Intent matters: Resolving the intentional versus incidental learning paradox in episodic long-term memory

Popov, Vencislav ; Dames, Hannah

Abstract: Decades of research have established that the intent to remember information has no effect on episodic long-term memory. This claim, which is routinely taught in introductory cognitive psychology courses, is based entirely on pure-list between-subjects designs in which memory performance is equal for intentional and incidental learning groups. In the current 11 experiments, participants made semantic judgements about each word in a list but they had to remember only words presented in a specific color. We demonstrate that in such mixed-list designs there is a substantial difference between intentionally and incidentally learned items. The first four experiments show that this finding is independent of the remember cue onset relative to the semantic judgment. The remaining seven experiments test alternative explanations as to why intent only matters in mixed-list designs but not in pure-list designs—inhibition of incidentally learned items, output interference, selective relational encoding, or selective threshold-shifting. We found substantial support for the threshold-shifting account according to which the intent to remember boosts item-context associations in both mixed- and pure-list designs; however, in pure-list between-subjects designs, participants in the incidental learning group can use a lower retrieval threshold to compensate for the weaker memory traces. This led to more extralist intrusions in incidental learning groups; incidental learning groups also showed a source memory deficit. We conclude that intent always matters for long-term learning, but that the effect is masked in traditional between-subjects designs. Our results suggest that researchers need to rethink the role of intent in long-term memory.

DOI: https://doi.org/10.1037/xge0001272

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001272

Intent Matters: Resolving the Intentional Versus Incidental Learning Paradox in Episodic Long-Term Memory

Vencislav Popov and Hannah Dames
Department of Psychology, University of Zurich

Decades of research have established that the intent to remember information has no effect on episodic long-term memory. This claim, which is routinely taught in introductory cognitive psychology courses, is based entirely on pure-list between-subjects designs in which memory performance is equal for intentional and incidental learning groups. In the current 11 experiments, participants made semantic judgements about each word in a list but they had to remember only words presented in a specific color. We demonstrate that in such mixed-list designs there is a substantial difference between intentionally and incidentally learned items. The first four experiments show that this finding is independent of the remember cue onset relative to the semantic judgment. The remaining seven experiments test alternative explanations as to why intent only matters in mixed-list designs but not in pure-list designs—inhibition of incidentally learned items, output interference, selective relational encoding, or selective threshold-shifting. We found substantial support for the threshold-shifting account according to which the intent to remember boosts item-context associations in both mixed- and pure-list designs; however, in pure-list between-subjects designs, participants in the incidental learning group can use a lower retrieval threshold to compensate for the weaker memory traces. This led to more extralist intrusions in incidental learning groups; incidental learning groups also showed a source memory deficit. We conclude that intent always matters for long-term learning, but that the effect is masked in traditional between-subjects designs. Our results suggest that researchers need to rethink the role of intent in long-term memory.

Keywords: deep processing, directed forgetting, episodic memory, incidental memory

> It is abundantly clear that what determines the level of recall or recognition of a word event is not intention to learn [. . .]; rather it is the kind of operations carried out on the items, that determines retention. (Craik & Tulving, 1975)

People go through most of their lives without explicitly intending to remember the events they experience, but despite this they have rich episodic memories. At the same time, learning and retention is often perceived as effortful in both daily life and in academic settings. These seemingly conflicting scenarios beg the question: Does the intention to remember matter for long-term memory (LTM)? Most existing research on intentional versus incidental memory suggests that the answer is no—when people process information in an elaborate and meaningful way, it does not matter whether they intend to learn the information or not. For example, free recall and recognition of words are just as good when people do not expect their memory to be tested later (Craik & Tulving, 1975; Hyde & Jenkins, 1969, 1973; Johnston & Jenkins, 1971; Mandler, 1967; Oberauer & Greve, 2022; Till et al., 1975). This result is obtained as long as people engage in so called "deep processing"—a semantic analysis of the meaning of the memoranda that leads to an elaborate memory trace connecting new information with existing knowledge (Craik & Tulving, 1975). The claim that intent does not matter for encoding in episodic LTM is widely accepted and routinely taught as part of university courses in cognitive psychology (see Appendix A; e.g., Anderson, 2015; Dudai, 2002; Groome, 2014; Hulstijn, 2003; Styles, 2005; Willingham & Riener, 2019). Here we provide evidence that intent does matter for episodic memory after all.

In their seminal article, Craik and Tulving (1975) argued that intent has no causal role in learning. According to them what determines memory performance is the nature of the operations performed during learning—if those operations are equated under intentional and incidental learning instructions, there should be no difference in memory performance. This is what most studies have found by comparing a group of participants that expected a memory test with another group that did not (Craik & Tulving, 1975;

Vencislav Popov https://orcid.org/0000-0002-8073-4199

The experimental software, data, and analysis code for all experiments are freely available at https://github.com/venpopov/intentional-incidental-ltm-paradox. The article is available as a preprint on Psyarxiv: https://psyarxiv.com/jf2en. The results reported here were presented during the 2021 Context and Episodic Memory Symposium—International.

We are grateful to Vanessa Vallesi for implementing Experiment 1 in *lab.js* and to Klaus Oberauer for thoughtful comments on a previous version of this article. This research was supported by a Grant from the Swiss National Science Foundation to Klaus Oberauer (Project 100014_192204) and by a Grant from the German Research Foundation (DFG; Grant RA1934/5-1) to M. Ragni.

Correspondence concerning this article should be addressed to Vencislav Popov, Department of Psychology, University of Zurich, Binzmühlestrasse 14/22, 8050 Zürich, Switzerland. Email: vencislav.popov@gmail.com

Hyde & Jenkins, 1969, 1973; Johnston & Jenkins, 1971; Mandler, 1967; Oberauer & Greve, 2022; Till et al., 1975). An effect of intent only seems to appear in between-subjects designs when there is no deep-processing orienting task (Block, 2009; Naveh-Benjamin et al., 2014).[1] In contrast to LTM, intentional encoding greatly benefits working memory (WM) performance, even when stimuli are processed deeply (Oberauer & Greve, 2022).

This apparent "intent dissociation" between LTM and WM could be explained by the purpose and functional properties of each of those systems. According to Oberauer and Greve (2022), because WM has very limited capacity, it needs a selective encoding policy—to prevent interference. In contrast, LTM does not suffer from such capacity limitations and as a result it encodes all attended information regardless of intention. Oberauer and Greve (2022) further argued that nonselective encoding in LTM is beneficial since one cannot always know in advance which memories will be relevant later.

Although this argument is compelling, LTM is not without its limitations. Even though its total storage capacity might be unlimited, there are limits on how much information can be stored in LTM within any given time. For example, LTM performance declines with each subsequent word presented for study (Murdock, 1962) and when stimuli are presented at a faster pace (Criss & McClelland, 2006; Malmberg & Nelson, 2003; Murdock, 1960). Furthermore, recent investigations into the mixed-list word frequency paradox have suggested that encoding information in LTM depletes a limited encoding resource that recovers gradually over time (Popov & Reder, 2020; Reder et al., 2007; also see Diana & Reder, 2006). Because of these encoding limitations, LTM also needs some selection mechanism that determines which information to prioritize during learning.

It is possible that prior studies have failed to reveal an effect of intent to learn on LTM performance due to their experimental designs. Typically, these studies have used between-subjects designs where all participants have to make some semantic judgment for each word, but only one group of participants is told to expect a subsequent memory test (Craik & Tulving, 1975; Hyde & Jenkins, 1969, 1973; Johnston & Jenkins, 1971; Mandler, 1967; Oberauer & Greve, 2022; Till et al., 1975). Consequently, these studies used a pure-list design and not a design where intentionally and incidentally learned items appear intermixed in the same list (mixed-list).[2] As we argue later in the present article, the selective nature of LTM encoding might be masked in a pure-list design. It is an open question whether an effect of intent to learn may appear in a mixed-list design in which participants are given instructions to remember or not a word on an item-by-item basis.

Two prior studies have investigated whether memory performance is better for intentionally than incidentally learned items in a within-subject design (Abel & Bäuml, 2019; Geiselman et al., 1983). In both studies, participants provided pleasantness ratings for some items on a list (i.e., "Judge" items) and remembered other items on the same list for a later test (i.e., "Learn" items). Even though both studies found better recall for the Learn relative to the Judge items, participants only gave pleasantness ratings for the Judge items, and not for the Learn items. Thus, intentionality instructions were confounded with the type of task. This type of confound was also present in early research with between-subjects designs, and it was criticized at the time by Saltzman (1953) and more recently by Block (2009). Thus, it is still unclear whether the intent to remember improves recall in within-subject designs.

Although it is widely accepted in the intentional versus incidental-learning literature that intent does not matter for episodic LTM, the parallel literature on Directed Forgetting reveals that item-specific memory instructions (remember vs forget this item) have large effects on subsequent recall (Bjork, 1972; Fawcett & Taylor, 2008; MacLeod, 1989; Popov et al., 2019). However, in this paradigm, it is typical to present items without an orienting task—participants read items and attempt to remember the items followed by "remember" cues, while attempting to forget the items followed by "forget" cues. Craik and Tulving (1975) argued that in the absence of a deep processing orienting task, intent boosts memory by inducing deeper processing. Thus, results from Directed Forgetting studies are not directly comparable with the intentional-versus-incidental memory literature for two reasons—they involve instructions to actively forget some items, and they do not typically require deep processing.

Here, we report 11 within-subject experiments that show a large effect of intent to remember in LTM despite deep semantic processing of all stimuli. In contrast to most prior studies (e.g., Craik & Tulving, 1975; Oberauer & Greve, 2022), we instructed participants on an item-by-item basis to remember only some words on a list for a subsequent memory test. Experiments 1 through 4 were exploratory, and they showed that intentional learning instructions lead to substantially better recall relative to an incidental baseline. These results establish the presence of a novel mixed-list paradox: intent helps memory in mixed-lists (within-subject) but not in pure-lists (between-subjects) designs. The remaining seven experiments characterize the boundary conditions of this effect and test a variety of explanations for this paradox. The solution proved to be surprisingly simple, yet theoretically important, because it is in stark contrast to the Craik and Tulving's (1975) framework. We conclude that the intent to remember significantly boosts item-context associations in both pure-lists and mixed-lists. Yet, the effects in pure-list between-subjects designs are masked by changes in retrieval thresholds between the intentional and incidental learning groups.

Overview of Experiments and General Method

The designs of the 11 experiments shared many aspects of the general methods described below. Appendix B provides a summary table of the differences between experiments. The experimental software, data, and analysis code are freely available at https://github.com/venpopov/intentional-incidental-ltm-paradox.

[1] For an extended discussion, see the Section "Prior Evidence for the Effect of Intent on Memory" in the General Discussion.

[2] Between-subject designs use pure-list designs where all list items are either intentionally or incidentally learned. In contrast, mixed-list experiments are operationalized within subjects. However, a pure-list design could also be achieved within subjects (e.g., by having the same participant learn one list incidentally and another list intentionally). Hence, any difference in memory performance for incidentally learned items between designs cannot be fully attributed to a difference between within- and between-subject manipulations but rather to a difference between pure vs. mixed lists. Between-subject and pure-list designs are confounded, so we refer to these as "pure-list between-subject" (and "mixed-list within-subject," respectively). Please note that the terms "pure-list," "list-wise," and "list-method" as well as the terms "mixed-list," "item-wise," and "item-method" are interchangeable here.

Table 1
Sample Sizes and Number of Excluded Participants for Each Experiment

		Exclusions					
Exp.	Initial *N*	Switching tabs	Expect test	Used help	Restarted experiment	Chance performance	Final *N*
1	64	9	3	6	0	0	46
2	58	4	3	5	0	0	46
3	65	7	2	3	0	0	53
4	61	14	4	2	0	0	41
5	184	4	3	0	0	0	177
6	59	3	3	4	0	0	49
7	70	6	4	4	1	6	49
8	40	6	5	3	0	0	26
9	60	5	3	3	0	0	49
10	167	0	12	6	0	0	149
11	171	0	16	2	0	0	153

Note. Initial *N* refers to the number of participants who completed the experiment. Final *N* refers to the number of participants included in the analyses after the exclusion criteria (see main text).

Experiment Administration and Participants

All experiments were programmed in *lab.js* (Henninger et al., 2019) and were administered online through Prolific. We defined the following exclusion criteria prior to running any experiments: (a) switching tabs/windows away from the experiment more than twice; (b) responding "Yes" to a postexperiment question "Did you expect to be tested on all items on the final test, regardless of whether you were supposed to remember them?"; (c) responding "Yes" to a postexperiment question "Did you use any aid (e.g., pencil-and-paper or computer file) to remember the words?"; (d) restarted the experiment; or (e) performed at chance (see Table 1 for the number of participants excluded in each experiment owing to these criteria). We ensured participants that their answers would not affect their compensation. We initially aimed to gather 200 participants for Experiment 1, because Oberauer and Greve (2022) found substantial Bayesian evidence for a null effect of intent in a between-subjects design with 100 participants per condition with twice as many observations per participant relative to our study. However, we stopped recruiting participants when after the first batch of 70 participants (of which we retained 46 participants after the applying our exclusion criteria), instead of a null effect, we found a substantial effect of intention (BF = 4.8×10^7). Based on this initial result, we aimed to gather 50 participants in the subsequent experiments, although the final numbers varied due to the exclusion criteria and failed data transmission for 5% to 10% of participants in each experiment. For these reasons, we recruited 70 participants per experiment, aiming to retain ~50 of them.

All experiments were conducted in accordance with the guidelines issued by the Ethics Committee of the Faculty of Arts and Social Science, University of Zurich. All participants agreed with an Informed Consent statement at the beginning of each experiment.

Material

The first 10 experiments used the same pool of items—180 high-frequency English words for concrete objects from the stimulus pool used by Popov et al. (2019). All words were five to seven letters in length.

Study Procedure

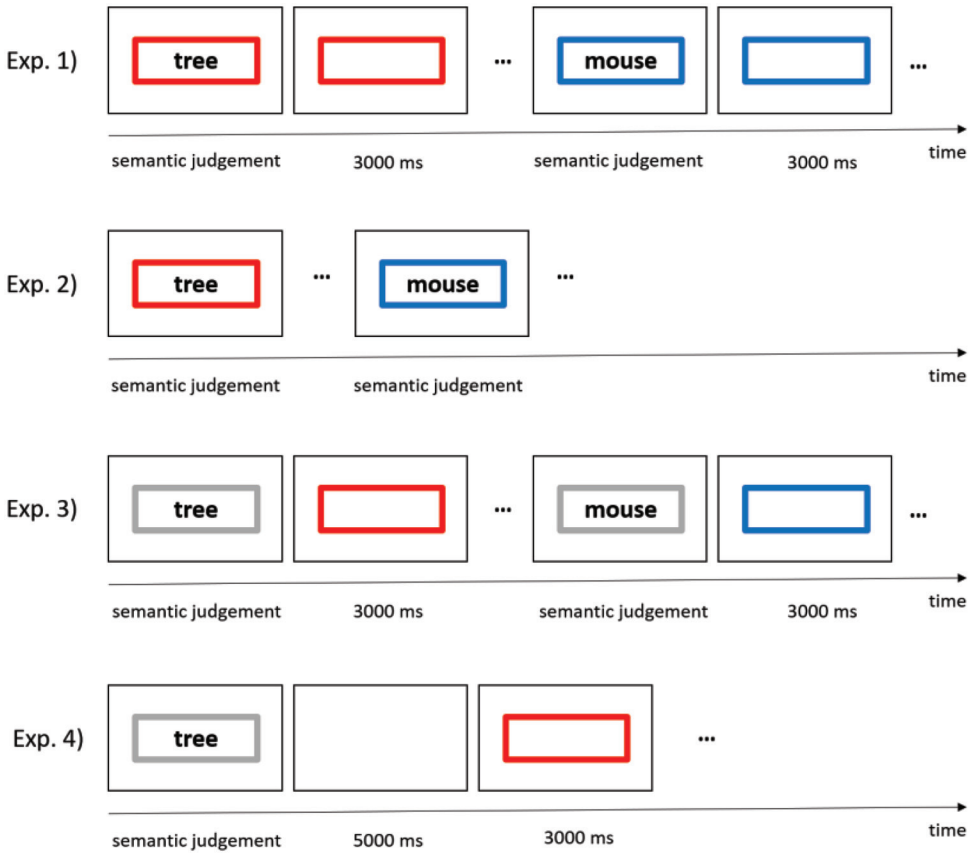
In Experiments 1–9, participants studied one to three lists of words or word pairs, presented one at a time in the middle of the screen. In the final two experiments, words were either presented at different locations on the screen or together with different real-world scenes. In Experiments 1–10, for each list and each participant, the presented words were randomly selected without replacement from the pool of 180 items. On each trial, participants performed an orienting task in which they had to judge the real-world size of the objects. The size-judgment task required (a) judging whether each item is larger or smaller than a soccer ball or (b) judging which of two items is larger than the other. The size-judgment task was self-paced with no time limit and words disappeared immediately after the participant's response. This task has been previously used in between-subjects designs, and it was sufficient to induce deep semantic processing and to prevent differences in memory between incidental and intentional learning conditions (e.g., Oberauer & Greve, 2022). In Experiment 11, instead of size-judgements, people had to estimate how often they might encounter each object in a particular real-world location.

In nine of the 11 experiments, half of the items in each list had to be remembered ("Remember" items), whereas the other half only had to be processed ("Process-only" items). The instruction for each item was cued by a change in the border color around the word—blue or red. The assignment of memory instructions ("Remember" or "Process-only") to colors was counterbalanced across participants in each experiment. The two types of items were randomly intermixed on each list (mixed-list design). The timing of the instructions for each item was varied across experiments—simultaneous with word onset, immediately after size-judgment, or 5 seconds after size-judgment (see Figure 1). One experiment used a between-subjects, pure-list manipulation, replicating Oberauer and Greve's (2022) LTM condition.

Distractor Task

Each study list was followed by a distractor task in which participants had to solve as many arithmetic equations as they could in 1 minute. These equations were of the form $(x + y)/z = ?$ and were drawn from the equations used by Oberauer and Greve

Figure 1
Illustration of the Study Procedure in Experiments 1–4



Note. Participants had to make a self-paced semantic judgment to each presented word (30 words/word pairs per list). Additionally, they were asked to remember only the words presented in red or in blue (counterbalanced) for the subsequent memory test. The only difference among the four experiments was the timing of the color cue. See the online article for the color version of this figure.

