Memory processes underlying long-term semantic priming

Christopher Was¹ · Dan Woltz² · Dale Hirsch³

Published online: 18 October 2018
© Psychonomic Society, Inc. 2018

Abstract

Two decades of research in semantic priming has provided substantial evidence for a distinction between short- and long-term semantic priming effects. Early models of cognition suggested a single mechanism to explain priming at short and long lags. Later models refuted this explanation and proposed that different mechanisms are necessary to account for different durations of priming effects. Two alternative explanations of long-term semantic priming effects have been proposed in the extant literature. The first explanation is that long-term semantic priming effects rely upon the incremental strengthening of abstract semantic memory representations. The second explanation is that long-term priming is the result of memory for prior cognitive operations. In two experiments, we used different semantic content – word meaning versus category membership – to investigate the mechanisms accountable for long-term semantic priming. Evidence from the two experiments suggests that long-term semantic priming effects are due to different memory processes for different semantic content. Long-term semantic priming of word meanings was dependent on strengthening abstract semantic memory representations and persistent priming of category membership was dependent on memory for prior cognitive operations.

Keywords Long-term semantic priming · Semantic priming

Introduction

When response speed and/or response accuracy is facilitated during the processing of words or pictures following prior processing of semantically related information (e.g., *bread-butter*) compared to following prior processing of unrelated information (e.g., *bread-tire*) the phenomenon is referred to as semantic priming (McNamara, 2005). Multiple models of semantic priming have been proposed to account for the facilitation described above, such as spreading activation models (Collins & Loftus, 1975; Quilian, 1967) and compound-cue models (Doshier & Rosedale, 1989; Ratcliff & McKoon, 1988). Although details regarding these models and discussion of the efficacy

of these models to account for priming effects exceeds the scope of the present manuscript, it is important to note that the early views of semantic priming held that priming effects were short-lived – on the scale of seconds or one intervening stimulus trial. Becker, Moscovitch, Behrmann, and Joordens (1997) and Joordens and Becker (1997) first demonstrated long-term semantic priming effects lasting several seconds and with several intervening trials. Importantly, their evidence demonstrated priming effects at longer lags than could be explained by then current theoretical models of short-term priming such as spreading activation (McNamara, 1992), compound cues (Doshier & Rosedale, 1989; Ratcliff & McKoon, 1988), and distributed networks (Masson, 1991). They suggested that greater semantic processing demands in their priming tasks accounted for the novel empirical evidence and proposed a distributed network model that incorporated persistent rather than temporary network changes from prime processing.

The initial network model proposed by Becker et al. (1997) contained a single mechanism to explain priming effects at short and long lags. In contrast, follow-up work by Joordens and Becker (1997) suggested that different model components were necessary to account for short- and long-term priming

✉ Christopher Was
cwas@kent.edu

¹ Psychological Sciences, Kent State University, Kent, OH, USA

² Educational Psychology, University of Utah, Salt Lake City, UT, USA

³ Educational Psychology, Kent State University, Kent, OH, USA

effects. This proposed distinction relates to a subsequent discussion by McNamara (2005), who questioned whether extant long-term priming evidence represented memory processes that differed from those underlying temporary semantic priming effects found in lexical decision and naming experiments.

Long-term category priming

In more recent years, researchers have demonstrated long-lasting semantic priming effects with different tasks that require a variety of semantic processing demands. One line of research investigated priming of category exemplars and features. Hughes and Whittlesea (2003) reported a series of experiments using a two-choice task in which participants selected a term that matched the category of a third term. For example, in Experiment 1B each prime was a category exemplar flanked by the category to which it belonged on one side and a non-matching category on the other (e.g., FRUIT – APPLE – BIRD). Participants were instructed to select the category to which the exemplar belonged as quickly as possible by pressing the appropriate key. The correct category was randomly presented to the right or left of the exemplar on each trial. Probe trials were highly similar, presenting both primed and unprimed probes. The difference between primed probes and their primes was the category exemplar (e.g., FRUIT – ORANGE – BIRD) and category labels were again randomly assigned to the left and right of the exemplar. Hughes and Whittlesea demonstrated long-term category priming effects under varied task conditions. Of interest here, in two experiments they failed to find otherwise robust priming effects when prime processing required a different category operation to the target (either category feature *PEEL* – *APPLE* – *FEATHER* or category label matching, *FRUIT* – *APPLE* – *BIRD*). Exemplar-to-feature matching primed subsequent exemplar-to-feature matching within the same category, and exemplar-to-category label matching primed subsequent matching of this type, but there was no measurable priming when prime and target processing differed in these match operations. Hughes and Whittlesea concluded that the results of Experiments 3, 4, and 5 could be explained by an overlapping operations account of memory effects originally proposed by Kolers (1973). In other words, the category priming was dependent on repeating the mental operations used to access category information and thus demonstrated priming via the repetition of mental operations.

Woltz and Was (2006, 2007) and Was (2010) also investigated long-term category priming, but their experiments differed from those by Hughes and Whittlesea (2003) in several ways. Of particular importance, the prime-target pairs that showed significant priming in studies by Hughes and Whittlesea always repeated one key stimulus word (e.g., *PEEL*-*APPLE*-*FEATHER* and *PEEL*-*ORANGE*-*HISS*). We

contend that the observed facilitation effects may have incorporated a degree of repetition priming of a stimulus word. In contrast, the experiments by Woltz and Was showed category priming across events without repeating any stimulus terms. In a typical experiment, participants were given a memory set consisting of words from two distinct categories (e.g., *ruby*, *daughter*, *diamond*, *uncle*) followed by a 2-s delay. After being cued to recall terms from one of the categories (e.g., *remember the relatives*), participants typed their first two letters of each of the to be recalled words using a standard computer keyboard. After a 6-s delay that included an instruction screen, they next performed target trials in which they decided if two new terms were from the same or different categories (e.g., *aunt brother*; *emerald sapphire*; *opal snail*; *mother soil*). These target trials were always preceded by two “warm up” trials. As indicated in these examples, the target trials included new category exemplars from both recalled and not-recalled categories, both being primed but with different degrees of prior processing. Target trials also included terms from a non-studied category (e.g., *century month*; *hour fudge*). Consistent with the idea that greater semantic processing demands lead to more robust semantic priming, same-category comparisons for both categories from the memory set showed long-lasting target trial facilitation, but this was greater for the recalled category that received additional processing in the memory load phase. These long-lasting priming effects were semantically mediated and without direct or repetition priming as words from the memory load were never presented in the category comparison task, only associates of the memory load items.

