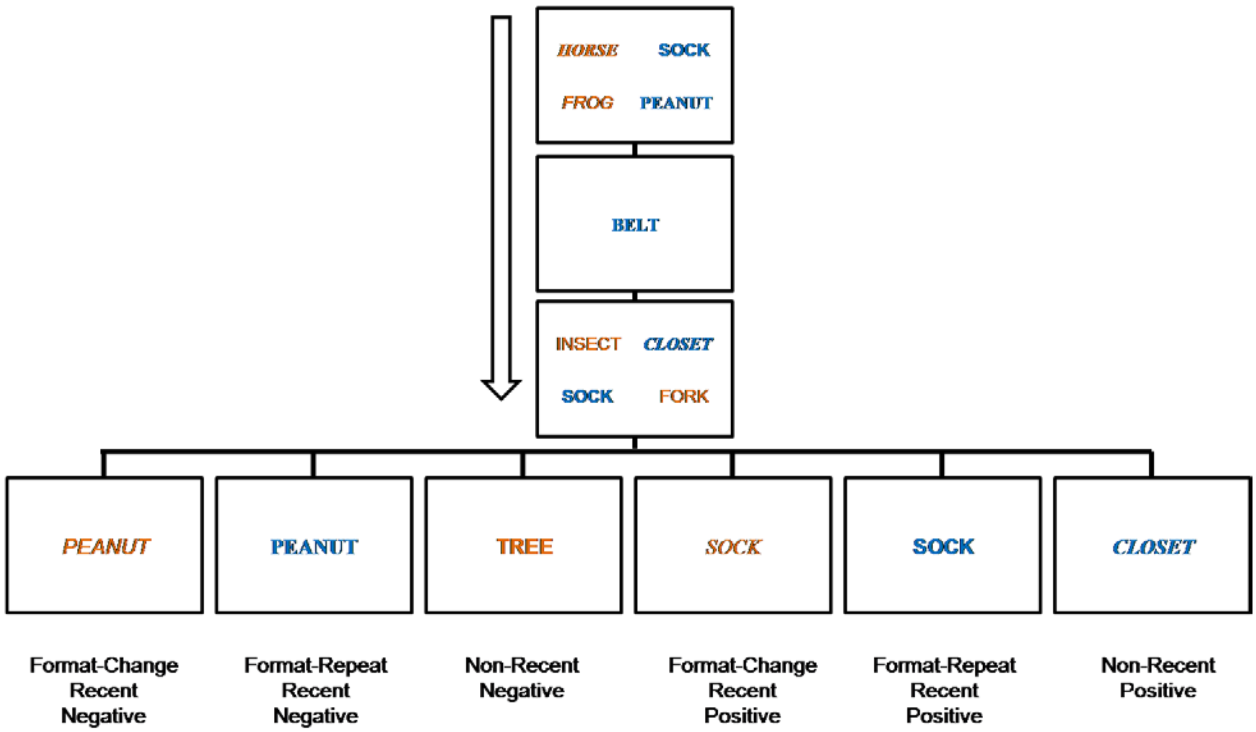

Figure 7. Sample trial sequences for Experiment 5. Perceptual similarity was manipulated between the first presentation of an item and its second presentation as a recent probe word, so that probes on format-change trials had very little perceptual similarity across presentations, while probes on format-repeat trials had exactly the same format across presentations.

Results and Discussion

Both median response times and accuracies for negative trials were analyzed using ANOVA with trial type (non-recent 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 seen in Figure 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 non-recent negatives, $F(2,34) = 15.26$, $p < .001$, $\eta^2_G = .07$, but there was no difference between format-repeat and format-change trials, $t < 1$.

Table 9. Average median response times (standard error) in milliseconds, and accuracy scores (standard error) by trial type for Experiment 5.

			"NO" response	"YES" response
Experiment 5	Non-Recent	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^2_G
3x1 ANOVA for negative RTs	15.26	< .001	.07
3x1 ANOVA for positive RTs	2.13	.14	.01
3x1 ANOVA for negative accuracies	15.74	< .001	.34
3x1 ANOVA for positive accuracies	2.07	.15	.06

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 in rejecting either type of recent negative, $F(2,34) = 15.74$, $p < .001$, $\eta^2_G = .34$, and correct rejection rates for format-repeat vs. format-change recent-negative trials were equivalent, $t < 1$. (Figure 6d).

In summary, changing the perceptual qualities of the recent probe from its first presentation did not alter the recent negatives 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 temporal characteristics of that item were important for responding to the test cue, as in STM trials. Conversely, changing non-temporal 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, 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 only tested 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 proactive interference 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 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 & b). 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 to non-recent 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 versus 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 to reject it.

Summary

How items are represented in STM and what factors lead to forgetting and interference are issues of long standing (see 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 & 4). Overall, however, our results suggest that to escape the past, make it irrelevant.

Author note

This chapter presents research conducted in collaboration with Marc G. Berman, John Jonides and Cindy Lustig. This work was supported by an NIMH grant to J.J. (MH 60655). We thank Patricia Reuter-Lorenz for comments on a previous version of this manuscript, and Madison Weisberg, Emily Kurzawa, and Matthew Lugo for their assistance with data collection. We also thank the University of Michigan Center for Statistical Consulting and Research for assistance with effect-size and power analyses during the revision process.

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 Bower, G.H. (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. & 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.

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 Berntson, G.G. &

Cacioppo, J.T. (Eds.), *Handbook of neuroscience for the behavioral sciences, Vol. 1*. Hoboken, N.J.: John Wiley & Sons, Inc.

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.

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.

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

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.

Chapter 3

Introduction

Proactive interference occurs whenever previous information interferes with current information processing. Traditional theories suggest that interference occurs because recent information retains relatively high amounts of activation strength (e.g., Wickelgren & Norman, 1966; Altmann & Gray, 2002), or that highly similar information creates interference (e.g., Johnson, 1933; Keppel & Underwood, 1962; Bunting, 2006). However, research presented in Chapter 2 suggests that there may be additional factors that influence whether or not a current item will be affected by proactive interference. Specifically, proactive interference was found when the current item was highly similar to a previously-seen item in a task-relevant stimulus dimension. However, when the task was changed to one in which that dimension of similarity was irrelevant (Experiments 1, 2, and 3) or when the dimension of similarity manipulated was not relevant to the task being performed (Experiments 4 and 5), proactive interference was not observed.

We hypothesized that an additional boundary condition existed for the creation of interference: that highly similar information could interfere only when it was along a task-relevant stimulus dimension. When completing any given task, a participant should check in memory only for the stimulus dimensions that are important for the task. When a previously-seen item is highly similar to a target item along a checked dimension, it will interfere, creating confusion and forgetting in short-term memory (STM). However, when a previously-seen item is highly similar to a target item along a dimension that is not checked, this similarity goes unnoticed, and proactive interference is not created. This allows similarity to exist along task-irrelevant dimensions without affecting performance, and also allows the same similarity to affect performance on some tasks (when it is relevant) but not others (when it is irrelevant; see Experiment 3, pg. 35).

However, it is possible that in some cases task-irrelevant dimensions of similarity may influence proactive interference. Experiment 2 of Chapter 2 examined whether similarity along the temporal dimension would create interference in a task-relevant (recent probes) or a task-irrelevant (perceptual-judgment categorization) context. Surprisingly, there was some evidence for recency-based interference even when this dimension of similarity was irrelevant (on the perceptual-judgment task). However, closer examination of the data suggested that this interference might only occur for those individuals who had difficulty making the task-relevant (perceptual) judgment (in particular, those who had difficulty differentiating italicized from non-italicized font).

Why might this happen? It is possible that when a task is very difficult, consulting only relevant dimensions of similarity does not lead to an obvious correct response; in these cases, individuals may consult task-irrelevant dimensions in order to identify which response is most likely correct. In the situation described above, this strategy would lead to proactive interference in the low-performing group even in the irrelevant task.

To further test the hypothesis that the results of Experiment 2 occurred because of difficulty with the italics judgment, we trained individuals on each type of perceptual-judgment before repeating the experimental task from Experiment 2. If difficulty with the italics judgment caused individuals to rely on task-irrelevant stimulus dimensions to make a response, training should eliminate the need to consult these task-irrelevant stimulus dimensions.

Methods

Participants

Forty-seven individuals (17 female, average age = 19.00 years, SD = 1.04) participated in this study. All individuals were recruited through the University of Michigan Subject Pool and received course credit for participation. Exclusion criteria were the same as in Chapter 2; for more details, see pg. 19. Three participants were excluded due to health conditions and/or current medications, four due to poor performance on the ERVT, four for failure to achieve at least 80% accuracy on STM and/or perceptual-judgment trials, and data from four participants were lost for technical reasons. Thirty-two healthy individuals (11 female) were included in the final analysis. These participants had ERVT scores ranging from 9.00 to 36.75, M = 17.70, SD = 6.66. They had an average age of 19.09 (SD = 1.17) years, and had completed an average of 12.59 years of formal education (SD = 0.84).

Design and Materials

All aspects of the research were approved by the Behavioral Sciences Institutional Review Board at the University of Michigan. With the exception of the training task (see below), all aspects of the design, materials, and procedure were identical to Experiment 2 of Chapter 2.

Stimuli were displayed in 16-point font using E-Prime 2.0 software (Psychology Software Tools, Inc.). Fifty percent of all words were displayed in italics and 50% in bold, with the exception of part 1 of the training task (see description below). Words in bold were also displayed in Copperplate Gothic Bold font; items not in bold were displayed in Copperplate Gothic Light font. For more details on how bold and italic qualities of the items were organized within the experimental task, see Experiment 2, Chapter 2 (pg. 30).

Training task.

The training task was broken into two sections; in the first section participants were trained on italics judgments, and in the second, participants were trained on bold judgments. Participants completed both sections of the training task before completing the experimental task. Part 1 of each section consisted of a block of 10 trials. In this block, participants saw a set of four words and had to determine which word was displayed with the relevant perceptual quality for that section (italics or bold font). A black fixation cross appeared for 2000 msec, followed by a red fixation cross for 1000 msec. After this the 4-word display appeared until the participant indicated that they were ready to respond by pressing the space bar. Finally, a cue appeared indicating that participants should press the button on the keyboard that corresponded with the first letter of the word they had identified as written in italics or bold font in the set of 4 words (see Figure 8a). Participants were able to repeat this portion of the training if they chose to do so.

