

Semantic relatedness can impair memory for item locations

Xinyi Lu, Mona J.H. Zhu, Evan F. Risko

University of Waterloo, Waterloo ON, Canada

This version of the article may not completely replicate the final authoritative version published: <https://doi.org/10.1007/s00426-023-01889-7>. It is not the version of record and is therefore not suitable for citation.

Send correspondence to:

Xinyi Lu

Department of Psychology

University of Waterloo

Waterloo, Ontario

Canada, N2L 3G1

email: xinyi.lu@uwaterloo.ca

While memory for semantically related items is improved over unrelated items in many cases, relatedness can also lead to memory costs. Here we examined how the semantic relatedness of words within a display influenced memory for their locations. Participants learned the locations of words inside grid displays; the words in a given display were either from a single category or were from different assorted categories. When a display containing words from a single category was compared to a scrambled display containing words from multiple categories, location memory performance was rendered worse, while word recall performance was significantly improved. Our results suggest that semantically structured spaces can both help and harm memory within the context of a location memory task. We hypothesize that relatedness can improve memory performance by increasing the likelihood that matching candidates will be retrieved, yet might worsen performance that requires distinguishing between similar target representations.

Keywords: semantic memory, location memory, category membership, spatial organization

Semantically related information tends to be better remembered than unrelated information. For example, a related set of words is usually recalled better compared to an unrelated set of words (Hunt & McDaniel, 1993; Lewis, 1971; Lu et al., 2022; Mandler, 1967; Puff, 1970). However, semantic relatedness can also, in certain tasks, lead to memory costs; for example, in comparing the learning of related words vs. unrelated words, relatedness can lead to increased false recall of unpresented words (Guerin & Miller, 2008; Lu et al., 2022; Roediger & McDermott, 1995). Both the benefits and costs of semantic relatedness have been proposed to be a result of the same underlying principle, that is, it promotes both *generalization* across and *interference* between semantically similar items (Kahana et al., 2022). Semantic relatedness, therefore, might be something of a “double-edged sword” in memory; depending on the situation, it could lead to benefits and/or costs, both within and across different tasks (Kahana et al., 2022). In the present investigation we examine a prediction that arises henceforth; namely, how learning semantically related (vs. unrelated) items will influence memory for their locations.

According to Kahana et al. (2022), semantic relatedness typifies the law of similarity in memory, a key principle around which memories are organized. Memory retrieval is often triggered by some similarity between the present and some past experience (Kahana, 2020; Surprenant et al., 2006). This idea has been formalized in models that conceive of retrieval as a search of memory in response to similar or matching cues at test (e.g., Kimball et al., 2007; Raaijmakers & Shiffrin, 1981). In these accounts, memory for a related list of words tends to be better than an unrelated list of words as the categorical information that can be extracted from the former serves as a powerful cue to access information during retrieval (Mandler, 1967; Tulving & Pearlstone, 1966). The retrieval of one item on the list can also cue the retrieval of other related items on the list, as illustrated by the observation that individuals tend to make transitions

among semantically related items in their free recall output (Bousfield et al., 1954; Romney et al., 1993), even when learning ‘unrelated’ lists without obvious semantic structure (Howard & Kahana, 2002). From a memory search perspective, candidates that match the given semantic cues are distinct compared to non-matching distractors, thus are more likely to be retrieved leading to a benefit in free recall. This benefit, however, goes hand in hand with increased similarity (decreased distinctiveness) between items *within* a category (Hunt, 2013; Hunt & Einstein, 1981). This effect is readily observed in participants’ tendency to retrieve unrepresented items that are nonetheless consistent with the semantic cues (e.g., believing that SLEEP was presented after studying BED, REST, AWAKE; Roediger & McDermott, 1995). That is, the generalizing effect of semantic relatedness facilitates the true recall of targets but also leads to the false recall of distractors (i.e., interference). In a free recall task, therefore, the effect of semantic relatedness is chiefly a boon to memory performance, but nonetheless comes with some costs to precision (i.e., an increase in false alarm rate).

Here we investigate the potential “double-edged” sword of semantic relatedness in the context of item location memory (Postma et al., 2008; Postma & De Haan, 1996). In contrast to a typical free recall task, in which participants are simply required to recall all items, a location memory task requires binding to-be-remembered items to specific locations (e.g., Hund & Plumert, 2003; Lu et al., under review). Thus, while we might expect increased inter-item semantic similarity to provide a boost to free recall (i.e., item memory), this same inter-item similarity might increase interference across item-location pair bindings (i.e., location memory). To date, there has been surprisingly little empirical work investigating the effect of inter-item semantic relatedness on location memory. In the following sections, we first discuss previous investigations into the effects of semantic relatedness on a similar task (memory for serial order),

before reviewing the handful of extant investigations into the effects of semantic relatedness on location memory.

Semantic Relatedness and Order Memory

One can draw parallels from the aforementioned location memory task to a serial recall or serial reconstruction task: a location memory task requires binding to-be-remembered items to specific spatial locations, while the latter tasks require binding to-be-remembered items to specific serial positions (Kowialiewski et al., 2023). While semantic similarity has been consistently found to facilitate item memory (i.e., increased items recalled), some studies have found that it is detrimental to order memory (i.e., increased rate of serial order errors; e.g., Saint-Aubin et al., 2005; Tse, 2009; Tse et al., 2011). Conceptualized within a “double-edged sword” framework of similarity and distinctiveness, semantic relatedness is thought to hinder order recall by increasing similarity among retrieval cues (i.e., interference). A strong version of this argument is that semantic similarity-based interference leads to a form of confusion known as an “interpretation problem”: when recalling an item, another item is erroneously recovered and recalled in the position of the targeted one (Tse, 2009). However, a number of studies have not found a cost of semantic relatedness on serial order (e.g., Saint-Aubin & Poirier, 1999; Guérard & Saint-Aubin, 2012; Neale & Tehan, 2007; Neath et al., 2022), although a recent meta-analysis by Ishiguro and Saito (2021) suggests that semantic similarity might have a small detrimental effect on order memory. Overall, the balance of evidence would suggest either a small cost or a null effect (but not a benefit) of semantic relatedness on order memory, in contrast to the robust finding of its benefit on item memory.

One possibility for the weak or perhaps even null effect of semantic similarity on order memory in previous studies is that these studies employ immediate serial recall tasks, as they are primarily concerned with questions of how item and order information are represented in short-term/working memory. In the traditional working memory model (Baddeley & Hitch, 1974), short-term memory is conceived as primarily relying on phonological representations (i.e., the phonological loop) while long-term memory relies on a semantic-based code. As this conception would suggest, acoustic similarity-based interference in short-term memory is well-established, while evidence for semantic similarity-based interference is equivocal (Baddeley, 1966a; 1966b). Conversely, there is a longstanding body of literature establishing robust semantic similarity-based interference in long-term memory tasks. For example, proactive interference refers to the phenomenon whereby, if a series of to be remembered items are presented for later recall, performance on the later items becomes progressively worse as interference from previous items is built up (Underwood, 1957). However, if the category of the to-be-remembered items is changed partway during the presentation of the list (for example, from flowers to games), then performance returns to near baseline level (i.e., release from proactive interference; Gardiner, Craik & Birtwistle, 1972).

In sum, while evidence for a cost of semantic relatedness in short-term memory for order is weak to none, there is robust evidence that semantic relatedness leads to interference in memory tasks that operate at longer time frames. We have highlighted two such examples: semantic relatedness can increase the false recall of related critical lures (e.g., Roediger & McDermott, 1995) and lead to the buildup of category-specific interference in recall (Gardiner, Craik & Birtwistle, 1972). In the current investigation, we examine the question of semantic

similarity-based interference in an item location memory task, at typical delays for investigating the influences of long-term memory.

Semantic Relatedness and Location Memory

As noted above, there are few existing studies investigating the effect of inter-item semantic relatedness on location memory. Early research established that pre-existing notions of semantic categories can have a biasing effect on our memory for location and distance. For example, we tend to think of post offices as belonging in the same vicinity as other commercial establishments like banks, as opposed to educational institutions like schools (Hirtle & Mascolo, 1986). As a result, we might remember a post office and a bank as being closer to each other than they actually are. Further evidence of the semantic biasing of location memory comes from a study by Hund and Plumert (2003) that asked participants to learn the locations of items arranged in either semantically homogenous quadrants (containing related items) or scrambled quadrants (containing unrelated items). They found that participants tended to remember objects as being closer together when they were semantically related than when they were scrambled. The biasing effect of semantic relatedness in these above examples is consistent with the notion of category generalization (Kahana et al., 2022), that is, the representations of same-category items became more similar to each other but also more distinct from other-category items. In the context of location memory, relatedness appears to have the effect of pushing same-category items closer together in space (Hund & Plumert, 2003).

Given the above, we might reasonably predict that semantic relatedness will lead to either benefits and/or costs to location memory performance depending on task demands (i.e., whether performance will be helped by generalization or hindered by interference). One demonstration of a benefit of inter-item semantic relatedness comes from a recent series of studies by Tomparý

and Thompson-Schill (2021). Participants were asked to study items that were distributed across a few on-screen regions, such that each region contained mostly items within the same category (e.g., animals: *lion*, *giraffe*, *raccoon*), though items from other categories could be present as well. Tomparry and Thompson-Schill (2021) found that location memory accuracy was greater for same-category items that were clustered in the same region, compared to same-category items located in a different region. For example, “raccoon” would be located with greater accuracy if it had originally been clustered with other animals, rather than if it had been within the household items cluster (e.g., *chair*, *bucket*, *bowl*). The authors proposed that since same-category items tended to be clustered near each other, participants were able to use the spatial cluster-category association and use the category label as a cue to retrieve the region associated with a particular item’s category (Tomparry & Thompson-Schill, 2021). This led to a benefit for items that were within their category-themed cluster, but not for items outside the cluster. In other words, in this particular task, generalizing a category’s item locations to a given region would benefit the items that were actually located in that region (consistent with the category-region association), but not items that were inconsistent with that region. In a similar series of studies, Lu et al. (under review) asked participants to learn the locations of items in a semantically partitioned display (i.e., an environment that was composed of four partitions, each containing items from a single category) as well as a scrambled display (where each partition contained an assortment of items from different categories). A semantically partitioned display, for example, could contain all clothing items (e.g., *shirt*, *dress*, *jacket*) in the top-left partition, and all office supplies (e.g., *stapler*, *notebook*, *ruler*) in the top-right partition, while a scrambled display would have an assortment of items scrambled across the four partitions. During test, participants were given each item one by one and asked to indicate its original location.