Of importance here, one experiment supported conclusions by Hughes and Whittlesea (2003) that category priming depends on repeating the same category matching operation in prime and target trials (Woltz & Was, 2007, Experiment 2). Memory set words could either be exemplars or features associated with a category (e.g., *poodle*, *biology*, *hound*, *astronomy* vs. *leash*, *theory*, *growl*, *laboratory*). Following recall of words from a cued category (*dogs* or *science*), target trials required comparisons either of new category exemplars or of features. Exemplar comparisons were facilitated only by exemplar memory sets, and feature comparisons were facilitated only by feature memory sets.

Thus, evidence from several category priming experiments support the role of memory for prior mental operations rather than abstract content in long-term semantic priming, and this is consistent with the notion that long-term semantic priming effects may represent underlying memory processes differently from those underlying short-term semantic priming, specifically the repetition of mental operations. In addition to the empirical evidence for operation-specific long-term priming, this interpretation makes intuitive sense because the demonstration of long-term effects appears to depend on tasks with increased semantic processing demands. It seems likely that

when priming tasks require more complex semantic operations in the prime trials, memory for these processing operations would partly underlie the subsequent facilitation in target processing. Physiological evidence (pupillometry) suggests just such a relationship. Papesh, Goldinger, and Hout (2012) found that after participants listened to a list of 80 items (40 non-words, 20 low-frequency words, and 20 high-frequency words), pupil diameter increased for correctly recognized items at testing as compared to items that were missed. They interpreted these findings to indicate that during the testing phase participants elaborated on the previously presented items, recreating the cognitive processes used during encoding. However plausible this account of long-term semantic priming is, it has been challenged by evidence from tasks that demand the evaluation of word and proposition meaning rather than category membership, as well as simple list learning tasks.

Long-term word and proposition priming

Tse and Neely (2005) reviewed five studies comprising 14 experiments reporting that eight of the 14 experiments demonstrated long-term semantic priming of critical items from a Deese/Roediger-McDermot (DRM) design. Tse and Neely conducted four experiments of their own that demonstrated a long-term semantic priming effect for the critical items. Experiment 4 is of particular interest as the results demonstrated a critical item long-term semantic effect exceeding an interval of 50 s. They proposed that the effects were due to the long-term semantic activation of the critical items. Relevant to the current study, the long-term priming was not dependent on repetition of either the item or the cognitive operations performed on the primes and targets. Tse and Neely (2007) extended these findings by eliminating the possibility of participants using intentional retrieval strategies, again demonstrating long-term semantic priming not dependent on repetition of items or operations, and independent of strategy usage.

Woltz (2010) first reported long-term semantic priming effects for word-meaning comparisons. In these experiments, participants were required to determine if two words presented simultaneously were synonyms or if they were unrelated. Meaning comparisons such as *moist damp* were primed by a related comparison such as *soggy wet*. Half the word comparisons had similar meanings as in this example, and half had different meanings (e.g., the comparison *ample enclose* primed by the comparison *enough surround*). Given findings from the category-matching tasks, it was assumed that memory for the meaning comparison operation would be at least partly responsible for the long-term priming effects. To test this, two experiments introduced an unpredictable subset of prime trials in which the meaning comparison between two words was omitted. In all trials, the first word of a pair was

presented with a delay of 750 ms before the second word was presented. This was assumed to allow for full meaning retrieval of the first word. In the subset of prime trials without meaning comparisons, the first word was followed by the same word (positive match) or a string of X's (negative match). Both words of a subsequent target trial in this trial condition were primed in this manner, with the assumption that both word meanings were activated but no meaning comparison operation was represented in memory. If priming effects were reduced or eliminated in this set of trials, the operation-based interpretation of long-term priming would be supported. In comparison to the operation-specific priming effects found in category tasks, the magnitude of priming was equivalent whether or not meaning comparisons were performed in the prime trials. This was observed at both short and longer prime-target lags. These results suggested that memory for abstract word meaning rather than processing operations was responsible for relatively long-term semantic priming of words not associated with well-structured categories.

A similar conclusion was drawn from subsequent priming experiments using short sentences that represented single propositions (Woltz, Sorensen, Indahl, & Splinter, 2015). They reported a series of experiments designed to contrast the roles of memory for abstract proposition meaning and memory for mental operations used to encode the meaning and select a response. Evidence from six experiments supported the role of abstract proposition meaning in long-term semantic priming, and there was no evidence for operation-based priming effects associated with the mental operations involved in making decisions about sentence syntax, meaning evaluation, or response choice.

Proposed resolution

The contrasting conclusions about memory processes instrumental in long-term semantic priming likely reflects differences in semantic memory organization associated with well-structured category information versus relatively unstructured word and proposition meaning. For common categories, category labels are associated with many exemplars and a set of core features shared by exemplars. Prime and target trials that require a comparison of either exemplars or features from the same category, repeat similar memory retrieval and decision operations within this semantic structure. For example, if a prime trial requires participants to decide if *pear* and *banana* are exemplars of the same category, the decision is likely made by a relatively automatic retrieval of the common category label *fruit* for both. A target trial that presents two new fruit exemplars, *cherry* and *mango*, for a membership comparison would repeat the same category label retrieval based on common exemplars. In a sense, this repeated retrieval operation represents a form of repetition rather

than semantic priming. As with other forms of repetition priming, facilitation from repeated category retrieval operations apparently can reflect a high degree of specificity. Repetition priming has been demonstrated across a variety of tasks and stimulus domains, and some experiments have shown facilitation when the target trial format differs from the prime trials (e.g., Biederman & Cooper, 1991, 2009). Nevertheless, considerable evidence suggests that facilitation from stimulus repetition can have a high degree of specificity with respect to stimulus orthographic features, associative relations (e.g., when the associations between target items are consistent from study to test), and response actions (Schacter, Dobbins, & Schnyer, 2004). Such specificity may explain why retrieval or activation of a category from the comparison of category features is not facilitated by prior activation of that same category from the comparison of exemplars.