(2022). The *x*, *y*, and *z* variables were one- or two-digit numbers, the operation in the parentheses was either addition or subtraction, and the second operation was either multiplication or division. The answer was always an integer that participants had to type.

Test Procedure

Immediately after the distraction task that followed each list, participants' memory for the list was tested via free recall, forced-choice recognition, cued-recall, or source-memory tests, depending on the experiment. All lists but the last one tested memory only for the Remember items but not for the Process-only items. The last list always tested memory for all items, regardless of the memory instructions.

Size-Ratings Validation

Because the conclusions of these studies depend crucially on whether participants performed the orienting task diligently, it was important to establish the accuracy of those ratings. We did not have existing ratings for the size of the objects prior to the experiments reported here. Instead, we used the ratings data from the first ten experiments to calculate the relative size of all objects, and then used those scores to determine whether each individual response was correct or not (see Appendix C).

Data Analysis

We used Bayesian statistics for the data analyses. Bayes Factors (*BF*s) enabled us to quantify the evidence in favor of the null as well as the alternative hypotheses. We calculated BFs using Bridge Sampling for comparing models that included the effect of interest to models that did not. *BF*s are reported in the direction of the favored model. A *BF* close to 1 means that both models are equally likely, whereas a $BF > 3$ is conventionally interpreted as moderate evidence and a $BF > 10$ as strong evidence in favor of the preferred model (Lee & Wagenmakers, 2013). We applied multilevel logistic (for recall probability) or linear (for response times) Bayesian regressions as implemented in the *brms* R-package (Bürkner, 2017), in which we included crossed random intercepts for subjects and items, as well as random subject slopes for the effect of interest. For the logistic models, the population-level regression coefficients had a weakly informative Student's *t* distribution prior that was zero-

centered with 3 degrees of freedom and a scale of 2.5 (Gelman et al., 2008). All models were run with three chains and 2,000 to 10,000 iterations per chain; half of the iterations were considered as "warm-up." The potential scale reduction factor indicated good convergence for all parameters ($\hat{R} < 1.01$).

Experiment 1

The goal of Experiment 1 was to determine whether free recall probability was higher for items that participants were instructed to remember relative to items on the same list that participants were instructed only to process.

Method

Participants

Sixty-four native English speakers aged 18–30 were recruited. After exclusions (see Table 1), the final sample consisted of 46 participants. Of those, 25 studied word pairs on each trial and 21 studied individual words.

Procedure

Participants studied three lists of 30 trials each. On each trial, half of the participants saw individual words presented in the middle of the screen, whereas the other half saw a pair of words presented one above the other. The between-subjects manipulation of studying individual words versus word pairs was of no theoretical interest. We included it to generalize our findings across more than one type of material and orienting task. When the stimuli were individual words, participants judged on each trial whether the word referred to an object that is larger or smaller than a football. When the stimuli were word pairs, participants judged which of the two words referred to the larger object. Participants pressed the upper arrow key (↑) for the response "larger" and the lower arrow key (↓) for the response "smaller." Each stimulus (word or word pair) was surrounded by either a red or a blue border. Half of the participants were instructed to remember only the words surrounded by a red border for a later memory test; the other half were instructed to remember only the words surrounded by a blue border. After participants made their size-judgment on each trial, the word/s disappeared, but the colored border remained on screen for another three seconds. Following a blank interval of 250 ms, the next trial began. During the first list of trials, participants saw reminders about the size-judgment task that were presented below the word/s on each trial. In addition, on trials that required participants to remember the words (Remember trials), there were instructions above the word saying, "Remember this word for the later test." These reminders were removed in Lists 2 and 3 to avoid distraction from encoding the words.

At the end of each list, participants performed the 1-minute distractor task described in the General Method section, which was then followed by a free recall test only for the words they were told to remember. Following the third list, we additionally asked participants to recall as many words as possible from the list regardless of whether they were asked to remember them or not. In each test, participants had 60 seconds to type as many words as they could remember, and they had to wait the full 60 seconds before they could continue onto the next list.

Results and Discussion

Figure 2 shows the main results about free recall probability. There was a strikingly large main effect of instructions to remember across all lists and both stimulus conditions. Despite the deep semantic processing induced by the size-judgment orienting task, free recall performance for Process-only items was overall near floor (M = 2.7%, 95% CI [1.8, 3.8]).[3] In contrast, free recall was much higher for the Remember items (M = 31%, 95% CI [25.4, 36.7]). This is not unexpected in Lists 1 and 2, where participants had to recall only the Remember words. However, in List 3 where participants had to recall all items regardless of the whether they were asked to remember them or not, performance was only slightly higher for Process-only items relative to the previous lists (M = 4.4%, 95% CI [2.2, 6.6]), and it was still substantially lower than performance for Remember items in List 3, BF = 4.8×10^7. As shown in the scatterplot, the distribution of individual participants' means in the two instruction conditions were almost completely separated. The results cannot be explained by failure to engage with the orienting task for Process-only words—size-judgment accuracy was high and did not differ between Process-only (M = 87.5%, 95% CI [79.6, 95.3]) and Remember words (M = 90.4%, 95% CI [83.2, 97.5]), BF_{null} = 3.3. Thus, in contrast to previous between-subjects studies (e.g., Oberauer & Greve, 2022), we found a substantial effect of intentional remembering on free recall, despite deep processing of all stimuli.

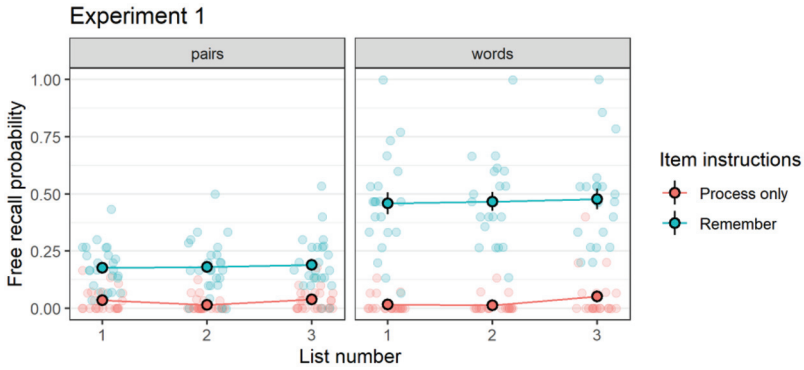
To better understand what people were doing during encoding, we next looked at the size-judgment response times (RTs), and at subsequent free recall probability as a function of size-judgment RTs. We removed outliers using a two-pass procedure. We first excluded trials with RTs faster than 500 ms or slower than 10,000 ms (1.7% of trials). We then removed trials on which the deviance of the RT relative to the median RT was greater than three times the median absolute deviation (7%; Leys et al., 2013). These criteria were established a priori and were the same for all experiments. As can be seen from Figure 3, participants took longer to judge the size of the Remember words (M = 2,184 ms, 95% CI [2,123, 2,245 ms]) than they took to judge the size of the Process-only words (M = 2,015 ms, 95% CI [1,962, 2,068 ms]), BF = 2.4×10^3. This finding replicates the between-subjects results of Oberauer and Greve (2022); however, in their experiment, the extra study time for the Remember group did not translate into better memory.

The previous finding made us wonder whether the total size-judgment time on each trial would predict free recall performance. For each participant, we split trials into ten equally sized bins based on the participant's size-judgment RT quantiles. Figure 4 shows the results of this analysis. For Process-only items, memory performance did not vary as a function of how long it

[3] All confidence intervals and standard errors reported in the text and the figures are between-subject confidence intervals and standard errors. For within-subject designs, it is common to apply a correction using Morey's (2008) standardization method, which removes between-subject variance. However, this and all other methods we are aware of, do not work well when the variance across conditions is drastically different. The between-subject variance for Remember items is several times larger than the variance for Process only items. Applying Morey's (2008) method actually leads to an inflation of the error bars in this case and is inappropriate. Because of this, we report between-subject error bars in all results and the corresponding figures.

Figure 2

Free Recall Probability in Experiment 1 as a Function of List Number, Instructions to Remember for Each Item, and Stimulus Type



Note. Error bars represent ±1 *SE*. The solid points represent the mean of each condition, whereas the fainter points represent individual participants' performance in each condition. See the online article for the color version of this figure.

took participants to make the size-judgment, BF_{null} = 4.4. However, for the Remember items, the longer it took participants to make the size-judgment, the better their subsequent memory for that item was, BF = 438.7. This result suggests that given instructions to remember, participants were performing some additional processing of the words before they made their size-judgment, and the extent of this additional processing predicted subsequent memory. It is noteworthy, however, that when we matched study RTs between the Process-only and Remember items (i.e., each quantile bin and the intercepts of both lines in Figure 4), there was still a large difference in memory between Process-only and Remember items. Thus, additional operations prior to the size-judgment cannot entirely explain the memory advantage for Remember items.

Experiment 2

It is possible that the extra 3 seconds postjudgment time in Experiment 1 allowed participants to continue elaborating Remember but not Process-only words. In the pure-list version of Oberauer and Greve (2022), there was only a 100-ms pause between words. To make our results more comparable, in Experiment 2 the instructions to Remember were only present during the size-judgment and the next word was presented immediately after the response. If extra processing during the postcue period was responsible for most of the effect in Experiment 1, then for the lowest RT quantile the difference between Remember and Process-only words (i.e., the intercept difference in Figure 4) should disappear.

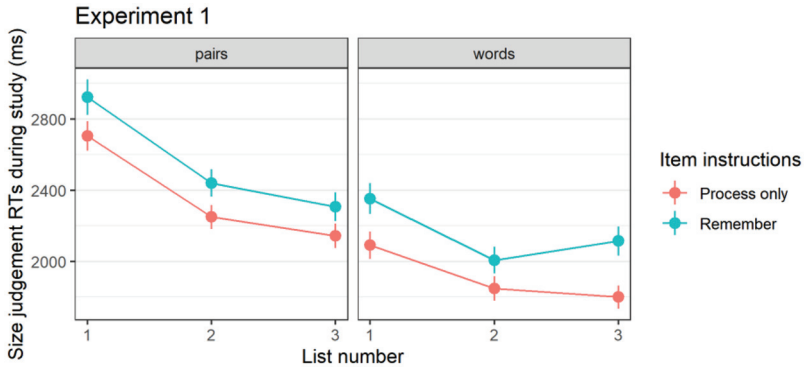
Method

Participants

Fifty-eight native English speakers aged 18–30 were recruited. After exclusions (see Table 1), the final sample consisted of 46 participants. Of those, 28 studied word pairs on each trial and 18 studied individual words.

Figure 3

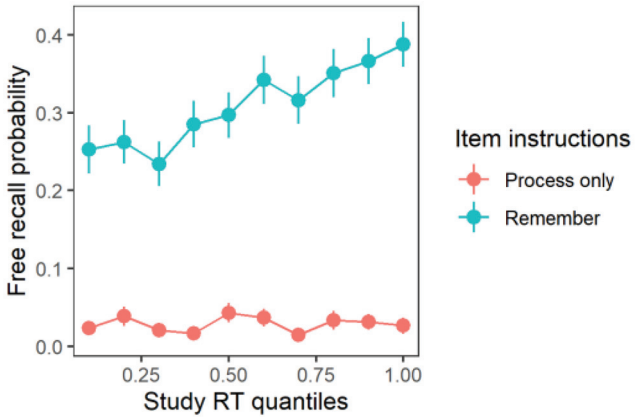
Size-Judgment Response Times (RTs) During Study as a Function of List Number, Stimulus Type, and Memory Instructions for Each Item in Experiment 1



Note. Error bars represent ±1 *SE*. See the online article for the color version of this figure.

Figure 4

Free Recall Probability as a Function of Size-Judgment Response Times (RTs) During Study in Experiment 1



Note. The *x* axis shows the RT quantiles, which were calculated separately for each participant. Quantiles were then split into 10 equally sized binds. Error bars show ±1 *SE*. See the online article for the color version of this figure.

Procedure

The procedure was identical to Experiment 1 except for that the interstimulus-interval between words on each list was reduced from three seconds to zero seconds.

Results and Discussion

We replicated the main patterns from Experiment 1 (see Figure 5). Memory for Process-only items was again near floor (M = 3.2%, 95% CI [2.4, 4.1]). Reducing the interstimulus-interval to zero seconds reduced free recall probability for the Remember items (M = 20.7%, 95% CI [16.7, 24.8]) relative to Experiment 1 (M = 31%, 95% CI [25.4, 36.7]). Nevertheless, memory for Remember items was still substantially higher than memory for Process-only items, BF = 3.6×10^6.

As can be seen from Figure 6, again, participants took longer to do the size-judgements for Remember items (M = 2,441 ms, 95% CI [2,370, 2,511 ms]) relative to Process-only items (M = 2007 ms, 95% CI [1,954, 2,060 ms]), BF = 4×10^6. The difference in RTs between the two instruction conditions (Δ = 433 ms) was nearly three times higher than the same difference in Experiment 1 (Δ = 159 ms). This was specifically attributable to an increase in RTs for Remember items (Δ = 257 ms), without a corresponding increase in RTs for Process-only items (Δ = −8 ms). Removing the opportunity to do selective additional processing for Remember words after the size-judgment likely caused participants to do some of that additional processing before making the size-judgment.

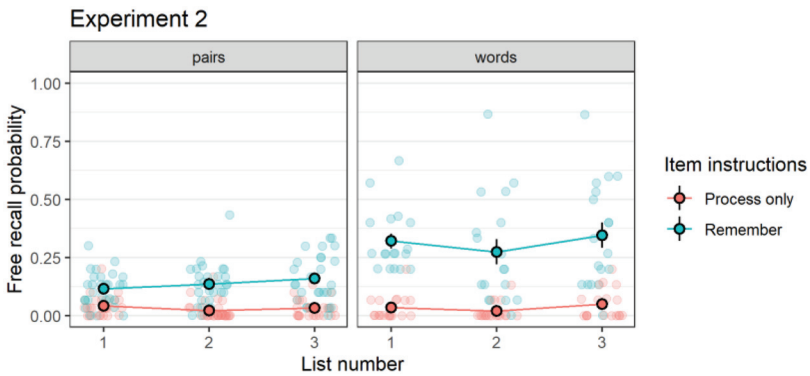
As in Experiment 1, recall probability increased as size-judgment RTs increased, but only for Remember items (see Figure 7). As we expected, removing the interstimulus-interval led to the decrease in the intercept difference between Remember and Process-only items, such that at the lowest RT quantiles the difference between Remember and Process-only items was much smaller (Δ = 6.5%) than in Experiment 1 (Δ = 24%).

Experiment 3

In Experiments 1 and 2, the instructions to Remember or to only Process each item were presented simultaneously with each word onset. This is problematic because the results of Experiment 1 and 2 could in part be explained by people having engaged differently with the size-judgment task depending on the memory instructions. To ensure that the processing of size-judgements would not differ between the Remember and Process-only conditions, in Experiment 3, each word/word-pair was first presented surrounded by a black border and only changed color (to indicate the memory instruction) after participants made their size-judgment. This way, any difference in memory strength between the Remember and Process-only conditions should reflect additional processing performed after the size-judgment was completed.

Figure 5

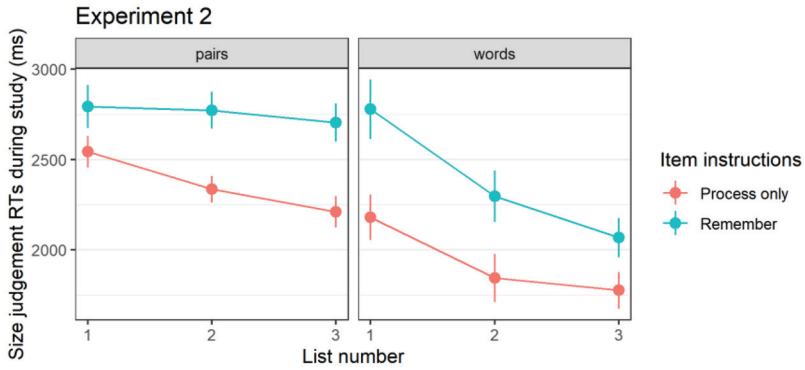
Free Recall Probability in Experiment 2 as a Function of List Number, Instructions to Remember for Each Item, Stimulus Type



Note. Error bars represent ±1 *SE*. The solid points represent the mean of each condition, whereas the fainter points represent individual participants' performance in each condition. See the online article for the color version of this figure.

Figure 6

Size-Judgment Response Times (RTs) During Study as a Function of List Number, Stimulus Type, and Memory Instructions for Each Item in Experiment 2



Note. Error bars represent ±1 *SE*. See the online article for the color version of this figure.

Method

Participants

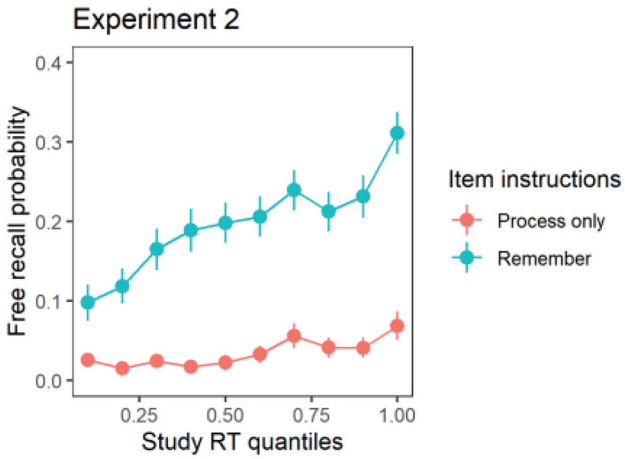
Sixty-five native English speakers aged 18–30 were recruited. After exclusions (see Table 1), the final sample consisted of 53 participants. Of those, 29 studied word pairs on each trial and 24 studied individual words.

Procedure

The procedure was identical to Experiment 1 except items were presented with a black border and border color, which indicated the memory instructions, only changed after participants' size-judgment.

Figure 7

Free Recall Probability as a Function of Size-Judgment Response Times (RTs) During Study in Experiment 2



Note. The *x* axis shows ten equally sized RT quantiles, which were calculated separately for each participant. Error bars represent ±1 *SE*. See the online article for the color version of this figure.