Critically, half the test trials were cued trials, on which participants were explicitly told which partition the item was located; the rest were uncued trials, on which participants were not given this information. Overall, Lu and colleagues found that the semantically partitioned display led to a location memory benefit over the scrambled display. However, this benefit was eliminated on cued trials, when participants were explicitly told which partition the item was in. Lu and colleagues proposed that the benefit was driven by increased distinctiveness between semantic category-themed partitions, allowing participants to associate the general category of items with a given region (e.g., tools are all in the top-right section). On uncued trials, this association would benefit memory to the extent that participants could at least place them in that region (rather than somewhere else in the full display). On cued trials where this effect was controlled for, however, the benefit of semantic relatedness was entirely eliminated.

If semantic relatedness increases distinctiveness between category clusters in space, but also increases similarity/interference across item-location pairs within a category cluster (Hunt, 2013; Hunt & Einstein, 1981), then we should expect to observe a memory cost in a task where the benefit of between-category distinctiveness is reduced or eliminated. The cued condition in Lu et al. (under review) provided a test of this hypothesis, as in this condition, knowing only the category-partition association would provide no benefit. However, Lu and colleagues found no semantic cost (or benefit) to location memory within a given partition. Thus, while Lu and colleagues found support for a benefit from increased distinctiveness between semantic partitions, they found no evidence of a cost for increased interference within a semantic partition. One possible reason for the null effect on within-partition memory obtained by Lu et al. is that, in the context of location memory, semantic relatedness does not increase similarity/interference across same-category items. Another possibility is that the null effect was due to the use of a

partitioned display, wherein the partition-cueing manipulation used by Lu and colleagues may not have been entirely successful in eliminating the benefit of between-category distinctiveness.

In the present investigation, we sought to clarify the existence of a cost or a benefit of semantic relatedness to location memory in an unpartitioned display, comparing a semantic grid display (containing items from a single category) to a scrambled grid display (containing items from different categories). We expected that this kind of display would more closely match past item memory research that has demonstrated robust benefits of semantic organization, wherein items are often drawn from either a list composed of items from a single category or not (Lewis, 1971; Lu et al., 2022; Puff, 1970). This design allowed us to examine the effect of semantic relatedness on location memory performance without the contribution of the partition-level (between-category) benefits.

Current Investigation

We report three pre-registered experiments (Experiment 1a: <https://osf.io/ncfg6>; Experiment 1b: <https://osf.io/8nxb6>; Experiment 2: <https://osf.io/m4phu>) comparing item location memory for a semantically homogenous display to a scrambled display. Display type (semantic vs. scrambled) was manipulated within-participants: in each experiment, participants completed two study-test blocks, one in which the display contained words all belonging to the same semantic category, and one in which the display contained a random array of words from different semantic categories. Within each block, participants were shown the target words in the display one by one in the study phase, and subsequently, performed an explicit location memory test where they were given each word one by one in random order and asked to indicate its location. In Experiments 1a and 1b, we included a final global recall task as a measure of item memory. We expected that the semantically homogenous display would result in an

improvement in item recall relative to the scrambled display, akin to how semantically blocked word lists lead to higher rates of recall than scrambled lists (e.g., Lewis, 1971; Lu et al., 2022; Puff, 1970). We predicted that the semantically homogenous display might result in a cost to location memory relative to the scrambled display, based on increased semantic similarity-based interference between the items (Saint-Aubin et al., 2005; Tse, 2009; Tse et al., 2011).

An additional aspect of interest in the current investigation was in the metacognitive consequences of interference. Previous studies have shown that participants experience semantically related word pairs as more fluent to process than unrelated word pairs (e.g., Undorf & Erdfelder, 2015), that is, they report that the related word pairs feel easier to process. However, as we anticipated semantic relatedness to lead to interference-based costs in the current location memory task, we wondered to what extent participants would be sensitive to this. Hence, we included a post-experiment exploratory probe, where we asked participants to rate the subjective difficulty of the semantic versus scrambled displays.

Experiments 1a & 1b

Methods

Experiments 1a and 1b were identical (1b was intended as a replication of 1a) and are described together.

Participants

We pre-registered a sample size of 90 participants for each experiment based on a power analysis estimating a small-medium effect size for within-subjects comparisons ($d_z = .30$, $\alpha = .05$, two-tailed, within-subjects t -test) using G*Power (Faul et al., 2007). This effect size is a conservative estimate compared to previous, comparable experiments (Tomparly & Thompson-

Schill, 2021; Lu et al., under review). We further anticipated this sample size to be a conservative estimate with respect to the generally more powerful mixed effects models we planned to use (Quené & van den Bergh, 2004). After removing participants who did not meet the pre-registered criteria, this sample size was achieved in Experiment 1a ($N = 90$; 62 women, 24 men, 2 other, 2 undisclosed, $M = 20.00$ years, $SD = 2.85$), while two additional participants had to be removed in Experiment 1b for previously participating in similar experiments ($N = 88$; 62 women, 26 men, $M = 20.21$ years, $SD = 3.69$). Participants were undergraduate students at the University of Waterloo participating for course credit and provided informed consent.

Materials

Stimuli were words from six categories of household objects (see Appendix). Each category contained ten possible exemplars. For each participant, one category was randomly chosen to populate the semantic (single category) grid; the other, scrambled (multiple category) grid was populated by randomly selecting two items from each of the remaining five categories. As each 5 x 5 grid had 25 clickable location squares, 10 of these contained items and 15 were empty squares. Item positions in each grid were pseudo-randomly assigned such that each row and column contained two items (cf. Siegel & Castel, 2018). The order of the semantic/scrambled grids was again randomized across participants. Examples of the two grids are shown in Figure 1 (note that participants never saw the grid with all items visible).

		PLATE	TOASTER		BASS	SCISSORS			
SPOON				PAN				CLARINET	DEODORANT
	TRAY	MUG					SWEATER	HAIRBRUSH	
KETTLE	WHISK				RULER				SNEAKERS
			SPONGE	SPATULA		PENCIL	PLIERS		

Figure 1. Semantic (left) and scrambled (right) item grids in Experiments 1a/1b.

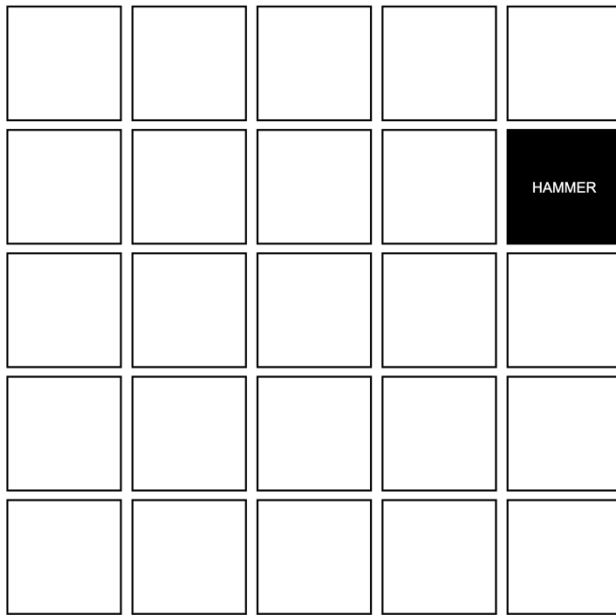
Procedure

After being given a brief overview of the experiment, each participant performed two blocks, each containing the encoding task, the location memory task, and a difficulty probe.

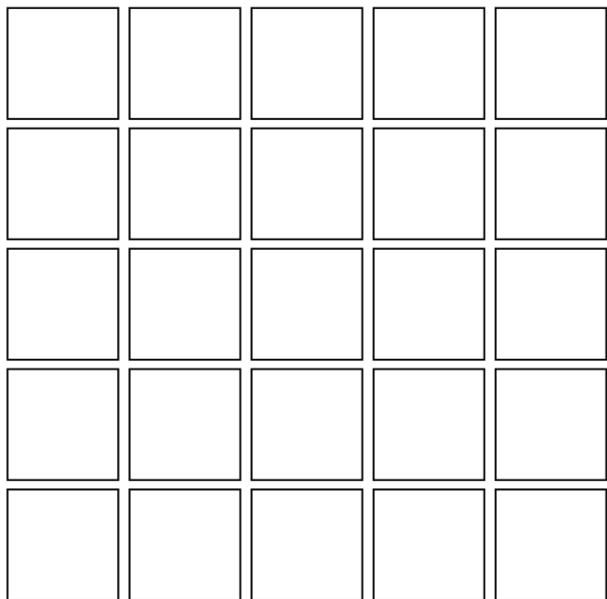
These tasks are described below.

Encoding Task. On each trial, only a single item in the grid was visible to the participant (see Figure 2). The visible item was presented for 5000 ms in white text on a black square. There were 10 trials for a single grid (one for each item, presented in random order) and the intertrial interval was 400 ms.

Location Memory Task. On each trial, participants were presented with the empty grid from the encoding task to the left. The target item was shown on the right. They were instructed to click on the square that corresponded to the original location of the target item. There were 10 trials for a single grid (one for each item, presented in random order) and no feedback was given.



Remember the object's location.



Where was this object?:

FLOSS

Figure 2. Encoding task (top) and location memory task (bottom).

Difficulty. Upon completion of the location memory task, participants were asked to answer, “How difficult was the location memory task you just did?” on a 7-point scale anchored by 1: Very Easy and 7: Very Difficult.

After participants completed both the semantic and scrambled blocks, they were asked to recall as many items as they could from both grids by typing them into an on-screen text field (i.e., global free recall task). Upon completion of this task, they were probed on (1) whether they noticed anything about the two grids; (2) what strategies they used to remember the locations of items. Participants were then asked to indicate which of the two grids they found to be easier. Finally, participants were asked to provide their age and gender and complete an attention and effort check questionnaire.