The shared meaning of two related words or proposition expressions lacks an organized structure of common exemplars with shared physical features and a common label that is typical of well-learned categories. Furthermore, the shared semantic features of both can be abstract in contrast to those of common category exemplars. For example, the overlapping semantic features of adjectives such as *moist*, *damp*, *soggy*, and *wet* are abstract compared to the shared features of objects representing a category such as *bird*. And as noted, no universal label exists for the semantic features or the synonyms that share them. Consequently, synonym decisions presumably reflect a relatively ill-defined evaluation of whether the degree of semantic overlap reaches some threshold, rather than the retrieval and evaluation of a well-established label. Consequently, no single operation such as a category label retrieval from either exemplars or features is repeated across prime and target trials, only evaluations of semantic overlap. Indeed, Hutchison (2003) reviewed the semantic priming literature to date and found that automatic short-term priming was due to association strength and, in the case of synonyms and antonyms, feature overlap. As such, facilitation due to repeated retrieval operations is likely to be minimal in a word-meaning task compared to that in a category-matching task, but facilitation due to reprocessing abstract semantic content may be greater. Consistent with this explanation, prior evidence suggests long-term semantic priming effects for word and proposition meaning comparisons depend on the incremental strengthening of abstract semantic memory representations rather than memory for prior processing operations (Woltz, 2010; Woltz et al., 2015).

Given this interpretation of previous evidence, we propose that the memory processes underlying long-term semantic priming effects differ as a function of the associative organization of semantic memory for the stimulus domain in use. The structure of category knowledge allows for facilitation in category matching due to repeated retrieval operations across prime and target trials. Memory representations of semantically related

words and propositions lack a common, well-defined structure, and, consequently, comparisons of these stimulus domains only show facilitation from activation of abstract meaning rather than structure-specific repeated operations.

We acknowledge another viable explanation for previous findings that support this assertion. The category and word or proposition experiments cited here differed in their methods of presenting prime and target phases of the priming tasks. The category priming experiments primed categories with the presentation of a memory set that required category matching to identify words for recall. Target trials differed from the priming events in requiring yes-no decisions as to whether two exemplars or features represented the same category. In the word and proposition priming experiments, prime and target trials were of the same format, typically requiring a selection of one of two terms to match the meaning of an initial term. Conceivably, these methodological differences could be responsible for the differing evidence regarding the role of memory for prime trial operations in target trial facilitation. In order to resolve these different explanations, we conducted two experiments that employed comparable priming procedures for category and word matching tasks.

Overview of experiments

Two priming experiments were designed to be as structurally similar as possible, with one containing category exemplar and feature stimuli and the other containing synonym and antonym stimuli. To assess the role of memory for prime trial operations in target trial facilitation, the evaluation operation was manipulated to be the same or different across these trials. In the category stimuli experiment, trials either required participants to identify a category exemplar or a category feature, and this was either consistent or inconsistent across the prime and target phases for a given category. In the word meaning experiment, trials required participants to identify either a word's synonym or its antonym, and again, this evaluation operation was either consistent or inconsistent across primes and targets. Although the operations manipulated in each experiment were necessarily different because of the two stimulus domains, the structural similarity between the two experiments afforded an improved evaluation of whether memory for the prime trial evaluation operation plays a different role in the priming effects within the two domains.

Experiment 1: Category priming

Method

Participants To determine sample size for Experiments 1 and 2 we reviewed the most relevant literature. A review of the work

by Woltz and Was (Was, 2010; Woltz, 2010; Woltz & Was, 2006, 2007) found that the mean sample size across eight experiments was 86, with a median of 79. Effect sizes comparing primed to unprimed trials in their experiments ranged from $\eta_p^2 = .17$ to $.71$. A power analysis using the G*Power 3 computer program (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a total sample of 87 participants would be needed to detect moderate to large effects ($\eta_p^2 = .17$) with 90% power using an F test with alpha at $.05$. We chose a sample of 90 participants for each experiment, and in Experiment 1, 90 undergraduate students (68 female, 22 male) participated for partial credit in an introductory educational psychology course. Ages ranged from 15 to 44 years, with a median age of 20 years.

Apparatus Participants performed the experimental task on PCs with 22 in SVGA monitors and standard keyboards. Programming of all tasks was completed with E-Prime® software (Schneider, Eschman, & Zuccolotto, 2002).

Materials Seventy-two category groups were created, with each group consisting of target stimuli and two sets of prime stimuli. Stimuli are available in the Online Supplementary Materials as well as on Open Science Framework. Target stimuli were comprised of three words: a category exemplar (e.g., *daisies*), another exemplar representing the same category (e.g., *marigolds*), and a word describing features common to most exemplars of the category (e.g., *stems*). The two prime stimulus sets also contained three words each, and these represented exemplars and features from the same category as the target stimuli (e.g., *tulips*, *orchids*, *petals*; *roses*, *carnations*, *nectar*). The word triplets used for first and second prime trials and target trials were fixed for all participants. To the extent possible, category exemplars within a triplet were selected to represent a moderate degree of the natural variability within the category. For example, the first prime exemplars for the category bird consisted of *eagles* and *canaries* rather than more similar exemplars such as *eagles* and *hawks*. In addition, we attempted to select exemplars and features for the initial prime stimuli that were easier to evaluate than those assigned to the second prime and the target. These two features of stimulus assignment were intended to maximize accuracy and priming effects.

Thirty of the 72 category groups were drawn from the category norms produced by Overschelde, Rawson, and Dunlosky (2004). The remaining 42 groupings did not correspond to published category norms because of the requirement that each category must have at least three relatively unique features that are common to all exemplars. For example, in contrast to the category flower described here, categories also contained in the category norms such as *carpenters' tools* do not have shared features that would be familiar to all participants.

