Title

The Effects of Repeated Sequential Context on Recognition Memory

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN

1069-7977

Authors

Kachergis, George
Cox, Gregory
Shiffrin, Richard

Publication Date

2013

Peer reviewed

The Effects of Repeated Sequential Context on Recognition Memory

George Kachergis, Gregory E. Cox, and Richard M. Shiffrin
{gkacherg, grcox, shiffrin}@indiana.edu
Department of Psychological & Brain Science / Cognitive Science Program
Bloomington, IN 47405 USA

Abstract

Many people have had the experience of knowing what song will play next on an album (even one heard only a few times). Conversely, many people fail to recognize an acquaintance encountered in an unfamiliar context. Associations can likely form simply because items appear nearby in time, and not only due to semantic similarity. Using surprise recognition testing, we examine the automatic storage of associations between successively encountered words on a list of incidentally studied words. Many modern memory models assume storage of such associations, but with little evidence as yet (e.g., Cox & Shiffrin, 2012; REM-II Mueller & Shiffrin, 2006). We find evidence for sequential associations, which are further improved by shared semantics or study context. We also find improved accuracy and response time for old words preceded by old words, and for new words preceded by new words—regardless of the previous response.

Keywords: recognition; episodic memory; temporal context; sequential association; priming

Introduction

We have all had the experience of knowing what song will play next on an album that we have listened to several times, even without having looked at the list of songs. Conversely, we have also had the experience of seeing an acquaintance in a new context and not immediately recognizing them. Without realizing it, we often form associations between co-occurring events in a context, and memory is strengthened if the context reoccurs. In general, associates stored together help us remember if they are present at test. When any event is experienced there are a host of potential associations that make up the context—when and where and with what other things did the event happen to me? The memory and its context, and the retrieval of both, are termed episodic memory. The current study investigates the formation and retrieval of one type of context: the other words in a presented sequence of words. A critical factor in this research is the existence of source confusion. For example, given a recognition test of a word, test word familiarity is partly governed by the familiarity of the previous test word. At both study and test we confuse features of nearby events. For example, Jacoby and Whitehouse (1989) found in a recognition experiment when unstudied words were preceded by a subliminal prime (50 or 35ms) of the same word, people were more likely to incorrectly endorse the word as a studied one (i.e., false alarm). When the prime was a different unstudied word than the target, false alarms decreased for the 50ms primes, but oddly not for 35ms primes. In contrast, for liminal primes (200 or 600ms), a studied prime decreases hits for a matching target, and an unstudied prime reduces false alarms to a matching target.

The ROUSE—Responding Optimally with Unknown Sources of Evidence—model of short-term priming (Huber, Shiffrin, Lyle, and Ruys, 2001) incorporates feature leakage from the prime to the target, leading to biased responses. However, ROUSE's decision rule has a discounting mechanism that attempts to correct for leakage: under-discounting explains why primed words are chosen after passive priming, and over-discounting accounts for foil preference after active priming. Although ROUSE was applied mainly to identity and orthographically similar primes, semantic priming and leakage of semantic features also occur, and all these features should (with some probability) be incorporated in the storage of the next few events, and in the test probe of the next few tests. This might suggest episodic-recognition context effects would match those in perceptual recognition, but Malmberg and Annis (2012) investigated sequential dependencies in recognition and found patterns that did not seem to match those found in perceptual experiments. We will investigate this issue in some detail in this research.

We examine the storage of associations between adjacent words in a studied list, and how memory for a studied word is affected at test by the presence of its study-list neighbor. When people expect a memory test, they will form explicit associations between nearby items using a variety of coding schemes. Since our main interest is in automatic and non-strategic storage and retrieval, we limit explicit associative strategies by using an incidental study task: participants make alternating pleasantness/animacy judgments at study.