In part 2 of each section of the training task, participants completed blocks of 36 trials where they viewed words one at a time (see Figure 8b). Each word was preceded by a black fixation cross for 2000 msec, followed by a red fixation cross for 1000 msec. Then the test word appeared on screen until participants made a response indicating whether the word was or was not displayed with the relevant quality for that section (italics or bold font). Responses were made on a standard computer keyboard using the ‘0’ and ‘1’ keys. The relationship between key and response was counterbalanced across participants, and they were instructed to perform all keypresses with the left and right index fingers.

After each block of 36 trials, participants were given feedback on their performance and rested for 15 seconds. A criterion accuracy score of $\geq$ 90% was required to advance to the next section of training or to the experimental task. Participants continued to complete blocks of trials until they achieved this criterion score.

In the sets of 4 words, one word always appeared in italics and one word always appeared in bold font. On half of these trials, the same word appeared in both italics and bold font. When words were seen one at a time, 25% were displayed in plain text, 25% in both italics and bold font, 25% in italics only, and 25% in bold font only.

Experimental task.

Following the training task, participants completed the experimental task. Trials followed the same design as that of Experiment 2 of Chapter 2 (see pg. 30), and consisted of two distinct types: STM trials and perceptual-judgment categorization trials (see Figure 1, pg. 23). Each STM trial consisted of a black fixation cross (for 2000 msec) followed by a red fixation cross (1000 msec). The target set of four words was then presented (2000 msec), followed by a second black fixation cross (3000 msec) before presentation of the probe word in the center of the screen. The probe word appeared for 2000 msec or until the participant made a keypress response (‘0’ or ‘1’) indicating whether it was (positive probe) or was not (negative probe) a member of the current memory set. The relationship between key and response followed the guidelines described above for the training task.

Perceptual-judgment trials proceeded in the same manner as STM trials, with one exception. Rather than a set of four words, a perceptual-judgment categorization prompt appeared for the same duration of time. The prompts were identical to those used in Experiment 2 of Chapter 2 (“ITALICS?” or “BOLD?”) and indicated which category dimension was relevant. As in STM trials, when the probe appeared participants made a keypress (‘0’ or ‘1’) to indicate their response.

As in Experiment 2 of Chapter 2, proactive interference was manipulated through recency. In recent trials, the probe was a member of the previous trial’s memory set; in non-recent trials, the probe had not been seen as either a probe or part of a memory set for at least 3 trials before the current trial. All trials were repeated exactly as in Experiment 2 of Chapter 2; for more details on trial organization, see pgs. 20 and 30.

A 2 (trial type: STM or category) X 2 (recency: recent, non-recent) X 2 (correct response: positive, negative) design was again used for this experiment. The order of block was counterbalanced across subjects using a Latin square design.

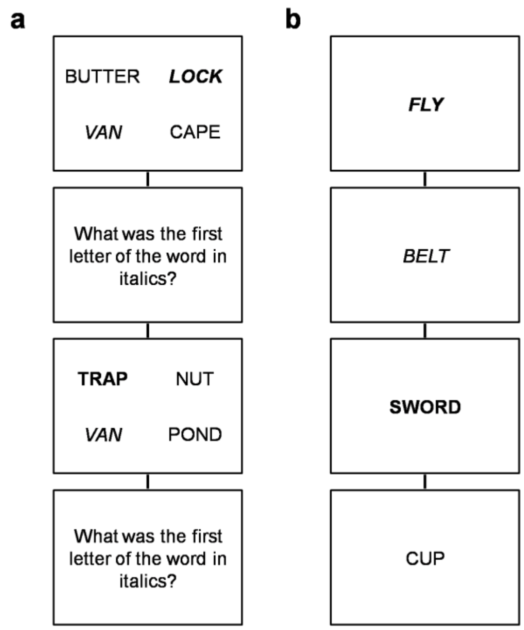

Figure 8. Sample trial sequence for parts 1 (a) and 2 (b) of the training task; fixation crosses have been removed for clarity and spacing concerns. In (a), the cue asked was about the first letter of the italics word (in the italics section) and about the bold word (in the bold section). Time progresses linearly from top to bottom.

Procedure

After providing written informed consent, participants were trained on the italics and bold judgments. They then completed practice on the experimental task before beginning the experiment. Practice consisted of four STM trials and four perceptual-judgment trials, two of each perceptual-judgment type. Participants were able to repeat practice if desired. Following practice, each participant completed four blocks of 64 trials, with 60 seconds of rest in between blocks.

Results and Discussion

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 was 1.87% 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 non-recent 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 11; for statistics on positive trials, see Table 12.

Both response time and accuracy measures were analyzed using a repeated-measures ANOVA design with two independent variables (recency: recent, non-recent; and trial type: STM, perceptual-judgment), followed by planned t-tests comparing recent and non-recent trials of the same general type (e.g., recent negative memory trials vs. non-recent negative memory trials). Because we used a repeated-measures design, effect sizes are reported in η^2_G (for more details on this measure of effect size, see pg. 24).

Table 11. Average median response times (standard error) in milliseconds, and accuracy scores (standard error) in percentages, by trial type for Chapter 3 analyses.

		Recent Negative	Non-Recent Negative	Recent Positive	Non-Recent Positive
Memory	RT	712.55(24.09)	623.02(20.56)	646.59(25.82)	694.06(28.28)
	Accuracy	96.29(0.75)	98.05(0.50)	95.22(0.79)	95.22(0.76)
Perceptual-Judgment	RT	586.22(22.62)	596.31(25.74)	556.58(24.14)	536.38(21.45)
	Accuracy	94.53(0.89)	94.34(0.85)	94.04(1.03)	92.58(0.70)

Training Task

As noted above, all participants completed the training task to a criterion 90% accuracy. All but two participants achieved the criterion score on their first block of 36 trials for both the italics and bold versions of the task; the remaining two achieved the criterion score on their second block.

Because participants scored so accurately on their first block of trials, training was assessed with response times. Faster times indicated that the skills of identifying words displayed in italics and bold font were successfully trained. This was indexed by the slope of the response time data over the course of the block, and evaluated using a one-sample t-test. Results indicated that for both italics (t(x) = -3.03, p < .05, d = -0.54) and bold (t(31) = -1.86, p = .07, d = -0.33) judgments, training significantly improved performance.

Response Time

As seen in Figure 9a, the standard proactive interference effect was found for STM trials (cf., Berman et al., 2009; Monsell, 1978) but there was no proactive interference effect on perceptual-judgment trials; interaction of recency and trial type, $F(1,31) = 58.41$, $p < .0005$, $\eta^2_G = .04$. Confirming this impression, planned t-tests revealed a significant difference between recent negative and non-recent negative STM trials ($t(31) = 9.68$, $p < .0005$, d = 1.71), but no difference between recent negative and non-recent negative perceptual-judgment trials, $t(31) = -1.12$, $p = .27$, d = -0.20.

Although not relevant to our theoretical question, for completeness we note a statistical main effect of trial type, with perceptual-judgment trials faster overall than STM trials, $F(1,31) = 41.88$, $p < .0005$, $\eta^2_G = .08$. There was also a statistical main effect of recency, with recent trials slower overall compared to non-recent trials, $F(1,31) = 38.24$, $p < .0005$, $\eta^2_G = .02$; as noted above this effect was driven by STM trials and did not occur on perceptual-judgment trials.

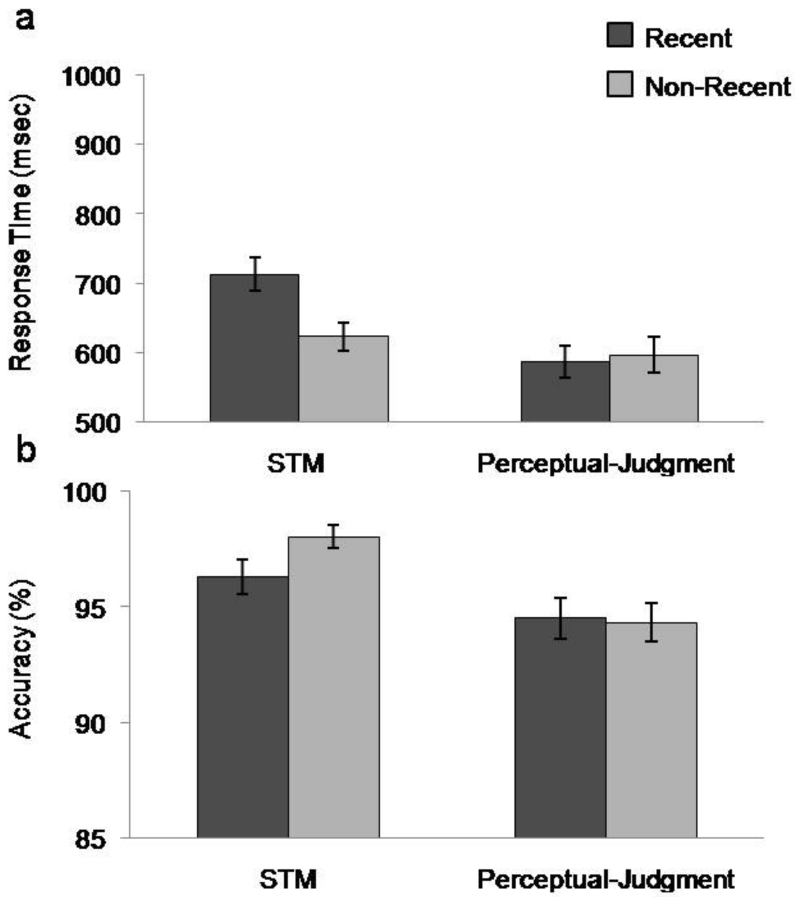

Figure 9. Average median response time (top panel) and accuracy (bottom panel) for negative trials. Error bars represent between-subjects standard error and should not be used for evaluating the significance of within-subjects comparisons.