A separate set of 64 stimulus groups was created to serve as practice examples and filler trials for separating prime and target trials. As with the experimental stimuli, these contained a primary category exemplar, a second exemplar, and a feature, all from the same category, but they consisted only of three words rather than the full prime and target groups.

Design and procedure The category-priming task was a 2×3 within-subject manipulation of target trial evaluation (exemplar vs. feature identification) and prime condition (primed by same evaluation, primed by different evaluation, and unprimed). In order to maximize the priming effects regardless of evaluation consistency, there were two prime trials preceding each primed target, and prime-target lags were relatively short for the investigation of long-term semantic priming of categories. On average, there was a lag of seven trials between the second prime trial and the corresponding target trial. Based on prior evidence, we expected the overall priming magnitude to be relatively large, and if priming is at least partly attributed to memory for priming operations rather than abstract category representation, we expected greater facilitation when prime and target operations were consistent.

Participants performed the experimental task in groups from one to four in a room consisting of individual computer carrels separated by sound-deadening panels. All task instructions were presented on the computer display and over headphones. The experimental session lasted approximately 45 min.

Each trial began with a string of three asterisks presented for 500 ms, followed by a blank screen for 500 ms. This was followed by the presentation of the trial stimuli, which remained visible until a keyboard response was made. The primary category exemplar term (e.g., *tulips*) was situated in the location of the previous asterisks, at the horizontal center and 7.5 cm from the display top; 3 cm below the primary exemplar was a phrase that defined the evaluation to be performed. For half the trials, the phrase was *are like* (or *is like* if the primary exemplar term was singular), and on the other half it was *have* (or *has*). The second category exemplar (e.g., *orchids*) and the category feature (e.g., *petals*) were presented in a single row of text 4 cm below the evaluation statement and approximately 12 cm apart from each other. Left and right arrows were shown immediately below the left and right words, respectively. These arrows indicated the response key to use in selecting the word that matched the primary exemplar given the evaluation statement. The left versus right position of the exemplar and feature terms was randomized in each trial. After the participant responded to the trial by pressing the right or left arrow key, a blank screen appeared for 1,500 ms before the asterisks appeared to begin the next trial.

At the beginning of the session, participants saw four example trials (two with *is/are like* and two with *has/have* evaluation phrases) with auditory instructions presented over headphones. Then, participants performed 16 practice trials

with accuracy feedback following each response. Final instructions before the experimental trials encouraged them to respond as quickly as possible while achieving at least 90% accuracy.

There were 12 experimental trial blocks, each with 18 trials. The first four trials were primes, one each for the four primed target trial conditions (exemplar-evaluation target primed by exemplar-evaluation primes; exemplar-evaluation target primed by feature-evaluation primes; feature-evaluation target primed by exemplar-evaluation primes; feature-evaluation target primed by feature-evaluation primes). Thus, in a randomized order, two of these primes contained the *is/are like* evaluation statement and two contained the *has/have* evaluation statement. The next two trials were fillers, with one of each evaluation type in a random order. The next four trials were more primes. These prime stimuli were from the same category stimulus group and used the same evaluation statements as the first set of primes, thus priming the same four target trials in the same manner. Again the order was randomized. Two more filler trials, one of each evaluation type in random order, preceded the final six trials of the block, which were targets. Two target trials, one of each evaluation

type, were unprimed in that their categories had not been seen in earlier prime trials. The remaining four targets corresponded to the four prime trial conditions. The six target trials were presented in a random order (see Fig. 1).

The 72 category stimulus groups were divided into six sets of 12. Assignment of stimulus set to the six target trial conditions was counterbalanced across the 90 participants so that each stimulus set was presented with equal frequency in all six conditions as recommended by McNamara (2005).

No individual trial feedback was presented during the 12 blocks. At the conclusion of each block, the participant's average response time (RT) and percentage correct were displayed. Participants were also shown their average performance data for all previous blocks to encourage

their monitoring of RT accuracy across blocks. Before each new block, participants were reminded to maintain an accuracy of at least 90% while responding as quickly as possible.

Results and discussion

Mean RT and percentage of errors are reported for all trial conditions. Because priming effects in this task have been

Fig. 1 Example of an experimental block

found in both of these measures, hypothesis tests will be reported for a single measure of rate of correct responses (RCR). RCR is calculated as (proportion of correct responses/(reaction time in ms/60,000)) and is interpreted as the number of correct responses per minute, as such higher values mean better performance. This measure has been used previously in related priming experiments (Woltz & Was, 2006, 2007; Woltz, 2010; Was, 2010; Woltz, et al., 2015) and has been demonstrated to be appropriate for incorporating and combining meaningful variance from both latency and accuracy (Vandierendonck, 2017). Priming effects have been shown to be larger using this measure compared to either RT or errors (Woltz, 2010; Woltz et al., 2015), and evidence indicates that it can adjust for different speed-accuracy tradeoffs under some task conditions (Sorensen & Woltz, 2015).

Table 1 contains mean error and RT data for all trial conditions. The top section of Table 1 presents data for prime trials. As seen in prime trial means, there were substantial trends for faster and more accurate performance on trials requiring an evaluation of category exemplars compared to features. In addition, responses to first primes tended to be faster and more accurate. Figure 2 presents the response rate means for the prime trials. There was a large main effect in the prime trials of evaluation type with greater response rates to prime trials requiring the identification of category exemplars, $F(1,84) = 57.32$, $MSe = 18.01$, $p \leq .001$, $\eta_p^2 = .41$. There was also a response rate difference between first and second prime trials. Response rates were lower in second primes, $F(1,84) = 10.82$, $MSe = 15.87$, $p = .001$, $\eta_p^2 = .11$. The effect of prime sequence did not differ by evaluation type, $F(1,84) = 1.02$, $p = .315$.

Table 1 also contains mean RT and percentage error data for target trials. As in the prime trials, participants responded more quickly with fewer errors to trials requiring a category exemplar evaluation. There were also notable trends suggesting faster and more accurate responses to primed compared to unprimed trials.