Results

Data from 16 participants in Experiment 1a and 40 participants in Experiment 1b were not analyzed according to the pre-registered criteria; they self-reported that they were not paying attention or did not give effort during the task in the post-study questionnaire. An additional 2 participants in Experiment 1b were excluded due to previously participating in related experiments. After all exclusions, this left a sample size of $N = 90$ in Experiment 1a and $N = 88$ in Experiment 1b. The analyses we report below were pre-registered unless stated otherwise. All analyses were conducted using R (R Core Team, 2019). Mixed-effect regressions were conducted using the *lme4* package (Bates et al., 2015). Categorical predictors (e.g., semantic vs. scrambled displays) were coded in the models using sum-contrasts. For the random effects structure, we began with a model containing by-participant and by-stimuli random intercepts; by-participant and by-stimuli random slopes for the effect of semantic/scrambled display were included only when doing so significantly improved the fit of the model (Bates et al., 2018; Matuschek et al., 2017). In cases where the initial model was singular (indicating possible overfitting), we removed either the by-participant or by-stimuli random intercept to reduce model complexity. Unless otherwise specified, all linear and logistic mixed-effects models were

run using the *bobyqa* optimizer. Given that degrees of freedom can be difficult to estimate accurately in mixed-effects models (Bates et al., 2015), approximated p -values using Wald z -statistics are provided, considering the relatively large number of observations in the current study. This was done using the *sjPlot* package. All data and analysis code are available at <https://osf.io/8xf6a/>.

Location Memory

In order to examine whether the semantic grid influenced location memory performance, we looked at (1) absolute location accuracy, defined by whether participants chose the correct target square or not; (2) Euclidean distance from the chosen location to the correct location on all trials.

Location Accuracy. A mixed-effects logistic regression revealed that there was a significant main effect of condition in Experiment 1a, such that location memory accuracy was lower in the semantic grid than the scrambled grid, $b = -0.18$, 95% CI [-0.28, -0.08], $z = 3.61$, $p < .001$. This effect was not significant in Experiment 1b, $b = -0.09$, 95% CI [-0.21, 0.03], $z = 1.37$, $p = .170$. Figure 3A shows the effect of semantic grid on location accuracy.

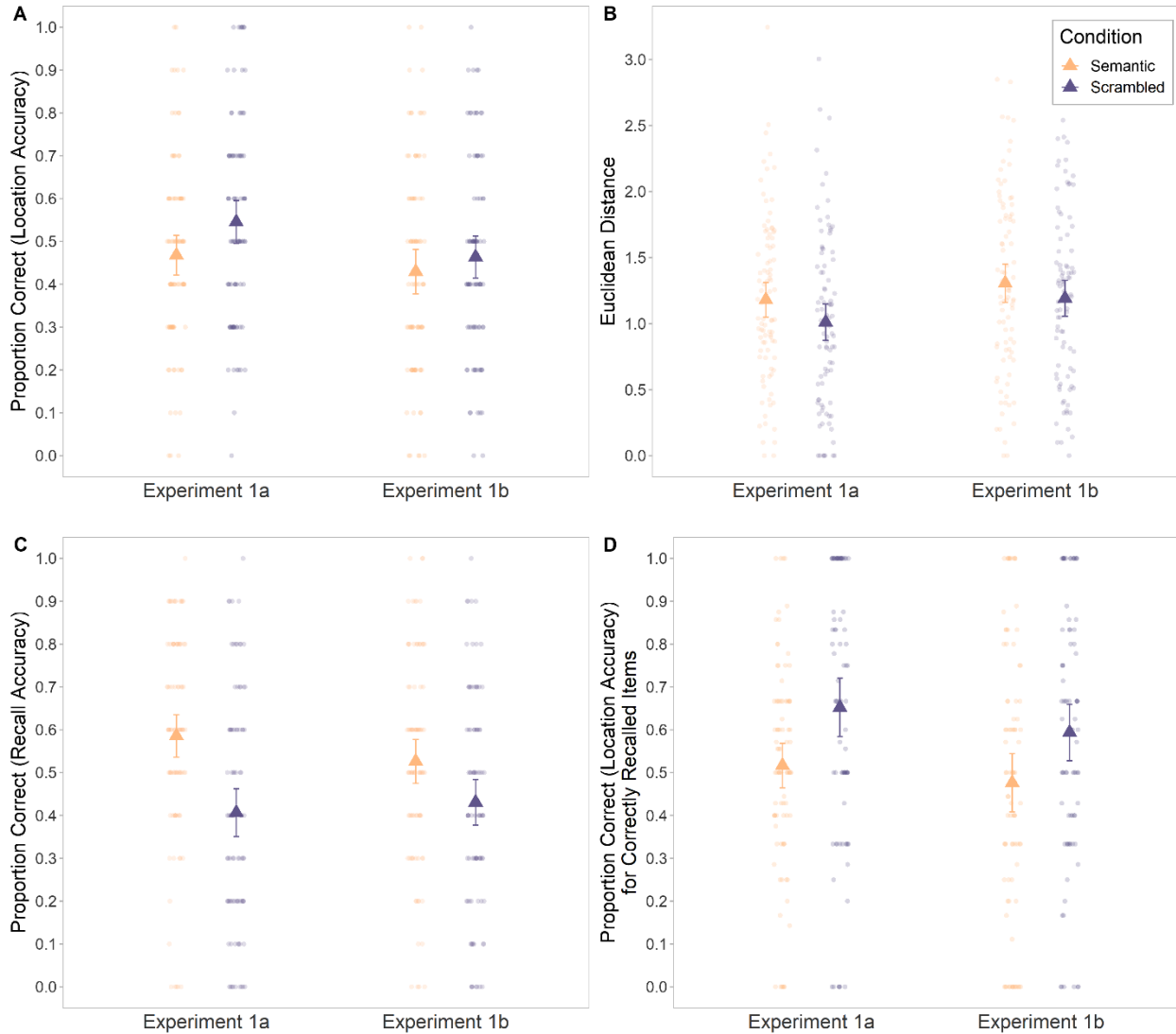


Figure 3. Mean (A) location accuracy, (B) Euclidean distance, (C) item recall accuracy, and (D) location accuracy for correctly recalled items only by semantic/scrambled display in Experiment 1a and 1b. Error bars represent 95% confidence intervals. Circles in the background represent each participant's average performance per condition.

Euclidean Distance. A mixed-effects linear regression revealed that there was a significant main effect of condition in Experiment 1a, such that mean Euclidean distance between selected and target locations was larger in the semantic grid than the scrambled grid, $b =$

0.09, 95% CI [0.03, 0.15], $t = 2.68$, $p = .004$. This effect was not significant in Experiment 1b, $b = 0.06$, 95% CI [-0.02, 0.14], $t = 1.45$, $p = .141$. Figure 3B shows the effect of semantic grid on Euclidean distance.

Item Memory

Item Recall Accuracy. The semantic grid significantly improved item recall performance in both experiments, Exp 1a: $b = 0.44$, 95% CI [0.29, 0.60], $z = 5.29$, $p < .001$; Exp 1b: $b = 0.19$, 95% CI [0.03, 0.36], $z = 2.27$, $p = .023$. Figure 3C shows the effect of semantic grid on recall accuracy.

Intrusion Rate. The overall recall intrusion rate was very low, Exp 1a: 2.8%; Exp 1b: 3.4%. We performed a post-hoc classification of intrusions into the following categories: (1) related or similar to a word in the scrambled display; (2) related or similar to a word in the semantic display; (3) unrelated intrusion. The raw intrusion counts per category were: Exp 1a: scrambled: 11, semantic: 9, other: 6; Exp 1b: scrambled: 11, semantic: 16, other: 3.

Location Memory Controlling for Item Memory

In line with our predictions, the semantic grid improved item memory performance relative to the scrambled grid. Therefore, in order to assess the effect of the semantic grid on location memory performance alone, we attempted to account for the semantic benefit to item memory by statistically controlling for item memory performance. We entered both item recall accuracy and condition (semantic vs. scrambled) into a mixed-effects linear regression predicting location memory accuracy. There was a main effect of item recall on location memory performance, Exp 1a (exploratory): $b = 0.77$, 95% CI [0.55, 0.99], $z = 6.92$, $p < .001$; Exp 1b (pre-registered): $b = 0.78$, 95% CI [0.56, 1.01], $z = 6.96$, $p < .001$, suggesting that location

memory accuracy was higher for items that were correctly recalled. Importantly, when controlling for item recall, the semantic grid was associated with reduced location memory performance relative to the scrambled grid, Exp 1a (exploratory): $b = -0.26$, 95% CI [-0.36, -0.15], $z = 4.83$, $p < .001$; this result was not significant in Exp 1b (pre-registered): $b = -0.12$, 95% CI [-0.25, 0.01], $z = 1.83$, $p = .067$.

Item Location Substitution Rate

In an exploratory analysis, we examined whether display type influenced how likely participants were to choose a location square containing another (incorrect) item when making an error. A mixed-effects linear regression showed that there was no difference between the semantic compared to scrambled grids, Exp 1a: $b = 0.04$, 95% CI [-0.13, 0.20], $z = 0.43$, $p = .667$; Exp 1b: $b = -0.01$, 95% CI [-0.14, 0.12], $z = 0.15$, $p = .885$.

Subjective Difficulty

We examined subjective difficulty in a series of exploratory analyses. We first conducted a mixed-effects linear regression on perceived task difficulty. As there was a significant effect of block, we report the model with condition and block, as well their interaction, as predictors. Participants reported that the semantic grid was more difficult than the scrambled grid, Exp 1a: $b = 0.09$, 95% CI [0.06, 0.12], $z = 6.23$, $p < .001$, Exp 1b: $b = 0.07$, 95% CI [0.04, 0.10], $z = 4.66$, $p < .001$. They also found the second block to be more difficult than the first, Exp 1a: $b = -0.14$, 95% CI [-0.17, -0.11], $z = 9.76$, $p < .001$, Exp 1b: $b = -0.09$, 95% CI [-0.12, -0.06], $z = 6.10$, $p < .001$. The interaction term was not significant, Exp 1a: $b = -0.22$, 95% CI [-0.47, 0.02], $z = 1.77$, $p = .076$, Exp 1b: $b = -0.09$, 95% CI [-0.30, 0.12], $z = 0.83$, $p = .407$.