Table 12. Statistical analyses and effect sizes for all Chapter 3 2x2 ANOVA comparisons.

	F value	*p* value	η^2_G
2x2 ANOVA for negative RTs			
Main effect of trial type	41.88	<.0005	.08
Main effect of recency	38.24	<.0005	.02
Trial type x recency interaction	58.41	<.0005	.04
2x2 ANOVA for positive RTs			
Main effect of trial type	50.75	<.0005	.16
Main effect of recency	2.73	.11	.002
Trial type x recency interaction	23.50	<.0005	.01
2x2 ANOVA for negative accuracies			
Main effect of trial type	10.20	<.01	.09
Main effect of recency	1.77	.19	.008
Trial type x recency interaction	1.59	.22	.01
2x2 ANOVA for positive accuracies			
Main effect of trial type	5.86	<.05	.04
Main effect of recency	0.85	.36	.006
Trial type x recency interaction	0.97	.33	.006

Accuracy

As seen in Figure 9b, perceptual-judgment trials were less accurate overall than STM trials, $F(1,31) = 10.20$, $p < .01$, $\eta^2_G = .09$. All other analyses were nonsignificant, suggesting that interference did not have an effect on accuracy, either in STM or perceptual-judgment trials. Planned t-tests between recent and non-recent STM ($t(31) = -1.96$, $p = .06$, d = -0.35) and perceptual-judgment ($t < 1$) trials indicated a trend towards significance in STM trials, suggesting that accuracy in STM trials may have been slightly influenced by interference. However, this effect was not strong enough to be observed in the omnibus test.

Difficulty

In addition to the omnibus tests described above, we examined whether there was a difference in experimental performance on italics and bold judgments after training. Italics judgments were responded to more slowly ($t(31) = 9.53$, $p < .0005$, d =1.69) and less accurately ($t(31) = -4.35$, $p < .0005$, d = -0.77) than bold judgments, suggesting that training did not successfully improve overall performance. However, only 8 participants out of the total 32 scored >5% less accurately on italics judgments when compared to bold judgments, indicating that this performance impairment was numerically small.

Because of the small number of participants meeting the >5% less accurate criterion (which was the criterion used to split participants into high- and low-accuracy groups in Experiment 2 of Chapter 2), a traditional median split based on overall accuracy on perceptual-judgment trials was used to split participants into high- and low- accuracy groups (see Table 13).

High-accuracy group.

Response time. The standard proactive interference effect was found for STM trials but not for perceptual-judgment trials (interference effect, $F(1,15) = 28.29$, $p < .0005$, $\eta^2_G = .05$); this was confirmed by planned t-tests indicating that recent STM trials ($t(15) = 5.75$, $p < .0005$, d = 1.44) but not recent perceptual-judgment trials ($t(15) = -0.60$, $p = 0.56$, d = -0.15) were influenced by proactive interference. For completeness, we note that recent trials also took longer overall compared to non-recent trials (*F*(1,15) = 11.95, $p < .005$, $\eta^2_G = .04$), and memory trials took longer than perceptual-judgment trials (*F*(1,15) = 37.76, $p < .0005$, $\eta^2_G = .16$).

Accuracy. There was a trend suggesting that interference may be found in STM but not in perceptual-judgment trials (interference effect, *F*(1,15) = 3.05, $p = .10$, $\eta^2_G = .06$); planned t-tests supported this conclusion, indicating that perceptual-judgment trials were not influenced by recent information (*t*(15) = 0.32, $p = .75$, d = 0.08), while STM trials were (*t*(15) = -2.39, $p < .05$, d = -0.60). The effect of recency itself was also a trend, (*F*(1,15) = 3.30, $p = .09$, $\eta^2_G = .04$), and there was no significant accuracy difference between STM and perceptual-judgment trials, (*F*(1,15) = 0.05, $p = .82$, $\eta^2_G = .001$). However, this may be because of the reduced variance found for the high-accuracy group, potentially due to ceiling effects on performance.

Low-accuracy group.

Response time. Proactive interference was found for STM but not perceptual-judgment trials (interaction effect, (*F*(1,15) = 29.15, $p < .0005$, $\eta^2_G = .03$). Further supporting this conclusion, STM trials were affected by proactive interference (*t*(15) = 8.13, $p < .0005$, d = 2.03), but perceptual-judgment trials were not (*t*(15) = -0.96, $p = .35$, d = -0.24). Recent trials were also slower overall than non-recent trials, *F*(1,15) = 34.60, $p < .0005$, $\eta^2_G = .02$, and STM trials were slower than perceptual-judgment trials, *F*(1,15) = 12.28, $p < .01$, $\eta^2_G = .05$.

Accuracy. Category trials were less accurate than memory trials, *F*(1,15) = 27.73, $p < .0005$, $\eta^2_G = .34$. All other comparisons were not significant, *F*s < 1.

Table 13. Average median scores (standard error) for performance on STM and perceptual-judgment tasks using a median split based on overall accuracy on perceptual-judgments.

		High Performers		Low Performers	
		Recent Negative	Non-Recent Negative	Recent Negative	Non-Recent Negative
Memory	RT	684.97(27.57)	601.19(20.09)	740.13(39.22)	644.84(35.79)
	Accuracy	95.31(1.31)	98.44(0.57)	97.27(0.69)	97.66(0.83)
Perceptual-Judgment	RT	554.13(22.06)	562.09(27.46)	618.31(38.61)	630.53(42.76)
	Accuracy	97.27(0.80)	96.88(0.86)	91.80(1.27)	91.80(1.17)

Conclusions

The results presented here replicated the pattern of results shown in previous chapters: Temporal recency led to interference when it was task-relevant (i.e., on a STM task) but not when recency was task-irrelevant (i.e., on a perceptual-judgment task). These data support the hypothesis that similarity must be along a task-relevant dimension in order to create interference within a given task.

The results of Experiment 2 of Chapter 2 contradicted this hypothesis somewhat; specifically it appeared that those participants who had difficulty making the perceptual judgment might begin to be influenced by the nominally-irrelevant temporal dimension of similarity, and to thus become vulnerable to recency-based interference. In the present study, participants were trained on the relevant perceptual judgments, and did not show recency-based interference on the perceptual-judgment task.

Comparing across the two experiments, the results are consistent with the idea that irrelevant (in this case, temporal) dimensions may begin to have an influence on participants' judgments if they have difficulty making judgments on the putatively relevant dimensions. However, especially given the cross-experiment nature of these comparisons, it is not possible to exclude alternative explanations. One such alternative is that the apparent interference in Experiment 2 Chapter 2 was a chance finding. Further investigation should directly manipulate the difficulty of stimuli, preferably in a within-subject design, to determine whether increased difficulty will prompt participants to rely on task-irrelevant dimensions of similarity when making their response. Putting aside the difficulty issue, the present experiment joins with the others to suggest that interference depends on competition along task-relevant dimensions.

Taken together, these results suggest that a dimension of similarity will create proactive interference primarily when it is relevant to the current task.

References

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

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.

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

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

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.

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

Wickelgren, W.A. & Norman, D.A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology, 3,* 316-347.

Chapter 4

Introduction

Proactive interference is a major source of forgetting in short-term memory (STM), and occurs when previously-processed information disrupts current processing. Historically, theories of activation strength and similarity have proposed the conditions necessary for the creation of proactive interference: activation strength theories suggest that information should be recently-processed to interfere (e.g., Altmann & Gray, 2002; Wickelgren & Norman, 1966), while similarity theories argue that such information must be similar to the target(s) of current processing in order to generate interference (e.g., McGeogh & McDonald, 1931; Underwood, 1945, 1957; Wickens, Born, & Allen, 1963). However, recency or similarity alone may not be enough to create proactive interference. In Chapters 2 and 3, we saw that neither recency nor similarity was sufficient to produce interference by itself. In fact, when similarity was along a task-irrelevant dimension, we saw that it did not influence the amount of proactive interference found on a STM task (Chapter 2, Experiments 4 and 5). In addition, recent information that interfered with a STM task did not interfere with category-judgment tasks, presumably because whether or not the item was recently presented was not relevant to judging its category membership (Chapter 2, Experiments 1, 2, and 3; Chapter 3).

While Chapters 2 and 3 suggest that a similar item will create interference only when similarity is along a task-relevant dimension, they do so by demonstrating a series of single dissociations between task-type and stimulus dimension. However, a single dissociation could lead to misleading conclusions. For a stronger test of our hypothesis, a double dissociation should be established between task-relevance and stimulus dimension.

In addition, these chapters rely heavily on the use of the recent probes task. Although the recent probes task is certainly widely-used in studies of interference in short-term memory, it is important to test whether the conclusions drawn from studies using this task will generalize to other STM and interference paradigms.

The n-back task may prove particularly helpful in this respect (Gevins, Bressler, Cutillo, Illes, Miller, Stern, & Jex, 1990; Gevins & Cutillo, 1993; Smith & Jonides, 1997). Often used within functional magnetic resonance imaging (fMRI) studies, the n-back has become a standard task in STM research. Classically, the n-back task requires participants to determine whether the item on the screen was presented n- positions earlier in the experiment. For example, in a 1-back task, participants evaluate whether the item presented is the same as the last item shown (e.g., the second C in the sequence A-B-C-C-D). Interference is created by lure trials, in which a current item does not match the n-back item, but matches an item in a neighboring position, such as the n+1 position (e.g., in a 1-back task, the second C in the sequence C-D-C). The match between the current item and the neighboring item leads to decreased accuracy and increased response times on these lure trials (e.g., Gray, Chabris & Braver, 2003; Szmalec, Verbruggen, Vandierendonck & Kemps, 2011).

Finding a double dissociation in the recent probes task was not possible because of the difficulty in manipulating the relevance of stimulus identity within the task. In contrast, the n-back task is ideal to test for a double dissociation between task-relevance and stimulus dimension. The n-back involves the comparison of a current item to an item presented n- positions earlier within the experiment; further, the basis for this comparison can vary based on instruction.

