Memory and Proactive Interference for spatially distributed items

Ansgar D. Endress
Department of Psychology, City, University of London, UK

Draft of December 10, 2019

Our ability to briefly retain information is often limited. Proactive Interference (PI) might critically contribute to these limitations, for example by reducing the items' temporal distinctiveness: For example, rejecting items in a recognition test is harder if they appeared on recent trials. In vision, recent results suggest that spatial information protects against PI: If an item is encoded using separate item-location combinations at different spatial locations, representations of item-location combinations will be more distinct than simple item-based representations and show less interference. Here, I show that PI is reliably observed for spatially distributed items except when it is weak. I show mathematically that, if observers encode simple items, the strength of PI is determined by the ratio between (a) the number of display items and (b) the total number of available items, but that, if observers encode item-location combinations, it is determined by the total number of available items only. The results support the item-based model. Finally, I analyze the effects of (1) the presence spatial information and (2) its predictiveness on memory and susceptibility to PI. While memory is impaired when items are spatially distributed, but unaffected by the predictiveness of spatial information, the susceptibility to PI is unaffected by either manipulation. Taken together, these results suggest that (1) in the absence of PI, we have a large capacity for storing information, (2) PI occurs between items rather than item-location combinations, and (3) the presence of spatial cues might change the task by providing retrieval cues for memory searches.

Keywords: Temporary Memory; Long-Term Memory; Working Memory; Short-Term Memory; Proactive Interference; Distinctiveness

The author thanks T. Makovski for helpful comments on an earlier draft of this manuscript. Data and analysis scripts are available at `https://figshare.com/s/4c4d52414702cc7a9286`. Please note that the URL will change in the final version of the manuscript.

1 Introduction

1.1 Memory limitations and proactive interference

Our ability to retain information over brief periods of time is often severely limited, with important consequences in domains ranging from language acquisition (e.g., Baddeley, Gathercole, & Papagno, 1998; Kam & Newport, 2009; Newport, 1990) to educational attainment (e.g., Gathercole, Pickering, Knight, & Stegmann, 2004) to fluid intelligence (e.g., Ackerman, Beier, & Boyle, 2005; Alloway &

Alloway, 2010; Engel de Abreu, Conway, & Gathercole, 2010; Fukuda, Vogel, Mayr, & Awh, 2010; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). However, the reasons for such limitations are debated.

To the extent that was is retained in memory is what is not forgotten, the traditional view is that memory loss is largely determined by interference from other memory items (e.g., Berman, Jonides, & Lewis, 2009; Keppel & Underwood, 1962; Lewandowsky, Geiger, & Oberauer, 2008; Lewandowsky, Oberauer, & Brown, 2009; Postman & Underwood, 1973; Wickens, Born, & Allen, 1963). In fact, susceptibility to Proactive Interference (that is, impaired learning of new information due to pre-existing memory representations) is closely related to Working Memory capacity (e.g., Engle, 2002; Jonides & Nee, 2006; Kane & Engle, 2000; Lustig, May, & Hasher, 2001; May, Hasher, & Kane, 1999; Nee, Jonides, & Berman, 2007; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Oberauer & Lin, 2017; Rosen & Engle, 1998; Rowe, Hasher, & Turcotte, 2010; Shipstead & Engle, 2013).

In a recent demonstration of the importance of the effects of Proactive Interference (PI), Endress and Potter (2014a) presented participants with rapid sequences of pictures of everyday objects; following each sequence, they viewed another picture and had to decide if it had been part of the sequence. When the pictures were trial-unique and never repeated across trials (in the *unique condition*), Endress and Potter (2014a) observed no memory capacity limitations: the probability of encoding any single one of the sequence items was relatively independent of the number of items in a sequence, at least for larger set-sizes. In a marked contrast, when items were drawn from a limited pool of items and reused across trials (in the *repeated condition*), performance was much lower and memory capacity estimates remained in the range previously reported.[1] Repeating items across trials creates PI; for example, it is relatively hard to reject a test item we have seen on recent trials, at least much harder than rejecting a test item that has never occurred at all. These results thus suggest that PI can massively impair memory performance.

PI might plausibly impair our ability to remember information over brief periods of time, especially because it has long been recognized that not only intra-experimental information might create PI, but also information learned previously in our life (e.g., Underwood & Postman, 1960). For example, we might need to type in a phone number, but have a lifetime of experience with number sequences (e.g., other phone numbers, credit card numbers, lottery numbers, . . .) that might plausibly interfere with our ability to learn a new number sequence. More generally, Endress and Szabó (2017) proved that the mere presence of interference from other items in memory mathematically guarantees limited memory capacities under fairly general conditions.