We also conducted a linear regression using participants' semantic vs. scrambled performance differences to predict their subjective difficulty differences. We found that their semantic-scrambled performance differences negatively predicted their semantic-scrambled difficulty rating differences, Exp 1a: $b = 2.52$, 95% CI [1.55, 3.50], $t = 5.14$, $p < .001$, Exp 1b: $b = 2.81$, 95% CI [2.02, 3.59], $t = 7.01$, $p < .001$. That is, participants who experienced a larger semantic performance cost also rated the semantic display as more difficult.

When explicitly asked to indicate which of the two displays was easier, participants were evenly split between the scrambled grid and the semantic grid, Exp 1a: 43 vs. 47, Exp 1b: 43 vs. 46 (one participant did not record a response). However, we found that these preferences were associated with differences in semantic-scrambled display performance, Exp 1a: $t(87.96) = 5.40$, $p < .001$, Exp 1b: $t(83.64) = 5.65$, $p < .001$. That is, participants who thought the scrambled display was easier also experienced a larger semantic performance cost (Mean semantic-scrambled performance difference: Exp 1a = -0.20, Exp 1b = -0.18), compared to participants who preferred the semantic display (Mean semantic-scrambled performance difference: Exp 1a = 0.03, Exp 1b = 0.10).

Our results suggested that participants had a bias towards the semantic display, such that experience with a larger semantic performance cost was needed to shift participants' preferences away from the semantic display. To confirm this, we conducted a logistic regression to identify the performance point at which participants were indifferent between the two displays. An unbiased participant who had zero performance difference between the two displays would therefore have a 50% chance of selecting either display. The semantic-scrambled performance difference was significantly predictive of display preference, Exp 1a: $b = 5.67$, 95% CI [3.23, 8.60], $z = 4.17$, $p < .001$; Exp 1b: $b = 5.48$, 95% CI [3.18, 8.24], $z = 4.28$, $p < .001$. Importantly,

the indifference point was associated with a negative semantic-scrambled performance difference (Exp 1a = -0.10, Exp 1b = -0.05), indicated that participants were biased towards selecting the semantic display as easier.

Discussion

Across two experiments, we found that participants performed worse (Experiment 1a) or the same (Experiment 1b) in a location memory task when presented with a display consisting of items from a single category compared to a scrambled display (consisting of items from different categories). Specifically, we found that (1) overall location accuracy either decreased in the semantic display or remained the same; (2) the average distance between the chosen location and the target location was either greater in the semantic display or remained the same. On the other hand, we found a significant benefit of the semantic display on item recall performance. When we statistically controlled for the benefit of item recall on location memory performance, we found evidence that the semantic display was associated with a reduction in location memory relative to the scrambled display; this result was significant in Experiment 1a but not in Experiment 1b. Altogether, these results provide evidence that a semantically related item display facilitated participants' ability to recall the items, but either reduced (Experiment 1a) or did not influence (Experiment 1b) their ability to correctly recall the item locations. That said, the results were less clear in Experiment 1b, and this was true for both item location memory and item memory. This might have reflected lower overall data quality, as suggested by the fact that many more participants in the sample had to be excluded due to self-reported inattention, and this may have inflated Type II error rate.

In an exploratory analysis, we compared item location substitution rate on participants' error trials for the two displays. We were motivated by the possibility that increased interference

in the semantically related item display would manifest in increased confusability between semantically related items—that is, an increased likelihood to confuse item locations with each other in the semantic display when making an error. However, the two displays did not differ in item location substitution rate, suggesting that participants were not more likely to confuse item locations with each other in the semantic display when making an error. If semantic relatedness indeed increased interference between same-category representations (as was clear in E1a), then it did not appear to do so via whole item substitutions/confusions.

Additional notable findings from the exploratory analyses of Experiments 1a and 1b were that (1) participants reported that the semantic display was more difficult than the scrambled display; (2) participants' objective semantic-scrambled performance predicted their subjective reports, such that participants who experienced a larger semantic performance cost also rated the semantic display as more difficult; and (3) in spite of the aforementioned two points, participants actually exhibited a bias *towards* the semantic display, such that experience with a larger semantic performance cost was needed to shift participants' perception that the semantic display was easier. That is, despite firsthand experience that the semantic display was associated with poorer performance, they nonetheless selected it as the easier display in higher proportions than would be expected from their objective performance. This could suggest that participants might have had some initial notion that the related display is easier, which was then modified after their experience performing the actual task.

Experiment 2

In Experiment 1a, our results suggested that there was a cost in location memory associated with the semantically homogenous display. A similar trend was found in Experiment

1b, though not statistically significant. The goal of Experiment 2 was to replicate the results of Experiment 1a/b with an improved experimental manipulation and a larger sample size.

In Experiment 2, we selected 11 categories from the Van Overschelde et al., (2004) norms as stimuli. We selected categories that were high in category potency (a measure of how many items participants could generate to each category label) so that participants would be likely to recognize each item as belonging with the given category. A full list of the items can be found in the Appendix. Another change made in Experiment 2 was to use 10 words that were all from 10 different categories in the scrambled display, rather than 2 words from 5 different categories as in Experiment 1a/b. That is, the scrambled display was a ‘truer’ scrambled display in the sense that none of the words belonged to the same category, thus serving as a stronger manipulation relative to the semantically homogenous display. Furthermore, we added a distractor task between the encoding task and the location memory task so as to eliminate possible contributions from short-term memory. Finally, we pre-registered the item location substitution rate analysis that had been exploratory in Experiment 1a/b.

In terms of subjective difficulty, we also aimed to conceptually replicate the finding that participants might have a bias towards preferring the semantically related display, in spite of objective performance evidence to the contrary. Rather than asking about difficulty, we asked participants which of the two types of displays they would prefer to learn for an upcoming task. We predicted that while their location accuracy performance would be superior in the scrambled display, participants would prefer the semantic display after controlling for objective performance.

Methods

Participants

We pre-registered a sample size of 130 to achieve 80% power based on a power analysis using Superpower (Lakens & Caldwell, 2021) using the estimates obtained from Experiments 1a and 1b ($M_1 = 0.45$, $M_2 = 0.51$, $SD = 0.23$, $r = 0.45$). After exclusions, we analyzed complete data from 129 participants (65 women, 62 men, 1 other, 1 unknown, $M = 40.25$ years, $SD = 13.35$). Participants were recruited from Prolific and were paid 1.90 GBP.

Materials

Stimuli were words from eleven categories adapted from Van Overschelde et al. (2004) and can be found in the Appendix. Each category contained ten possible exemplars. For each participant, one category was randomly chosen to populate the semantic (single category) grid; the other, scrambled (multiple category) grid was populated by randomly selecting one items from each of the remaining ten categories. As in Experiments 1a and 1b, item positions in each grid were pseudo-randomly assigned such that each row and column contained two items (cf. Siegel & Castel, 2018), and the order of the semantic/scrambled grids was again randomized across participants.

Procedure

The main experimental procedure was identical to Experiment 1a and 1b except with the addition of a distractor task between the encoding and location memory tasks in each block. Each participant performed two blocks, with each block containing the encoding task, a distractor task, and the location memory task.

On each two-minute distractor task, participants were given a series of simple arithmetic statements (e.g., $1+1=2$) and had to indicate whether each statement was True or False. Each statement was presented for up to 10 seconds, or until participants made a response.

After participants completed both the semantic and scrambled blocks, they were asked to indicate (1) which of the two types of displays they would prefer to learn for an upcoming task, and (2) what strategies they used to remember the locations of items. Finally, participants were asked to provide their age and gender and complete an attention and effort check questionnaire.

Results

Data from 4 participants were not analyzed according to the exclusion criteria set in the pre-registration; they either did not adequately complete the arithmetic distractor task, and/or self-reported that they were not paying attention or did not give effort during the task, leaving a final sample size of $N = 129$. Data and analysis code are available at <https://osf.io/8xf6a/>. Mixed-effects models were used to examine each variable of interest (e.g., logistic for location accuracy and linear for distance) with display (semantic/scrambled) as a fixed effect; random effects structure was determined via model comparison as previously described in Experiment 1.

Location Memory

In order to examine whether the semantic grid influenced location memory performance, we examined (1) absolute location accuracy, defined by whether participants chose the correct target square or not; and (2) Euclidean distance from the chosen location to the correct location on all trials.

Location Accuracy. A mixed-effects logistic regression revealed that there was a significant main effect of condition, such that location memory accuracy was lower in the

semantic grid than the scrambled grid, $b = -0.12$, 95% CI $[-0.23, -0.01]$, $z = 2.15$, $p = .031$.

Figure 4 shows the effect of display type on location accuracy.