Interference can be created in two independent stimulus dimensions by matching a stimulus dimension in the current item to the same dimension in the item n+1 back. Further, these two stimulus dimensions can be deemed task-relevant or task-irrelevant based on what comparison participants are instructed to perform.

In this study, we expand our understanding of the role of task-relevance on the creation of proactive interference by observing its influence within a 1-back task. Further, we attempt to create a double dissociation between the dimension of similarity used to create interference (color and number information) and the task-relevance of that dimension of interference (comparing items on either color or number identities). Unlike the experiments in Chapters 2 and 3, these manipulations will show that similarity along two dimensions (color and number identity) can and will interfere when they are relevant, but when these same dimensions are irrelevant, they will not interfere.

Methods

Participants

Sixty-two individuals (40 female, average age = 18.87 years, SD = 1.00) participated in this study. All individuals were recruited through the University of Michigan Subject Pool and received course credit for participation. Exclusion criteria were identical to those used in Chapters 2 and 3. A total of three participants were excluded due to health conditions and/or current medications, three participants were excluded due to <80% accuracy, one participant was excluded due to failure to follow instructions, three participants were excluded due to poor performance on the ERVT, and data from two participants were lost for technical reasons. Fifty healthy individuals (32 female) were included in the final analysis. These participants had ERVT scores ranging from 9.00 to 32.25, M = 16.39, SD = 5.03. They had an average age of 18.88 (SD = 1.06) years, and had completed an average of 12.64 years of formal education (SD = 0.90).

Participants were assigned to receive color instructions or number instructions by a Latin Square design. T-tests did not indicate any significant differences between groups on ERVT score, age, or education (*t*s < 1). Those receiving color instructions (N = 25, 17 female) had ERVT scores ranging from 12.00 to 15.00, M = 12.72, SD = 4.91. They had an average age of 19.00 (SD = 1.29) years, and had completed an average of 12.72 years of formal education (SD = 1.02). Individuals receiving number instructions (N = 25, 15 female) had ERVT scores ranging from 9.00 to 32.25, M = 16.33, SD = 5.26. They had an average age of 18.76 (SD = 0.78) years, and had completed an average of 12.56 years of formal education (SD = 0.77).

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 40-point bold Arial font using E-Prime 2.0 software (Psychology Software Tools, Inc.).

Each trial consisted of a black fixation cross appearing for 250 msec followed by the test item for 500 msec. Each test item was then backmasked by a multicolored display made up of overlapping straight and curved lines for 250 msec. This was followed by another black fixation cross, which appeared for 2000 msec.

Responses to each test item were made by a ‘0’ or ‘1’ keypress to indicate a ‘yes’ or ‘no’ response; these response mappings were counterbalanced across participants. Participants were instructed to perform all keypresses with the left and right index fingers. Although each test item appeared for only 500msec, responses were recorded for any keypress made within 2000 msec of the onset of the test item.

Two stimulus dimensions (color and number) were manipulated in this experiment. Task-relevance of color and number were manipulated between subjects through instruction. The color instruction group was told to compare the current item's color with the color of the item seen 1-back, while the number instruction group was told to compare the items based on number information only.

As in the recent probes task, all critical trials were negative trials. Negative trials consisted of four types: No-Interference, Number-Lure, Color-Lure, and Double-Lure. In addition, there were three types of positive trials: Number-Yes, Color-Yes, and Double-Yes (see Figure 10a for a sample trial sequence).

Interference was created by manipulating the color and/or number of the current item compared to the item presented 2-back (for a similar manipulation, see Szmalec et al., 2011). All interference trials were negative trials; participants should respond 'no' because the relevant dimension (color or number) did not match the item presented 1-back. However, these trials do match the item 2-back on the color and/or number dimension, creating the potential for interference. Number-Lure trials created interference only along the number identity dimension, Color-Lure trials created interference only along the color identity dimension, and Double-Lure trials created interference along both number and color dimensions. Importantly, when a trial was considered a lure in only one stimulus dimension (e.g., Number-Lure and Color-Lure trials), the item did not match the 1-, 2-, or 3-back item on the second stimulus dimension. For example, on a Color-Lure trial, the item did not match the number of the item presented 1-, 2-, or 3-back.

In addition, No-Interference trials were negative trials which did not match the 1-, 2- or 3-back item on either the color or number dimension.

Lure trials made up 24% of the total number of trials, while No-Interference trials made up 38% of all trials. Half of all interference trials manipulated color interference and half manipulated number interference; one-third of all interference trials manipulated both number and color interference.

‘Yes’ responses fell into three categories: Color-Yes, Number-Yes, and Double-Yes trials. Color-Yes trials matched the 1-back item on color dimension only, Number-Yes trials matched the 1-back item on number dimension only, and Double-Yes trials matched the 1-back item on both dimensions. Interference was not manipulated on these trials; because of this, the stimulus dimension not matching the previous item (i.e., the number identity dimension on Color-Yes trials) did not match the 1-, 2-, or 3-back item.

To maintain an equal number of examples of each dimension of interference, 6 numbers (1 through 6) and 6 colors were chosen for use as stimuli in the experiment. Colors were chosen such that they spanned the color wheel, with 80% saturation and 50% lightness on the HSL color point representation scale. This consistency ensured that all colors were equally bright and could easily be seen against a gray background. In addition, colors were chosen which were clearly representative of standard color names (pink, purple, yellow, green, blue, and orange).

A response bias was created in this task such that 25% of trials were positive. Trials of each type were evenly distributed across blocks. Within each block, trials were pseudo-randomly distributed with the constraint that no more than 3 responses of one type occurred in a row. The order of block was counterbalanced across subjects using a Latin square design.

a

Stimulus	Trial Type
3	No-Interference
4	No-Interference
3	Double-Lure
1	Color-Yes
1	Number-Yes
2	No-Interference
5	No-Interference
6	Color-Lure
5	Number-Lure
5	Double-Yes

b

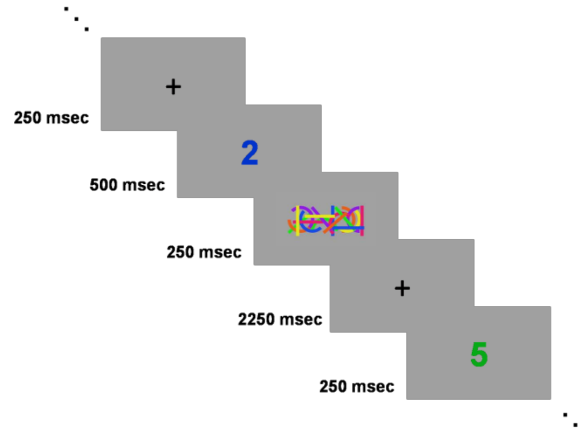

Figure 10. Sample trial sequence and trial types (a) and presentation sequence (b) for trials in the interfering 1-back task. Time progresses from top to bottom (a and b) and also from left to right (b). No-Interference, Color-Lure, Number-Lure, and Double-Lure trials all required a 'no' response because they did not match the number or the color of the item 1-back. Double-Yes trials required a 'yes' response because they matched the item 1-back. Color-Yes and Number-Yes trials required a different response from each group of participants; the group attending to color should respond with a 'yes' to Color-Yes trials and a 'no' to Number-Yes trials; this pattern should be reversed for the group attending to number.

Procedure

After providing written informed consent, all participants completed practice on the task before beginning the experiment. Practice consisted of 20 trials, 25% of which were positive trials. Participants were able to repeat practice if desired. Following practice, each participant completed five blocks of 65 trials, with 30 seconds of rest in between blocks.

Results and Discussion

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 was 1.86% 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 was indexed by the contrast between No-Interference and lure trials, and so these were the focus of our analyses. For completeness, means and standard deviations for all negative trial types are given in Table 14; for statistics on positive trials, see Table 15.

A mixed design was used with one between-subjects factor (instruction type: color or number) and two within-subjects factors, color interference (yes vs. no) and number interference (yes vs. no). Effect sizes are reported in generalized eta squared (η^2_G) values.

Table 14. Average median response times in milliseconds, and accuracy scores in percentages by trial type for all negative 1-back trial types. Standard errors are reported in parentheses.

		No-Interference	Number-Lure	Color-Lure	Double-Lure
Color Instructions	RT	489.02(29.57)	503.96(31.18)	530.20(34.61)	548.56(39.05)
	Accuracy	96.37(0.55)	96.16(0.91)	92.64(1.30)	88.96(1.47)
Number Instructions	RT	536.64(44.02)	567.68(43.48)	544.12(44.68)	595.94(51.19)
	Accuracy	96.00(0.38)	89.12(1.54)	96.00(0.65)	91.20(1.82)

Response Time

As seen in Figure 11, color interference (*F*(1,48) = 16.54, $p < .0005$, $\eta^2_G = .006$) and number interference (*F*(1,48) = 27.02, $p < .0005$, $\eta^2_G = .006$) both affected response times. Number interference only affected the group with number-relevant instructions (*F*(1,48) = 4.92, $p < .05$, $\eta^2_G = .001$); for color interference, this relationship trended towards significance, (*F*(1,48) = 2.80, $p = .10$, $\eta^2_G = .001$). Planned t-tests revealed that there was no difference between groups for any trial type, (*t*s(48) $\leq$ 1.23, *p*s $\geq$ .22), suggesting that the increased power of adding Double-Lure trials in the omnibus test allowed for a significant interference effect in relevant trials to be observed. In addition, planned t-tests indicated that there was no difference between Double-Lure and Color-Lure trials for the color instruction group (*t*(24) = 1.64, $p = .11$) and no difference between Double-Lure and Number-Lure trials for the number instruction group (*t*(24) = 1.55, $p = .13$), ruling out the possibility that an increased amount of interference occurred for Double-Lure trials.

For completeness we note that all other interactions were not significant, (*F*s $\leq$ 1.17, *p*s $\geq$ .28). In addition, there was no overall difference between groups, (*F*(1,48) = 0.64, $p = .43$).