Results and Discussion

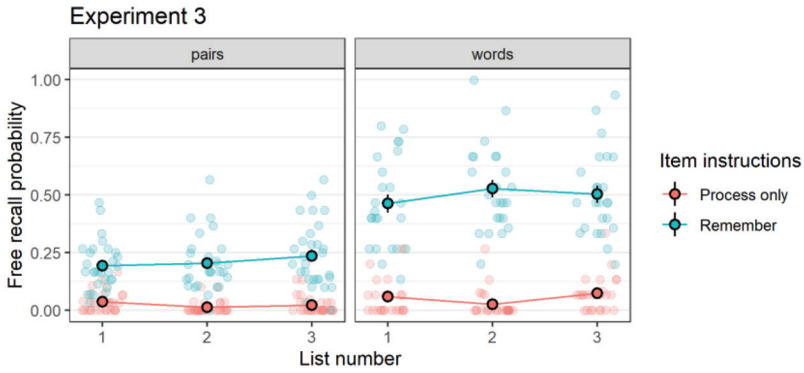
The results were very similar to what we found in Experiments 1 and 2 (see Figure 8). Even though participants did not know whether they were supposed to remember an item before they judged its size, memory for Process-only items was again drastically lower (M = 3.7%, 95% CI [2.5, 4.9]) relative to memory for Remember items (M = 34.3%, 95% CI [29.1, 39.4]), BF = 9.01×10^{13}. Size-judgment responses were highly accurate for both Remember and Process-only items (M = 94.7%, BF_{null} = 3.9), indicating that participants processed all items diligently. Thus, the results of the previous experiments cannot be explained by assuming that people engaged with the size-judgment task differently depending on the memory instructions.

Experiment 4

In Experiment 4 we tested whether delaying the Remember cue by 5 seconds after the size-judgment response would eliminate the difference between the Remember and Process-only condition. One possible explanation for the results in Experiment 3 is that participants may have interpreted the lack of memory instructions for Process-only items as a reason to actively interrupt their consolidation into LTM to prevent interference with the recall of Remember items. Recent evidence suggests that consolidation is not a ballistic process, as previously thought (Ricker & Hardman, 2017), but that it can be actively interrupted under the right conditions (Overkott, 2020). Previous research on Directed forgetting has shown that delaying a Forget instruction by up to 6 seconds eliminates the Directed Forgetting effect (Hourihan & Taylor, 2006). By analogy, we expected that if the absence of memory instructions for process-only items causes participants to interrupt their consolidation, then delaying the memory instructions for each item by five seconds after the item's offset would eliminate the recall difference between Remember and Process-only items, as consolidation would likely have completed by then. Alternatively, if people do additional processing only when the Remember instruction appears, delaying the instruction by 5 seconds

Figure 8

Free Recall Probability in Experiment 3 as a Function of List Number, Instructions to Remember for Each Item, and Stimulus Type



Note. Error bars represent ±1 *SE*. The solid points represent the mean of each condition, whereas the fainter points represent individual participants' performance in each condition. See the online article for the color version of this figure.

should not eliminate the difference in recall between Remember and Process-only items.

Method

Participants

Sixty-one native English speakers aged 18–30 were recruited via Prolific. After exclusions (see Table 1), the final sample consisted of 41 participants. Of those, 22 studied word pairs on each trial and 19 studied individual words.

Procedure

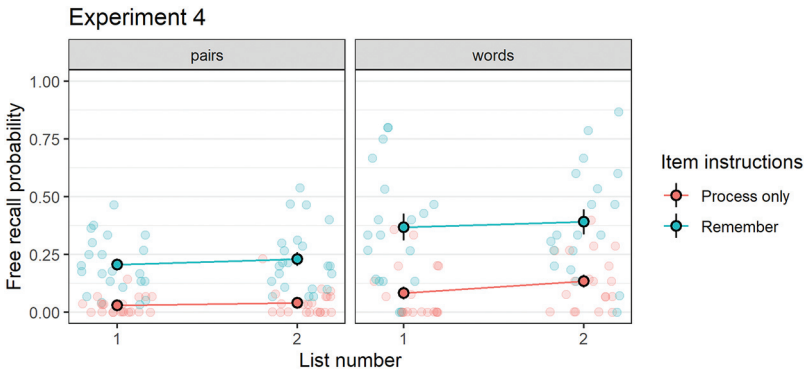
The procedure was identical to Experiment 3 except for the following difference. The border around the stimuli remained black for additional 5 seconds after the word offset. Then, the border color changed to blue/red and remained on screen for 2 seconds before the next trial began. The duration of the memory cue was reduced from 3 to 2 seconds to minimize the overall length of the experiment. For the same reason, there were only two rather than three memory lists, and the surprise test came after List 2.

Results and Discussion

Delaying the memory instructions for each item by 5 seconds increased recall for Process-only items (M = 8%, 95% CI [4.9, 11]) relative to the previous experiments (see Figure 9). However, this increase was small, and there remained a large difference between Process-only and Remember items (M = 30.2%, 95% CI [24, 36.4]), BF = 3.7×10^6. The analyses of size-judgment RTs replicated the lack of difference between Remember and Process-only items that we found in Experiment 3, BF_{null} = 12.5. We also replicated Experiment 3's finding about the lack of correlation between size-judgment RTs for individual items and their subsequent free recall probability, BF_{null} = 20.1. Because consolidation likely had been completed by the time the memory cue appeared in Experiment 4 and we still

Figure 9

Free Recall Probability in Experiment 4 as a Function of List Number, Instructions to Remember for Each Item, and Stimulus Type



Note. Error bars represent ±1 *SE*. The solid points represent the mean of each condition, whereas the fainter points represent individual participants' performance in each condition. See the online article for the color version of this figure.

found much better recall for Remember than Process-only items, interrupting consolidation of Process-only items cannot explain the results from the previous experiments.

Discussion of Experiments 1–4: The Intentional Versus Incidental Learning Paradox

The first four experiments were conclusive—the intent to remember some words on a mixed-list leads to substantially better free recall performance for those words relative to words participants were not instructed to remember. This occurred even though each word was processed deeply using a classical size-judgment orienting task. This result contrasts starkly with most prior studies that have manipulated intent with between-subjects designs using pure-lists (Craik & Tulving, 1975; Hyde & Jenkins, 1969, 1973; Johnston & Jenkins, 1971; Oberauer & Greve, 2022; Till et al., 1975). As far as we are aware, this is the first demonstration of such a dissociation for the effect of intentional vs incidental encoding between pure- and mixed-lists (or between- and within-subject designs, respectively). This dissociation is quite puzzling, and we are terming it the "Intentional Versus Incidental Learning Paradox." If there was nothing more to remembering than deep processing (Craik & Tulving, 1975), and intent to remember did not matter, why is it that we find such a huge effect of intent in our studies using mixed-lists?

Additional processing time for Remember words may provide a simple explanation for the difference between Remember and Process-only words reported in the present Experiments. Remember words could be processed longer in all experiments, either because of prolonged size-judgment RTs in Experiments 1 and 2, or because of selective processing during the postcue interstimulus interval in Experiments 1, 3, and 4. Our analyses revealed that at the item level size-judgment RTs predicted subsequent free recall performance for Remember, but not for Process-only words and that most of the difference between Remember and Process-only words can be accounted for by this relationship (e.g., Figure 7). Thus, the intent to remember is causing additional selective processing of Remember words over and above the semantic analysis induced by the orienting size-judgment task.

However, simply saying that this is a processing time effect is not a satisfying explanation for two reasons. First, in Oberauer and Greve's (2022) pure-list between-subjects experiments, although the intentional learning group spent longer processing the words relative to the incidental learning group, there was no difference in free recall performance between the two groups. Second, on a more conceptual level, "longer processing times" is not a mechanistic explanation—the question is what is happening during these prolonged processing times that leads to better memory for Remember words in mixed-list but not in pure-list designs.

The most striking part of these findings was that free recall performance for Process-only items was near floor in all four experiments. Differences in processing times might be able to explain the overall difference between Remember and Process words, but they cannot explain why recall for Process-only words was so abysmal. This was unexpected, because these are words that participants have processed actively—for each word, they had to retrieve its real-world referent from LTM and evaluate its size relative to another object. This type of relational encoding usually produces high levels of free recall even when people do not expect to be tested later (Craik & Tulving, 1975; Oberauer & Greve, 2022).

There are several possible explanations for the poor performance of Process-only items. One possibility is that the intent to remember does not lead to better memory, but that participants interpret the lack of instructions to remember the Process-only items as instructions to actively forget them. This would make our task more akin to the item-method Directed Forgetting paradigm. One dominant theory to explain the item-method Directed Forgetting effect is the active inhibition theory (Fawcett & Taylor, 2008, 2012; Geiselman et al., 1983). If participants interpret the Process-only instructions as Forget instructions, then this would explain why pure-list between-subjects studies find no effect of intent whereas we do find it in a mixed-list within-subject design. Let us assume that deep semantic processing is sufficient for establishing a strong memory trace and that intent does not add anything on top of that. In between-subjects designs using pure-lists, the participants in the incidental learning group have no reason to inhibit memory for the words. In contrast, in our within-subject paradigm, Remember and Process-only items are mixed in a single list, and participants expect that only Remember items will be tested. Thus, they might attempt to inhibit or prevent the formation of a stable memory trace for Process-only items to reduce interference for Remember items.

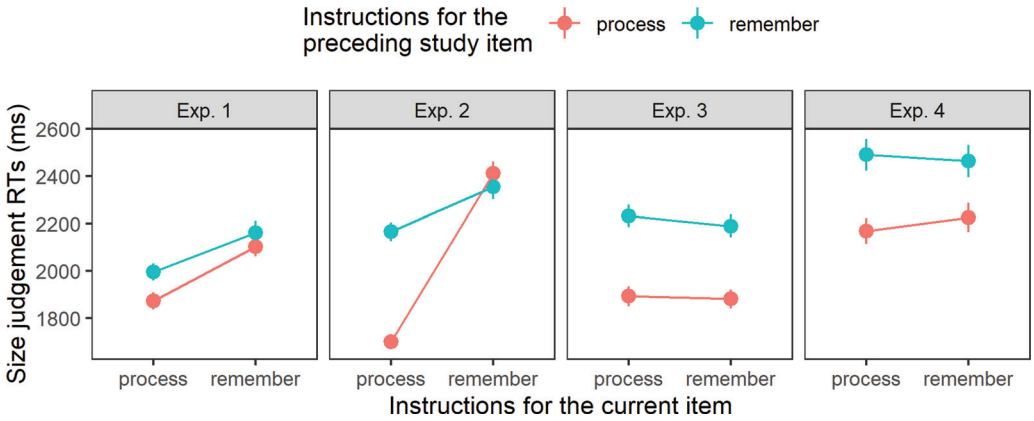
Putting aside the fact that inhibition accounts of Directed Forgetting are highly controversial (MacLeod et al., 2003), inhibition would fail to explain several of the patterns we uncovered. As we noted, free recall probability increases as the size-judgment RTs increase, but only for Remember items. This is direct evidence that some additional processing *during the encoding of Remember* items boosts memory trace strength above the strength that is created by the orienting task itself. Furthermore, inhibition is conceptualized as an active process that takes time: In support of active inhibition, Fawcett and Taylor (2008) found that participants were slower to respond to probes in a secondary task if those probes appeared after To-be-forgotten relative to To-be-remembered items. If inhibition occurs for Process-only items, we expect to see similar effects in our task—slower size-judgment RTs for trials that are preceded by Process-only trials relative to trial preceded by Remember trials. As can be seen from Figure 10 we found the exact opposite pattern in all four experiments—slower size-judgment RTs if the preceding study trial was a Remember rather than a Process-only trial (also see Popov et al., 2019; who found a similar effect on accuracy in Directed Forgetting). Based on these considerations, an inhibition account of the Intentional Versus Incidental Learning Paradox is unplausible.

Alternatively, the absence of memory instructions for Process-only items might have caused participants to interrupt their consolidation into LTM to reduce interference for Remember items. Experiment 4 did not find support for this possibility. Delaying the memory instructions by five seconds after the size-judgment RTs reduced the difference between Remember and Process-only items, but only slightly. It does not seem likely that consolidation could be interrupted after such a prolonged encoding period, because it is likely to have been completed by then.

Another possibility is retrieval inhibition\output interference during recall (Basden et al., 1993; Criss et al., 2011; Smith, 1971; Wilson et al., 2019). As people recall information from LTM, it typically becomes more and more difficult to recall additional

Figure 10

Size-Judgment Response Times (RTs) in Experiments 1, 2, 3, and 4 as a Function of the Memory Instructions for the Current and the Preceding Study Trial



Note. Error bars show ±1 *SE*. See the online article for the color version of this figure.

information even if it was encoded strongly in the first place. The usual explanation for this effect is that each successfully recalled item creates a new memory trace, which causes interference with memory traces created during encoding (Criss et al., 2011). If we presume that in our experiments participants create similarly strong traces for both Remember and Process-only items, but that they always begin recall with the Remember items, this might create enough output interference to prevent the subsequent retrieval of Process-only items. We tested and discarded this output interference possibility in Experiment 5.

Another possibility is that participants engage in selective relational encoding of Remember items only (Basden et al., 1993). Free recall is guided by interitem organizational processes, such as semantic or temporal associations between different items on a study list (Healey, 2018; Kahana, 1996). These associations, created during study, are often used to guide memory search during recall and they determine the success of retrieval. It is possible that in our mixed-list experiments, participants create strong memory traces for both Remember and Process-only items, but that they engage in relational processing only for the Remember items. That is, when a Remember item appears, they attempt to relate it to previous Remember items, but not to previous Process-only items. This would create strong interitem associations between Remember items, which might cause them to retrieve one another during recall. Combined with output-interference, this might be sufficient to explain our results. We tested and discarded this selective relational encoding possibility in Experiments 6 and 7.

Readers who are most interested in the final explanation of the paradox, and not in these alternative hypotheses, could skip directly to section Proposed Solution of the Paradox: Selective Threshold-Shifting.

Experiment 5

The goal of Experiment 5 was to test an alternative explanation for the effect of intent in the mixed-list design, according to which the effect occurs at retrieval—output interference. Retrieving some information from memory often impedes the ability to retrieve further information (Criss et al., 2011; Smith, 1971). A now standard explanation of this result is that each recall event creates an additional memory trace in memory, which increases interference (Criss et al., 2011; although see Osth et al., 2018).

It is possible that in the mixed-list recall tasks presented above, participants begin recalling the Remember items first, and because of that the Process-only items become less and less accessible due to output interference. This account might explain why the effect of intent is different for mixed-lists and pure-lists experiments; however, on its own it would struggle to explain differences in study RTs, their relationship to recall probability, and the source memory findings of Experiment 10 and 11. Nevertheless, it might still play a supporting role in producing the Intentional Versus Incidental Learning Paradox.

To test this possibility, we used the design of Experiment 1, where colors were presented from the word onset and remain on screen for 3 seconds after the size-judgment. However, we asked participants to remember all items regardless of color. In addition, they were instructed to remember each item-color association because their memory for the item-color associations would be tested. At test, we asked them first to recall all items that were presented in blue (or red, counterbalanced), and then to recall all words from the other color. Since there are no Remember/Process-only instructions, this will be a pure test of output interference. Item-color binding strengths should be equal for the two different color conditions, so any difference at test as a function of recall order must be attributable to output interference. If performance for the second tested color is at floor, this would be evidence that the entire Remember/Process-only effect in our mixed-list experiments could be attributed to output interference. If performance for the second tested color is lower than that of the first tested color, but substantially above floor level, then output interference might play a supporting, but not complete, role in producing the Intentional Versus Incidental Learning Paradox.

Method

Participants

One hundred eighty-four English speakers aged 18–30 were recruited via Prolific. After exclusions (see Table 1), the final sample consisted of 177 participants that were split into three

groups—recall Blue first ($N = 50$), recall Red first ($N = 64$) and a control group ($N = 63$).

Procedure

The procedure was similar to Experiment 1, except participants were instructed to remember all words (thus, regardless of color), and there was only one list of individual words to study. Participants had to remember each word and the color of the border surrounding it. At test, an input field appeared with a border matching one of the two colors. Participants were told to recall only words from that particular color. After 30 seconds, the border changed color and participants were now told to recall items from that color instead. The control group was asked to recall all items regardless of color.

Results and Discussion

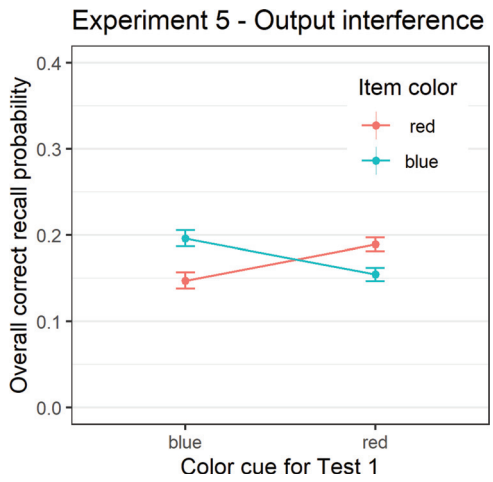
There was weak evidence for output interference—the proportion of recalled words associated with the second tested color was lower than that of words associated with the first tested color, even though participants did not know which color they would have to recall first (see Figure 11). However, this difference was relatively small ($\Delta = 4\%$, 95% CI [1.7, 6.5]), and the evidence in support of the difference was weak ($BF = 2.5$). This difference was much smaller than the difference between Remember and Process-only items in Experiment 1 ($\Delta = 43.3\%$). As such, output interference might play a small role in the Intentional Versus Incidental Learning Paradox using mixed-lists, but it cannot account for all of the effect.

Experiment 6

Experiments 6 and 7 tested one more alternative hypothesis. In the Directed Forgetting literature, some have suggested that participants use relational processing to selectively associate Remember words to one another (e.g., Basden et al., 1993). This relational encoding then facilitates retrieval of other Remember items due to interitem associations. Whereas Basden et al. made that argument for the list-method Directed Forgetting task, it is conceivable that the same occurs in our mixed-list intentional versus incidental learning experiments. The goal of Experiments 6 and 7 was to induce relational processing of all items on the list and to test whether that will reduce the difference between Remember and Process-only items. To achieve this, we changed the nature of the size-judgment task. Rather than judging each word's size relative to a football, participants had to indicate whether the current word represented an object larger or smaller than the previous word on the list (regardless of the memory instructions for each of the words). Although this type of local relational encoding is not the same as higher-level relational encoding strategies such as sentence or story generation, it is still a form of relational encoding that should significantly reduce the effect.