Figure 3 presents the response rate means for the target trials. As seen in this figure, there was a main effect for target trial evaluation type, $F(1,84) = 30.85$, $MSe = 4.96$, $p \leq .001$, $\eta_p^2 = .27$. There was also an overall priming effect represented by the contrast of both primed conditions with the unprimed trials, $F(1,84) = 62.18$, $MSe = 17.76$, $p \leq .001$, $\eta_p^2 = .43$. The overall priming effect did not differ by target trial evaluation type, $F(1,84) < 1$. Of primary importance, the response rate of primed trials differed as a function of the evaluation operation in the prime trials, $F(1,84) = 34.53$, $MSe = 23.35$, $p \leq .001$, $\eta_p^2 = .29$. As seen in Fig. 2, the magnitude of priming was greater when the evaluation operation of the target trial was the same as that in the prime trial. Furthermore, this difference was not moderated by target trial evaluation, $F(1,84) < 1$.

The observed impact of evaluation consistency on priming magnitude supports previous evidence that long-lasting category priming partly or wholly reflects memory for the processing operations used to make prime trial category decisions. Though facilitation was notably larger when prime and target trials required the same category evaluation, regardless of whether this was in identifying category exemplars or common features, there was a significant priming effect, albeit smaller, when prime and target operations differed, $F(1,84) = 16.27$, $MSe = 22.26$, $p < .001$, $\eta_p^2 = .16$. This is evident in Fig. 3. This differs from earlier findings by Woltz and Was (2007), and suggests that some facilitation reflects a generalized increase in availability of category information regardless of category operation.

Experiment 2

The goal of this experiment was to provide a conceptual replication of the first experiment using word meaning rather than category evaluations. The design of this experiment was virtually identical to the first experiment. Trials presented a primary word, and two alternatives. Experiment 2 differs in that to manipulate the comparison operation in this stimulus

Table 1. Mean response time and errors of Experiment 1 by trial condition

	Response errors (%)				Response latency (ms)			
	Exemplar evaluation		Feature evaluation		Exemplar evaluation		Feature evaluation	
Trial type	M	SD	M	SD	M	SD	M	SD
Prime trials								
1st Prime	4.7	5.2	7.1	6.5	1,809	404	1,889	443
2nd Prime	4.3	5.1	6.7	6.7	1,874	462	2,005	504
Target trials								
Unprimed	6.9	8.5	8.7	8.7	2,188	562	2,280	683
Primed same	4.5	6.4	5.7	6.9	1,960	506	2,072	560
Primed different	6.1	8.9	9.5	9.2	2,107	611	2,139	569

Fig. 2 Mean response rate for prime trials of Experiment 1. Error bars represent within-subject 95% confidence intervals

domain, one alternative was a synonym and one was an antonym, and trials instructed participants to identify either the alternative that had a similar or the alternative that had an opposite meaning. For example, the target trial consisting of *moist, damp, dry* could have had the instruction to choose the word that is either similar to or the opposite of *moist*. This would have been primed by trials with similar semantic content (e.g., *wet, soggy, arid*), and could have had either instruction for the evaluation operation. Again, the critical test was whether priming magnitude depended on the same evaluation operation being performed on prime and target trials.

Method

Participants Ninety undergraduate students (70 female and 20 male) participated in the experiment and received partial credit in an introductory educational psychology course. Ages ranged from 18 to 43 years, with a median age of 21 years.

Fig. 3 Mean response rate for target trials of Experiment 1 by comparison operation and prime condition. Error bars represent within-subject 95% confidence intervals

Apparatus Participants performed the experimental task on PCs with 17 in SVGA monitors and standard keyboards. Programming of all tasks was completed with E-Prime® software (Schneider, Eschman, & Zuccolotto, 2002).

Materials Similar to Experiment 1, 72 stimulus groups were created, with each group consisting of target stimuli and two sets of prime stimuli. Stimuli are available in the Online Supplementary Materials as well as on Open Science Framework. Target stimuli were comprised of three words: a primary word, a synonym, and an antonym (e.g., *tranquil, placid, chaotic*). Similarly, the two sets of prime stimuli each contained three words, and these had similar meanings to the three target words (e.g., *peaceful, serene, hectic; calm, still, frenzied*).

A separate set of 88 stimuli was created to serve as practice examples and filler trials to separate prime and target trials. These consisted only of three words: a primary word, a synonym, and an antonym, rather than the full prime and target group.

Procedure Participants performed the experimental task in groups from one to five in a room consisting of individual computer carrels separated by sound-deadening panels. All task instructions were presented on the computer display and over headphones. The experimental session lasted approximately 45 min.

The experimental task was nearly identical to that of Experiment 1 other than the stimuli and corresponding instructions. The only procedural difference was that each of the 12 blocks began with two additional filler trials, one of each evaluation type in a random order. These served as warmup trials before the first primes and therefore the addition of the two filler trials at the beginning of each trial did not increase the lag between the second prime and the target. Warmup trials were not possible in Experiment 1 because appropriate category stimuli were exhausted in creating the experimental and filler stimulus conditions.

Results and discussion

Table 2 contains mean error and RT data for all trial conditions. The top section of Table 2 presents data for the prime trials. As seen in prime trial means, there were substantial trends for faster and more accurate performance on trials requiring an evaluation of similar compared to opposite meanings. In addition, responses to first primes tended to be faster and more accurate. Figure 4 presents the response rate means for the prime trials. There was a large main effect in the prime trials of evaluation type with greater response rates to prime trials asking to identify words with a similar meaning, $F(1,84) = 147.60$, $MSe = 5.07$, $p \leq .001$, $\eta_p^2 = .64$. There was also a response rate difference between first and second prime trials.