PI might also plausibly explain the fixed and limited memory capacity estimates in one of the most prominent demonstrations of memory limitations — Luck and Vogel's (1997) change detection paradigm (see also Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Fukuda et al., 2010;

[1] In terms of memory capacity estimates, Endress and Potter (2014a) observed "capacities" of up to 30 items for sequences of 100 items, derived from Cowan's (2001) two-high-threshold formula. However, given that the accuracy in the memory test was relatively constant across set-sizes, the capacity estimates depended on the set-size. For example, an accuracy of 75% yields a capacity estimate of 5 for set-size 10 and a capacity estimate of 10 for set-size 20. However, one would expect participants to perform at ceiling as long as the set-size remains below the capacity estimate (e.g., 30 for a set-size of 100 items), which was clearly not the case. As a result, capacity estimates are not necessarily meaningful in these experiments. This does not seem to be a specific problem of Endress and Potter's (2014a) experiments, as performance decreases even within the putative memory capacity range in other visual Working Memory experiments as well (e.g., Schneegans & Bays, 2016).

Hartshorne, 2008; Jiang, Olson, & Chun, 2000; Lin & Luck, 2012; Makovski, 2016; Makovski & Jiang, 2008; Pertzov & Husain, 2014; Rouder et al., 2008; Schneegans & Bays, 2016; Treisman & Zhang, 2006; Zhang & Luck, 2008, among many others). In this paradigm, observers view an array of objects (e.g., colored squares); after a delay, they view another array of objects (or a single object in some versions of the paradigm) and have to report whether or not the test array changed with respect to the sample array. The accuracy in these tasks can be converted to memory capacity estimates (e.g., Cowan, 2001; Rouder et al., 2008). The resulting capacity estimates are typically low (see e.g. Bays & Husain, 2008; Bays et al., 2009; Zhang & Luck, 2008, for more sophisticated analyses that are not necessary for the current purposes).

Critically, at least in principle, memory performance in this paradigm might suffer greatly from PI, as an extremely limited set of items is reused over many trials; for example, in Luck and Vogel's (1997) experiment, just 7 colors were used over hundreds of trials. Surprisingly, however, the effects of PI in this paradigm seem to be fairly weak (e.g., Hartshorne, 2008; Lin & Luck, 2012; Makovski & Jiang, 2008). These authors compared performance in trials with strong PI from immediately preceding trials and in trials with weaker background PI, where PI was still substantial because the same pool of items was reused in all trials, but less strong than when items had occurred in immediately preceding trials. Results showed little additional PI when items had occurred in immediately preceding trials.[2]

Such results contrast markedly not only with Endress and Potter's (2014a) results, but also with other Working Memory paradigms such a the recent probes task, where participants view a sample array of items and have to make a decision as to whether items in a test array were part of the sample array. Performance is usually impaired when nearby trials share items, irrespective of whether the items are letters (e.g., D'Esposito, Postle, Jonides, & Smith, 1999; Jonides & Nee, 2006), words (e.g., Craig, Berman, Jonides, & Lustig, 2013) or line drawings of animals (e.g., Loosli, Rahm, Unterrainer, Weiller, & Kaller, 2014).

1.2 Is verbal memory more susceptible to PI than visual memory?

What might explain such differences? One possibility is that visual memory is independent from other forms of memory and has different properties. In fact, it is widely accepted that visual working memory is independent of verbal working memory (e.g., Baddeley, 1996; Cortis Mack, Dent, & Ward, 2018; Cowan, Saults, & Blume, 2014; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000). If Endress and Potter's (2014a) task as well as the recent probes task (D'Esposito et al., 1999; Jonides & Nee, 2006) fall on the verbal side of this spectrum, while change detection tasks fall on the visual side, it is at least possible that verbal WM might be more susceptible to PI than visual WM.

If this possibility is correct, it would severely restrict the use of visual Working Memory for general cognitive processing. In fact, most of Endress and Potter's (2014a) experiments used visual material in the form of meaningful pictures. Such pictures are likely encoded in terms of their conceptual meaning rather than its visual properties (e.g., Potter, 1976; Potter, Staub, & O'Connor, 2004; Potter, 2010), which is independent of verbal processes in turn (Endress & Potter, 2012). If change detection experiments differ from other memory experiments in

[2]Even more dramatically, in a large scale Web-based study, Hartshorne (2008) observed a memory capacity of only 2.75 on the first trial of experiment, where PI could not yet have occurred. However, given that his displays consisted of only 4 items, memory capacity estimates (e.g., Cowan, 2001; Rouder et al., 2008) are mathematically limited to 4, and a capacity of 2.75 corresponds to an accuracy of 84%. It is thus possible that Hartshorne (2008) would have observed larger capacity estimates if he had used larger set-sizes.