It is not a priori clear whether participants can perform both tasks (size-judgment and selective remembering of Remember items) in this paradigm. Because they must judge the size of the current word relative to the preceding word, they might ignore the Remember/Process-only instructions entirely. Alternatively, they might focus on the Remember instructions, and then perform poorly at the size-judgment task. Thus, finding a null effect between Remember and Process-only items would not be very informative, as it might indicate that participants simply ignored the memory instructions and focused on the size-judgment task. However, finding that Remember > Process-only items together with highly accurate size-judgements would mean that (a) selective local relational encoding cannot explain the results of the previous experiments and (b) during study participants still have memory for the preceding item even after Process-only instructions, and that the subsequent lack of memory in the recall test is not due to inhibition, but due to selective encoding of item-context associations for Remember items.

Figure 11
Overall Free Recall Probability in Experiment 5



Note. Proportion of recalled items associated with each color depending on whether participants were asked to recall the blue or the red color items first before recalling the items from the other color. Error bars reflect ±1 *SE*. See the online article for the color version of this figure.

Method

Participants

Fifty-nine native English speakers aged 18–30 were recruited via Prolific. After exclusions (see Table 1), the final sample consisted of 49 participants.

Procedure

The procedure was similar to Experiment 3 with the following exceptions. Only individual words were presented for study. The size-judgment task was changed in the following way. For every word (except the first one) that appeared during the study list, participants had to indicate whether its referent's size was smaller or larger than the referent size of the word on the immediately preceding trial. After each size-judgment, the word disappeared and the border surrounding it changed to either blue or red. Participants were instructed to remember only words from one of those colors. The test was free recall.

Results and Discussion

First, it is crucial to know whether people were able to perform the relational size-judgment task. The scoring of this task is not trivial, as the words were not selected to be clearly different in size but varied on a continuous dimension. We evaluated accuracy by determining the proportion of "The current object is larger than the previous object" responses based on the relative size difference between each set of consecutive words (see Appendix C for details about how these were calculated). If participants were able to perform the relational size-judgment task we should see a greater proportion of "Larger" responses as the current object size increases relative to the previous object (with ~50% responses when the objects are of similar size); conversely, we should see a greater proportion of "Smaller" responses as the current object size decreases relative to the previous object. Figure 12 indicates that participants were highly accurate in making these relational size-judgements, and more importantly, with similar accuracy regardless of whether the previous item was a Remember or a Process-only item. When the difference between the two objects sizes was maximal (i.e., ±2 on our scale, see Appendix C), 92.5% (preceding word was Process-only) and 93.1% (preceding word was Remember) of the responses were consistent with the size differences direction (either Larger or Smaller). Thus, even though participants were not told to remember Process-only words for the later test, they were able to maintain them in WM long enough so that they could judge the subsequent word's size relative to them.

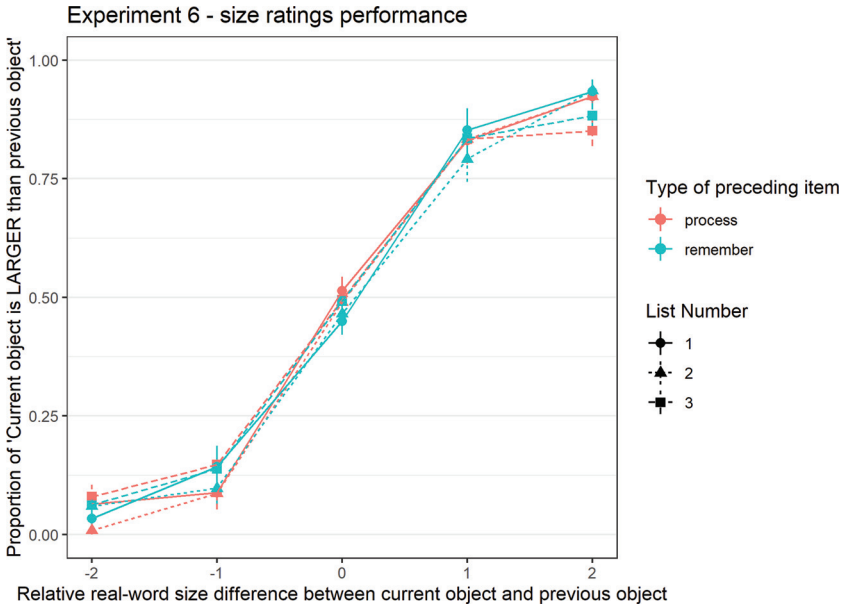
Despite this relational encoding task, the free recall findings closely replicated Experiment 3 (see Figure 13). Memory for Process-only items on List 3 was once again drastically lower (M = 11.5%, 95% CI [8.4, 14.6]) relative to memory for Remember items (M = 37.5%, 95% CI [32.3, 42.8]), BF = 1.9×10^7. The recall probability for Process-only items was slightly higher than the corresponding condition in Experiment 3 (Process-only, individual words, List 3: M = 7.2%, 95% CI [4.1, 10.4]) suggesting that the relational encoding task might have provided a little boost to Process-only items, but this boost was relatively small compared with the overall difference between Process-only and Remember items.

These results indicate that selective relational encoding of Remember words to each other cannot explain the effects of the previous experiments, as inducing relational encoding of all words did not substantially reduce the intent effect. It is also worth noting that in the current task each word, including Process-only words, was processed twice—once as the word whose size was judged relative to the previous word and once as the referent to whom the subsequent word size was judged. Thus, even this double semantic processing was insufficient to prevent the intentional memory benefit, which provides even stronger evidence for the claim that intent matters over and above deep semantic processing. One could ask whether this type of relational encoding is representative of natural higher-level relational encoding strategies such as sentence generation. Although the task is obviously different, here each word is linked to the ones before and after it, creating a continuous chain of associations (e.g., mouse was smaller than tree, which was smaller than a planet, which was larger than a truck, etc.). Just like a sentence generated from studied stimuli, such a continuous chain of associations could potentially be used to help retrieval.

Experiment 7

The difference in free recall between Remember and Process-only items was not eliminated by the relational encoding task in

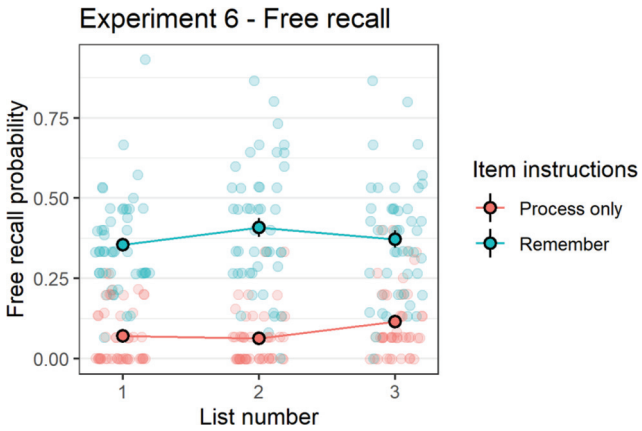
Figure 12
Size-Judgment Accuracy in Experiment 6



Note. Proportion of larger responses as a function of relative size-difference between the current and previous object (see Appendix C for more information about the calculation of these differences). Error bars represent ±1 *SE*. See the online article for the color version of this figure.

Figure 13

Free Recall Probability in Experiment 6 as a Function of List Number and Instructions to Remember for Each Item



Note. Error bars represent ±1 *SE*. The solid points represent the mean of each condition, whereas the fainter points represent individual participants' performance in each condition. See the online article for the color version of this figure.

Experiment 6, but it was reduced relative to the otherwise identical Experiment 3. It is possible that the relational encoding task helped form more interitem associations among all items, but that Participants were not using these associations to guide retrieval. Experiment 7 tested this idea by using the same relational encoding task as in Experiment 6 while changing the final test on List 2 from free recall to cued recall (this cued-recall test took significantly longer, so we reduced the number of lists from three to two). On each test trial after List 2, one item from the list was presented as a cue, and participants had to recall the two items it was compared with during the size-judgment task. Only items studied in even serial positions were presented as cues, so that people were never asked to recall one of the items presented as cues during the test. Cues could have been either Remember or Process-only during study, and the same applied to the target items. People did not expect the final test to be cued-recall, which ensured that their encoding strategy would be the same as in Experiment 6.

Method

Participants

Seventy native English speakers aged 18–30 were recruited via Prolific. After exclusions (see Table 1), the final sample consisted of 49 participants.

Procedure

The procedure was identical to Experiment 6, except for the following differences. There were two lists instead of three. List 2 was followed by a surprise cued-recall test in which every word presented in an even serial position during List 2 was shown as a cue (randomized order). Cues were presented one at a time and participants had to recall the two words which the cue word's size was compared with (the immediately preceding and following word during the study phase). Participants were told to try and recall all words regardless of their instructions during the study phase.

Results and Discussion

Relational size-judgements were highly accurate (see Figure 14). When it comes to the recall performance, first we treated the cued-recall test as a free recall test and scored each recalled item from the list as correct, regardless of whether it was provided as response to the correct cue. This lenient scoring allowed us to test whether providing cues during test reduces the overall difference between Remember and Process-only items. As can be seen from Figure 15, providing cues at test significantly boosted recall for Process-only items relative to Experiment 6 (∼23% in the current experiment versus 11.5% in Experiment 6), and also boosted the recall of Remember items (∼44% in the current experiment versus 37.5% in Experiment 6). However, a substantial difference between Remember and Process-only items remained, BF = 2.2×10^5. Thus, inducing relational encoding across Remember and Process-only items and providing cues during test is not sufficient to eliminate the effect of intent in experiments using a mixed-list design. Therefore, the effect of intent is not simply an effect at retrieval.

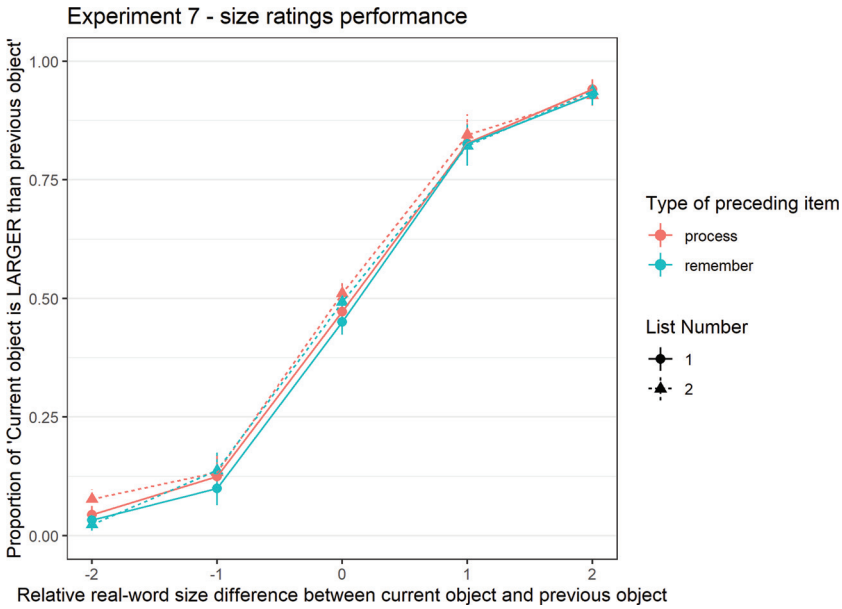
Because memory instructions were randomized for each item, cues and targets could fall into four conditions—both were instructed to be Remembered, both were Process-only, or only one of each was to be Remembered, and the other was Process-only. We next analyzed how cued-recall accuracy depended on the instructions for the cue and the target (with strict coding where an item was considered correct only if it appeared immediately next to the cue during study). The results of this analysis are shown in Figure 16. There was a significant interaction between the instructions for the cue and the target, BF = 13.6. The main finding was that both the cue and the target had to have Remember instructions for intent to boost cued-recall performance. Because this is a cued-recall task, correct performance depends on retrieving the episodic association formed between the cue and the target. Thus, the current results suggest that intent primarily boosts binding memory rather than item memory—otherwise we would have observed better recall of Remember targets regardless of the instructions given for the cue. We return to this point in the next section.

Proposed Solution of the Paradox: Selective Threshold-Shifting

Experiments 5–7 did not support output interference and selective relational encoding as explanations for the Intentional vs Incidental learning paradox. In the Section Discussion of Experiments 1–4 we already explained why we do not believe inhibition of Process-only items to be a viable account. So, what are we left with? The final explanation we considered is that regardless of the experimental design, deep semantic processing is not sufficient to create a strong memory trace and that the trace can be strengthened further by the intent to remember, contrary to the claims by Craik and Tulving (1975) and Oberauer and Greve (2022). Additionally, a shift in retrieval-thresholds could be masking the memory strength difference in pure-list between-subjects experiments. We expand on this idea below.

Figure 14

Size-Judgment Accuracy in Experiment 7



Note. Proportion of LARGER responses as a function of relative size-difference between the current and previous object (see Appendix C for more information about the calculation of these differences). Error bars represent ±1 *SE*. See the online article for the color version of this figure.

A LTM trace is a combination of item information and a binding between an item and the experiential context (Howard & Kahana, 2002; Popov & Reder, 2020). Most models of episodic memory assume that in free recall tests, the study/list context is used to spread activation to all items in memory. Thus, the activation of an item during recall depends on two factors—its current activation and the boost it receives through its binding with the experiential context. Furthermore, successful retrieval occurs when an item's activation passes some retrieval threshold (Howard & Kahana, 2002; Osth et al., 2021; Osth & Farrell, 2019; Polyn et al., 2009; Popov & Reder, 2020). It is possible that deep semantic processing increases only item strength, but as suggested by Experiment 7's results, it might be insufficient to create strong item-item or item-context bindings. The intent to remember is required to boost that item-context binding strength (see Figure 17).

If that is the case, why is there no effect of intent in pure lists between-subjects designs? Prior research on recognition memory has suggested that people can strategically change their retrieval thresholds depending on the overall memorability of the list (Hirshman, 1995; Singer, 2009; Verde & Rotello, 2007). For example, Hirshman (1995) examined yes–no recognition memory for pure- and mixed-lists of strong and weak items (i.e., items studied for 0.4 s or 2 s). Using a signal-detection theory analysis, he found that people used a lower retrieval threshold for pure weak lists relative to mixed-lists and pure strong lists. The same process can potentially occur in between-subjects manipulations of intent—the incidental learning group might use a lower retrieval threshold compared with the intentional learning group, to compensate for their weaker memories. This could in principle lead to equal free recall probability despite different memory strengths in the two groups. In contrast, changing thresholds by condition is not possible when Remember and Process-only items are mixed in a single list. We tested the threshold-shifting account in Experiments 8–11, which provided evidence consistent with it.

Experiment 8

In the preceding sections we proposed that deep semantic processing might only boost item memory, but that it does not create strong item-item or item-context bindings. In Experiment 8, we used a surprise item recognition task instead of a free recall task in the critical third list. If deep semantic processing boosts item memory, then, in contrast to the free recall tasks in Experiments 1–4, we should observe recognition rates for Process-only words that are significantly above chance. As we used relatively short lists, we also expect the difference between Remember and Process-only words to be small, because item activation would be similar for both and sufficient for recognition performance. To ensure that any observed effects in this experiment are not attributable to a different encoding strategy, the recognition test was a surprise.

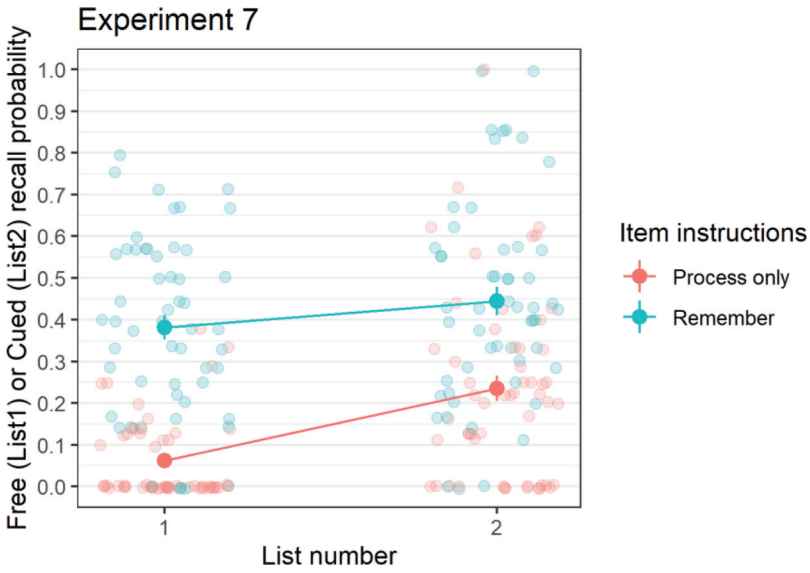
Method

Participants

Forty native English speakers aged 18–30 were recruited via Prolific. After exclusions (see Table 1), the final sample consisted of 26 participants. The sample size in this experiment was

Figure 15

Free Recall (List 1) and Cued Recall (List 2) Probability in Experiment 7 as a Function of the Memory Instructions for Each Item



Note. Cued recall performance was scored leniently—a recalled item was considered correct as long as it was from the same list, regardless of whether it was associated with the specific cue provided at test. Error bars represent ±1 *SE*. See the online article for the color version of this figure.

substantially smaller than the previous experiments, because most participants were at ceiling, and we did not see the need to continue collecting data.

Procedure

The experiment was similar in most aspects to Experiment 3. The word-pairs group was eliminated, and all participants studied individual words. Lists 1 and 2, as well as the study and distraction phase of List 3 were identical to those used in Experiment 3. List 3 was followed by a surprise two-alternative-forced-choice item recognition test instead of a free recall test. The 30 words presented during List 3 were paired with 30 new words. The new words were randomly selected for each participant from the 90 unused words from the total pool. The thirty old-new pairs were presented in random order, each pair shown in the middle of the screen. Half of the old words were presented on the left side of the screen, while the other half were presented on the right side of the screen. For each pair participants had to press the left or the right arrow button to indicate which of the two words they had studied during List 3. There was no time limit on the recognition decisions. When participants made their response, a blank screen appeared for 1 s followed by the next pair of words.