Evidence for temporal associations have been found in recall following explicit attempts to remember. Participants are serially shown individual, unrelated words (e.g. 'crow', 'bottle', 'house', …) and then asked to recall words from the list in any order. Given that a participant recalls a word (e.g. 'bottle'), the next word they recall is very likely to be the next word that was presented (e.g. 'house'; Kahana, 1996). In recognition tests, participants are shown words one at a time, some from the studied list, and some new, and asked to indicate those studied. A positive recognition response is thought to occur when the test word seems sufficiently familiar, via a fast and automatic parallel search of memory, or when its study event is recalled explicitly, typically via a slow and strategic process (Malmberg, Holden & Shiffrin 2004). Models such as REM and TCM explicitly have a role for word context. We seek to understand such effects when study is incidental.

Experiment 1

This study explores the automatic formation and retrieval of associations in recognition memory between temporally proximal events. Specifically, we varied the relation of two successive words at study for incidental judgments, and explored the effect when words related to these were tested successively, each for separate judgments of presence during study. For example, if "banana" is followed by "chair" at study, is "chair" recognized better or differently at test when preceded by a test of "banana"? The words in this example are semantically unrelated, but some of the adjacent words were made to be semantic associates.

The conditions we used included identical repeats, i.e. the same successive words at study and test, the case probably most likely to produce recognition benefits. In another condition the context word itself does not repeat, but its meaning does: The forward migration of matching semantic features at both study and test could produce improved recognition. In addition, meaning could be altered by the meaning of a recent word. For example, *bank* might be encoded as an earthen side if preceded by *river*, but encoded as a monetary institution if preceded by *money*. Table 1 shows examples of each condition, as well as the possible features that the preceding word (cue) may contribute to the target word at test: Familiarity (F) if the cue was a studied word; Semantics (S) if the cue is semantically related to the target; and Context (C), if the cue was also the target's study neighbor.

Cue Type	Study	Test	F	S	C
Same, Related	cash bank	cash bank	1	1	1
Same Sense	cash bank	robber bank	1	1	0
Different Sense	cash bank	river bank	1	1*	0
Different, Unrelated	cash bank	sloth bank	1	0	0
Same, Unrelated	sloth bank	sloth bank	1	0	1
Unrelated	sloth bank	glass bank	1	0	0
Novel	… bank	*lamp* bank	0	0	0

Table 1: Features (**F**amiliarity, **S**emantics, **C**ontext) that the cue may contribute to the target at test in each condition. *=related in the lexicon.

Subjects

Participants were 57 undergraduates at Indiana University who received course credit for participating.

Stimuli & Procedure

We selected 40 common polysemous words (e.g., diamond) and their two strongest forward associates for each meaning (e.g., ruby/emerald and spade/ace) from the free association norms (Nelson, McEvoy, and Schreiber, 1998). For each participant, the 40 polysemes are assigned randomly to one of five conditions. In the Same, Related (SR) condition, the strongest associate of the dominant meaning is presented just prior to the polyseme at both study and test (e.g. *ruby–diamond*). In the Same, Unrelated (SU) condition, an unrelated word is presented just prior to the polyseme at both study and test (e.g. *lawn–diamond*). The remaining conditions all have the strongest associate of the dominant meaning immediately prior to the polyseme at study, and a word that was studied elsewhere presented prior to the polyseme at test. In the Same Sense (SS) condition, a different associate from the same meaning was presented prior to the polyseme at test (e.g. *emerald–diamond*). In the Different Sense (DS) condition, the strongest associate from the other meaning was presented prior to the polyseme at test (e.g. *spade–diamond*). Finally, in the Different, Unrelated (DU) condition, an unrelated word was presented prior to the polyseme at test (*fruit–diamond*), for comparison to SR.

These 40 pairs of words are shuffled among 80 common filler words to compose a study list of 160 words. At study, each word was rated for either animacy or pleasantness, in an alternating fashion, in order to induce belief that this was the primary task and to reduce explicit encoding of successive words in identical ways. Each word was presented for 900ms, followed by 2,000ms of prompting for a response (which was not recorded), followed by 800ms of blank screen before the next word was presented. After the study list was completed, participants were instructed that they would now perform a recognition test for the words they had just studied. The 160 studied words were randomly shuffled among 160 new words for surprise yes/no recognition testing. In order to reduce the use of strategic and explicit recollection we required participants to respond to the old-new test task within 700ms. Slow responses elicited a "Too slow!" feedback message. Feedback on correctness was given on each test trial in Experiment 1.