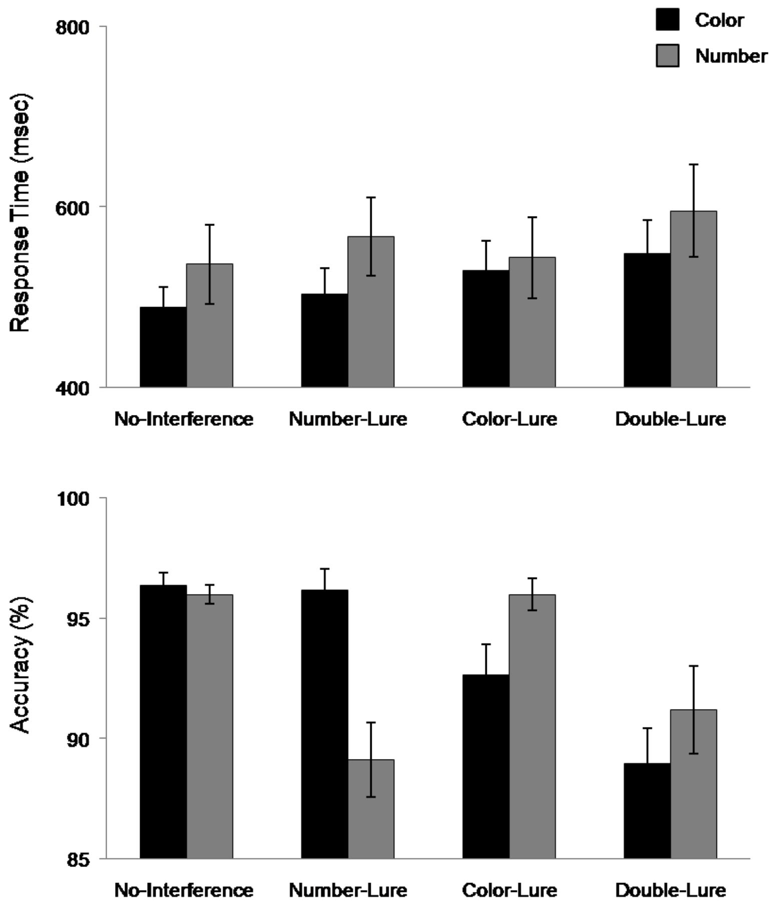

Figure 11. Average median response times (a) and accuracy (b) for negative trials. Error bars represent between-subjects standard error.

Table 15. Statistical analyses and effect sizes for the 1-back task.

2x2x2 ANOVA for negative RTs	*F* value	*p* value	η^2_G
Main effect of color interference	16.54	<.0005	.006
Color x instruction interaction	2.80	.10	.001
Main effect of number interference	27.02	<.0005	.006
Number x instruction interaction	4.92	.03	.001
Color x number interaction	1.17	.28	<.001
Color x number x instruction interaction	0.60	.44	<.001
Main effect of instruction type	0.64	.43	.01
2x3 ANOVA for positive RTs			
Main effect of trial type	0.68	.51	<.001
Main effect of instruction type	1.11	.30	.02
Trial type x instruction interaction	1.04	.36	<.001
2x2x2 ANOVA for negative accuracies			
Main effect of color interference	8.52	<.01	.04
Color x instruction interaction	18.41	<.0005	.07
Main effect of number interference	24.93	<.0005	.10
Number x instruction interaction	6.25	.02	.03
Color x number interaction	0.34	.56	.001
Color x number x instruction interaction	5.41	.02	.01
Main effect of instruction type	0.16	.69	.002
2x3 ANOVA for positive accuracies			
Main effect of trial type	2.71	.07	.02
Main effect of instruction type	0.23	.63	.003
Trial type x instruction interaction	30.36	<.0005	.19

Accuracy

Color interference ($F(1,48) = 8.52$, $p < .01$, $\eta^2_G = .04$) and number interference ($F(1,48) = 24.93$, $p < .0005$, $\eta^2_G = .10$) both affected accuracy. Both types of interference only affected the group with relevant instructions (color by instruction interaction, $F(1,48) = 18.41$, $p < .0005$, $\eta^2_G = .07$; number by instruction interaction, $F(1,48) = 6.25$, $p < .05$, $\eta^2_G = .03$). There was also a significant 3-way interaction, $F(1,48) = 5.41$, $p < .05$, $\eta^2_G = .01$, indicating that Double-Lure trials may have had a stronger effect than the Color-Lure trials for the color instruction group.

Planned t-tests revealed that adding the irrelevant dimension as a second lure decreased accuracy for the color group when compared to relevant lure trials (Color-Lure and Double-Lure trials, $t(24) = -3.07$, $p < .01$, d = -1.16) but did not affect the number group (Number-Lure and Double-Lure trials, $t(24) = 1.03$, $p = .31$, d = 0.21) suggesting that number interference affected responses even when this information was irrelevant, but color interference affected responses only when relevant to the task. As described below, this suggests that number identity information may be difficult to ignore, potentially because of automatic Stroop-like processing.

For completeness we note that the interaction between color and number interference was not significant ($F < 1$). In addition, there was no difference between groups, ($F(1,48) = 0.16$, $p = .69$).

Conclusions

While color and number information each created interference within this 1-back task, their influence was predominantly within a task-relevant context, supporting the hypothesis that a similar item will only create interference when it is along a task-relevant dimension. Color information created interference only for individuals who were instructed to respond to color information. Number interference, however, created interference not only for the individuals instructed to respond to number information, but also to some extent for individuals responding to color information. Specifically, when the color group responded to trials with both color and number interference, they were even less accurate than they were for trials with only color interference.

Why might this occur? One possibility is that this is simply a chance finding: Notably, the color instruction group had equivalent performance on No-Interference and Number-Interference trials (as did the number instruction group on No-Interference and Color-Lure trials). This, along with the failure to find an exaggerated interference effect in the response-time data and the weight of the evidence across our experiments, fits with the idea that similarity along task-relevant dimensions is critical for interference to occur. However, if one assumes that the effect is real, a possible explanation is that it may occur due to the relative integrality of the two stimulus dimensions manipulated. In particular, number information may be more integral to an item's identity than color information. This integrality may make it more difficult to treat the stimulus dimension as task-irrelevant, even if instructions declare it to be so. This integrality makes number information more difficult to ignore when it is nominally-irrelevant to the task at hand.

The potential of integrality to interfere with task-relevance is an intriguing one. Follow-up studies should be conducted to more thoroughly investigate how integrality can affect proactive interference in STM; for instance, if some stimulus dimensions are more integral to an item’s identity than others, these dimensions may have more pervasive effects on proactive interference than stimulus dimensions which are more easily separated from an item’s identity.

References

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

Gevins, A.S., Bressler, S.L., Cutillo, B.A., Illes, J., Miller, J.C., Stern, J., & Jex, H.R. (1990). Effects of prolonged mental work on functional brain topography. *Electroencephalography and clinical Neuropsychology, 76,* 339-350.

Gevins, A.S., & Cutillo, B.A. (1993). Spatiotemporal dynamics of component processes in human working memory. *Electroencephalography and clinical Neuropsychology, 87,* 128-143.

Gray, J.R., Chabris, C.F., & Braver, T.S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6*, 316-322.

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

Smith, E.E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33,* 5-42.

Smith, E.E., Jonides, J., & Koeppe, R.A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6,* 11-20.

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 137-151.

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.

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.

Wickelgren, W.A. & Norman, D.A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology, 3,* 316-347.

Chapter 5

Discussion

Historically, proactive interference has been viewed in fairly simple terms: it has been described either as the result of residual activation strength (e.g., Anderson & Spellman, 1995) or as the result of similarity between two items (e.g., Crowder, 1976; Johnson, 1933; Keppel, 1968; McGeogh & McDonald, 1931; Underwood, 1957; Wickens, 1970). Current research suggests that this simplicity may be exaggerated, and that multiple factors together determine whether proactive interference will or will not be created.

Traditional theories of proactive interference, for instance, would argue that interference is found on the recent probes task because when a recently-seen item (in the context of a memory set) returns as a probe item, responses are slowed because of the activation strength from its initial presentation or because of the high degree of similarity between the two presentations of the item (e.g., the item's identity match).

Current Findings

While these explanations are appealingly simple, they do not predict the findings presented in Chapters 2, 3, and 4. Proactive interference created by temporal recency in the recent probes task does not create interference in semantic-judgment or perceptual-judgment categorization tasks (Chapter 2, Experiments 1, 2, and 3), and it is not affected by visually distinct perceptual similarities such as color and font (Chapter 2, Experiments 4 and 5). These data lead us to the conclusion that an additional factor must be considered in the creation of proactive interference: task-relevance. We hypothesized that when similarity is along task-relevant dimensions, it creates interference, but when it is not relevant, it does not interfere.

Further, tasks that are difficult may rely on task-irrelevant dimensions of similarity to create interference. In Chapter 2, individuals with difficulty identifying italicized words used the task-irrelevant temporal dimension of similarity to make perceptual-judgments (Experiment 2, pg. 32). However, when participants are first trained on perceptual-judgment tasks, irrelevant information is no longer used to make these judgments (Chapter 3, pg. 74). This suggests that difficulty with the task is an additional factor in the creation of interference.

Additional support for this hypothesis can be found in Huang, Kahana and Sekuler (2009), who find that task-irrelevant stimulus attributes affect perception and interference in STM. Importantly, this study used difficult-to-separate stimulus attributes as the task-relevant and task-irrelevant stimuli (vertical and horizontal components of a 2-D Gabor stimulus), suggesting that when a perceptual discrimination is difficult, task-irrelevant components of the stimulus may create interference.

Chapter 4 extends the themes of this dissertation in two important ways. First, it provides some evidence that the basic principles tested here generalize across tasks (Sternberg recent probes and n-back). Second, it allowed us to demonstrate a double dissociation within the same experiment. When an item was similar to the target in color but not digit-identity, it only caused interference if participants were making their judgments based on color matches. Conversely, items similar to the target in digit-identity but not color only caused interference for those participants making judgments on the basis of digit-identity.