Table 2. Mean response time and errors of Experiment 2 by trial condition

	Response errors (%)				Response latency (ms)			
	Similar evaluation		Opposite evaluation		Similar evaluation		Opposite evaluation	
Trial type	M	SD	M	SD	M	SD	M	SD
Prime trials								
1st Prime	5.8	5.4	6.1	5.7	2,490	718	2,883	783
2nd Prime	7.4	6.1	9.4	6.7	2,735	741	3,225	936
Target trials								
Unprimed	9.8	9.7	11.4	9.7	2,900	874	3,345	1006
Primed same	9.6	9.2	10.0	9.1	2,856	879	3,271	1104
Primed different	10.0	8.8	11.6	9.7	2,758	804	3,270	961

Responses rates were lower in second primes, $F(1,84) = 154.96$, $MSe = 6.08$, $p \leq .001$, $\eta_p^2 = .65$. The effect of prime sequence did not differ by evaluation type, $F(1,84) < 1$.

Table 2 also contains mean RT and percentage error data for the target trials. As in the prime trials, participants responded more quickly with fewer errors to trials requiring a similar meaning evaluation. There were also trends suggesting faster responses to primed compared to unprimed trials. Figure 5 presents the response rate means for the target trials. As suggested by the RT and error data, there was a large main effect for target trial evaluation type, $F(1,84) = 83.09$, $MSe = 4.0$, $p \leq .001$, $\eta_p^2 = .50$. There was also an overall priming effect represented by the contrast of both primed conditions and the unprimed trials, $F(1,84) = 5.97$, $MSe = 10.30$, $p = .017$, $\eta_p^2 = .07$. The overall priming effect did not differ by target trial evaluation type, $F(1,84) < 1$. Of primary importance, there was no evidence that the response rate of primed trials differed as a function of consistency of evaluation type across primes and targets, $F(1,84) < 1$. This result did not differ by target trial evaluation type, $F(1,84) = 1.24$, $p = .231$.

Fig. 4 Mean response rate for prime trials of Experiment 2. Error bars represent within-subject 95% confidence intervals

General discussion

Both experiments demonstrated a form of semantic priming that persisted over an average of seven unrelated, intervening trials. These priming effects conform to the conceptual definition of semantic or indirect priming, because primed target trials in both experiments shared no content words with their corresponding prime trials. In Experiment 1, corresponding prime and target trials contained exemplars and features from the same category. In Experiment 2, corresponding prime and target trials contained synonyms and antonyms. Participants responded more quickly, $F(1,178) = 97.84$, $MSe = 453639$, $p < .001$, $p = .36$, and with fewer errors, $F(1,178) = 16.27$, $MSe = 15.60$, $p < .001$, $p = .08$, to the category comparisons of Experiment 1 compared to the word meaning comparisons of Experiment 2. In addition, priming effects measured in response rate were larger in the category domain, $F(1,178) = 17.99$, $MSe = 4.48$, $p < .001$, $p = .09$. These differences correspond to findings reported in previous research with similar priming tasks (see Woltz & Was, 2006, 2007; Was, 2010;

Fig. 5 Mean response rate for target trials of Experiment 2 by evaluation type and prime condition. Error bars represent within-subject 95% confidence intervals

Woltz, 2010). We attribute these differences to the concreteness and familiarity of category exemplars and features relative to antonyms and synonyms.

One outcome in the prime trial data of both experiments might seem counter to expectations. That is, mean response rate was lower on the second compared to the first primes, despite the expectation that the first prime should facilitate processing on the second priming trial. The experiment was not designed to evaluate priming effects within the prime trials, and prime stimuli were confounded rather than counterbalanced with respect to order. In selecting first and second prime trial content for each stimulus group, our intent was to identify the simplest comparison as the first prime in an effort to facilitate accurate semantic processing of both primes. In support of our assumption that the longer responses of the second prime reflect content differences rather than strategic processes, several previous experiments using a similar task have found facilitation in semantically related word-meaning comparisons at prime-target lags ranging up to five trials (see Woltz, 2010).

The goal of these experiments was to evaluate the role of memory for prior operations performed on the semantic content. In attempts to reconcile evidence of semantic priming effects that persist beyond the range of intervening events explained by temporary semantic priming theories, it has been suggested that such persistence might reflect different memory mechanisms such as the representation of mental operations rather than content (McNamara, 2005; Woltz & Was, 2006, 2007; Was, 2010). Our findings support the conclusion that semantic priming of category comparisons is partly operation-specific, and this is similar to findings by Woltz and Was (2007). In contrast, semantic priming of word comparisons appeared to be unaffected by the type of meaning evaluation operation. This is consistent with findings by Woltz (2010) and Woltz et al. (2015). The current evidence supporting the difference in memory representations affecting priming is important because the priming tasks were designed to be as similar as possible, while the previous evidence came from different tasks that could have been responsible for the difference in findings.

Why is operation-based priming unique to the domain of well-learned categories? Several differences may partly explain these findings. First, as noted earlier, words representing category exemplars and features are embedded within semantic concepts that are relatively well structured and have a common label. That is, the category concept of *bird* has numerous exemplars that are commonly referred to with the category label of *bird*, and most exemplars share the same common features such as beak, wings, feathers, and nest, which are closely associated with category label. In contrast, words that have a related meaning but do not represent exemplars of a commonly recognized category, lack the same well-defined semantic structure of categories.

Chaffin and Glass (1990) developed an empirical taxonomy of semantic relations that yielded five basic relationships. These relationships include similarity (e.g., synonyms), class inclusion (e.g., exemplar-category relations), and part-whole (attribute-category relations). Germane to the current study, Chaffin and Glass found RTs were faster for identifying true hierarchical relationships (i.e., class inclusions) than for true equivalent relationships (i.e., synonyms). They interpreted these results as demonstrating relative complexity for analyzing equivalent relationships as compared to hierarchical relationships.