Results & Discussion

Of the 60 subjects, 12 were removed for having a mean accuracy not significantly above chance (.522). Of the remaining responses, 2.8% were removed for being faster than 150ms. The remaining 13,397 responses were analyzed using mixed-effects logistic regression, which is more appropriate than ANOVAs for analyzing accuracy (Jaeger, 2008). As regressors, we used the features that the cue may contribute to the target (see Table 1): Familiarity, Semantics, and Context. The logistic regression (see Table 2) shows that each of the three factors increase the odds of recognizing the target, with Semantics being the strongest cue (OR=2.16), followed by Context (OR=1.60), and finally Familiarity (OR=1.09).

Factor	Coefficient	*Z*	*odds*	p-value
(Intercept)	0.40	9.67	1.49	<.001
Familiarity	0.09	2.52	1.09	=.01
Context	0.47	4.95	1.60	<.001
Semantics	0.77	9.57	2.16	<.001

Table 2: Logistic regression coefficients for Experiment 1.

Shown by condition in Figure 1, participants were most likely to respond old to old items in the SR condition, followed by the SS and DS conditions, then the SU condition, and finally the DU condition. The SR and SU findings imply that automatically encoded temporal context affects recognition, although we cannot say how much of the effect is due to a bias shift vs. a performance shift (because the design did not have equivalent conditions of cuing preceding new trials). We note in particular that the presence of a semantic relationship between the polyseme and the previous word at study (DU) or at study and test (SS, DS) increases the probability of giving an old response. From these results, it is clear that automatic associations are formed between both related and unrelated temporally proximal items. We also infer that familiarity accruing to the preceding test item tends to make the next test word seem familiar.

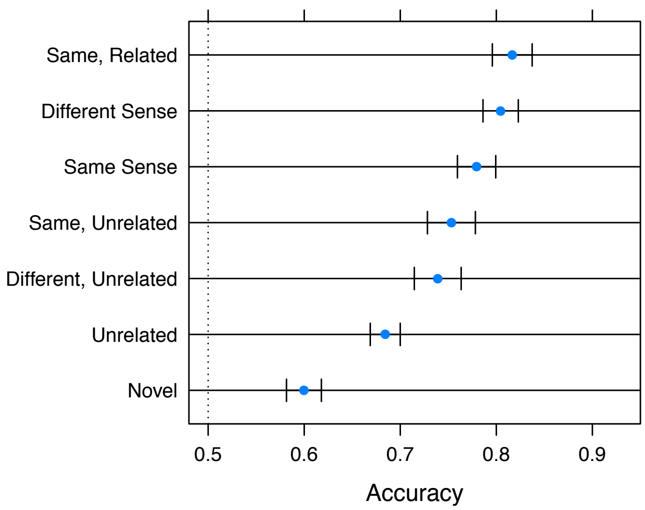

Figure 1: "Hits', p(old|old), for polysemous conditions in Exp. 1 (with feedback).

Experiment 2

In Experiment 1, in contrast to traditional recognition memory experiments, we provided corrective feedback after each response at test. It may be that participants used the feedback signal from the previous trial to classify their feeling of familiarity and strategically used it on the next trial in any of several ways. Thus, in Experiment 2 we did not provide accuracy feedback at test.

Subjects

Participants were 57 undergraduates at Indiana University who received course credit for participating.

Stimuli & Procedure

The same stimuli and procedure were used in Experiment 2, except at test there was no accuracy feedback given.

Results

Of the 57 subjects, 4 were removed for having a mean accuracy not significantly above chance (.522). Of the remaining responses, 1.8% were removed for being faster than 150ms. The remaining 15,275 responses were analyzed using multilevel logistic regression. As in Experiment 1, we found positive effects of Semantics (OR=1.84), Context (OR=1.51), and Familiarity (OR=1.25; see Table 4). Thus, we have evidence for all of these three cues influencing the proximal trial, with and without feedback, when responses are limited to within 700ms. In both experiments, semantics had the strongest effect, followed by context, and then familiarity.