One qualification to this dissociation is that when participants were making judgments based on color, items similar to targets on *both* color and digit-identity dimensions caused more interference than those only similar on the color dimension. That is, although similarity along the (nominally irrelevant) digit-identity dimension did not cause interference in isolation, it was associated with increased interference if presented in conjunction with (relevant) similarity on the color dimension. It is not entirely clear why this occurs, or why the inverse pattern (color similarity amplifying interference based on digit-identity similarity) was not observed. One possible explanation is that the integrality of different stimulus dimensions to an item's identity can play a role in the effects of proactive interference along that dimension. If a stimulus dimension is more integral to an item's identity, it may be more difficult to treat that information as task-irrelevant, allowing it to create interference, especially in situations where other task-relevant stimulus dimensions are also creating interference. In this case, if an item has already begun to attract attention and is slow to be rejected because of similarity on the color dimension, the opportunity may arise for digit-identity information to intrude.

Importantly, in each of these experiments the effects of temporal similarity followed the same pattern as other dimensions of similarity, demonstrating that it could create proactive interference when it was relevant, but not impacting performance when it was irrelevant. It is also important to consider that interference can also be created through semantic similarity, without the need for recent presentation (e.g. Atkins & Reuter-Lorenz, 2008). In light of these findings, it may be valuable to revisit the currently-accepted definitions of interference (i.e. proactive vs. semantic); for example, if temporal similarity has no special place in creating interference, perhaps it is unnecessary to create a separate label for interference along this dimension.

In summary, our findings are consistent with a more complex view of the mechanisms behind proactive interference: that several factors will affect whether an item with interfere with current information processing, including whether the item is similar along task-relevant dimensions. Additional factors may include whether the relevant task is difficult to perform, and whether the item is similar along a dimension that can be processed automatically.

Mechanisms of Proactive Interference

Based on these findings, it is clear that activation strength and similarity theories are not comprehensive explanations of why and how proactive interference occurs. However, each theory can contribute something to our overall understanding of how proactive interference works. The data presented here support a modified version of similarity-based competition theories: Interference is created when items are similar along a task-relevant dimension.

It is important to consider not only what factors are involved in the creation of proactive interference, but also the mechanisms by which proactive interference is created. Indeed, these two concepts are deeply connected to one another, and considering one may help to understand the other. For example, a better working understanding of the mechanisms behind proactive interference can help us to determine whether a particular factor (e.g., task-relevance) might realistically be involved in the process of creating interference, and if it is involved, how it might do so.

To consider how the mechanisms behind proactive interference can explain the results found here, we need to return to unitary models of memory (Cowan, 2001; Oberauer, 2002). These models posit that current information processing occurs within the focus of attention, while recently-processed items exist in the region of direct access.

How might activation strength fit into this model? The term ‘activation strength’ may be used here not as a ubiquitous descriptor for an item, but instead as a reference to an item’s representational integrity within the focus of attention or region of direct access.

Representational integrity may also be explained by using the concept of a noisy code, consistent with signal detection theory (Green & Swets, 1966).

While an item remains in the focus of attention, it retains a high degree of integrity, or in signal detection theory terms has a high signal-to-noise ratio. An item moving from the focus to the region of direct access would naturally retain a high degree of representational integrity (activation strength) because of its recent processing. It is this high activation strength that will allow it to potentially interfere with an item in the focus, provided that these items share task-relevant dimensions of similarity.

As other items are processed by the focus of attention, noise increases and the distinctiveness of the item in the region of direct access is reduced. This process would result in a reduced activation strength (lower signal-to-noise ratio), and the item would no longer be able to create interference. It is important to note that this reduction in signal-to-noise ratio would likely coincide with an exit from the region of direct access. Items within the region of direct access would have a higher signal-to-noise ratio, allowing them to compete with highly similar item(s) within the focus, causing forgetting and confusion during current processing.

As noted in the introduction (pg. 4), the features coded into STM for each item (Nairne, 2002; Oberauer & Lange, 2008) would be compared when they are task-relevant, leading to the pattern of results found in the experiments presented earlier.

The Neural Basis of Proactive Interference

While an understanding of the mechanisms behind proactive interference creation is important, it is equally important to consider how this model might be instantiated neurally. When an item enters the focus of attention, a pattern of neurons fire together to form this representation (Hebb, 1949). This pattern of representation is distributed across many cortical areas (McClelland, McNaughton, & O'Reilly, 1995; Munk, Linden, Muckli, Lanfermann, Zanella, Singer & Goebel, 2002) and can be considered a representation of the many features that make up the item. For example, information pertaining to an item's color may be coded in the visual cortex, while information about its semantic category may be coded in the temporal lobes. In addition, information about the time-of-presentation of the item is coded by the pattern's oscillatory frequency, or their pattern of firing over time (Brown, Neath, & Chater, 2007; Brown, Preece, & Hulme, 2000; Jacobs, Kahana, Ekstrom, & Fried, 2007; Lustig, Matell, & Meck, 2005).

We have hypothesized that an item will interfere when it is similar to a currently-processed item along task-relevant dimension(s). Information about stimulus dimensions is coded based on which cortical neurons are included in the representation of the item. Importantly, we have argued that time is just another dimension of similarity, and temporal similarity, when relevant to the task, will have the same impact on proactive interference as similarity along any other task-relevant dimension. Information about this temporal information is coded by the oscillatory frequency of the item's neural representation, as noted above.

How might this information be combined to explain the mechanisms behind proactive interference? It is possible that when two items are compared along task-relevant dimensions, their discriminability along these dimensions is evaluated. From the neural perspective described above, two highly similar items may use more overlapping cortical representations compared with two highly dissimilar items. For temporal similarity, two items presented close together in time may have more similar oscillatory frequencies (See discussion by Lustig et al., 2005). This overlap or similarity in frequency would reduce the distinctiveness of the items, causing them to become confusable and allowing the recent item to interfere. Items similar along a task-irrelevant stimulus dimension would not interfere: although they are similar, this dimension of the item should not be considered when making the comparison, and so this stimulus dimension would be unable to affect the overall discriminability between the two items. Thus, proactive interference should occur when it is difficult to discriminate between two competing items, as evaluated by task-relevant (and possibly highly-integrated, difficult-to-ignore) dimensions of similarity.

Future Directions

As noted above, task-relevance is not the only additional factor to consider when creating proactive interference; both the difficulty of the task being performed (Chapter 3) and the integrality of the dimension on which items are similar (Chapter 4) may also be factors. Future experiments should investigate the role of integrality on proactive interference, including further studies of what dimensions are more integral to an item's identity. In addition, experiments should look into differences in proactive interference and information processing on easy vs. difficult trials to determine whether the task-relevance factor influences these trials similarly. The roles of similarity and task-relevance in proactive interference should also be further investigated to determine in what contexts these factors will influence proactive interference.

Why do we forget? The story may not be simple, but it is becoming more clear: to avoid being confused by recent information, we should try to make it irrelevant.

References

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

Brown, G.D.A., Neath, I, & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*, 539-576.

Brown, G.D.A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127-181.

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

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

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. Oxford, England: John Wiley.

Hebb, D.O. (1949). *The organization of behavior*. New York: Wiley.

Huang, J., Kahana, M.J., & Sekuler, R. (2009). A task-irrelevant stimulus attribute affects perception and short-term memory. *Memory and Cognition, 37*, 1088-1102.

Jacobs, J., Kahana, M.J., Ekstrom, A.D., & Fried, I. (2007). Brain oscillations control timing of single-neuron activity in humans. *The Journal of Neuroscience, 27*, 3839-3844.

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

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.

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 Berntson, G.G. & Cacioppo, J.T. (Eds.), *Handbook of neuroscience for the behavioral sciences, Vol. 1.* Hoboken, N.J.: John Wiley & Sons, Inc.

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.

McClelland, J.L., McNaughton, B.L., & O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419-457.

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

Munk, M.H.J., Linden, D.E.J., Muckli, L., Lanfermann, H., Zanella, F.E., Singer, W., & Goebel, R. (2002). Distributed cortical systems in visual short-term memory revealed by event-related functional magnetic resonance imaging. *Cerebral Cortex, 12*, 866-876.

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.

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

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

Appendix A

Chapter 2 Difficulty Analyses

Experiment 2

The accuracy results in Experiment 2 of Chapter 2 (pg. 32) 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 based on 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 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 $F(1,17) = 8.86$, $p< .01$, $\eta^2_G= .02$. T-tests on STM trials revealed a significant difference between recent and non-recent 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 ($F(1,17) = 12.45$, $p < .01$, $\eta^2_G = .02$) and category ($F(1,17) = 33.22$, $p < .001$, $\eta^2_G = .07$) 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 both recency ($F(1,18) = 8.27$, $p = .01$, $\eta^2_G = .02$) and category (10.98, $p < .01$, $\eta^2_G = .06$) main effects 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 subjects. These results complement those of Experiment 2 – that is, in Experiment 2, low-accuracy subjects showed an effect of the nominally-irrelevant temporal dimension on the perceptual category judgment; in Experiment 4, low-accuracy subjects show an effect of the nominally-irrelevant perceptual category on the temporal (recency) judgment. As there was no category judgment in this experiment, median splits on accuracy were based on overall accuracy in negative trials (both recent and non-recent).

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 ($F(1, 15) = 1.90$, $p = .19$, $\eta^2_G = .06$) or accuracy ($F < 1$). Both response time and accuracy measures showed main effects of recency (both $p < .05$, $\eta^2_G > .10$) and no effects of color mis/match (both $F < 1$).

Low-accuracy group

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

Discussion

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 16; 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 judgment, and separating those individuals who had greater disparity between accuracy on those judgments from individuals who succeeded similarly in both, based on 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 subjects 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 on 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 16. Response times (msec) and accuracies (% correct) for negative trials of high- and low-accuracy groups in Experiments 1, 3, 4, and 5. Standard errors are presented in parentheses. Note: for Experiment 5, Non-Recent values are equivalent. This is due to the experimental design, where non-recent values are not classified according to format-repeat or format-change distinctions; values are repeated in this table for ease of comparison to recent values.