In our view, these differences are likely to explain the exclusive role of operation memory in category priming. In a prime trial, if a participant is required to evaluate whether roses and carnations are exemplars of the same category, it is likely that the category label *flower* is activated and the decision process evaluates whether both are exemplars of that category. If a second prime trial requires the evaluation of whether tulips and orchids are exemplars of the same category, the evaluation of these exemplars relative to the category label is repeated. Subsequent presentation of a target trial requiring membership evaluation of daisies and marigolds would again repeat the membership evaluation for the same category. The repetition of this operation performed relative to the same category is likely to play a substantial role in the target trial facilitation compared to unprimed category trials. Target trials in the category task that require a different evaluation operation from that in the prime trials show less facilitation. In the current example, evaluating whether stems are features of daisies does not necessarily require the activation of the category label *flower*; it requires the evaluation of physical features of a category exemplar. As such, previous repetitions of evaluating the correspondence of category labels can offer less benefit to this target trial. On the other hand, repeating physical feature evaluations of category exemplars that share the physical features can provide greater facilitation to this target trial. This could in part explain the difference in priming magnitude between Experiments 1 and 2.

The repetition of distinct evaluation operations across prime and target trials is more difficult to conceive in the word-meaning task of Experiment 2. Deciding that *placid* rather than *chaotic* has a similar meaning to *tranquil* does not occur with reference to an overarching category label with which the synonyms are exemplars. The evaluation may depend on an assessment of the degree to which the words have overlapping semantic features (as explained by Hutchison, 2003), but semantic features are relatively abstract compared to the physical features of category exemplars. Having previously been exposed to prime trials requiring a decision that *peaceful* has a similar meaning to *serene*, and *calm* has a similar meaning to *still*, presumably activates common semantic representations, but no common operation refers to a label of the semantic concept. Nor is there a common operation of

evaluating concrete, physical features as was the case with category exemplars; the shared semantic features of words are abstract and vary somewhat from synonym to synonym. Similarly, trials that require the selection of words with opposite meanings depend on discriminations of abstract rather than concrete, physical features. Consequently, it appears that semantic priming in this task does not depend on memory for the specific evaluation operation performed on the semantic features. Instead, it reflects the shared semantic content that is activated during the evaluations.

Furthermore, the complexity described by Chaffin and Glass (1990) could possibly describe the differential facilitation effects between the two experiments. Becker, Moscovitch, Behrmann, and Joordens (1997) proposed that for long-term semantic priming to occur a considerable amount of semantic processing must transpire. Relevant to the current study, if equivalency comparisons (i.e., synonyms and antonyms) require a different type of semantic processing due to the complexity of the relationships, it is not surprising that in Experiment 2 we found non-significant differences between same and different evaluations – suggesting that the priming was due to the strengthening of the abstract representations.

It is also possible that the mental operations underlying antonym and synonym decisions are highly similar compared to those underlying feature and category decisions. They might not only be similar, but it is possible that participants could have used the same mental operation for synonym and antonym decisions by selecting a “match the definition” strategy. On each trial, participants might have identified the synonym and for dissimilar decision chosen the other item. We do not have evidence to support this conclusion, but it is clear that the synonym/antonym decisions likely represent similar cognitive operations and thus the difference in priming magnitude between these operations is non-existent.

Our interpretation of the different involvement of memory for processing operations in the two experiments refers to the different structure of semantic knowledge in words that represent common categories and those that do not. Nevertheless, we acknowledge several other differences between these two experiments that could contribute to the findings. First, the manipulation of content evaluations was not isomorphic across experiments. We made a concerted effort to equate the two experimental tasks on all procedural elements, but it was not possible to utilize the same evaluations in the two content domains. We rejected the idea of using surface feature evaluations that could have been comparable across domains. Our focus was on semantic priming and persistent semantic priming effects appear to depend on substantial semantic processing demands. Instead, we identified evaluations within each domain that were distinct and could be differentiated with relatively simple instructions within each trial, but that pertained to core semantic attributes of the stimuli. Despite

these efforts, it is possible that the nature of the operations used in Experiment 1 lent themselves to priming differences more than those used in Experiment 2, irrespective of the semantic content.

Second, the domain of category exemplars and features necessarily consists of concrete nouns. Many nouns do have synonyms and antonyms, but the number of synonyms and antonyms required to create a word-meaning priming task equivalent to the category priming task necessitated the use of adjectives and adverbs that often have several words with similar or opposite meanings. Consequently, the stimuli used in the two experiments differed considerably in concreteness.¹ Differential priming effects due to level of concreteness have been demonstrated in previous research. In an attempt to test the different organizational frameworks theory of Crutch and Warrington (2005, 2010), Ferré, Guasch, García-Chico, and Sánchez-Casas (2015) found that when primes and targets were associated semantic priming was present for both concrete and abstract words. In contrast, when primes and targets were semantically related but not associated, priming effects were only present when the words were concrete. It is tempting to state that the difference in concreteness could in part explain our findings, but Ferré et al. (2015) used a lexical decision task as a measure of short-term priming, whereas our task is a long-term semantic priming task, requiring greater processing, and – although not as robust as the category stimuli – did result in the long-term semantic priming of semantically related, abstract words.

Despite some competing explanations for our findings related to inherent task differences when comparing semantic priming in category and word-meaning domains, we favor a knowledge structure interpretation. That is, operation-specific facilitation is uniquely found in the category domain because of the hierarchical structure of that knowledge. Common categories are linked by a category label, and shared category features partly define category membership. Because of this well-learned structure, comparisons between either exemplars or features involve a form of evaluation that is repeated with new exemplars or features from the category. Synonyms and antonyms have no such hierarchical structure. Their relationships are not defined by a superordinate label, and the meaning overlap consists of shared semantic features that are not well-learned attributes defining membership in a recognizable group. Consequently, there can be no repeated retrieval and comparison of a well-learned label to facilitate subsequent evaluations. Nor can there be repeated recognition and comparison of defining characteristics. Instead, we believe that long-term semantic priming for meaning comparisons

¹ A systematic comparison of word concreteness values and word frequency values could not be conducted because many of the category member stimuli were proper nouns not found in normative databases (e.g., Danube, Potomac, Rhine, Rio Grande, Nile, Amazon).

depends on increments of strength in abstract semantic representations that partially overlap for synonyms and antonyms.