Factor	Coefficient	*Z*	*odds*	p-value
(Intercept)	0.40	7.03	1.49	<.001
Familiarity	0.22	5.77	1.25	<.001
Context	0.41	4.00	1.51	<.001
Semantics	0.61	7.17	1.84	<.001

Table 4: Logistic regression coefficients for Experiment 2.

Figure 2 shows the probability of a "hit" (old to an old test item) by condition for the polysemous manipulations in Exp. 2, which look much like those in Exp. 1. The only qualitative difference is that Same Sense was higher than Different Sense in Exp. 2, whereas a trend in the opposite direction was found in Exp. 1. Even given this difference, the experiments—with and without feedback—had much the same results.

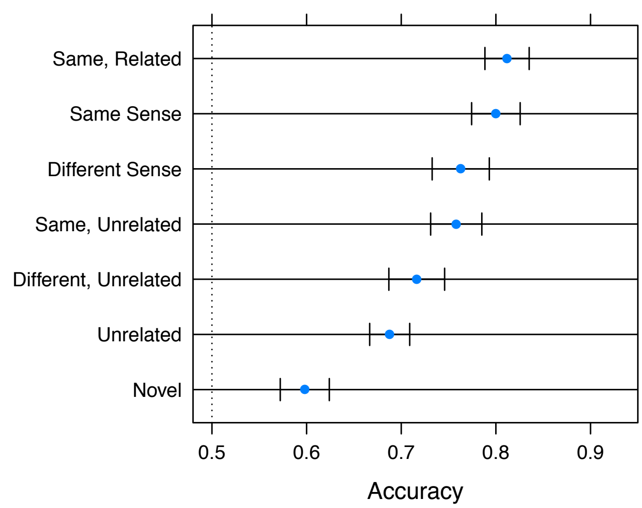

Figure 2: "Hits" for polysemous conditions in Exp. 2 (without feedback).

The previous results, from both studies, were those for the carefully balanced conditions. However there also many filler items that were studied and tested, and many new items tested. Analyses of these items and their sequential effects are taken up in the next section.

Further Sequential Analysis

Analyses of the filler words and new words showed a more general sequential context effect. We analyzed all of the data in both experiments using mixed-effects logistic regression, trying to predict correct responses (old for old, new for new) as a function of the current item's oldness, the previous item's oldness, the correctness of the response to the previous item, and feedback (i.e., Experiment).

Factor	Coeff	*Z*	p-value
(Intercept)	0.29	4.31	<.001
Prev. Correct	0.37	5.39	<.001
Previous Old	0.13	1.53	=.13
Current Old	0.37	4.60	<.001
Feedback	0.16	1.73	=.08
PrevCorr*PrevOld	-0.58	-0.86	=.39
PrevCorr*CurOld	-0.66	-4.19	<.001
PrevOld*CurOld	0.08	0.67	=.51
PrevCorr*Feedback	-0.12	-1.37	=.17
PrevOld*Feedback	-0.52	-4.49	<.001
CurrOld*Feedback	-0.47	-4.30	<.001
PrevCorr*PrevOld* CurrOld	1.06	6.88	<.001
PrevCorr*PrevOld* Feedback	0.50	3.43	<.001
PrevCorr*CurrOld* Feedback	0.26	1.85	=.07
PrevOld*CurrOld* Feedback	1.00	6.01	<.001
PrevCorr*PrevOld* CurrOld*Feedback	-0.89	-4.22	<.001

Table 3. Coefficients for accuracy in both experiments.

Being correct on the previous trial increases the odds of being correct on the current trial (previous: M_{corr} = .64, M_{incorr} = .61). The odds of being correct on the current trial also increase if the previous trial was an old (i.e. studied) word rather than a new (i.e. unstudied) word (prev old M = .64, prev new M = .61). There is no significant effect of the current item's familiarity. There was a significant interaction of previous correctness and the current item type, showing that if a new cue was misidentified as old, subjects were much worse at the current trial (.59 vs. .64). Most strikingly, there was a significant interaction of the cue's and target's familiarity: old targets were more likely to be identified after an old cue (Old|Old = .73, Old|New = .56)—regardless of the response to the cue—and new items were similarly more likely to be correctly identified as new after a new cue (New|New = .64, New|Old = .54).