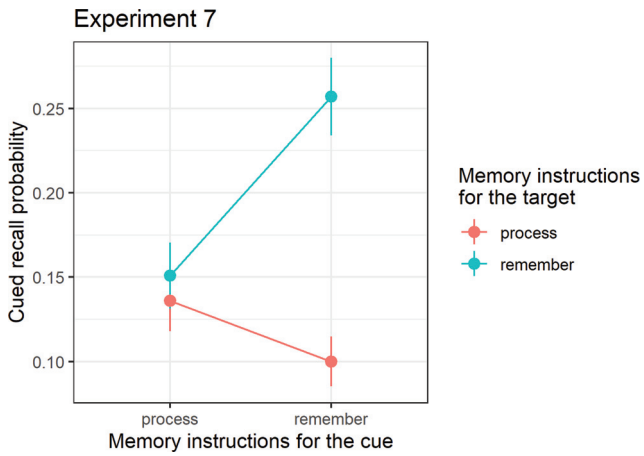
Results and Discussion

There was near ceiling forced-choice recognition performance for both Remember (M = 97%, 95% CI [94.4, 99.5]) and Process-only words (M = 92.3%, *95%* CI [88.5, 96.1]). The difference between the two conditions was small (Δ = 4.7%), but significant, BF = 8.0. As can be seen from Figure 18, free recall performance during List 1 and 2 was similar to that we observed in Experiment 3.

This near-ceiling performance for Process-only words is in stark contrast to the near-floor free recall performance in Experiments 1–4. In fact, nine of the 26 participants had perfect recognition performance for both Remember and Process-only words. Excluding those six participants decreased slightly the mean recognition rates for both Remember (M = 95%, 95% CI

Figure 16

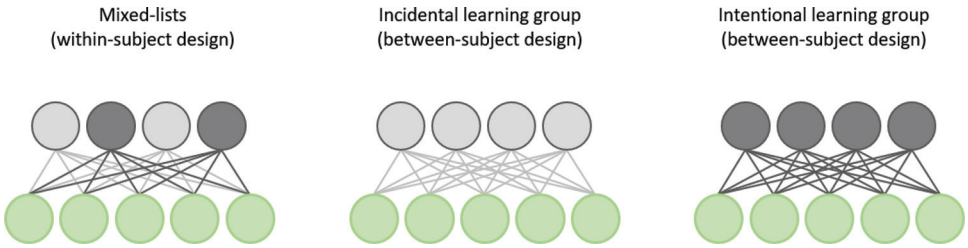
Cued-Recall Accuracy in Experiment 7 as a Function of the Memory Instructions for the Cue and for the Target



Note. Error bars represent ±1 *SE*. See the online article for the color version of this figure.

Figure 17

Example of Memory Representations for Intentionally and Incidentally Learned Items in Within- and Between-Subject Designs According to the Threshold-Shifting Account



Note. Green nodes (bottom nodes) represent a context layer to which all items (gray nodes/upper nodes) are bound (gray lines). In both within- and between-subject designs, the item-context bindings are stronger for intentionally learned items (darker gray lines), which results in higher activation for those items during retrieval (darker gray fill of gray nodes). In mixed lists, the retrieval threshold for all nodes must be fixed, and thus intentionally remembered items are much more likely to be retrieved. In between-subject designs, the incidental learning group can set a lower retrieval threshold to compensate for the lower activation of item nodes. See the online article for the color version of this figure.

[91.6, 99]) and Process-only words (M = 88.2%, 95% CI [83.4, 93.1]), and slightly increased the difference between them (Δ = 7.1%). Nevertheless, performance for both remained high even when we excluded the participants with perfect recognition rates.

These results are consistent with the notion that the semantic orienting task was sufficient to greatly increase item activations in LTM, which lead to near ceiling recognition rates for Process-only words despite near floor free recall probability for them. Because the experiment was otherwise identical to Experiment 3, and participants did not expect the recognition test, these results cannot be explained by a different encoding strategy between the free recall and forced-choice recognition conditions.

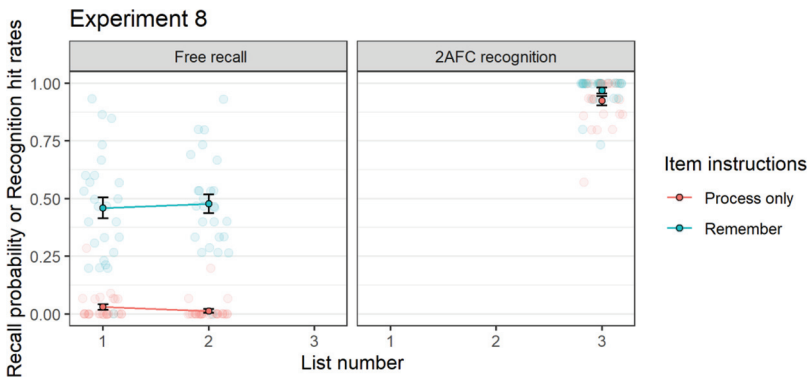
Still, we observed a small difference in recognition rates between Remember and Process-only words. This difference could be explained by dual-process theories of recognition. Since both familiarity (item memory) and recollection (item-context binding memory) contribute to recognition decisions (Popov & Reder, 2020; Reder et al., 2000; Yonelinas, 2002), the small differences could be attributable to different levels of recollection. This would be consistent with our proposal that the semantic orienting tasks boost primarily item memory, but that the intent to remember is required to create or to strengthen the item-context bindings. The question remains why free recall performance in pure-list between-subjects designs shows no difference between the intentional and incidental encoding groups.

Experiment 9

There are many differences in the procedure and design of our mixed-list within-subject experiments and the typical pure-list between-subjects experiments previously reported in the literature. It is possible that the differing results could be attributable not to the pure-list versus mixed-list difference but rather to differences

Figure 18

Free Recall and 2AFC Recognition Accuracy in Experiment 8 as a Function of List Number and Memory Instructions



Note. 2AFC = two-alternative forced choice. Error bars represented ±1 *SE*. The solid points represent the mean of each condition, whereas the fainter points represent individual participants' performance in each condition. See the online article for the color version of this figure.

in some other trivial details of the experiments. In Experiment 9, we replicated the lack of an effect of intent on free recall in a pure-list between-subjects design while keeping all other procedural details identical to Experiment 3.

Method

Participants

Sixty native English speakers aged 18–30 were recruited via Prolific. After exclusions (see Table 1), the final sample consisted of 24 participants in the incidental learning group and 25 participants in the intentional learning group.

Procedure

The experiment was similar in many aspects to Experiment 3. The word-pairs group was eliminated, and all participants studied individual words. Both groups were asked to judge the size of each word as it was presented on the screen. The intentional learning group was told that their memory for all words will be tested after each list. The incidental learning group was given no instructions about a future memory test, but they received a surprise free recall test after the final third list of words. As in Experiment 3, each word was surrounded by either a blue or a red border. Even though the colors had no functional purpose in this experiment, we kept them to maximize similarity between our previous mixed-list experiments and this one.

Results and Discussion

We successfully replicated prior between-subjects studies (Craik & Tulving, 1975; Oberauer & Greve, 2022) and found that free recall performance did not differ overall[4] between the incidental learning group (M = 16.6%, 95% CI [12.9, 19.7]) and the intentional learning group (M = 16.3%, 95% CI [13.2, 19.3]; Figure 19, left panel, BF_{null} = 132).[5] This result is in stark contrast to the mixed-list intent manipulation in Experiment 3. Because the only difference between this experiment and Experiment 3 was the mixed-list within- versus pure-list between-subjects manipulation, the results are indeed attributable to this design difference.

The right panel of Figure 19 shows that even though there were no differences in recall performance, the intentional learning group took much longer to judge the size of the objects (M = 1,582 ms, 95% CI [1,543, 1,621 ms]) relative to the incidental learning group (M = 1,346 ms, 95% CI [1,314, 1,377 ms]), $BF = 6.3 \times 10^6$. This dissociation between the effect of intent on free recall and size-judgment RTs matches Oberauer and Greve (2022) findings. Just like for Remember items in mixed-list designs (see Figure 3), the intentional learning group is doing additional processing of Remember items but without it helping their recall.

How can we explain the dissociation between mixed-list and pure-list experiments? We suggest that deep semantic processing boosts item memory, but that the intent to remember is crucial for forming or strengthening item-context bindings in both mixed-list within-subject and pure-list between-subjects designs. Furthermore, we argue that the incidental learning group uses a lower retrieval threshold to compensate for the weaker memory traces as predicted by the threshold-shifting account. As a result, free recall is equal between the intentional and the incidental learning group despite differences in memory strengths.

This explanation accounts for all of the free recall and size-judgment RT findings presented so far (for an extended discussion, see the General Discussion), but how can we test it further? For the current pure-list between-subjects experiment, the threshold-shifting account predicts that despite the equal levels of correct free recall, in List 3, the incidental learning group should incorrectly recall more items that were not presented on the list (i.e., extralist intrusions) relative to the intentional learning group. This will occur because using a lower retrieval threshold should cause more items from LTM to pass the threshold and to be incorrectly recalled.

This is exactly what we found—there were more than twice as many extralist intrusions on List 3 for the incidental learning group (1.77 words, 27% of all recalled items) relative to the intentional learning group (.71 words, 13% of all recalled items), BF = 23. This prediction should also hold for previous pure-list between-subjects experiments, and we applied the same analysis to the delayed recall conditions of Experiment 4 and Experiment 7 by Oberauer and Greve (2022).[6] In Experiment 4, Oberauer and Greve presented lists of either three, five, or seven words each to three different groups of participants—the "Remember" and the "Incidental" groups were identical to the groups in our experiment. The third "Forget group" was similar to the incidental learning group, but they were told that after each list they should try to forget all words to clear their mind for the next list. Experiment 7 had a similar design, but it used five longer lists of 25 words each. We again found support for our prediction—in both experiments, there were substantially more extralist intrusions for the Incidental and the Forget groups relative to the Remember group. The left panel of Figure 20 shows the number of intrusions in Experiment 4 as a function of participant group and list-length.

The right panel shows the number of recalled words on the final List 5 of Experiment 7 as a function of which list the recalled word was originally presented on. Although all three groups correctly recalled a similar number of words from the current List 5, the Incidental and the Forget groups incorrectly recalled on average two words from each of the preceding four lists. When we add all intrusions together, the Incidental and Forget groups had just as many intrusions (6.01 words recalled from previous lists) as they did correct recalls (5.47 words recalled from the current list). These results show that people in the Incidental/Forget groups recall overall more

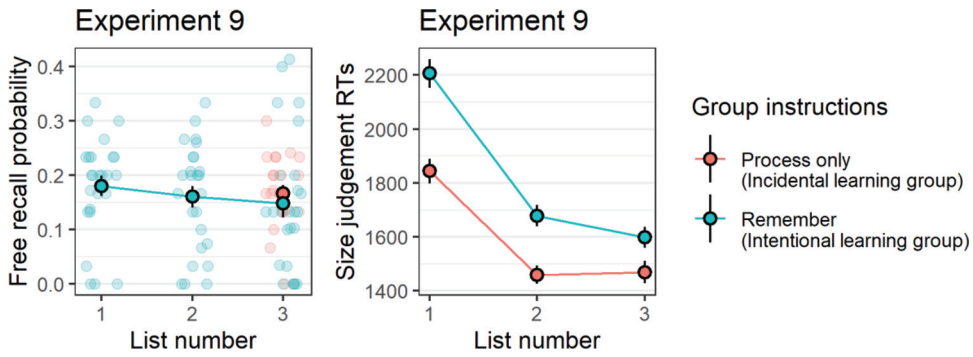
[4] This comparison collapsed performance over all three lists for the intentional learning group. It could be argued that the most appropriate comparison is just between List 3 performance, because at that point participants in both groups have spent an equal amount of time doing the experiment. Alternatively, it could be argued that we should compare free recall performance on List 1 in the intentional group with List 3 in the incidental group, as those are the first tests each group has experienced. Both arguments are valid to a degree, but as can be seen from Figure 19, free recall probability changed very little across the three lists in the intentional learning group (a decrease from 18.5% to 14.8%), thus, ultimately it does not matter which lists we compare.

[5] This experiment presented the same number of words as in the previous experiments—30 per list. Whereas in the previous mixed-list experiments participants were asked to remember only half of the words on each list, in the current experiment the intentional learning group had to remember all 30 words. This explains why the percentage of correctly recalled words in the current experiment is less than half of that from Experiment 3.

[6] We focused only on these two experiments, because they presented multiple lists even for the Incidental and Forget groups, which allowed us to look at the number of intrusions from prior lists.

Figure 19

Performance in Experiment 9 as a Function of List Number and Between-Subject Memory Instructions



Note. Left, Free recall probability in Experiment 9 as a function of list number and between-subject memory instructions. Error bars represent ±1 *SE*. The solid points represent the mean of each condition, whereas the fainter points represent individual participants' performance in each condition. Right, Size-judgment response times (RTs) in Experiment 9 as a function of List number and between-subject memory instructions. See the online article for the color version of this figure.

words, but most of those recalls are intrusions, which fits nicely with the idea that these groups use a lower retrieval threshold.

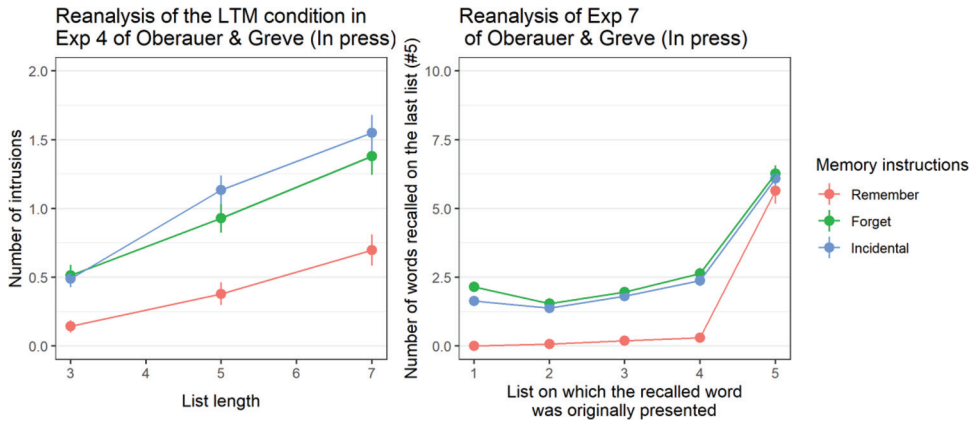
Experiment 10

We argued that deep semantic processing is not sufficient to create strong item-context bindings and that those are boosted by the intent to remember. If this is the case, then a key test for the threshold-shifting account would be to compare Incidental and Intentional memory in pure and mixed lists on a memory task in which participants cannot adjust their retrieval thresholds. The prediction of the threshold-shifting account is that in such a task memory will be better for intentionally learned items in both pure and mixed lists.

We tested this question in Experiment 10 using a 4AFC source memory task in which people had to recall the screen location of studied words. In contrast to the free recall and item recognition tests, performance in this source memory task cannot be done purely based on item-activation, and participants have to find the word's stored location in LTM through the item-context binding. If no item-context bindings are created for Process-only words, we should see near-chance accuracy for the source memory test. However, if item-context bindings are created for both types of words,

Figure 20

Reanalysis of Recall Performance in Oberauer and Greve (2022)



Note. Left, Reanalysis of the delayed recall condition of Experiment 4 by Oberauer and Greve (2022). Number of extralist intrusions as a function of between-subject memory instructions and list length. Right, Reanalysis of Experiment 7 by Oberauer and Greve (2022). The number of words recalled during the test for the final fifth list, as a function of which list the recalled word was originally presented on. Words recalled from Lists 1–4 on the test of List 5 are considered extralist intrusions. Error bars show ± *SE*. See the online article for the color version of this figure.

but are boosted by the intent to remember, we should see above chance performance for both types of items but with better performance for Remember relative to Process-only items.

We directly compared a mixed-list within-subject and a pure-list between-subjects design for the source memory task. We predicted that whereas free recall will only show differences in the mixed lists, the source-memory task should show differences between Process-only and Remember items in both pure-list and mixed-list conditions.

Method

Participants

One hundred sixty-seven native English speakers aged 18–30 were recruited via Prolific. After exclusions (see Table 1), the final sample consisted of 41 participants in the incidental learning group, 35 participants in the intentional learning group, and 73 participants in the within-subject group.

Procedure

Each word on the study lists appeared in one of the four quadrants of a 2 × 2 matrix. Participants were presented with two lists of 32 words each. Words appeared equally often in each of the four quadrants, and the position of words on each trial was randomized across participants. As in Experiment 3, participants judged the size of each word; immediately after the size-judgment, the border around the cell in which the word had appeared turn either blue or red. There were three groups of participants—a mixed-list within-subject group, who were instructed to remember the word and what location it appeared in for only one of the colors; an intentional learning pure-list between-subjects group, who were instructed to remember all words and their locations; and a pure-list between-subjects incidental learning group, who were not told anything about the purpose of the colors, and for whom the final memory test was a surprise. For the intentional-learning group and the mixed-list within-subject group, each list was followed by a free recall test and a source-memory test. For the incidental-learning group the free recall and source memory tests appeared as a surprise only after the second (final) list. The source memory test was identical in List 2 for all groups, except for List 1: For the intentional learning group, all words for the preceding list were presented in a random order and participants had to recall the location of each word. For the mixed-list within-subject group, in List 1 only the words in the Remember color were tested; in List 2 all words were tested as a surprise. For the incidental learning group, there was a test only after List 2 and all words were tested in it. All other aspects of the study procedure and the free recall tests were identical to Experiment 3.

Results and Discussion

We replicated the main dissociation between mixed-list and pure-list designs when it comes to free recall performance (Figure 21, left). As in Experiment 3, for the mixed-list within-subject condition there was a large difference in free recall probability on the last list between Process-only words (M = 7.7%, 95% CI [5.7, 9.7]) and Remember words (M = 31.4%, 95% CI [27.4, 35.4]), $BF = 7.21 \times 10^6$. As in Experiment 9, for the pure-list between-subjects condition there was no difference in free recall probability between the incidental learning group (Process-only, M = 13.2%, 95% CI [11, 15.5]) and the intentional learning group (Remember, M = 15.6%, 95% CI [12.8, 18.4]), BF = 9.6. The interaction between design type (mixed-list within-design vs. pure-list between-subjects) and memory instructions was significant, $BF = 1.78 \times 10^6$.

In line with the threshold shift account (see Figure 21, right), source memory accuracy was above chance (25%) for all conditions (M = 47.5%, 95% CI [44.8, 50.6]), despite near floor performance in free recall for the mixed-list within-subject Process-only condition. There was a main effect of memory instructions—better source memory for Remember (M = 55%, 95% CI [50.9, 59.1]) relative to Process-only items (M = 42.4%, 95% CI [38.9, 45.9]), $BF = 4.2 \times 10^5$. In contrast to the free recall results, this difference was similar for both mixed-list within- (Δ = 12.3%) and pure-list between-subjects (Δ = 12.7%) designs, with no interaction between design type and memory instructions, BF = 520.8.