Although we feel the provided explanations adequately describe the source of the priming effects demonstrated in the two experiments, there are other possible alternative explanations in the priming effects. One explanation for the priming effects in our two experiments could be strategic processing. Specifically, prime rehearsal might have played a role in the priming effects. In our experiments, four of the six target trials were related to the primes and the targets occurred within an average of seven trials of the second prime. This could be considered a high relatedness proportion (RP: Neely, Keefe & Ross, 1989) and may have led participants to create an expectancy – the active generation of possible stimuli for the upcoming target (McNamara, 2005) – thus facilitating responses on target trials. Thus, in Experiment 1 after a number of 18-trial blocks, subjects may have realized that if they attempted to rehearse words from Primes 1 and 2, or attempted to predict what other exemplars might appear in the Targets, it would help them respond more quickly and accurately during target presentation. Using the data from Experiment 1, we conducted four paired-samples t-tests to determine if the priming effects in the second half of the experiment were larger than in the first half of the experiment, reflecting strategic processing. Each of the t-tests compared RCS on target trials for matched evaluations (category evaluation primed by category evaluations, and feature evaluation primed by feature evaluation) compared to unprimed trials of the same evaluation. Two of the t-tests were conducted using the mean RCS from the first six trials of the experiment and the other two were conducted using the means of the last six trials. All priming effects were significant, but the key finding was a comparison of effect sizes. For the category evaluations, the priming effects were larger for the first six trials, $t(89) = 6.42, p < .001, d = .57$, compared to the last six trials, $t(89) = .86, p < .005, d = .20$. Comparison of the feature evaluations produced similar effect sizes for the first half of trials, $t(89) = 3.79, p < .001, d = .37$, and the second half $t(89) = 4.99, p < .001, d = .39$. This suggests that strategic processing did not lead to larger priming effects as the experiment progressed.

To conclude, the evidence from these experiments supports a view that long-term semantic priming reflects different memory processes for different semantic content. The suggestion by McNamara (2005) that long and short-term semantic priming effects might reflect different memory mechanisms may be accurate when contrasting models of short-term semantic priming with the operation-specific explanation for long-term category priming. Even so, the explanation for long-term semantic priming in word-meaning comparisons is similar to that those for short-term priming. Although mechanisms like temporary activation must be replaced with incremental strengthening, both refer to semantic memory rather than memory for operations. Supporting the possibility that

common memory mechanisms underlie short- and long-term semantic priming of word meanings, Woltz (2010) found a systematic, negatively accelerated decline of priming magnitude across prime-target lags that represents both short- and long-term semantic priming. A full resolution of this issue will require further evidence of semantic priming effects stemming from a greater variety of stimuli and task demands across a range of prime-target lags.

References

- Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: a computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1059.
- Biederman, I., & Cooper, E.E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, 20(5), 585–593.
- Biederman, I., & Cooper, E.E. (2009). ‘Evidence for complete translational and reflectional invariance in visual object priming’: Authors’ update. *Perception*, 38(6), 809–817.
- Chaffin, R., & Glass, A. (1990). A comparison of hyponym and synonym decisions. *Journal of Psycholinguistic Research*, 19(4), 265–280.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407.
- Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, 128, 615–627.
- Crutch, S. J., & Warrington, E. K. (2010). The differential dependence of abstract and concrete words upon associative and similarity-based information: Complementary semantic interference and facilitation effects. *Cognitive Neuropsychology*, 27, 46–71.
- Doshier, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General*, 118(2), 191.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Ferré, P., Guasch, M., García-Chico, T., & Sánchez-Casas, R. (2015). Are there qualitative differences in the representation of abstract and concrete words? Within-language and cross-language evidence from the semantic priming paradigm. *The Quarterly Journal of Experimental Psychology*, 68(12), 2402–2418.
- Hughes, A. D., & Whittlesea, B. W. A. (2003). Long-term semantic transfer: An overlapping-operations account. *Memory and Cognition*, 31(3), 401–411.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, 10(4), 785–813.
- Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1083.
- Kolers, P. A. (1973). Remembering operations. *Memory & Cognition*, 1(3), 347–355.
- Masson, M. E. J. (1991). A distributed memory model of context effects in word identification. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 233–263). Hillsdale, NJ: Lawrence Erlbaum Associates.
- McNamara, T. P. (1992). Priming and constraints it places on theories of memory and retrieval. *Psychological Review*, 99(4), 650–662.

- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.
- Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1003.
- Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, 83(1), 56–64.
- Quillian, M. R. (1967) Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, 12, 410–430.
- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, 95(3), 385.
- Schacter, D.L., Dobbins, I.G., & Schnyer, D.M. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nature Reviews Neuroscience*, 5(11), 853–862.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide* (Version 1.1). Pittsburgh: Psychology Software Tools Inc.
- Sorensen, L. J., & Woltz, D. J. (2015). Transforming response time and errors to address tradeoffs in complex measures of processing speed. *Learning and Individual Differences*, 40, 73–83.
- Tse, C. S., & Neely, J. H. (2005). Assessing activation without source monitoring in the DRM false memory paradigm. *Journal of Memory and Language*, 53(4), 532–550.
- Tse, C. S., & Neely, J. H. (2007). Semantic and repetition priming effects for Deese/Roediger—McDermott (DRM) critical items and associates produced by DRM and unrelated study lists. *Memory & Cognition*, 35(5), 1047–1066.
- Vandierendon, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior research methods*, 49(2), 653–673.
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, 50(3), 289–335.
- Was, C. A. (2010). The persistence of content-specific memory operations: Priming effects following a 24-hour delay. *Psychonomic Bulletin & Review*, 17(3), 362–368.
- Woltz, D. J. (2010). Long-term semantic priming of word meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6), 1510.
- Woltz, D. J., Sorensen, L. J., Indahl, T. C., & Splinter, A. F. (2015). Long-term semantic priming of propositions representing general knowledge. *Journal of Memory and Language*, 79, 30–52.
- Woltz, D. J. & Was, C. A. (2006) Activation in long-term memory during and after attention focus in working memory. *Memory and Cognition*, 34(3), 668–684.
- Woltz, D. J., & Was, C. A. (2007). Available but Unattended Conceptual Information in Working Memory: Temporarily Active Semantic Knowledge or Persistent Memory for Prior Operations? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(1), 158–168.