Figure 3 displays correct rejection of unstudied (New) items and recognition of studied (Old) items as a function of the previous trial's familiarity and response correctness for Exp. 1 (with feedback). Figure 4 displays the same information for Exp. 2.

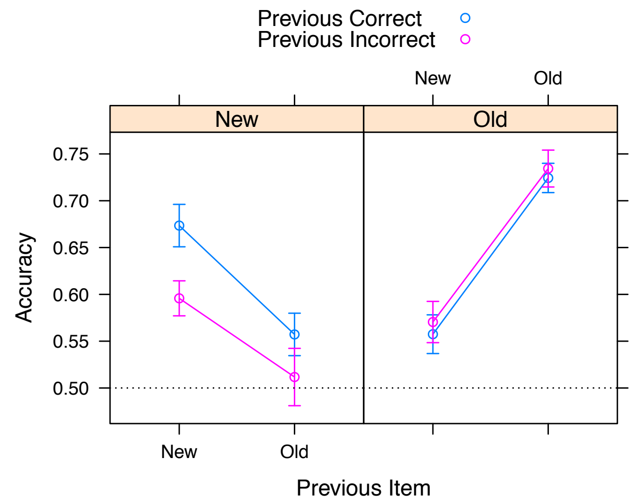

Figure 3: Proportion of correct responses for unstudied (New) items and studied (Old) items by panel, broken down according to the studied/unstudied status of the item on the previous trial, as well as the correctness of the response on the previous trial. Note that New|(Previous New) items are more likely to be correctly rejected than New|(Previous Old), regardless of the correctness of the response on the previous trial. Similarly, Old|Old accuracy is greater than Old|New.

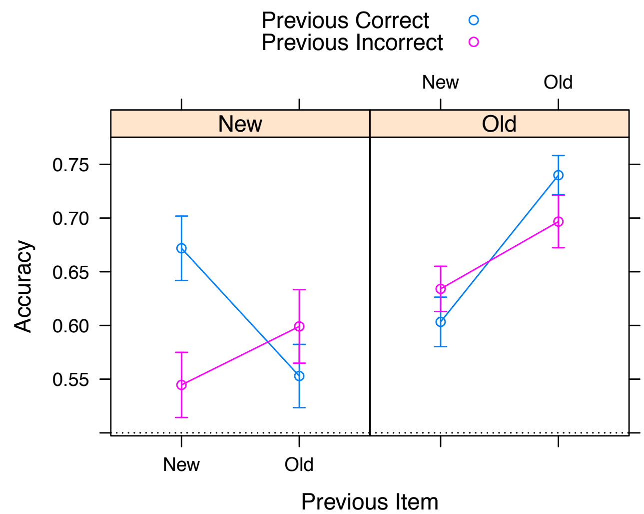

Figure 4: Without feedback, almost the same pattern is evident: Old|Old responses are more accurate than Old|New responses, regardless of correctness on the previous trial. New|(New,Correct) responses are better than New|(Old,Correct), but New|(New,Incorrect) trends lower than New|(Old,Incorrect), breaking the pattern.

We also investigated the 18,023 correct response times using log-linear mixed-effects regression. Shown in Table 4, there was a significant main effect of the previous item's oldness (Previous Old), and a significant interaction of previous oldness with current oldness (PrevOld*CurOld). The mean correct RT when the previous item was old was 506ms vs. 504ms when the previous item was new. When the current item is new, Ss were faster after new items (504ms) than old items (526ms). When the current item is old, Ss were faster after old items (490ms) than new items (507ms). This corroborates the accuracy fluency finding, showing an advantage when the current item is the same oldness as the previous item. There was also a marginally significant interaction of previous response correctness, previous oldness, and feedback.