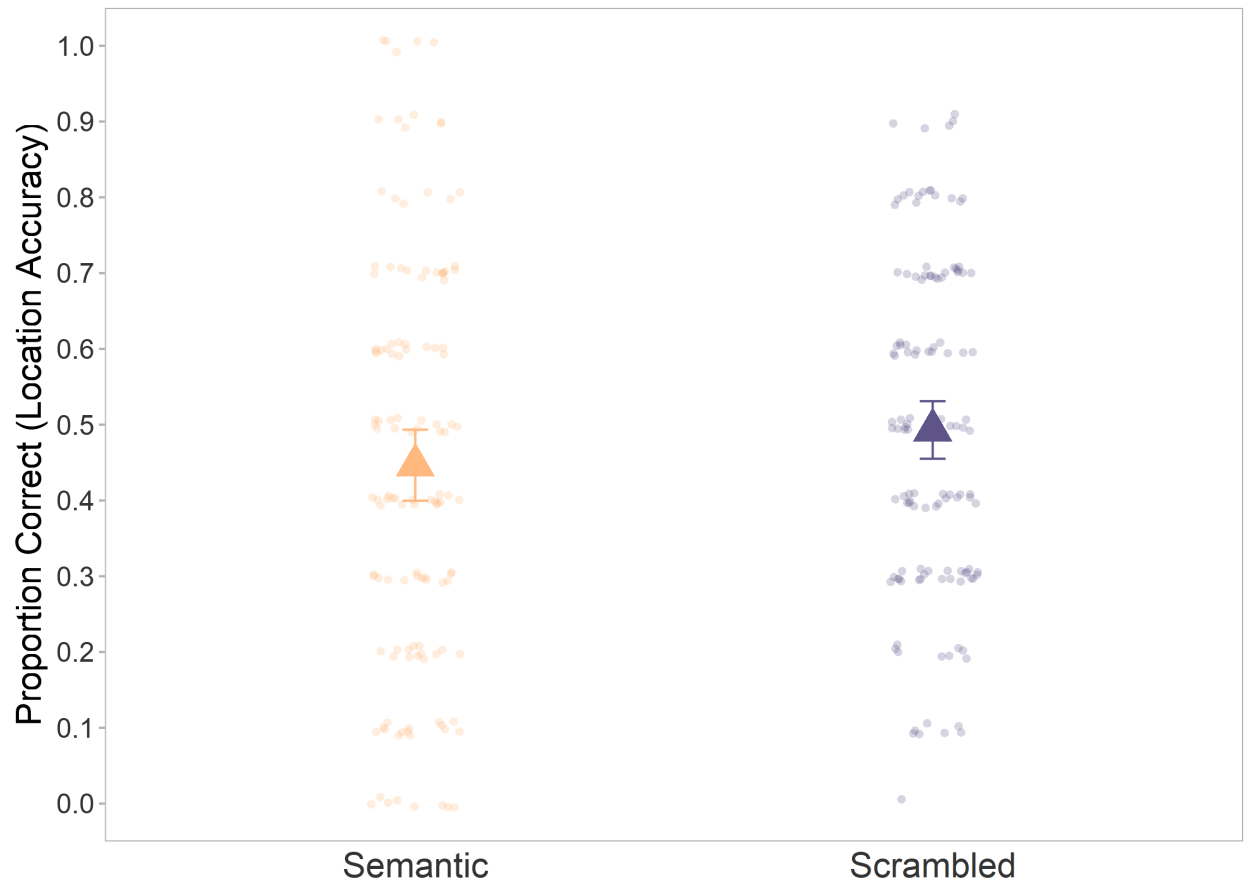


Figure 4. Mean location accuracy by semantic/scrambled display in Experiment 2. Error bars represent 95% confidence intervals. Circles in the background represent each participant's average performance per condition.

Euclidean Distance. A mixed-effects linear regression revealed no significant main effect of condition, $b = 0.05$, 95% CI $[-0.02, 0.11]$, $t = 1.39$, $p = .166$.

Item Location Substitution Rate

A mixed-effects logistic regression showed that participants were more likely to substitute another item location when making an error in the semantic compared to scrambled grid, $b = 0.11$, 95% CI [0.01, 0.22], $z = 1.98$, $p = .048$.

Participant Choice

When explicitly asked to choose one of the two display types for a future task, more participants preferred the semantic display than the scrambled display (76 vs. 53). This difference was significantly different from chance, $\chi^2(1) = 4.10$, $p = .043$. We found that these two groups of participants (semantic preference group vs. scrambled preference group) were associated with differences in semantic-scrambled display performance, $t(121.83) = 7.12$, $p < .001$. That is, participants who preferred the scrambled display had experienced a larger semantic performance cost (Mean semantic-scrambled performance difference = -0.21), compared to participants who preferred the semantic display (Mean semantic-scrambled performance difference = 0.07). We conducted a logistic regression to identify the performance point of indifference between the two displays. The semantic-scrambled performance difference was significantly predictive of display preference, $b = 5.39$, 95% CI [3.50, 7.61], $z = 5.17$, $p < .001$. The indifference point was predicted by a semantic-scrambled performance difference of -0.14, suggesting that participants by and large preferred the semantic display, but experience with a large semantic cost was able to shift this preference.

Discussion

In Experiment 2, we replicated the finding that participants performed worse in a location memory task when presented with a display consisting of items from a single category compared to a scrambled display consisting of items from different categories. Specifically, we found that

overall location accuracy was significantly decreased in the semantic display; the average distance between the chosen location and the target location did not differ between the two displays. Together with Experiment 1a/b, these results provide evidence that a semantically related item display reduced the ability to correctly recall the item locations. These results are consistent with the notion that semantic relatedness between items on a list renders them more similar to each other and less distinct, thereby increasing interference at retrieval. While we did find that participants were more likely to substitute another item location when making an error in the semantic compared to scrambled grid, this effect was not robust across experiments.

In terms of participants' preferences, we found that more participants chose the semantic display than the scrambled display when explicitly asked to choose one of the two display types for a future task. This finding is surprising given that participants performed significantly *worse* in this display. While participants by and large preferred the semantic display, experience with a large semantic cost was apparently able to shift this preference, as participants who selected the scrambled display tended to be those who had experienced a larger semantic performance cost. Our results suggest that participants might have a bias towards the semantic display, and their preferences were only partially rooted in objective performance considerations.

Experiment 3

In Experiment 2, we replicated the finding that participants performed worse in a location memory task when presented with a display consisting of items from a single category compared to a scrambled display consisting of items from different categories. Together with Experiment 1a/b, these results provide evidence that a semantically related item display reduced the ability to correctly recall the item locations. However, the evidence for a whole item confusion-based account was not robust: while we did find a significant effect in Experiment 2, such that

participants were more likely to substitute the location of one item for another in the semantic display, the size of the effect was small, and this effect was not found in Experiment 1a and 1b.

The goal of Experiment 3 was to investigate an alternative hypothesis for the cost of semantic relatedness, that participants may have put in less effort into studying the item locations in the semantically homogenous display as they believed this type of display to be easier. Previous studies have shown that participants harbor beliefs that related items are easier to remember; for example, when given a description of a hypothetical memory experiment, participants predict that they will remember more related word pairs (e.g., cow-milk) compared to unrelated word pairs (e.g., fish-pen; Mueller et al., 2013). Given a self-paced study task, participants also tend to spend less time studying related word pairs compared to unrelated word pairs, though they still demonstrate better memory for the related than unrelated pairs (e.g., Castel et al., 2007; Mueller et al., 2016). In Experiment 1a and 1b, we found that participants exhibited a bias towards the semantic display: despite having just experienced that this display was associated with poorer performance, they nonetheless selected it as the easier display in higher proportions than would be expected from their objective performance. In Experiment 2, we also found that more participants selected the semantic display than the scrambled display when explicitly asked to choose one of the two display types for a future task, again, despite firsthand experience that the semantic display was associated with poorer performance. These results could suggest that participants might have had some initial notion that the semantically related items display was easier to learn, which could have led them to put less effort into studying these items, and/or putting in more effort into studying the scrambled items to compensate. To investigate this “metacognitive loafing” hypothesis, Experiment 3 used a self-paced version of the location memory learning task, where participants were allowed to freely

vary the amount of time they spent studying the location of each item as it was presented. We predicted that if the metacognitive loafing hypothesis was correct, then participants would spend less time studying the items in the semantic display compared to the scrambled display, and that the cost to location memory associated with the semantic display would be mediated by this reduced study time.

Methods

Participants

We pre-registered a sample size of 220 to achieve 80% power based on a power analysis using Superpower (Lakens & Caldwell, 2021) using the estimates obtained from Experiment 2 ($M_1 = 0.44$, $M_2 = 0.49$, $SD = 0.24$, $r = 0.42$). After exclusions, we analyzed complete data from 220 participants (88 women, 121 men, 3 other, 8 unknown, $M = 38.05$ years, $SD = 12.74$). Participants were recruited from Prolific and were paid 1.90 GBP.

Materials

Stimuli were words from eleven categories of household items and can be found in the Appendix. Each category contained ten possible exemplars. For each participant, one category was randomly chosen to populate the semantic (single category) grid; the other, scrambled (multiple category) grid was populated by randomly selecting one item from each of the remaining ten categories. As in previous experiments, item positions in each grid were pseudo-randomly assigned such that each row and column contained two items (cf. Siegel & Castel, 2018), and the order of the semantic/scrambled grids was again randomized across participants.

Procedure

The main experimental procedure was identical to Experiment 2 except that the encoding task was self-paced. On each encoding trial, participants were presented with a single visible item in a grid. Participants were told that they were to remember each item's location as it was presented, and that they could control how much time they spent study each item. Participants were instructed to click on a button labeled 'Continue' or anywhere on the grid to advance to the next trial when they felt they had completed studying an item. There were 10 trials for a single grid (one for each item, presented in random order) and the intertrial interval was 400 ms.

After completing both the semantic and scrambled blocks, participants were asked to provide their age and gender and complete an attention and effort check questionnaire.

Results

Data from 8 participants were not analyzed according to the exclusion criteria set in the pre-registration; they either did not adequately complete the arithmetic distractor task, and/or self-reported that they were not paying attention or did not give effort during the task. Due to an error in condition assignment, an additional 35 participants that were collected in excess of the stopping rule were excluded, leaving the final pre-registered sample size of $N = 220$. Data and analysis code are available at <https://osf.io/8xf6a/>. Mixed-effects models were used to examine each variable of interest (e.g., logistic for location accuracy and linear for distance) with display (semantic/scrambled) as a fixed effect. As outlined in our pre-registration, we compared these models to the models that included the logarithm of study time as a fixed factor. Random effects structures were determined via model comparison.

Study Time

Study Time. A mixed-effects linear regression revealed that there was a significant main effect of condition, such that study times were shorter in the semantic grid than the scrambled grid, $b = -0.04$, 95% CI [-0.07, -0.00], $t = 2.06$, $p = .039$.