			High Performers		Low Performers	
			Recent Negative	Non-Recent Negative	Recent Negative	Non-Recent 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)

Appendix B

Inverse Efficiency Score Analyses

As noted in Chapters 2 and 3, median response times to stimuli were used as the primary measure of interest in the recent probes task; however, accuracy was also often vulnerable to interference effects. In addition, it is possible that the relationship of interference to response time and accuracy performance may be different in the 1-back task. Because of these considerations, we calculated the Inverse Efficiency Score (IES; Bruyer & Brysbaert, 2011) and used this measure of interference to more effectively compare interference across experiments. Negative trials continue to be the trials relevant to proactive interference; for completeness, we report average IES scores in Table 17 (Chapter 2) Table 19 (Chapter 3) and Table 21 (Chapter 4) and statistics for positive trials in Table 18 (Chapter 2) Table 20 (Chapter 3) and Table 22 (Chapter 4).

Chapter 2

Experiment 1

As seen in Figure 12, proactive interference was found for STM trials (cf., Berman et al., 2009; Monsell, 1978) but not for semantic-judgment trials; interaction of recency and trial type, $F(1,28) = 19.30$, $p < .0005$, $\eta^2_G = .04$. Confirming this impression, post-hoc t-tests revealed a significant difference between recent negative and non-recent negative STM trials ($t(28) = 5.18$, $p < .0005$, d = 0.96); for semantic-judgment trials, there was a trend in the opposite direction, suggesting that recent information if anything improved efficiency in performance, $t(28) = -1.79$, $p = .08$, d = -0.33.

Although not relevant to our theoretical question, for completeness we note that semantic-judgment trials were less efficient than STM trials, $F(1,28) = 75.00$, $p < .0005$, $\eta^2_G = .38$. There was also a trend towards recent trials being less efficient than non-recent trials ($F(1,28) = 3.39$, $p = .08$, $\eta^2_G = .006$); this is likely due to the crossover effect described above.

Experiment 2

Proactive interference reduced efficiency scores for STM but not for perceptual-judgment trials, interaction $F(1,36) = 9.78$, $p < .01$, $\eta^2_G = .01$. Planned t-tests indicated that recent responses were less efficient than non-recent responses for STM ($t(36) = 5.44$, $p < .0005$, d = 0.90) but not for perceptual-judgment ($t(36) = 0.86$, $p = .39$, d = 0.14) trials. In addition, recent trials were overall less efficient than non-recent trials ($F(1,36) = 21.42$, $p < .0005$, $\eta^2_G = .03$), and STM trials were less efficient than perceptual-judgment trials ($F(1,36) = 17.89$, $p < .0005$, $\eta^2_G = .03$).

Table 17. Average IES (Inverse Efficiency Score) with standard errors by trial type for Chapter 2, Experiments 1-5.

		Recent Negative	Non-Recent Negative	Recent Positive	Non-Recent Positive
Experiment 1	Memory	767.16(16.13)	689.77(18.11)	710.48(21.49)	706.20(21.93)
	Semantic-Judgment	931.36(33.58)	965.94(33.06)	969.36(36.34)	935.11(30.56)
Experiment 2	Memory	737.19(23.66)	653.61(17.27)	718.93(28.58)	719.00(22.49)
	Perceptual-Judgment	647.35(28.04)	634.10(25.97)	669.39(31.71)	671.20(27.65)
Experiment 3	CC	917.82(46.26)	888.36(42.29)	977.82(45.35)	948.63(41.88)
	CM	734.82(30.92)	667.84(24.92)	631.44(27.18)	675.81(31.20)
	MC	926.51(45.62)	903.97(43.85)	1034.73(58.81)	995.07(45.09)
	MM	757.80(29.93)	624.18(22.49)	625.14(27.29)	619.35(23.36)

Experiment 4	Color-Match	709.85(25.45)	592.43(15.48)	--	--
	Color-Mismatch	686.59(23.72)	602.25(17.40)	--	--
Experiment 5	Format-Repeat	727.83(24.79)	632.32(20.79)	676.80(24.24)	677.31(23.24)
	Format-Change	718.90(27.15)		673.44(27.02)	

Experiment 3

The identity of the previous trial type influenced the size of the difference in interference effects between STM and semantic-judgment trials (previous trial type x current trial type x recency interaction, $F(1,29) = 4.17$, $p = .05$, $\eta^2_G = .002$. In addition, there was an interaction between recency and current trial type, $F(1,29) = 23.92$, $p < .0005$, $\eta^2_G = .009$, indicating STM trials were affected by interference, while semantic-judgment trials may not have been affected. Planned t-tests confirmed this by showing that STM trials (MM trials, $t(29) = 7.93$, $p < .0005$, d = 1.45; CM trials, $t(29) = 3.19$, $p < .01$, d = 0.58) but not semantic-judgment trials (CC trials, $t(29) = 1.22$, $p = .23$, d = 0.22; MC trials, $t(29) = 0.99$, $p = .33$, d = 0.18) were affected by interference. For completeness we note that recent trials were less efficient than non-recent trials ($F(1,29) = 34.16$, $p < .0005$, $\eta^2_G = .02$) and semantic-judgment trials were less efficient than

STM trials ($F(1,29) = 42.08$, $p < .0005$, $\eta^2_G = .22$); all other analyses were nonsignificant, Fs < 1.7, ps > .20.

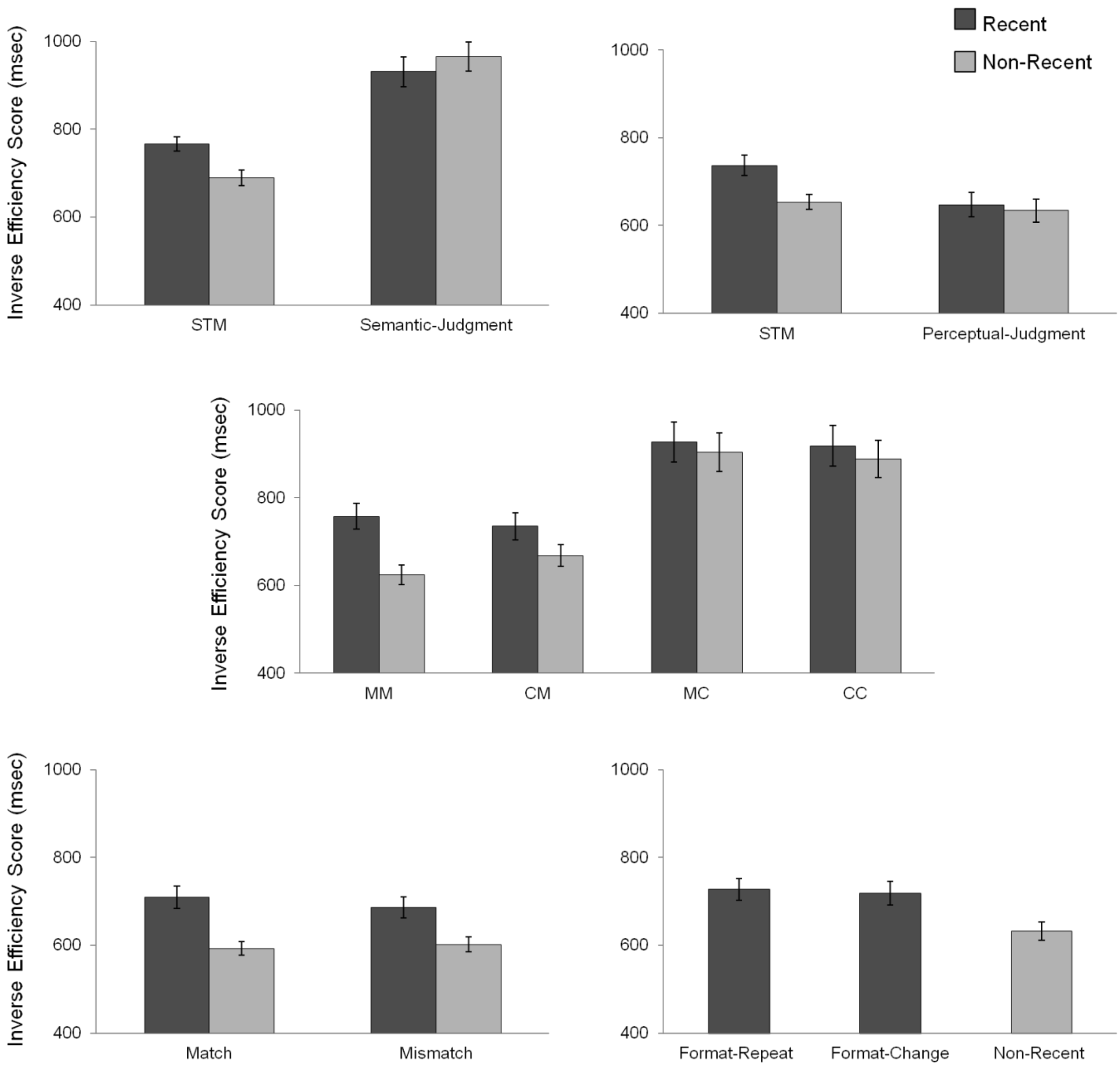

Figure 12. Inverse efficiency scores for negative trials in Chapter 2, Experiments 1-5.

Experiment 4

Match trials were more affected by recent information than mismatch trials ($F(1,31) = 5.90$, $p < .05$, $\eta^2_G = .005$). Planned t-tests indicated that recency did affect both match ($t(31) = 8.17$, $p < .0005$, d = 1.44) and mismatch ($t(31) = 7.29$, $p < .0005$, d = 1.29) trials, and that recency affected match trials more than mismatch trials ($t(31) = 2.15$, $p < .05$, d = 0.40). As noted in Chapter 2 (pg. 33), this interaction may be driven by individuals with low-accuracy scores, supporting the hypothesis that more difficult judgments may be affected by task-irrelevant dimensions of interference.

For completeness, we note that there was a significant effect of recency, $F(1,31) = 82.19$, $p < .0005$, $\eta^2_G = .16$, but no overall effect of trial type, $F(1,31) = 1.09$, $p < .30$, $\eta^2_G = .001$.

Table 18. Statistical analyses and effect sizes for IES in Chapter 2, Experiments, 1-5.