In summary, we found support for the idea that the intentional versus incidental learning paradox in free recall can be explained by a change in retrieval thresholds in the pure-list between-subjects group. Despite replicating the difference in mixed-list within-subject and pure-list between-subjects designs for free recall, the source memory task indicates that in both types of designs the item-context bindings were significantly stronger for the intentionally remembered items.

Experiment 11

In Experiment 10, participants in the intentional learning conditions had to process both the word and the location, while participants in the incidental learning conditions did not have to do anything with the location. Thus, intent to remember and depth of processing of the relevant source information was confounded. To address this issue, in Experiment 11 we changed the source memory task in the following way. Each word was presented together with a different real-world scene (e.g., a beach, a barn, a kitchen, etc.). In all conditions, participants had to indicate how frequently they expected to encounter the object in that location on a scale from 1 (*rarely or never*), 2 (*sometimes*) to 3 (*often or always*). This way, participants had to process deeply the relation between each object and each scene even in the incidental conditions. Thus, any effect on source memory performance should reflect the effect of intent on item-context bindings. In contrast to Experiment 10's 4AFC location test, we used a 3AFC scene selection test in which participants had to select between an old, a lure or a new scene image in response to the cue word.

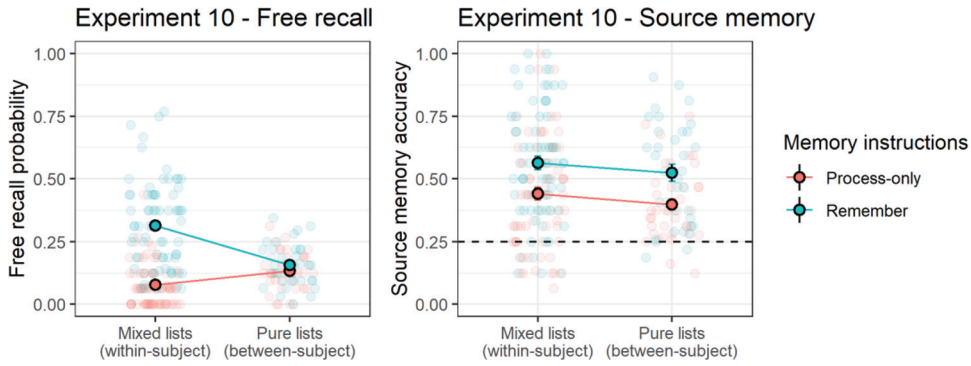
Method

Participants

One hundred seventy-one native English speakers aged 18–30 were recruited via Prolific. After exclusions (see Table 1), the final sample consisted of 38 participants in the incidental learning group, 37 participants in the intentional learning group, and 78 participants in the within-subject group.

Figure 21

Free Recall (Left) and Source Memory (Right) Performance in the Final List of Experiment 10 as a Function of Within- Versus Between-Subject Designs and Type of Memory Instructions



Note. Error bars represent ±1 *SE*. The solid points represent the mean of each condition, whereas the fainter points represent individual participants' performance in each condition. The horizontal dashed line represents chance level in the 4AFC task (25%). See the online article for the color version of this figure.

Procedure

Each word on the study lists appeared above an image of a real-world scene. Participants were presented with two lists of 32 words each. On each trial, participants had to judge how often they might encounter the object in that scene on the following scale: 1 (*rarely or never*), 2 (*sometimes*) or 3 (*often or always*); immediately after the semantic judgment, the border around the word and the image turned either blue or red. All other aspects of the study procedure and the free recall tests were identical to Experiment 10. The source memory test was 3AFC—each of the studied words was presented above three images from the same scene category—the image presented with the word during study ("old"), an image presented with a different word during study ("lure"), or a new image not seen during the study list. The images were labeled 1, 2, or 3 and participants had to press the corresponding keyboard key. Each image type (old, lure or new) appeared equally often in the three positions during the test.

Stimuli

The scenes were exemplars of eight distinct categories—forest, beach, barn, street, kitchen, conference room, inside of a car or a bathroom. There were sixteen exemplar scenes per category for a total of 128 unique images.[7] The images were selected from two existing databases of scenes used in previous memory experiments (Bylinskii et al., 2015; Konkle et al., 2010); some of the images of car interiors were selected from Pixabay (https://pixabay.com/de/). All image dimensions were standardized to be 256 by 256 pixels. For each participant, half of the images from each category were randomly selected to be presented during study; the other half were used as new images during the source memory test. During test, each studied image was presented once as an old image and once as a lure for a different cued word.

For each scene category we generated a list of eight object words that were congruent with the scene (e.g., "toothbrush" for the bathroom category) for a total of 64 unique words. During study, half of the images from each category were presented with a congruent word; the other half of the images were presented with a word for a different category that was incongruent with the presented scene. To ensure that each word served equally often as congruent or incongruent across participants, we generated two counterbalanced lists of word-scene combinations (see Appendix D). Half of the scene categories were presented on List 1, whereas the other half were presented on List 2; these categories were selected randomly for each participant.

Results and Discussion

We replicated all results from Experiment 10 (see Figure 22). For the mixed-list *within-subject* condition there was a large difference in free recall probability on the last list between Process-only words (M = 14.4%, 95% CI [11.8, 17]) and Remember words (M = 40.8%, 95% CI [36, 45.6]). For the pure-list between-subjects condition there was no difference in free recall probability between the incidental learning group (Process-only, M = 21.2%, 95% CI [17.8, 24.5]) and the intentional learning group (Remember, M = 23.9%, 95% CI [20.6, 27.3]). The main effect of intent on recall was significant, $BF = 3.86 \times 10^{22}$, and so was the interaction between design type (mixed-list within-design vs. pure-list between-subjects) and memory instructions, $BF = 6.17 \times 10^{6}$.

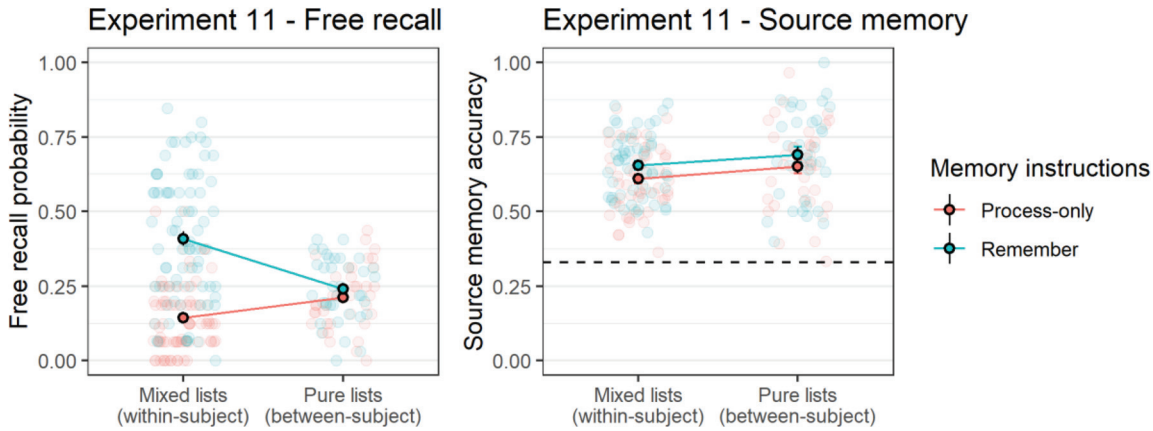
For source memory, there was a main effect of memory instructions on the probability of selecting the "Old" image—better source memory for Remember (M = 66.8%, 95% CI [63.6, 70.3]) relative to Process-only items (M = 62.5%, 95% CI [59.3, 65.7]). The Bayes Factor provided moderate evidence for the effect of intent on source memory (BF = 3.86). In contrast to the free recall results, this difference was similar for both mixed-list within- (Δ = 4.6%) and pure-list between-subjects (Δ = 4.0%) designs, with no interaction between design type and memory instructions, BF_{null} = 21.1.

It should be noted that the effect of intent on source memory was reduced from 12.6% in Experiment 10 to 4.5% in the current experiment (BF = 3.54 in favor of the interaction between experiment and memory instructions). As we discussed in the motivation

[7] The images are available at https://github.com/venpopov/intentional-incidental-ltm-paradox/blob/main/materials/scenes.zip.

Figure 22

Free Recall (Left) and Source Memory (Right) Performance in the Final List of Experiment 11 as a Function of Within- Versus Between-Subject Designs and Type of Memory Instructions



Note. Error bars represent ±1 *SE*. The solid points represent the mean of each condition, whereas the fainter points represent individual participants' performance in each condition. The horizontal dashed line represents chance level in the 3AFC task (33%). See the online article for the color version of this figure.

for Experiment 11, in Experiment 10 there was a confound between memory instructions and depth of processing for the location information. The significant reduction of the intent effect on source memory in this experiment reveals that part of the effect in Experiment 10 was indeed attributable to the presence of this confound; nevertheless, even after eliminating the confound, a significant effect of intent on source memory was observed for both mixed and pure lists, even though there was no difference in free recall in the pure list groups. The critical prediction of the threshold-shifting account was that a difference should exist in source memory even for pure-lists groups; however, without explicit computational modeling of the underlying processing, the account does not predict how large this difference should be. More importantly, the account predicts that the intent effect should be of similar size in both mixed- and pure-lists groups, because intentionally encoding information strengthens bindings equally regardless of the research design, which is what we found. Thus, although the current results are generally consistent with the threshold-shifting account, it cannot be concluded that this account is sufficient to explain the absolute size of the effects we discovered, and further computational modeling work is required to examine this question.

General Discussion

Eleven experiments consistently demonstrated a previously unknown dissociation concerning the effect of intent on recall from long-term episodic memory—the intent to remember does not matter for correct recall in between-subjects designs using pure-lists, but it improves recall substantially in mixed-lists within-subject designs, despite deep semantic processing of all items. Owing to decades of research with the pure-list between-subjects design (Craik & Tulving, 1975; Hyde & Jenkins, 1969, 1973; Johnston & Jenkins, 1971; Oberauer & Greve, 2022; Till et al., 1975), it was previously thought that intent has no effect on LTM (see Appendix A), but here we have demonstrated that this conclusion is false. We were quite surprised by these findings, as our original purpose was to replicate the lack of an intent effect using a mixed-list design. What can explain this dissociation, which we have termed the Intentional Versus Incidental Learning Paradox? Based on these experiments and current models of episodic LTM (e.g., Gillund & Shiffrin, 1984; Polyn et al., 2009; Popov & Reder, 2020), we propose the following explanation:

- Encoding information into episodic LTM occurs by increasing the activation of studied items and by creating bindings between the studied items and the context in which they have appeared.
- Deep semantic processing substantially boosts item activation, regardless of the intent to remember, but on its own creates weak item-context bindings. Because item recognition depends mostly on evaluating item activation, there is only a small difference in hit rates for intentionally and incidentally learned items (Experiment 8, Figure 18).
- The intent to remember motivates additional processing that increases item-context binding strengths, regardless of the research design (i.e., mixed-list within-subject or pure-list between-subject); this additional processing is reflected in prolonged study RTs for intentionally learned items in all experiments presented here, both in mixed-list and pure-list designs (Figures 3, 6, and 19). The idea that intent leads to additional processing also explains why study-RTs for intentionally learned items, but not for incidentally learned items, predict subsequent recall probability. Finally, it explains why study RTs are increased for items that follow intentionally learned items in mixed-lists (Figure 10)—participants might continue to strengthen the binding of the previous item when the current item appears, or might take extra time in order for their encoding resources to recover (Popov & Reder, 2020; Popov et al., 2019; Popov et al., 2021).
- Free recall is initiated by using the list context to spread activation to items in LTM based on the strength of their

bindings with the context. Items whose activation passes a certain retrieval threshold are retrieved and recalled.

- During free recall, intentionally learned items receive more activation from the context than incidentally learned items, owing to their stronger item-context bindings, regardless of the research designs.
- In mixed-list designs, in which intentionally and incidentally learned items are mixed in each list, a single retrieval threshold must be used, which causes greater free recall of intentionally learned items (Experiments 1–4, 6–7, 10, and 11). In pure-list between-subject experiments, participants in the incidental learning group can lower their retrieval threshold to compensate for the lower overall activation of studied items. This leads to equal free recall probability for the two groups (Experiments 9, 10, and 11; also Craik & Tulving, 1975; Hyde & Jenkins, 1969, 1973; Johnston & Jenkins, 1971; Oberauer & Greve, 2022; Till et al., 1975); however, it also causes more extra-list intrusions to occur for the incidental learning group because the lower threshold allows items activated on prior lists to pass over the threshold (Figure 20). Thus, differences in the group threshold in pure-list between-subject designs mask the difference in memory strength for intentionally and incidentally learned items.
- In source memory tasks (Experiments 10 and 11), where each word is presented as a probe during test and participants must retrieve the location in which it was studied, performance depends only on the strength of the word-location (i.e., item-context) bindings. Since this is a forced-choice task, thresholds play no role—one simply picks the alternative that receives the highest activation. Thus, performance in this task is a purer measure of item-context binding strength. As a result, in Experiments 10 and 11, we observed the predicted higher performance for intentionally learned relative to incidentally learned items for both mixed-list within- and pure-list between-subject designs (Figure 21, right, and Figure 22, right). This occurred even though the pure-list between-subject group once again showed no effect of intent on the free recall test that preceded the source memory task (Figure 21, left, and Figure 22, left).

This framework provides a complete qualitative account of the data patterns presented here, although explicit computational modeling is required to determine whether it is necessary or sufficient to capture the absolute quantitative differences across designs and test types (for a recent example of modeling a similar emotional list-composition paradox with a new version of the Context Retrieval & Maintenance model, see Talmi et al., 2019). This is particularly important considering the smaller effect of intent on source memory (∼5%) relative to the effect on free recall (∼25%; Experiment 11). It is not straightforward to compare these estimates, because free recall and 3AFC source memory testing differ significantly in the retrieval processes required (generating information from memory vs choosing one of several alternative responses). Thus, although the overall pattern of results is consistent with the threshold-shifting account, further computational modeling work is required to evaluate it fully.

This framework builds on existing models of free recall; what is novel here is applying this account to intentional vs incidental memory providing the insight that the commonly accepted claim that intent does not benefit LTM is false. This claim, based on the seminal work of Mandler (1967), Hyde and Jenkins (1969), and Craik and Tulving (1975), is routinely taught as part of introductory courses in cognitive psychology (see Appendix A). Our results suggest that researchers need to rethink the role of intent in LTM. Although we agree with Craik and Tulving (1975) that intention, just by existing, does not boost learning, it is clear from our results that intention causes additional processing on top of the deep processing orienting task, and that this processing can be unmasked in mixed-list free recall designs or in source memory forced-choice tasks in which thresholds plays no role in performance.

We do not currently know what the nature of this additional processing is, but one possibility is suggested by the Source of Activation Confusion model (SAC; Popov & Reder, 2020; Reder et al., 2007; also see Diana & Reder, 2006). Popov and Reder (2020) presented evidence that encoding information in LTM depletes a limited encoding resource that recovers gradually over time. They argued that the strength of episodic representations depends on the amount of available resources during encoding, and that spending more of the resource to encode some items leaves less of it available for processing subsequently presented information. As suggested by Oberauer and Greve (2022), by default the episodic memory system needs to encode all incoming information, because one cannot always know which experiences would be relevant for the future. However, combined with the resource assumptions of SAC, this argument also implies that encoding resources must be optimally distributed across experiences; otherwise, some experiences would be encoded poorly if others are strengthened maximally. Thus, it seems reasonable to suggest that by default, in the absence of the intent to learn, encoding in episodic memory does not fully deplete the available processing resources, but distributes them equally across sequential episodes. Sometimes, however, people can have reasonable expectations that a particular experience is more important to remember for the future and they can dedicate more resources to strengthen the underlying memory representation. This suggestion is consistent with findings by Popov et al., 2019, who showed that people can strategically control resource allocation in an item-based directed forgetting task and that intentionally remembering one item on a list hurts memory for items presented immediately after it. In summary, we propose that the effect of intent on long-term memory might be mediated by the control of limited encoding resources—intentional learning redirects more resources to a specific memory representation, strengthening it above baseline at the cost of impairing memory for subsequently presented information (Popov et al., 2019).

Our results add to existing research suggesting that people have intentional control over strengthening binding memory (e.g., in a Directed Forgetting setting; Bancroft et al., 2013; Hockley et al., 2016; but see Burgess et al., 2017). Nevertheless, in Experiments 10 and 11, source memory for Process-only items was above chance. This finding suggests that, with or without the intention to remember, some amount of context information is stored—a finding that is consistent with previous work (Bancroft et al., 2013; Burgess et al., 2017; Hockley et al., 2016; Hourihan et al., 2007) and Malmberg and Shiffrin's (2005) "one-shot" context storage hypothesis according to which a fixed amount of context information is stored

incidentally (regardless of strategic decision to remember that information).

A pressing question about the threshold-shifting account remains—why do people lower their thresholds only in some conditions (incidental learning group in pure-list between-subjects designs) and not others (intentional learning group in pure-list between-subjects designs, or everyone in mixed-list designs)? After all, it would seem like a good strategy for everyone to use a lower threshold allowing all studied items, including those incidentally studied, to surpass the retrieval threshold. The answer is "proactive interference." As we showed in Experiment 9 and the reanalysis of Oberauer and Greve's (2022) data, using a lower threshold in the incidental learning group leads to substantially more extralist intrusions. It could be argued that one goal of a memory system is to maximize correct retrievals while minimizing false alarms. Because there is some overlap in activation distributions between items that are relevant and irrelevant for retrieval, setting an optimal retrieval threshold always requires a compromise between the number of correct retrievals and false alarms. Our data suggest that people change their retrieval threshold based on their estimated memory strength to correctly retrieve at least some of the previously presented items at the cost of retrieving more intrusions.