Location Memory

Location Accuracy. A mixed-effects logistic regression revealed that there was a significant main effect of condition, such that location memory accuracy was lower in the semantic grid than the scrambled grid, $b = -0.10$, 95% CI [-0.19, -0.01], $z = 2.29$, $p = .022$. When $\log(\text{study time})$ was included in the model, longer study times were associated with increased location accuracy, $b = 0.53$, 95% CI [0.42, 0.64], $z = 9.59$, $p < .001$, while the effect of condition was marginally significant, $b = -0.08$, 95% CI [-0.16, 0.00], $z = 1.93$, $p = .054$. Figure 5A shows the effect of semantic grid on location accuracy.

Euclidean Distance. A mixed-effects linear regression revealed a significant main effect of condition, such that participants chose further away from the target location in the semantic grid than the scrambled grid, $b = 0.05$, 95% CI [0.01, 0.10], $t = 2.24$, $p = .025$. When $\log(\text{study time})$ was included in the model, longer study times were associated with shorter distances from the target, $b = -0.28$, 95% CI [-0.34, -0.23], $z = 9.78$, $p < .001$, and the effect of condition was marginally significant, $b = 0.04$, 95% CI [0.00, 0.09], $z = 1.91$, $p = .056$. Figure 5B shows the effect of semantic grid on Euclidean distance.

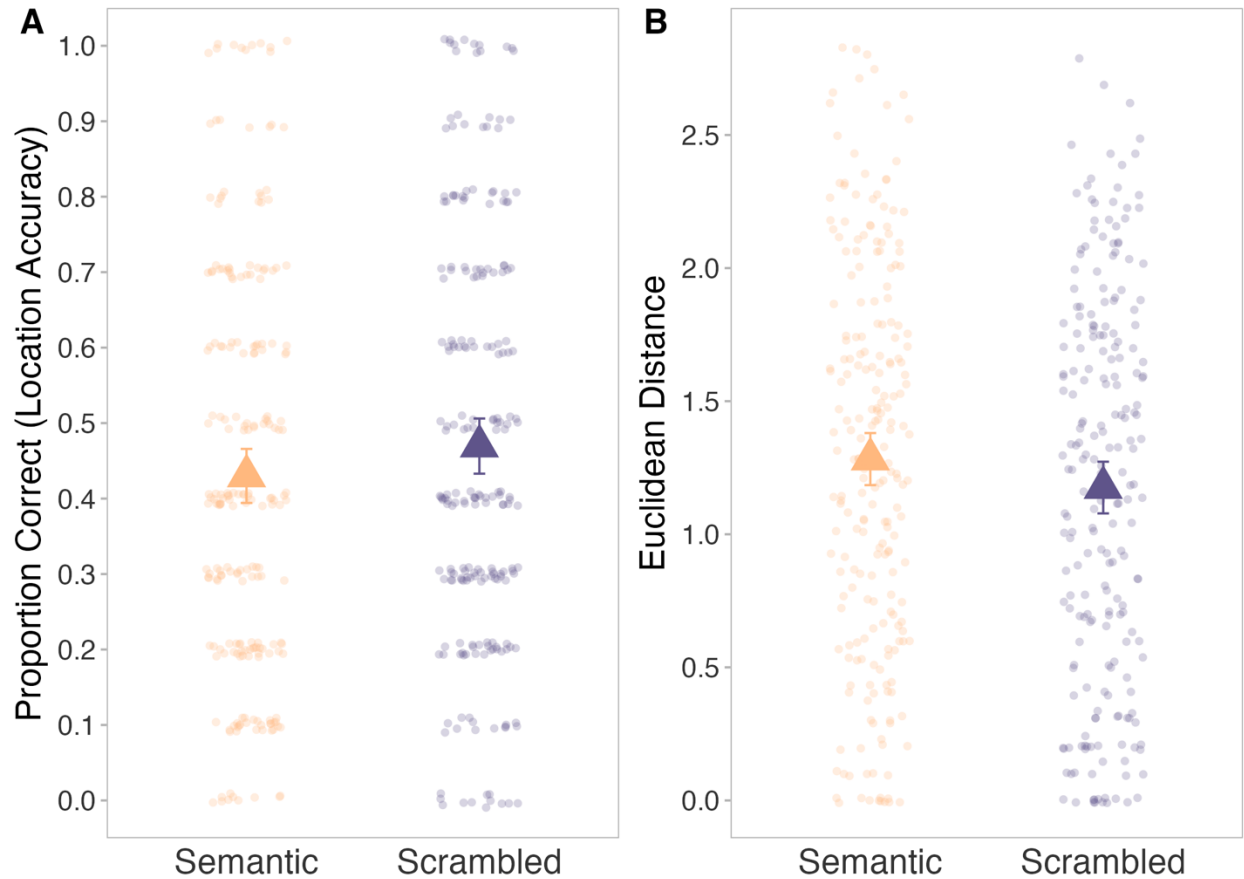


Figure 5. Mean (A) location accuracy, and (B) Euclidean distance by semantic/scrambled display in Experiment 3. Error bars represent 95% confidence intervals. Circles in the background represent each participant's average performance per condition.

Study Time Mediation Analysis

We performed a Bayesian mediation analysis using the `bmlm` package in R (Vuurde, 2016) with default priors (Normal(0, 1000) for regression coefficients, and Cauchy(0,50) for subject-level standard deviations). Display type (relatedness) was set as the predictor variable, $\log(\text{study time})$ as the mediating variable, and location accuracy as the outcome variable. 95% credible intervals were computed from the posterior distribution of the model parameters using

Markov chain Monte Carlo (MCMC) procedures with 10,000 iterations. Mirroring the earlier analyses, the semantic display type was predictive of shorter study times, $b = -0.07$, 95% CI [-0.14, -0.00], and increased study time had a positive effect on location accuracy, $b = 0.26$, 95% CI [0.11, 0.41]. However, the effect of display type on location accuracy was not mediated by study time: the mediated effect was negligible, $b = -0.03$, 95% CI [-0.09, 0.03]. On the other hand, there was a direct effect of display type on location accuracy, with the semantic display type predictive of poorer location accuracy, $b = -0.18$, 95% CI [-0.36, -0.00]. A parallel exploratory mediation analysis using Euclidean distance as the outcome variable found similar results: the semantic display type was predictive of shorter study times, $b = -0.07$, 95% CI [-0.14, -0.00], and longer study times were associated with shorter distances to the target, $b = -0.13$, 95% CI [-0.21, -0.05]. The effect of display type on Euclidean distance was, again, not mediated by study time, $b = 0.01$, 95% CI [-0.02, 0.04], while there was a direct effect of display type on Euclidean distance, with the semantic display type leading to participants choosing further distances from the target, $b = 0.10$, 95% CI [0.00, 0.19]. Figure 6 shows the estimated regression coefficients for the mediation models.

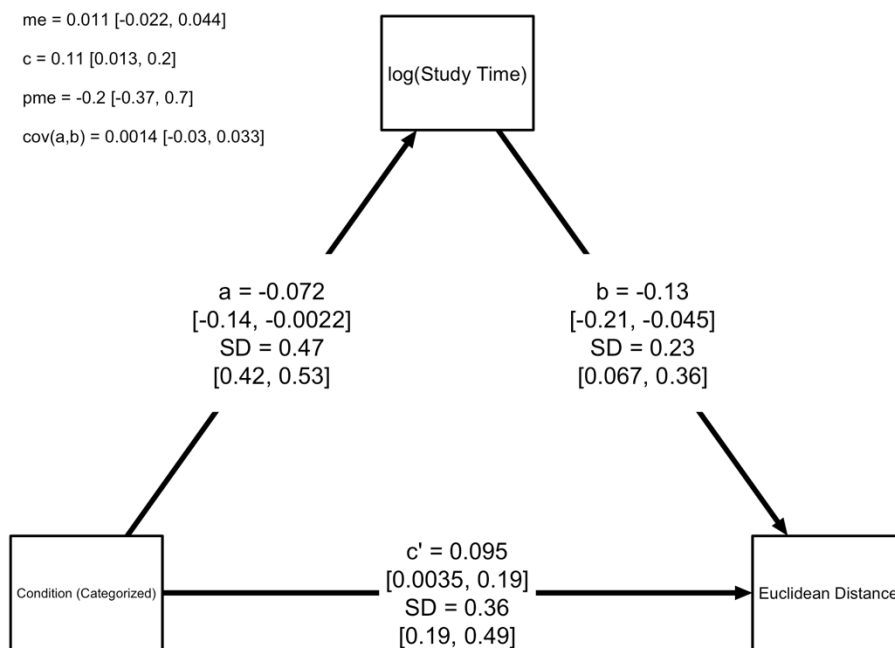
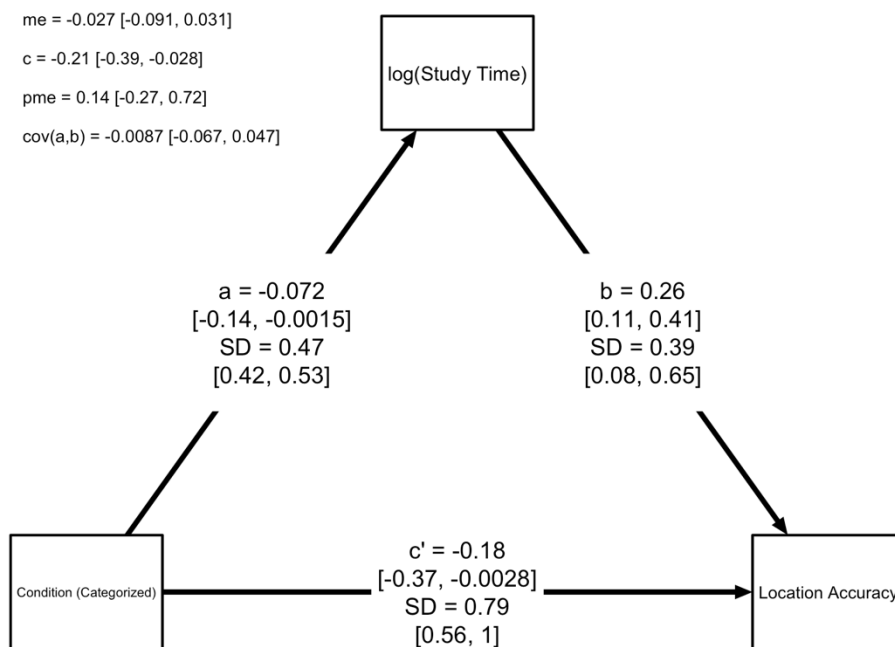


Figure 6. Estimated standardized regression coefficients for the relationship between condition and location accuracy as mediated by study time (top) and the relationship between condition

and Euclidean distance as mediated by study time (bottom). me = mediated effect, c = total effect, c' = direct effect, pme = proportion mediated effect.