Factor	Coeff	*t*	p-value
(Intercept)	497.26	65.24	<.001
Prev. Correct	6.05	1.52	=.13
Previous Old	21.44	4.20	<.001
Current Old	-3.20	-0.69	=.49
Feedback	0.07	0.01	=.99
PrevCorr*PrevOld	-10.05	-1.58	=.11
PrevCorr*CurOld	4.31	0.73	=.47
PrevOld*CurOld	-27.43	-3.97	<.001
PrevCorr*Feedback	-4.04	-0.75	=.45
PrevOld*Feedback	-3.77	-0.53	=.59
CurrOld*Feedback	5.12	0.80	=.42
PrevCorr*PrevOld* CurrOld	3.28	0.38	=.70
PrevCorr*PrevOld* Feedback	15.97	1.83	=.07
PrevCorr*CurrOld* Feedback	3.88	0.48	=.63
PrevOld*CurrOld* Feedback	-6.73	-0.71	=.48
PrevCorr*PrevOld* CurrOld*Feedback	-12.98	-1.09	=.27

Table 4. Coefficients for correct RTs in both experiments.

In summary, in an incidental-study recognition memory task with fast responding, we found that the oldness of the prior tested word affects the response time and accuracy on this word. When the current test word is studied, having seen a studied word on the previous trial makes you, on average, faster and more accurate on the current trial – regardless of your response on the previous trial. The accuracy effect happened with and without feedback, so the responses cannot merely be driven by feedback. Seeing a studied word reinstates context features from the study list, and those features contribute to the correct recognition on this trial. For unstudied items preceded by other unstudied items, there is no reinstated context from the previous trial to discount. The need for discounting may explain why correct responses for unstudied items preceded by studied items were drastically slower than for unstudied items preceded by unstudied items.

Discussion

In two recognition memory experiments with time-limited responses—limiting the role of recollection—we found evidence that associations form between incidentally-studied items. Although oldness and semantics can also serve to increase the likelihood of correct recognition, enhanced recognition due purely to sequential context was also observed.

Context Effects

Roughly additive priming effects were found for oldness (familiarity), semantics, and sequential context. Although many models could account for one or even two of these effects straightforwardly, additional assumptions would be required to account for all three. We begin by making a common assumption in memory modeling that study events are encoded as a set of features and that recognition decisions are made on the basis of a comparison of a probe—also consisting of a set of features—to each stored trace in memory with "old" responses given if this comparison is strong enough (e.g., Hintzman, 1988; Murdock, 1992; Shiffrin & Steyvers, 1997; Nelson & Shiffrin, in press). We make the further assumption that some features sampled during the preceding trial are able to "leak" into the probe features present on the current trial. The same leakage is assumed to occur at study, with features of recent items being present in short term memory during the encoding of a subsequent item, and hence joining that item's stored trace (implemented by Nelson & Shiffrin, in press).

Thus, because the memory trace contains some features from the preceding study item, preceding it by the same item at test leads to a stronger match between the test probe and the memory trace. The same account explains the positive but smaller priming when the preceding test item is semantically related—some of the semantic features overlap, but not the many physical features that also overlap when identity priming is used.

Priming due to oldness or semantics independent of sequential context requires yet more modeling assumptions, for which we turn to the dynamic model of recognition of Cox & Shiffrin (2012). This model was able to account for the Jacoby-Whitehouse illusion by assuming, as we have thus far, that primes (in this case, previous test items) contribute some features to the current test probe, at least initially (see Cox, Lewis & Shiffrin, under review, for more details). As more features are sampled and added to the probe, its match to memory evolves over time. If the probe begins with no features at all, the match to memory tends to go down slightly with the first few features sampled, regardless of whether the test item was studied or not. This is because, even if the test item is a target, it will tend not to match most of the other studied items and these mismatches outweigh the single target match until a sufficient number of features are sampled. Thus, after a few features have been sampled, the match for a target test will tend to increase while the match for a foil test will tend to continue to decrease. If, however, a few features are present at the start of the trial, these initial negative steps are avoided for both targets and foils, leading to a bias to say old. This bias is proportional to the similarity between the prime and the test item. Thus, an old unrelated item will lead to a slight bias, and a semantically related item will lead to a larger bias, as observed in the present studies.