Alternative Explanations

Although our framework provides an account for all data patterns reported here, there exist possible alternative explanations for at least some parts of our results. For example, in between-subjects experiments, reduced list-differentiation in the incidental learning (as compared with the intentional learning) group could alternatively explain the increased number of extraintrusions for that group as we observed in Experiment 9. Participants in the intentional learning group may engage in distinct, elaborative encoding strategies such as sentence or story generation for each list. In contrast, for the incidental learning group encoding operations do not change from one list to another, making it much more difficult to differentiate the list contexts. Although reduced list-differentiation is a plausible mechanism, on its own it is insufficient to explain why the number of correct recalls is equal in between-subjects experiments but drastically different in within-subject experiments. In contrast, our explanation which combines weaker item-context bindings plus retrieval threshold changes accounts for the entire data pattern, and parsimony leads us to prefer it instead. Nevertheless, we encourage future work to investigate list-differentiation as a potential contributor to extralist intrusions.

Another alternative explanation concerns the dissociation on free recall accuracy for between-subjects and within-subject designs. In between-subjects designs, both groups of participants may engage in the deepest possible processing to answer the size-judgements. Instead, in within-subject designs, participants may engage in the minimal required processing to provide size-judgements for the Process-only items, and in deeper processing to provide size-judgements for the Remember items. According to this proposal, intent does not matter for memory, and intent is confounded with depth of processing in within-subject designs. Although this proposal would potentially explain why we see an intent benefit in free recall for within- but not for between-subjects designs, it is inconsistent with several other aspects of the data. First, in Experiment 3 participants did not know whether an item would have to be remembered until they provided a size-judgment, thus they could not engage in different depth of processing for those judgements. Second, if participants did engage in deeper processing to perform the size-judgements for Remember items, we would expect them to be more accurate in their responses. This was not the case in Experiments 1 and 2, where they knew in advance whether each item was a Remember or Process-only item. Third, if participants engaged in equally deep processing of the size-judgements in both groups of the between-subjects Experiment 9, we would expect that their size-judgements RTs would be equal between the two groups. In contrast, just like in within-subject experiments, people spent more time in judging the size in the intentional relative to the incidental learning group, even though that did not lead to more correct recalls. Finally, in the within-subject Experiments 6 and 7, participants performed relative-size judgements, where they had to respond if the current word represented an object larger than the previous object. Because Remember and Incidental items were randomly intermixed, participants sometimes made the size judgment for a Remember word in reference to the preceding Process-only word. Thus, if they perform deeper processing of the size-judgements for Remember words, this should also extend to the Process-only items that served as a referent for them, leading to better memory for Process-only items that were in the same size-judgment as Remember items. Figure 16 (in which we analyzed cued-recall probability as a function of instructions to remember for both the cue and the target) shows that this is not the case—recall of Process-only targets was not better when the cue was a Remember item. In summary, different depth of processing the size-judgements in within-subject designs cannot explain many of the findings we presented.

Prior Evidence for the Effect of Intent on Memory

As we noted in the Introduction, most studies have shown that intent has no effect on memory in between-subjects designs (Craik & Tulving, 1975; Hyde & Jenkins, 1969, 1973; Johnston & Jenkins, 1971; Oberauer & Greve, 2022; Till et al., 1975); however, there are some exceptions. For example, Block (2009) showed in five experiments that recognition memory for pictures was better for the intentional learning group relative to the incidental learning group in a between-subjects design, seemingly in contrast to prior work with image recognition (Bower & Karlin, 1974). Unfortunately, all five of Block's experiments have a design flaw that was originally pointed out by Craik and Tulving (1975) for other studies, which we already discussed in the Introduction. In Block's experiments, there was either no orienting task (Experiments 1 and 5; Block, 2009), or the orienting task was designed in such a way that it might have detracted attention away from the stimuli that were to be tested (Experiments 2–4). When there is no orienting task, depth of processing is not equated between the intentional and incidental groups which precludes meaningful interpretation of the results because the intentional learning group might have engaged in deeper processing (Craik & Tulving, 1975). The same issue applies to a recent between-subjects study by Naveh-Benjamin and colleagues (2014), in which the incidental learning group only had to passively listen to the words. In Block's Experiments 2–4, which did include an orientation task, participants saw images of several different categories, and they had to count the number of car images. However, their memory was tested only for one of the other categories—faces, birds, and so forth As Bock notes on p. 673, "it is possible that the cover task [. . .] resulted in diminished attention to stimuli other than cars (especially in the

incidental condition)." Thus, we again do not know whether the effects reported in this paper are due to intention to remember.

Another study by Sahakyan and Delaney (2010) also showed a benefit for the intentional learning group in a between-subjects design. In this study, participants were asked to provide pleasantness judgements for two lists of items, with half of the participants being told that their memory will be tested later. In addition, half of the participants in each group also received an instruction to forget List 1 before List 2 was presented. After List 2, participants were asked to recall all items from both lists, regardless of the memory instruction. While there was a small recall benefit for the intentional learning group for List 1 items (7%), there was no effect of intent for List 2 items, and no overall effect for the two lists combined. Crucially, Sahakyan and Delaney (2010) asked participants to recall the two lists on separate sheets, and the authors used a strict coding procedure in which a word was counted as correctly recalled only if it was recalled on the appropriate sheet. This scoring procedure requires accurate list-discrimination, which requires the retrieval of the relevant encoding context. As we have shown, item-context bindings are impaired under incidental learning instructions, which could explain why these authors found a benefit of intent in a between-subjects design while most other such studies did not.

Finally, as we discussed in the Introduction, two prior studies demonstrated better recall for intentionally learned items in a within-subject design like ours (Abel & Bäuml, 2019; Geiselman et al., 1983). As we noted, however, in both studies intentionality instructions were confounded with the type of task. This type of confound was also present in early research on intentionality in between-subjects designs, and it was heavily criticized at the time by Saltzman (1953) and more recently by Block (2009). Furthermore, participants in Geiselman's study were explicitly told "do not learn" the Judge items, which could be interpreted as instructions to forget them.

In summary, whether an effect of intent can be clearly interpreted depends crucially on equating depth of processing between the incidental and intentional conditions (Craik & Tulving, 1975; Saltzman, 1953). Our study builds on this prior literature by clearly demonstrating effect of intent in a within-subject design even when depth of processing is equated.

Dissociation Between LTM and WM

It is a long-standing debate in the field whether LTM and WM are structurally and functionally distinct memory systems (Atkinson & Shiffrin, 1968; Brown et al., 2007; Crowder, 1993; Oberauer, 2002; Oberauer & Greve, 2022). One type of argument that has been put forward to defend the distinction between these systems is that of behavioral dissociations—for example, proactive interference affects retrieval from LTM, but not retrieval from WM (Oberauer et al., 2017). Recently, Oberauer and Greve (2022) suggested that intent could be another dissociating factor. In their experiments using a pure-list between-subjects design, they found that the intent to remember significantly improved WM performance, while it had no effect on LTM performance. They reasoned that this dissociation is due to the different functional properties of each system—WM is limited in capacity and requires a mechanism that determines what information is stored and what is not. In contrast, LTM is unlimited in capacity and because it cannot always be determined in advance what information might be useful for the future, all attended information is stored equally well in LTM, regardless of intent. Our results show conclusively that this intent dissociation between WM and LTM does not hold—the lack of an intent on LTM performance is an artifact of the pure-list between-subjects design. Furthermore, in our reanalyses of Oberauer and Greve's data, presented in the discussion of Experiment 9, we found that once you take into account extralist intrusions, a clear effect of intent in LTM is revealed. Although here we do not take a position on whether WM and LTM are distinct systems or not, we argue that the apparent intent dissociation provides no evidence for this distinction either way.

Conclusion

It is widely believed that the intent to learn does not matter for episodic LTM; we showed that this conclusion is false. The evidence indicates that intentional learning always leads to stronger memory representations in episodic memory, but that this effect might be masked in typical between-subjects designs. This potentially occurs because weaker unintentionally formed memories could be retrieved by using a lower recall threshold, which however also leads to more frequent extralist intrusions. We tentatively suggest that intent redirects more processing resources to the relevant memory representations, which strengthens them at the cost of leaving less resources for processing subsequent information. To avoid such trade-offs, we believe that by default, in the absence of intent, the episodic memory system encodes all incoming information at a lower baseline level. Our results suggests that we need to substantially rethink the role of intent in episodic memory.

Context

The experiments reported here were motivated by a recent paper by Oberauer and Greve (2022). Oberauer and Greve (2022) argued, similarly to Craik and Tulving (1975), that intent to remember plays no functional role in episodic memory encoding, as long as information is processed deeply. They further argued that this is a rational property to have for a system that is unlimited in its total capacity. Our previous work on item-wise Directed Forgetting (Popov et al., 2019) indicates that intentionally encoding information in memory depletes a limited resource. Thus, we originally wanted to replicate the lack of an intent effect in a mixed-list design to test whether intentionally learned items deplete more resources, despite not showing better recall. We were surprised to find in the first experiment that rather than replicating the lack of an effect of intent, we found a substantial difference between intentionally and incidentally learned items. From that point on, the goal of the project changed completely, because now we wanted to understand why intent shows an effect in mixed-list but not pure-list designs.

References

Abel, M., & Bäuml, K.-H. T. (2019). List-method directed forgetting after prolonged retention interval: Further challenges to contemporary accounts. *Journal of Memory and Language*, *106*, 18–28. https://doi.org/10.1016/j.jml.2019.02.002

Anderson, J. R. (2015). *Cognitive psychology and its implications* (8th ed.). Worth Publishers.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*, 89–195. https://doi.org/10.1016/S0079-7421(08)60422-3

Bancroft, T. D., Hockley, W. E., & Farquhar, R. (2013). The longer we have to forget the more weremember: The ironic effect of postcue duration in item-based directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 691–699. https://doi.org/10.1037/a0029523

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 603–616. https://doi.org/10.1037/0278-7393.19.3.603

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & , E. Martin (Eds.), *Coding processes in human memory* (pp. 217–235). Winston.

Block, R. A. (2009). Intent to remember briefly presented human faces and other pictorial stimuli enhances recognition memory. *Memory & Cognition*, *37*(5), 667–678. https://doi.org/10.3758/MC.37.5.667

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, *103*(4), 751–757. https://doi.org/10.1037/h0037190

Brown, G. D., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Burgess, N., Hockley, W. E., & Hourihan, K. L. (2017). The effects of context in item-based directed forgetting: Evidence for "one-shot" context storage. *Memory & Cognition*, *45*(5), 745–754. https://doi.org/10.3758/s13421-017-0692-5

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*(Part B), 165–178. https://doi.org/10.1016/j.visres.2015.03.005

Coin, C., & Tiberghien, G. (1997). Encoding activity and face recognition. *Memory*, *5*(5), 545–568.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language*, *55*(4), 447–460. https://doi.org/10.1016/j.jml.2006.06.003

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*(4), 316–326. https://doi.org/10.1016/j.jml.2011.02.003

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition*, *21*(2), 142–145. https://doi.org/10.3758/BF03202725

Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 805–815. https://doi.org/10.1037/0278-7393.32.4.805

Dudai, Y. (2002). *Memory from A to Z: Keywords, concepts, and beyond*. Oxford University Press.

Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition*, *36*(6), 1168–1181. https://doi.org/10.3758/MC.36.6.1168

Fawcett, J. M., & Taylor, T. L. (2012). The control of working memory resources in intentional forgetting: Evidence from incidental probe word recognition. *Acta Psychologica*, *139*(1), 84–90. https://doi.org/10.1016/j.actpsy.2011.10.001

Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General*, *112*(1), 58–72. https://doi.org/10.1037/0096-3445.112.1.58

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default priordistribution for logistic and other regression models. *The Annals of Applied Statistics*. *2*(4), 1360–1383.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67. https://doi.org/10.1037/0033-295X.91.1.1

Groome, D. (2014). *An introduction to cognitive psychology: Processes and disorders* (3rd ed.). Psychology Press.

Healey, M. K. (2018). Temporal contiguity in incidentally encoded memories. *Journal of Memory and Language*, *102*, 28–40. https://doi.org/10.1016/j.jml.2018.04.003

Henderson, J. (2005). *Memory and forgetting*. Routledge.

Henninger, F., Shevchenko, Y., Mertens, U., Kieslich, P. J., & Hilbig, B. E. (2019). *lab.js: A free, open, online study builder*. PsyArXiv. https://doi.org/10.31234/osf.io/fqr49

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 302–313. https://doi.org/10.1037/0278-7393.21.2.302

Hockley, W. E., Ahmad, F. N., & Nicholson, R. (2016). Intentional and incidental encoding of item and associative information in the directed forgetting procedure. *Memory & Cognition*, *44*(2), 220–228. https://doi.org/10.3758/s13421-015-0557-8

Hourihan, K. L., Goldberg, S., & Taylor, T. L. (2007). The role of spatial location in remembering and forgetting peripheral words. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *61*(2), 91–101. https://doi.org/10.1037/cjep2007010

Hourihan, K. L., & Taylor, T. L. (2006). Cease remembering: Control processes in directed forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1354–1365. https://doi.org/10.1037/0096-1523.32.6.1354

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756492.ch12

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, *82*(3), 472–481. https://doi.org/10.1037/h0028372

Hyde, T. S., & Jenkins, J. J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior*, *12*(5), 471–480. https://doi.org/10.1016/S0022-5371(73)80027-1

Johnston, C. D., & Jenkins, J. J. (1971). Two more incidental tasks that differentially affect associative clustering in recall. *Journal of Experimental Psychology*, *89*(1), 92–95. https://doi.org/10.1037/h0031184

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109. https://doi.org/10.3758/BF03197276

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551–1556. https://doi.org/10.1177/0956797610385359

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

MacLeod, C. M. (1989). Directed forgetting affects both direct and indirect tests of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 13–21. https://doi.org/10.1037/0278-7393.15.1.13

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). Opposition to inhibition. *Psychology of Learning and Motivation*, *43*, 163–214. https://doi.org/10.1016/S0079-7421(03)01014-4

Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, *31*(1), 35–43. https://doi.org/10.3758/BF03196080

Malmberg, K. J., & Shiffrin, R. M. (2005). The "One-Shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 322–336. https://doi.org/10.1037/0278-7393.31.2.322

Mandler, G. (1967). Organization and memory. *Psychology of Learning and Motivation*, *1*, 327–372. https://doi.org/10.1016/S0079-7421(08)60516-2

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482–488. https://doi.org/10.1037/h0045106

Murdock, B. B., Jr. (1960). The immediate retention of unrelated words. *Journal of Experimental Psychology*, *60*(4), 222–234. https://doi.org/10.1037/h0045145

Naveh-Benjamin, M., Guez, J., Hara, Y., Brubaker, M. S., & Lowenschuss-Erlich, I. (2014). The effects of divided attention on encoding processes under incidental and intentional learning instructions: Underlying mechanisms? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *67*(9), 1682–1696. https://doi.org/10.1080/17470218.2013.867517

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411

Oberauer, K., & Greve, W. (2022). Intentional remembering and intentional forgetting in working and long-term memory. *Journal of Experimental Psychology: General*, *151*(3), 513–541. https://doi.org/10.1037/xge0001106

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 1–22. https://doi.org/10.1037/xlm0000302

Osth, A. F., & Farrell, S. (2019). Using response time distributions and race models to characterize primacy and recency effects in free recall initiation. *Psychological Review*, *126*(4), 578–609. https://doi.org/10.1037/rev0000149

Osth, A. F., Jansson, A., Dennis, S., & Heathcote, A. (2018). Modeling the dynamics of recognition memory testing with an integrated model of retrieval and decision making. *Cognitive Psychology*, *104*, 106–142. https://doi.org/10.1016/j.cogpsych.2018.04.002

Osth, A. F., Reed, A., & Farrell, S. (2021). How do recall requirements affect decision-making in free recall initiation? A linear ballistic accumulator approach. *Memory & Cognition*, *49*(5), 968–983. https://doi.org/10.3758/s13421-020-01117-2

Overkott, C. (2020). *Consolidation in visual working memory: How does it operate and can it be facilitated?* [Doctoral dissertation, University of Zurich]. https://doi.org/10.5167/uzh-191183

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review*, *127*(1), 1–46. https://doi.org/10.1037/rev0000161

Popov, V., Marevic, I., Rummel, J., & Reder, L. M. (2019). Forgetting is a feature, not a bug: Intentionally forgetting some things helps us remember others by freeing up working memory resources. *Psychological Science*, *30*(9), 1303–1317. https://doi.org/10.1177/0956797619859531

Popov, V., So, M., & Reder, L. M. (2021). Memory resources recover gradually over time: The effects of word frequency, presentation rate, and list composition on binding errors and mnemonic precision in source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0001072

Postman, L. (1964). Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 145–201). Academic Press.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 294–320. https://doi.org/10.1037/0278-7393.26.2.294

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a double-edged sword: A computational model of the encoding/retrieval trade-off with familiarity. *Psychology of Learning and Motivation*, *48*, 271–312. http://linkinghub.elsevier.com/retrieve/pii/S0079742107480070

Ricker, T. J., & Hardman, K. O. (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology: General*, *146*(11), 1551–1573. https://doi.org/10.1037/xge0000346

Sahakyan, L., & Delaney, P. F. (2010). Item-specific encoding produces an additional benefit of directed forgetting: Evidence from intrusion errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1346–1354. https://doi.org/10.1037/a0020121

Saltzman, I. J. (1953). The orienting task in incidental and intentional learning. *The American Journal of Psychology*, *66*(4), 593–597. https://doi.org/10.2307/1418955

Singer, M. (2009). Strength-based criterion shifts in recognition memory. *Memory & Cognition*, *37*(7), 976–984. https://doi.org/10.3758/MC.37.7.976

Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 400–408. https://doi.org/10.1016/S0022-5371(71)80039-7

Styles, E. (2005). *Attention, perception and memory: An integrated introduction*. Psychology Press. https://doi.org/10.4324/9780203647554

Talmi, D., Lohnas, L. J., & Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychological Review*, *126*(4), 455–485. https://doi.org/10.1037/rev0000132

Till, R., Johnston, C., & Jenkins, J. (1975). Effects of orienting tasks and instructions about associative structure on free recall and clustering. *Bulletin of the Psychonomic Society*, *6*(4), 349–351. https://doi.org/10.3758/BF03333197

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition*, *35*(2), 254–262. https://doi.org/10.3758/BF03193446

Willingham, D. T., & Riener, C. (2019). *Cognition: The thinking animal*. Cambridge University Press. https://doi.org/10.1017/9781316271988

Wilson, J. H., Kellen, D., & Criss, A. H. (2019). Mechanisms of output interference in cued recall. *Memory & Cognition*, *48*, 51–68. https://doi.org/10.3758/s13421-019-00961-1

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

(*Appendices follow*)

Appendix A

Evidence for Widespread Belief That Intent Does Not Matter

The belief that intent to learn does not affect long-term memory is widespread, both in research articles, as well as textbooks on cognition and memory. The following table provides some (nonexhaustive) evidence for this widespread belief.