Item Location Substitution Rate

An exploratory mixed-effects logistic regression showed that participants were not more likely to substitute another item location when making an error in the semantic compared to scrambled grid, $b = 0.02$, 95% CI [-0.06, 0.10], $z = 0.47$, $p = .635$.

Discussion

In Experiment 3, we replicated the finding that participants performed worse in a location memory task when presented with a display consisting of items from a single category compared to a scrambled display consisting of items from different categories. Specifically, we found that overall location accuracy significantly decreased in the semantic display, and the average distance between the chosen location and the target location was greater in the semantic display. While participants allocated more time to studying the items from different categories compared to the items from a single category, this did not account for the cost of relatedness on location memory performance. The effect of display type (item relatedness) on location accuracy was not mediated by study time, and nor was its effect on Euclidean distance. Together with the earlier experiments, these results provide evidence that semantically related items lead to a decrease in participants' ability to correctly recall item locations. However, we did not find that participants were more likely to substitute another item location when making an error in the semantic compared to scrambled grid, suggesting that the semantic display did not increase the likelihood that participants confused whole item locations with each other.

General Discussion

Across four experiments, we demonstrated that participants tended to perform worse in a location memory task when presented with a display consisting of items from a single category compared to a scrambled display consisting of items from different categories. In Experiment 1a, while semantic relatedness increased item recall performance, it decreased location memory accuracy; in Experiment 1b, the latter result was numerically in the same direction but was not significant. In Experiment 2, using a stronger condition manipulation and a different set of item categories, we replicated the finding that semantic relatedness significantly decreased location memory accuracy. In Experiment 3, we replicated the cost of semantic relatedness in location memory accuracy as well as in Euclidean distance with a self-paced study task and a different set of item categories. While we found that semantic relatedness reduced study time, we did not find that cost of semantic relatedness on location memory was mediated by this reduced study time. Taken together, our results demonstrate a robust cost of semantic relatedness to item location memory, in contrast to its benefit to item recall memory. Thus, our results are an example of the “double-edged” sword of semantic relatedness in memory: depending on the situation, semantic relatedness can lead to benefits and/or costs, both within and across different tasks (Kahana et al., 2022; Nelson et al., 2013).

We have considered two possible explanations for the cost of semantic relatedness to location memory. According to the semantic interference hypothesis, while increased semantic similarity across items (generalization) increases the likelihood of outputting an item in free recall, it also increases interference across these items, reducing performance when discrimination among the to-be-remembered items is required. Unlike a free recall task, correct performance in a location memory task is contingent upon successful discrimination between items. A strong version of this argument is that semantic similarity-based interference increases

the likelihood that another item is erroneously recovered and recalled in the position of the targeted one (Tse, 2009). We investigated the confusion-based hypothesis by examining the item location substitution rate (the likelihood of substituting a different item's location for the target location) for the semantic and scrambled displays. However, across four experiments, we found little evidence that participants were more likely to substitute another item location when making an error in the semantic compared to scrambled grid (nonsignificant Experiment 1a, 1b and 3, and a small effect in Experiment 2). Overall, our results would suggest that the semantic display did not appear to increase the likelihood that whole item locations would be confused with each other. However, the results do not entirely preclude a role for interference: for example, increased semantic-based competition at output might reduce the likelihood that any one item's location is recalled, which could increase location errors but not item substitution errors specifically. Future research should further examine the role of semantic similarity-based interference and its potential role in the cost to location memory.

A second hypothesis we considered was a metacognitive loafing hypothesis: that the cost might be driven by participants spending less time and/or effort studying the items in the semantic display compared to the scrambled display. In Experiment 3, we investigated this possibility by allowing participants to self-pace their study time during the location memory task. While we did find that the semantic display was associated with decreased study time, the semantic cost to location memory was not mediated by study time, and we continued to observe a direct effect of display type on location memory. Our results cannot entirely rule out the possibility that metacognitive loafing might play a role in the cost of semantic relatedness to location memory: self-paced study time may not fully capture participants' experiences of subjective effort. However, to the extent that participants did spend less time studying the items

in the semantic display, we did not find that the reduced study time was able to explain the cost to location memory observed in that condition. Thus, across four experiments we have established the existence of a cost to location memory in semantically homogenous displays, as well as tested two potential mechanisms, seemingly ruling out strong versions of both as compelling explanations for that cost. We anticipate that future explorations into the mechanisms behind the cost will prove to be productive avenues for further research.

Another contribution of the current work is with respect to participants' subjective preferences and ratings. In Experiments 1a and 1b, participants rated the semantic display as more difficult than the scrambled display. These judgments likely reflected the objective location memory performance differences between the two conditions. While semantically related word lists are usually judged to be easier to learn than unrelated word lists in standard list learning tasks (Hourihan & Tullis, 2015; Matvey et al., 2006), objective performance also tends to be better with the related lists, so participants' subjective judgments do not conflict with objective performance. In the current experiments, however, the semantically related set of words was associated with poorer performance. Participants were indeed sensitive to their performance to some degree: in Experiment 1a/1b, they rated the semantic display as more difficult than the scrambled display, and their subjective ratings of difficulty were predicted by their objective performance. However, participants also judged the semantic display to be the easier display in higher proportions than would be expected from their objective performance. In Experiment 2, after controlling for objective performance, we found that participants tended to prefer the semantic display to the scrambled display when explicitly asked to choose one of the two display types for a future task. Thus, participants appear to have a metacognitive bias towards the semantic display—that is, they perceive the semantically related item display to be easier

(Experiment 1a/b) or indicate it as preferred (Experiment 2) more than they objectively should, based on performance considerations. Our results suggest that people may have some sort of a pre-existing notion that semantically related items are easier to learn in general, though experience with a large semantic cost was able to shift participants' preferences to some degree. Consistent with the idea that participants exhibit a semantic preference/bias, in Experiment 3, we found that participants spent less time studying the item locations in the semantic display compared to the scrambled display, suggesting that they may anticipate these items to be easier to learn. This is remarkable given that we avoided soliciting participants' preferences in this experiment to avoid leading/biasing them in one direction or another. We anticipate that relatedness effects on location memory will prove a fruitful avenue to examine potential metacognitive biases.

Finally, what implications might the current results have for location memory in the wild? In real-life environments, we often place items according to functional and/or semantic considerations (e.g., tools are all kept together in the shed, while clothes are kept in the closet). While this type of placement might benefit our ability to recall the items themselves, our ability to recall specific item-location associations appears to be rendered worse. While the current studies used carefully controlled word stimuli in a sparse grid, future investigations may investigate whether the results generalize across more ecologically valid contexts (e.g., using real objects in physical space, or virtual items in virtual reality space).

Conclusion

While semantic relatedness tends to benefit memory, it may also lead to memory costs in certain contexts. Here, we have shown that the semantic relatedness of words within a display had the effect of increasing item recall performance but decreasing location memory accuracy.

Our results are consistent with semantic relatedness serving as a “double-edged sword” in memory. Relatedness improves memory recall by increasing the likelihood of target retrieval, but worsens location memory performance which requires distinguishing between similar target representations.

Acknowledgements

We would like to acknowledge and thank Jocelyn Cheng and Aisling Sampson for their assistance in stimuli preparation and data coding.

Funding

This work was supported by a Discovery Grant (#04091) from the Natural Sciences and Engineering Research Council of Canada (NSERC), an Early Researcher Award from the Province of Ontario (#ER14-10-258), funding from the Canada Foundation for Innovation and Ontario Research Fund (#37872) and from the Canada Research Chairs (#950-232147) program to E.F.R, and Alexander Graham Bell Canada Graduate Scholarships from the NSERC to X.L. and M.J.H.Z.

Ethics declarations

Conflict of interest: The authors declare that they have no conflict of interest.

Ethics approval: The research protocols in this study were approved by the University of Waterloo Research Ethics Board (#42214).

Consent to participate: All participants gave written informed consent.

Open practices statement: The data and R scripts for these experiments have been made available on a third-party archive (<https://osf.io/8xf6a/>).

References

- Baddeley, A. D. (1966a). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, 18(4), 362-365.
- Baddeley, A. D. (1966b). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, 18(4), 302-309.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious Mixed Models. *ArXiv:1506.04967 [Stat]*. <http://arxiv.org/abs/1506.04967>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
<https://doi.org/10.18637/jss.v067.i01>
- Bousfield, W. A., Sedgewick, C. H. W., & Cohen, B. H. (1954). Certain Temporal Characteristics of the Recall of Verbal Associates. *The American Journal of Psychology*, 67(1), 111–118. <https://doi.org/10.2307/1418075>
- Castel, A. D., McCabe, D. P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, 14, 107-111.
- 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. <https://doi.org/10.3758/BF03193146>

- Gardiner, J. M., Craik, F. I., & Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 778-783.
- Guérard, K., & Saint-Aubin, J. (2012). Assessing the effect of lexical variables in backward recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 312.
- Guerin, S. A., & Miller, M. B. (2008). Semantic organization of study materials has opposite effects on recognition and recall. *Psychonomic Bulletin & Review*, *15*(2), 302–308.
<https://doi.org/10.3758/PBR.15.2.302>
- Hirtle, S. C., & Mascolo, M. F. (1986). Effect of semantic clustering on the memory of spatial locations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 182–189. <https://doi.org/10.1037/0278-7393.12.2.182>
- Hourihan, K. L., & Tullis, J. G. (2015). When will bigger be (recalled) better? The influence of category size on JOLs depends on test format. *Memory & Cognition*, *43*(6), 910–921.
<https://doi.org/10.3758/s13421-015-0516-4>
- Howard, M. W., & Kahana, M. J. (2002). When Does Semantic Similarity Help Episodic Retrieval? *Journal of Memory and Language*, *46*(1), 85–98.
<https://doi.org/10.1006/jmla.2001.2798>
- Hund, A. M., & Plumert, J. M. (2003). Does Information About What Things Are Influence Children’s Memory for Where Things Are? *Developmental Psychology*, *39*(6), 939–948.
<https://doi.org/10.1037/0012-1649.39.6.939>
- Hunt, R. R. (2013). Precision in Memory Through Distinctive Processing. *Current Directions in Psychological Science*, *22*(1), 10–15. <https://doi.org/10.1177/0963721412463228>

- Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, 20(5), 497–514.
[https://doi.org/10.1016/S0022-5371\(81\)90138-9](https://doi.org/10.1016/S0022-5371(81)90138-9)
- Hunt, R. R., & McDaniel, M. A. (1993). The Enigma of Organization and Distinctiveness. *Journal of Memory and Language*, 32(4), 421–445.
<https://doi.org/10.1006/jmla.1993.1023>
- Ishiguro, S., & Saito, S. (2021). The detrimental effect of semantic similarity in short-term memory tasks: A meta-regression approach. *Psychonomic Bulletin & Review*, 28(2), 384–408.
- Kahana, M. J. (2020). Computational Models of Memory Search. *Annual Review of Psychology*, 71(1), 107–138. <https://doi.org/10.1146/annurev-psych-010418-103358>
- Kahana, M. J., Diamond, N. B., & Aka, A. (2022). *Laws of Human Memory* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/aczu9>
- Kowialiewski, B., Krasnoff, J., Mizrak, E., & Oberauer, K. (2023). Verbal working memory encodes phonological and semantic information differently. *Cognition*, 233, 105364.
- Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM Model of False Recall. *Psychological Review*, 114(4), 954–993. <https://doi.org/10.1037/0033-295X.114.4.954>
- Lakens, D., & Caldwell, A. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, 4(1), 251524592095150. <https://doi.org/10.1177/2515245920951503>
- Lewis, M. Q. (1971). Categorized lists and cued recall. *Journal of Experimental Psychology*, 87(1), 129–131. <https://doi.org/10.1037/h0030179>

- Lu, X., Kelly, M. O., & Risko, E. F. (2022). The gist of it: Offloading memory does not reduce the benefit of list categorisation. *Memory*, *30*(4), 396–411.
<https://doi.org/10.1080/09658211.2021.1989465>
- Lu, X., Zhu, M. J. H., & Risko, E. F. (Under Review). Semantic partitioning facilitates memory for object location through category-partition cueing.
- Mandler, G. (1967). Organization and Memory. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (Vol. 1, pp. 327–372). Academic Press.
[https://doi.org/10.1016/S0079-7421\(08\)60516-2](https://doi.org/10.1016/S0079-7421(08)60516-2)
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.
<https://doi.org/10.1016/j.jml.2017.01.001>
- Matvey, G., Dunlosky, J., & Schwartz, B. (2006). The effects of categorical relatedness on judgements of learning (JOLs). *Memory*, *14*(2), 253–261.
<https://doi.org/10.1080/09658210500216844>
- Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, *20*, 378-384.
- Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about memory?. *Quarterly Journal of Experimental Psychology*, *69*(4), 781-799.
- Neale, K., & Tehan, G. (2007). Age and redintegration in immediate memory and their relationship to task difficulty. *Memory & Cognition*, *35*, 1940-1953.

- Neath, I., Saint-Aubin, J., & Surprenant, A. M. (2022). Semantic relatedness effects in serial recall but not in serial reconstruction of order. *Experimental Psychology*, *69*(4), 196.
- Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., & Bruza, P. D. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory & Cognition*, *41*(6), 797–819. <https://doi.org/10.3758/s13421-013-0312-y>
- Poirier, M., & Saint-Aubin, J. (1995). Memory for Related and Unrelated Words: Further Evidence on the Influence of Semantic Factors in Immediate Serial Recall. *The Quarterly Journal of Experimental Psychology Section A*, *48*(2), 384–404.
<https://doi.org/10.1080/14640749508401396>
- Postma, A., & De Haan, E. H. F. (1996). What Was Where? Memory for Object Locations. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 178–199.
<https://doi.org/10.1080/713755605>
- Postma, A., Kessels, R. P. C., & van Asselen, M. (2008). How the brain remembers and forgets where things are: The neurocognition of object–location memory. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1339–1345.
<https://doi.org/10.1016/j.neubiorev.2008.05.001>
- Puff, C. R. (1970). Role of clustering in free recall. *Journal of Experimental Psychology*, *86*(3), 384–386. <https://doi.org/10.1037/h0030189>
- Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*(1–2), 103–121.
<https://doi.org/10.1016/j.specom.2004.02.004>

- R Core Team. (2019). *R: A language and environment for statistical computing* (3.6.0) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. <https://doi.org/10.1037/0033-295X.88.2.93>
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. <https://doi.org/10.1037/0278-7393.21.4.803>
- Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting Clustering From Semantic Structure. *Psychological Science*, *4*(1), 28–34. <https://doi.org/10.1111/j.1467-9280.1993.tb00552.x>
- Saint-Aubin, J., Ouellette, D., & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials. *Psychonomic Bulletin & Review*, *12*(1), 171–177. <https://doi.org/10.3758/BF03196364>
- Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information?. *The Quarterly Journal of Experimental Psychology: Section A*, *52*(2), 367-394.
- Siegel, A. L. M., & Castel, A. D. (2018). Memory for Important Item-Location Associations in Younger and Older Adults. *Psychology and Aging*, *33*(1), 30–45. <https://doi.org/10.1037/pag0000209>
- Surprenant, A. M., Neath, I., & Brown, G. D. A. (2006). Modeling age-related differences in immediate memory using SIMPLE. *Journal of Memory and Language*, *55*(4), 572–586. <https://doi.org/10.1016/j.jml.2006.08.001>

- Tomparry, A., & Thompson-Schill, S. L. (2021). Semantic influences on episodic memory distortions. *Journal of Experimental Psychology. General*, *150*(9), 1800–1824.
<https://doi.org/10.1037/xge0001017>
- Tse, C.-S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory*, *17*(8), 874–891.
<https://doi.org/10.1080/09658210903376250>
- Tse, C.-S., Li, Y., & Altarriba, J. (2011). The effect of semantic relatedness on immediate serial recall and serial recognition. *Quarterly Journal of Experimental Psychology*, *64*(12), 2425–2437. <https://doi.org/10.1080/17470218.2011.604787>
- Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 381–391.
[https://doi.org/10.1016/S0022-5371\(66\)80048-8](https://doi.org/10.1016/S0022-5371(66)80048-8)
- Underwood, B. J. (1957). Interference and forgetting. *Psychological review*, *64*(1), 49.
- Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*, *43*, 647-658.
- 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. <https://doi.org/10.1016/j.jml.2003.10.003>
- Vuorre, M. (2016). *bmlm: Bayesian Multilevel Mediation* (1.3.12) [Computer software].
<https://cran.r-project.org/package=bmlm>

Appendix

Word lists used as stimuli

Experiments 1a and 1b

wearable	tools	toiletries	office	kitchen	instruments
sneakers	cutters	hairbrush	eraser	spatula	saxophone
necklace	pliers	deodorant	binder	muffin	banjo
sweater	screwdriver	toothpaste	clipboard	kettle	clarinet
sunglasses	scissors	lotion	ruler	sponge	harp
socks	wrench	floss	pencil	mug	violin
gloves	axe	comb	calculator	toaster	flute
jeans	drill	razor	notebook	spoon	trumpet
shorts	hammer	towel	envelope	tray	drum
shirt	pocketknife	mouthwash	folder	whisk	piano
belt	bolt	soap	printer	pan	bass

Experiment 2

animal	building	carpentry	clothing
bear	apartment	chisel	hat
cat	cabin	drill	jacket
cow	condo	hammer	pants
deer	dorm	nail	shirt
dog	house	ruler	shoes
elephant	hut	sander	shorts
horse	mansion	saw	skirt
lion	shack	screw	socks
pig	tent	screwdriver	sweater
tiger	trailer	wrench	underwear

fruit	reading	furniture	kitchen
apple	article	bed	bowl
banana	book	chair	fork
grape	journal	couch	knife
kiwi	letter	desk	ladle
orange	magazine	dresser	pan
peach	newspaper	lamp	plate
pear	novel	loveseat	pot
pineapple	pamphlet	sofa	spatula
strawberry	textbook	stool	spoon
watermelon	website	table	whisk

time	flavoring	relative
century	butter	aunt
day	garlic	brother
decade	ketchup	cousin
hour	mustard	father
millisecond	onions	grandfather
minute	pepper	grandmother
month	salt	mother
second	spices	niece
week	sugar	sister
year	vanilla	uncle

Experiment 3

wearables	tools	toiletries	office
sneakers	cutters	hairbrush	eraser
necklace	pliers	deodorant	binder
sweater	screwdriver	toothpaste	clipboard
sunglasses	scissors	lotion	ruler
socks	wrench	floss	pencil
gloves	axe	comb	calculator
jeans	drill	razor	notebook
shorts	hammer	towel	envelope
shirt	pocketknife	mouthwash	folder
belt	bolt	soap	printer

kitchen	instruments	toys	sports
spatula	saxophone	chessboard	basketball
plate	banjo	dice	bicycle
kettle	clarinet	doll	frisbee
tongs	harp	kite	helmet
mug	violin	puzzle	jumprope
toaster	flute	scrabble	shuttlecock
spoon	trumpet	slinky	skateboard
tray	drum	uno	skis
whisk	piano	lego	trampoline
pan	bass	checkers	volleyball

furniture	pantry	cleaning
table	pasta	mop
couch	rice	vacuum
lamp	milk	detergent
nightstand	sugar	rag
shelf	flour	duster
cabinet	bread	disinfectant
sofa	cereal	dustpan
stool	salt	broom
carpet	pepper	bleach
armchair	chips	sponge