	Negative IES			**Positive IES**		
	F value	*p* value	η^2_G	*F* value	*p* value	η^2_G
2x2 ANOVA for Experiment 1						
Main effect of trial type	75.00	<.0005	.38	91.69	<.0005	.40
Main effect of recency	3.39	.08	.006	2.82	.10	.004
Trial type x recency interaction	19.30	<.0005	.04	1.73	.20	.003
2x2 ANOVA for Experiment 2						
Main effect of trial type	17.89	<.0005	.03	5.65	.02	.02
Main effect of recency	21.42	<.0005	.03	0.01	.94	<.001
Trial type x recency interaction	9.78	<.01	.01	0.01	.92	<.001
2x2x2 ANOVA for Experiment 3						
Main effect of recency	34.16	<.0005	.02	0.72	.40	<.001
Main effect of previous trial type	0.02	.90	<.001	0.81	.38	.001
Main effect of current trial type	42.08	<.0005	.22	136.73	<.0005	.41
Recency x previous trial type interaction	1.10	.30	.001	1.21	0.28	.001
Recency x current trial type interaction	23.92	<.0005	.009	5.07	.03	.004
Previous x current trial type interaction	1.65	.21	.001	9.61	<.01	.01
3-way interaction	4.17	.05	.002	0.72	.41	.001
2x2 ANOVA for Experiment 4						
Main effect of trial type	1.09	.30	.001	--	--	--
Main effect of recency	82.19	<.0005	.16	--	--	--
Trial type x recency interaction	5.90	.02	.005	--	--	--
3x1 ANOVA for Experiment 5						
Main effect of trial type	22.12	<.0005	.16	0.04	.96	<.001

Experiment 5

There was a main effect of trial type, $F(2,34) = 22.12$, $p < .0005$, $\eta^2_G = .16$, suggesting that proactive interference had an effect on performance. Planned t-tests indicated that non-recent trials were more efficient than both format-repeat ($t(17) = 6.23$, $p < .0005$, d = 1.47) and format-change ($t(17) = 5.39$, $p < .0005$, d = 1.27) trials, but format-repeat and format-change trials did not differ ($t(17) = 0.55$, $p = .59$, d = 0.13), showing that perceptual qualities did not affect the impact of proactive interference in this recent probes task.

Chapter 3

Table 19. Average IES by trial type for Chapter 3. Standard errors are reported in parentheses.

	Recent Negative	Non-Recent Negative	Recent Positive	Non-Recent Positive
Memory	743.33(27.82)	637.11(22.67)	681.57(28.67)	731.07(31.77)
Perceptual-Judgment	622.74(26.08)	636.02(30.01)	596.70(28.99)	580.11(23.29)

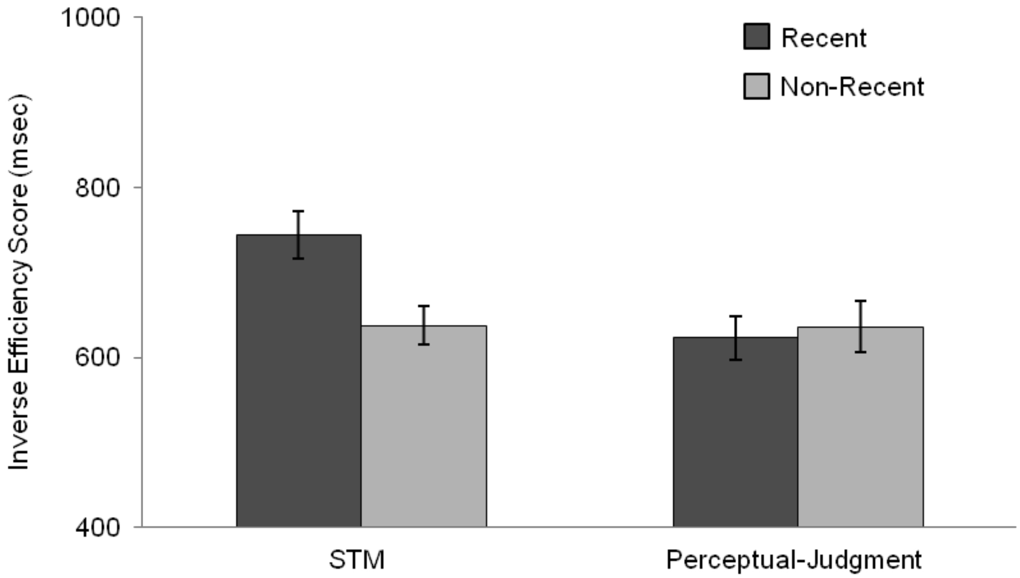

Figure 13. Inverse efficiency scores for negative trials in Chapter 3.

Inverse Efficiency Score

Proactive interference reduced efficiency scores for STM but not for perceptual-judgment trials, interaction $F(1,36) = 30.21$, $p < .0005$, $\eta^2_G = .04$ (see Figure 13). Planned t-tests indicated that recent responses were less efficient than non-recent responses for STM ($t(36) = 7.63$, $p < .0005$, d = 1.35) but not for perceptual-judgment ($t(36) = -0.92$, $p = .36$, d = -0.16) trials. In addition, recent trials were overall less efficient than non-recent trials ($F(1,36) = 26.33$, $p < .0005$, $\eta^2_G = .02$), and STM trials were less efficient than perceptual-judgment trials ($F(1,36) = 16.04$, $p < .0005$, $\eta^2_G = .04$).

Table 20. Statistical analyses and effect sizes for IES in Chapter 3.

	Negative IES			**Positive IES**		
	F value	*p* value	η^2_G	*F* value	*p* value	η^2_G
2x2 ANOVA						
Main effect of trial type	16.04	<.0005	.04	33.24	<.0005	.12
Main effect of recency	26.33	<.0005	.02	2.30	.14	.003
Trial type x recency interaction	30.21	<.0005	.04	10.83	<.01	.01

Chapter 4

Table 21. Average IES by trial type for Chapter 4; color instruction and number instruction groups are reported separately. Standard errors are reported in parentheses.

	No-Interference	Number-Lure	Color-Lure	Double-Lure	Color-Yes	Number-Yes	Double-Yes
Color Instructions	507.11(23.58)	523.30(27.92)	572.12(33.52)	619.06(43.28)	639.00(46.89)	526.20(27.04)	591.33(28.06)
Number Instructions	559.58(46.23)	641.19(50.05)	568.10(47.67)	661.03(59.19)	568.13(44.58)	665.63(43.83)	633.24(45.48)

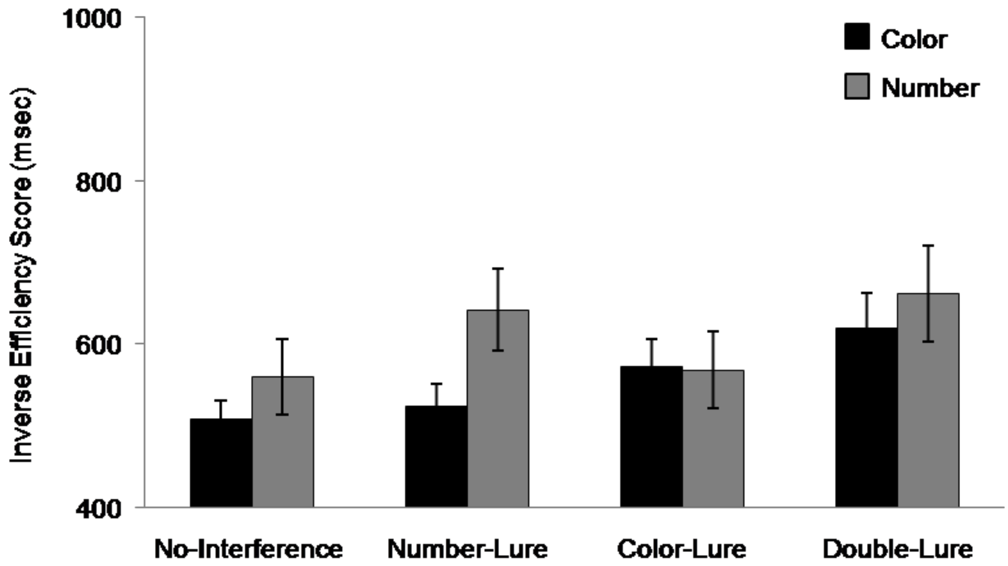

Figure 14. Inverse efficiency scores for negative trials in Chapter 4.

Inverse Efficiency Score

Color interference ($F(1,48) = 18.64$, $p < .0005$, $\eta^2_G = .01$) and number interference ($F(1,48) = 30.63$, $p < .0005$, $\eta^2_G = .02$) both affected efficiency scores (see Figure 14). Number interference only affected the group with number-relevant instructions ($F(1,48) = 6.73$, $p < .05$, $\eta^2_G = .004$), and color interference only affected the group with color-relevant instructions ($F(1,48) = 9.14$, $p < .01$, $\eta^2_G = .006$).

For completeness we note that all other interactions were not significant, (Fs ≤ 1.53, ps $\geq .22$). In addition, there was no overall difference between groups, ($F(1,48) = 0.80$, $p = .38$).

Table 22. Statistical analyses and effect sizes for IES in Chapter 4.

	F value	*p* value	η^2_G
2x2x2 ANOVA for negative IES			
Main effect of color interference	18.64	<.0005	.01
Color x instruction interaction	9.14	<.01	.006
Main effect of number interference	30.63	<.0005	.02
Number x instruction interaction	6.73	.01	.004
Color x number interaction	1.53	.22	.001
Color x number x instruction interaction	0.33	.57	<.001
Main effect of instruction type	0.80	.38	.02
2x3 ANOVA for positive IES			
Main effect of trial type	0.35	.63	.001
Main effect of instruction type	14.52	<.0005	.009
Trial type x instruction interaction	0.50	.48	.05

Summary

Results from analyses of the inverse efficiency score (IES) overwhelmingly support the conclusions made in Chapters 2, 3, and 4: that items will interfere with one another to create proactive interference when they are highly similar to one another along a task-relevant dimension.

References

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (ies) a better dependent variable than the mean reaction time (rt) and the percentage of errors (pe)? *Psychologica Belgica*, *51,* 5-13.