Old/New Effects

This mechanism is qualitatively consistent with the observed effects of oldness and correctness of the previous trial in the no-feedback condition. If the preceding item is new, it will tend to contribute features that do not match anything on the list, minimizing the similarity not just with the current test item, but with all the traces in recent memory, leading to a lower tendency to respond “old” than if the preceding item had been old. All else being equal, if we assume that the decision on the preceding trial reflects the quality of evidence provided by the test item on the preceding trial, the effect of the oldness of the previous test item should interact with correctness. For example, if the previous trial was a false alarm, then although the previous item was new, it had to contain enough old features to lead the participant to judge it as old. This account then predicts that the effect of the oldness of the previous test item on the current trial is mainly a function of whether the participant *thought* the previous item was old, manifesting as a crossover interaction between oldness and correctness on the previous trial.

This is the pattern observed in the no-feedback condition, and is consistent with the idea that there is little or no discounting (a la ROUSE; Huber, *et al.*, 2001) of previous item features in that condition. This interaction is absent from the feedback condition, however: one is still more likely to make an “old” response when the previous item was old, but correctness does not have a large effect on responses to old items; rather, correctness only seems to affect responses to new items, with incorrect responses on the previous trial leading to an overall bias to respond “old” on the current trial. In terms of ROUSE’s discounting mechanisms, these data suggest that participants might engage in discounting when the previous trial was incorrect, but they only discount new features. One problem with this account, of course, is that it is unclear whether “old” and “new” features can be identified and differentially discounted. Another problem is that there is no clear reason why participants would only discount new features since doing so only leads to more errors.

An alternative explanation in terms of response criteria—e.g., requiring more evidence to respond after an error—does not hold up either, since that would predict increased accuracy after an error, the opposite of what is observed here. In short, although current models of memory might account for most of the results reported here, the old/new effects in the feedback condition seem to require additional mechanisms that will require further research to elucidate.

Further questions include: What is the effect of using lures that were not studied, but are semantically related to the polyseme, as cues? If an associate of the nondominant meaning is the cue at study, does it still provide an advantage? If the associate is presented after the polyseme at study, is the association still formed?

In the world, things that occur nearby in time (or space) are often related, and if these relations can be remembered they may prove important. Having shown that automatic associations are formed—even between unrelated items—in recognition memory, much work remains to be done to determine how these associations are represented in memory, and what other forms of context they capture.

Acknowledgments

Thanks to Brendan Johns for helpful discussions.

References

Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. Topics in Cognitive Science, *4*(1), 135-150.

Huber, D. E., Shiffrin, R. M., Lyle, K.B., and Ruys, K.I. (2001). Perception and preference in short-term word priming. *Psychological Review 108*, 149-182.

Jacoby, L. L. and Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology 118*, 126-135.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition 24*, 103-109.

Malmberg, K. J., & Annis, J. (2012). On the relationship between memory and perception: Sequential dependencies in recognition testing. Journal of Exp. Psychology: General, *141*(2), 233-259.

Malmberg, K. J., Holden, J. E., and Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psych.: Learning, Memory, and Cognition*, **30**, 319-331.

Mueller, S. T. & Shiffrin, R. M. (2006). REM II: A model of the developmental co-evolution of episodic memory and semantic knowledge. Proceedings of the International Conference on Learning and Development.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). University of South Florida word association, rhyme, and word fragment norms. www.usf.edu/FreeAssociation/

Nobel, P. A. and Shiffrin, R. M., (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psych.: Learning, Memory, & Cognition, 27*, 384-413.

Ratcliff, R. (1978). A Theory of mental retrieval. *Psychological Review*, 95, 59-108.

Schwartz, G., Howard, M. W., Jing, B., and Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. Psych. Science, *16*, 898-904.

Shiffrin, R. M. and Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychological Bulletin & Review. 4*(2), 145-166.

Turvey, M. (1973). On peripheral and central processes in vision: Inferences from an information processing analysis of masking with patterned stimuli. *Psychological Review, 80*, 1-52.