Source	Type	Quote
Anderson, J. R. (2015). *Cognitive psychology and its implications* (8th ed.). Worth Publishers.	Textbook	"...an important finding that has been proved over and over again in the research on intentional versus incidental learning: Whether a person intends to learn or not really does not matter" p. 144
Groome, D. (2014). *An introduction to cognitive psychology: Processes and disorders* (3rd ed.). Psychology Press, Taylor & Francis Group.	Textbook	"even when we are deliberately trying to learn something we cannot improve on semantic processing. This finding is entirely consistent our experience of learning in real life. You can probably remember in considerable detail many of the things you did earlier today, despite the fact that you did not at any point say to yourself 'I must try hard to remember this'" pp. 168−169
Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Blackwell Publishing Ltd.	Handbook	"The presence or absence of an intention to learn does not figure as a theoretical construct in any current theory of human cognition" p. 353
Styles, E. (2005). *Attention, perception and memory: An integrated introduction*. Psychology Press.	Textbook	"What was astonishing was that the participants who had not been told to learn the words recalled them just as well as those who had been intending to learn, in all conditions." p. 252
Willingham, D. T., & Riener, C. (2019). *Cognition: The thinking animal.* Cambridge University Press.	Textbook	"Wanting to remember something doesn't help your memory. All that matters to your memory is whether you do the deep or the shallow processing" p. 216
Henderson, J. (2005). *Memory and forgetting*. Routledge.	Textbook	"it is a consistent finding that people's performance on tests is as reliable for material which they intend to commit to memory (intentional learning) as it is if they learn material as a result of processing they happen to do as part of an apparently irrelevant task (incidental learning)." p.31
Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268.	Article	"It is abundantly clear that what determines the level of recall or recognition of a word event is not intention to learn [. . .]; rather it is the kind of operations carried out on the items, that determines retention." p. 290
Postman, L. (1964). Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 145−201). Academic Press.	Chapter	"there is little or no reason to maintain a conceptual distinction between intentional and incidental learning" p. 193
Coin, C., & Tiberghien, G. (1997). Encoding activity and face recognition. *Memory, 5*(5), 545−568.	Article	"the similarity of the results obtained with intentional and incidental conditions seems clearly demonstrated" p. 549
Block, R. A. (2009). Intent to remember briefly presented human faces and other pictorial stimuli enhances recognition memory. *Memory & Cognition, 37(*5), 667–678.	Article	"Taken together, these and other studies had a chilling effect on research concerning intent to remember. As a result, during the past several decades, few researchers have investigated differences between incidental and intentional conditions" p. 668
Oberauer, K., & Greve, W. (2022). Intentional remembering and intentional forgetting in working and long-term memory. *Journal of Experimental Psychology: General, 151*(3), 513–541. https://doi.org/10.1037/xge0001106	Article	"by default—in the absence of any intention to remember or forget—the information we attend to and process is maintained in episodic LTM [. . .] episodic LTM is generally insensitive to assessments of future relevance at the time of experience and just maintains a record of our experiences regardless of our intentions [. . .] A beneficial effect of intentional remembering on recall is found if, and only if, a substantial part of recall performance can be contributed by working memory." p. 23

(*Appendices continue*)

Appendix B

Summary of Experimental Design

Exp.	Lists	Size-judgment instructions	Experimental design	Memory instructions	Color instruction timing		Test type	Comments
					Onset	Offset		
E1	3	"Larger or smaller than a soccer ball"?	Mixed-list (within-subject)	Remember blue items	Word onset	3 sec after response	Free recall	Allows for extra processing after size-judgment
E2	3	"Larger or smaller than a soccer ball"?	Mixed-list (within-subject)	Remember blue items	Word onset	Response	Free recall	No extra processing after size-judgment
E3	3	"Larger or smaller than a soccer ball"?	Mixed-list (within-subject)	Remember blue items	Response onset	3 sec after response	Free recall	Participants don't know during size-judgment whether they should remember the word
E4	2	"Larger or smaller than a soccer ball"?	Mixed-list (within-subject)	Remember blue items	5 seconds after response	after 2 sec duration	Free recall	Word must be maintained in WM for 5 seconds before instruction appears
E5	1	"Larger or smaller than a soccer ball"?	Mixed-list (within-subject)	Remember all items and the border color that follows them	Response onset	3 sec after response	Free recall. Group 1 recalls all items. Group 2 has to recall all red items first, then all blue items	Tests for output interference. Color is not associated with R or F instructions, but looking if recalling one color first reduces memory for items in the other color
E6	3	"Larger or smaller than the ***previous word***?"	Mixed-list (within-subject)	Remember blue items	Response onset	3 sec after response	Free recall	Goal was to induce relational encoding of sequential items and to ensure people have to remember even the "process only" items after the cue offset.
E7	2	"Larger or smaller than the ***previous word***?"	Mixed-list (within-subject)	Remember blue items	Response onset	3 sec after response	every odd word is given as a cue and people have to recall the words to which they compared it to	Same as E6, but goal was to see if the traces are completely inaccessible or whether recall can be boosted by providing inter-item cues from the list
E8	3	"Larger or smaller than a soccer ball"?	Mixed-list (within-subject)	Remember blue items	Response onset	3 sec after response	List 1 and 2 - free recall. List 3 - 2AFC recognition	People expect a free recall test, so study strategy must be the same as in E3
E9	3	"Larger or smaller than a soccer ball"?	Pure-list/ between-subject (Incidental learning / Intentional learning / mixed learning)	Remember all items / Process all items / Remember blue items	Response onset	3 sec after response	Free recall	Replicating Oberauer and Greve (2022), while keeping color cues to keep timing and procedure similar to item-wise experiments. The remember all, and Process-only groups were not told anything about the colors
E10	2	"Larger or smaller than a soccer ball"?	Pure-list/ between-subject (Incidental learning / Intentional learning / mixed learning)	Remember all items+locations / Process all items / Remember blue items+locations	Response onset	3 sec after response	Free recall followed by 4AFC source memory task	Words presented in one of four quadrants
E11	2	"How frequently do you expect to find this item in this location"?	Pure-list/ between-subject (Incidental learning / Intentional learning / mixed learning)	Remember all items+locations / Process all items+locations / Remember blue items+locations	Response onset	3 sec after response	Free recall followed by 3AFC source memory task for scenes	Words presented together with scenes

(*Appendices continue*)

Appendix C

Validation of Size-Ratings

All experiments reported in this article used a semantic orienting task during encoding, which asked participants to compare the real-world size of objects referred to by each word presented during the study phase. This size-judgment task had three variants across experiments:

1) Is this object larger or smaller than a football? (when individual words were presented one by one)
2) Which of the two objects is larger? (when word pairs were presented, one word pair per trial)
3) Is this object larger or smaller than the previous object? (when individual words were presented in the relational encoding task in Experiments 9 and 10).

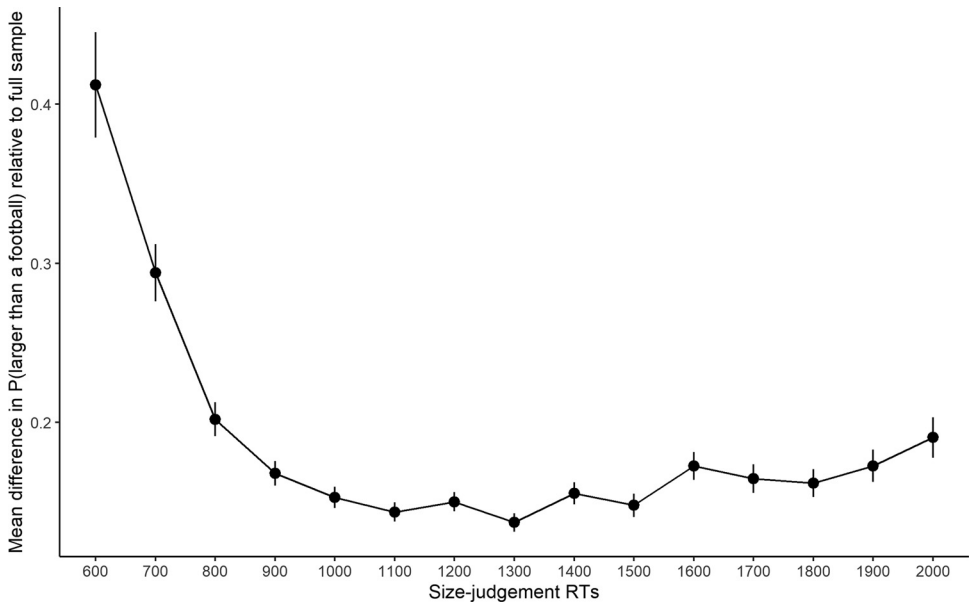
The 180 words were selected from the pool of words used by Popov et al. (Popov et al., 2021). We did not have size-ratings for these words, neither relative to the referent (football) nor to each other. To evaluate performance on the orienting task, we used the size ratings data from Experiments 1–5 and 8–10 to calculate norms for the relative size of objects. We only used the data from Variant 1 in which people were presented with single words (Is this object larger or smaller than a football?). For each of the 180 words, we computed the percentage of people across all 8 experiments who responded that it was larger than a football. We excluded responses that were faster than 900 ms, because as Figure C1 illustrates, size-judgements faster than 900 ms deviated significantly from the mean responses, whereas the function was stable for responses slower than 900 ms. Figure C2 shows that the average proportion of "larger than a football" responses for each of the 180 words formed a bimodal distribution, in which most words were either clearly larger than a football (proportions close to 1) or smaller than a football (proportions close to 0), with a few ambiguous words receiving intermediate proportion of "larger" responses.

Based on this distribution, we classified objects with P (larger) $\geq$ .80 as being larger than a football, and objects with P(larger) $\leq$ .20 as being smaller than a football. These labels were then used to determine whether responses to individual words in the experiments were correct or not.

Evaluating the accuracy of relational responses in Experiment 6 and Experiment 7 (e.g., responses to the question "Is this word larger or smaller than the previous word") required an extra step. Although we do not have ratings for the relative size of objects, we can again use the average proportion of P(larger than a football) responses to infer the relative size of objects. We split words into three groups of increasing size. Words with P(larger-than-a-football) $\leq$ .20 were scored as having size 0, words with P(larger-than-a-football) between .20 and .80 were scored as having size 1, and words with P(larger-than-a-football) $\geq$ .80 were scored as having size 2. Then, for each possible pair of words, we calculated the relative difference in size by subtracting the assigned size (0, 1, or 2) of the two objects. Thus, every pair of words could vary in relative size from -2 (object A is much smaller than object B) to $+2$ (object A is much larger than object B). As shown in Figure 12 and Figure 14, the proportion of "Object A is larger than object B" responses increased gradually as a function of the relative size scale described above.

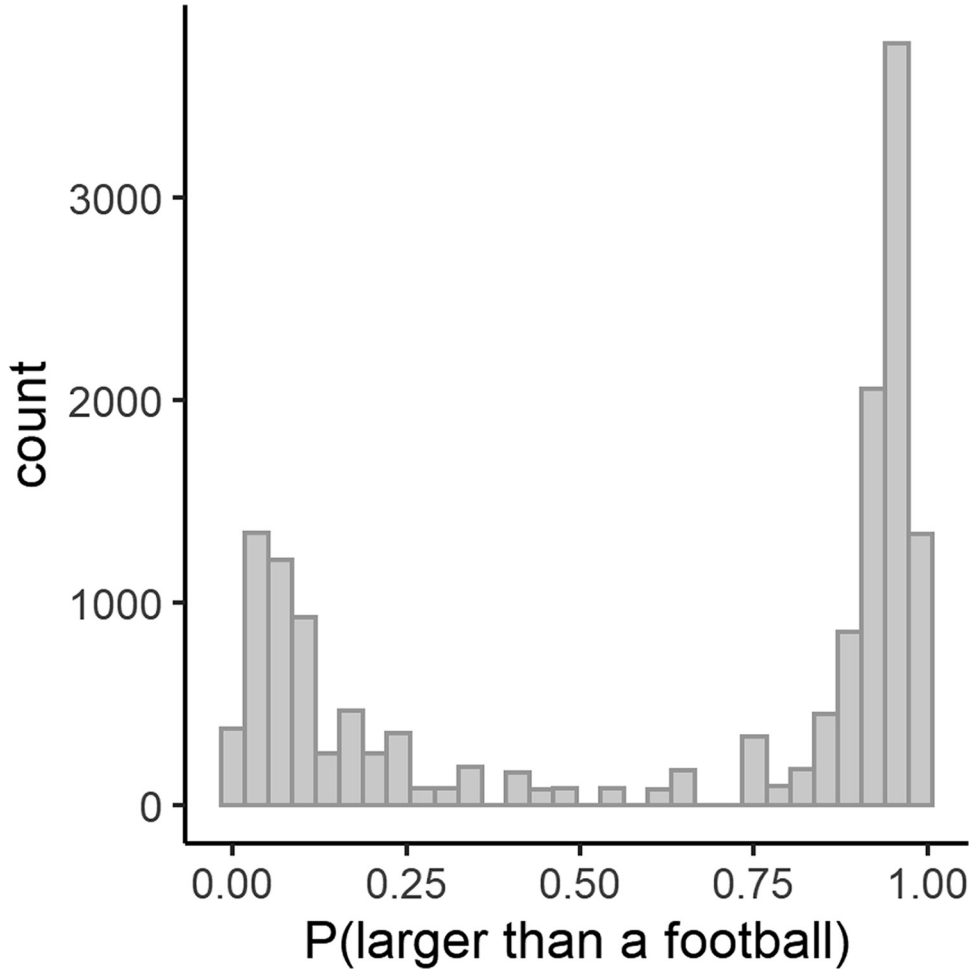
Figure C1

Deviation in Mean P(Larger Than a Football) Responses of Individual Responses Relative to the Mean Proportion as a Function of Size-Judgment RTs



(*Appendices continue*)

Figure C2
Histogram of P(Larger Than a Football) for Each Word



(*Appendices continue*)

Appendix D

Counterbalanced Stimuli Lists for Experiment 11

Table D1
Stimuli Used in Experiment 11

Scene category	Congruence	List A	List B	Congruence A	Congruence B
barn	congruent	tractor	hose	2.81	2.29
barn	congruent	sack	ladder	2.14	2.21
barn	congruent	horse	chicken	2.65	2.65
barn	congruent	rope	mouse	2.19	2.46
barn	incongruent	wine	cafe	1.42	1.19
barn	incongruent	hairspray	shampoo	1.08	1.19
barn	incongruent	candy	folder	1.18	1.19
barn	incongruent	lotion	aspirin	1.18	1.27
bathroom	congruent	toothbrush	lotion	2.84	2.64
bathroom	congruent	shampoo	razor	2.84	2.7
bathroom	congruent	plunger	scale	2.59	2.33
bathroom	congruent	aspirin	hairspray	2.15	2.41
bathroom	incongruent	freezer	map	1.08	1.1
bathroom	incongruent	bread	shoe	1.09	1.5
bathroom	incongruent	briefcase	bike	1.07	1.09
bathroom	incongruent	document	deer	1.12	1.06
beach	congruent	coral	shell	2.47	2.85
beach	congruent	boat	crab	2.66	2.71
beach	congruent	bikini	ball	2.66	2.09
beach	congruent	shark	snorkel	2.24	2.46
beach	incongruent	ladder	church	1.14	1.12
beach	incongruent	scale	tractor	1.18	1.05
beach	incongruent	belt	presenter	1.09	1.38
beach	incongruent	tire	pan	1.24	1.12
car	congruent	map	tire	2.26	2.44
car	congruent	pedal	belt	2.85	2.28
car	congruent	shoe	lipstick	2.09	1.8
car	congruent	shovel	blanket	1.16	1.59
car	incongruent	mouse	bench	1.12	1.21
car	incongruent	bird	horse	1.22	1.12
car	incongruent	fountain	coral	1.02	1.07
car	incongruent	crab	plunger	1.09	1.16
conference	congruent	laptop	document	2.84	2.78
conference	congruent	folder	briefcase	2.74	2.67
conference	congruent	presenter	poster	2.64	2.15
conference	congruent	magazine	candy	1.93	1.5
conference	incongruent	child	sack	1.26	1.14
conference	incongruent	dog	boat	1.07	1.07
conference	incongruent	ball	knife	1.16	1.23
conference	incongruent	mud	string	1.14	1.25
forest	congruent	mushroom	bird	2.44	2.86
forest	congruent	deer	mud	2.49	2.78
forest	congruent	bike	lake	1.97	2.01
forest	congruent	bear	tent	2.33	2.21
forest	incongruent	hose	pedal	1.22	1.3
forest	incongruent	razor	broom	1.08	1.06
forest	incongruent	poster	magazine	1.14	1.15
forest	incongruent	lipstick	toothbrush	1.12	1.09
kitchen	congruent	knife	bowl	2.9	2.7
kitchen	congruent	pan	freezer	2.86	2.81
kitchen	congruent	broom	bread	2.26	2.66
kitchen	congruent	string	wine	1.7	2.51
kitchen	incongruent	taxi	bus	1.03	1.1
kitchen	incongruent	lake	bear	1.04	1.12
kitchen	incongruent	snorkel	shark	1.05	1.09

(*table continues*)

(*Appendices continue*)

Table D1 (*continued*)

Scene category	Congruence	List A	List B	Congruence A	Congruence B
kitchen	incongruent	tent	shovel	1.05	1.16
street	congruent	bus	taxi	2.55	2.6
street	congruent	cafe	dog	2.27	2.24
street	congruent	church	fountain	1.96	1.64
street	congruent	bench	child	2.2	2.24
street	incongruent	bowl	mushroom	1.19	1.25
street	incongruent	chicken	rope	1.21	1.36
street	incongruent	blanket	bikini	1.29	1.19
street	incongruent	shell	laptop	1.12	1.46

Note. The columns Congruence A and Congruence B provide the average congruence ratings from all participants in Experiment 11 for Lists A and B. The ratings were provided on the following scale: How often do you expect to encounter this object in this location? 1 (*rarely or never*), 2 (*sometimes*) or 3 (*often or always*).

Received October 18, 2021
Revision received June 2, 2022
Accepted June 4, 2022 ■