their sensitivity to PI due to modality differences, change detection experiments likely target fairly low-level visual representations that are unlikely to participate in the general cognitive processes that are presumed to be subserved by Working Memory (e.g., Engle, 2002; Conway, Kane, & Engle, 2003; Cowan, 2005; Fukuda et al., 2010; Süß et al., 2002) — unless the critical processes involved in visual Working Memory really reflect attentional encoding rather than memory per se (Tsubomi, Fukuda, Watanabe, & Vogel, 2013; see also Fukuda & Vogel, 2009, 2011)

A second and related possibility is that different Working Memory experiments make different demands on different aspects of Working Memory, and that different aspects of Working Memory differ in how susceptible they are to PI. For example, Oberauer et al. (2000) evaluated no less than 23 Working Memory tasks that differed in the extent to which they relied on storage of information, transformation of information, supervision, coordination as well as in the modality in which the memory items were presented. It is thus possible that different aspects of Working Memory might be differently susceptible to PI. However, while this possibility is plausible, it is unlikely to account for the relative insensitivity of change detection experiments to PI, given that the superficially similar recent probes task (e.g., D'Esposito et al., 1999; Jonides & Nee, 2006) tends to show reliable PI effects.

A third possibility is that the presence of high background PI limits the observable effects of *additional* PI. For example, Shoval, Luria, and Makovski (2019) showed that PI effects are reduced when items come from a single category (such as faces or houses) compared to when they come from multiple categories. Given that the fact of using items from the same general category is well known to create PI within just a few trials (see, among many others, Carroll et al., 2010; Gardiner, Craik, & Birtwistle, 1972; Hopkins, Edwards, & Cook, 1972; Jitsumori, Wright, & Shyan, 1989; Kincaid & Wickens, 1970; Wickens et al., 1963; Wickens, 1970), Shoval et al.'s (2019) single category conditions compared a high PI condition to a condition with even higher PI. Given that items in change detection experiments typically come from an extremely limited set of items, the presence of strong background PI might thus reduce the observable effects of additional PI.

A fourth possibility is that change detection experiments elicit specific processing strategies as memory items are spatially distributed. Such strategies might take two forms: First, participants might use the spatial locations of the items to make them more distinctive across trials (Makovski, 2016). Second, they might encode entire spatially organized arrays of items as objects to be stored in memory. I will now discuss the first of these possibilities and return to the second one in the discussion.

1.3 PI and spatial organization

Makovski (2016) proposed that the relative insensitivity to PI of change detection experiments can be explained if objects are automatically bound to their spatial locations (e.g., Jiang et al., 2000; Makovski & Jiang, 2008; Pertzov & Husain, 2014; Treisman & Zhang, 2006; Udale, Farrell, & Kent, 2017). For example, if a dog at Location 1 is encoded separately from the same dog at Location 2, the item-location combinations are (more) trial-unique and thus reduce PI across trials.

Such a strategy is rather plausible. In fact, it is well known that even *imagined* spatial information improves memory. For example, the method of loci was known in Greek and Roman times, and involved (mentally) placing memory items (the elements of a speech for orators or the items in a memory list for participants in a memory experiment) in spatial locations such as rooms in a house (Yates, 1966). Interestingly, the resulting memory advantage is not specific to an imagined spatial context, but can also be obtained by imagining a temporal context such as autobiographical information or well rehearsed routines (Bouffard, Stokes, Kramer, & Ekstrom, 2018).

Further, in some experiments, spatial information seems to be encoded automatically. Participants confuse features of items that share spatial locations. For example, if they have to remember the orientation of a colored bar, the orientations of bars sharing their spatial positions tend to get confused; in contrast, no such confusion arises when items share non-spatial features such as color (Pertzov & Husain, 2014).

That being said, observers predominantly encode the *relative* spatial positions of objects (e.g., Jiang et al., 2000; Treisman & Zhang, 2006; Udale et al., 2017). In fact, in change detection experiments using colored shapes as stimuli, an effect of spatial congruency between the sample and the test array emerges only when *all* objects are presented during test, but not when a single test object is shown (e.g., Treisman & Zhang, 2006; Udale et al., 2017). Further, spatial information seems to be linked to entire objects rather than object features. For example, in Treisman and Zhang's (2006) and Udale et al.'s (2017) experiments, participants had to report feature changes irrespective of which other feature they were bound to. If the sample array contained a red square and a blue triangle, a test array containing a red triangle and a blue square would require a "no change" response because all features were already present in the sample array. (In contrast, an orange square would elicit a "change" response, as no sample object was orange.) When new feature combinations (e.g., the red triangle) are shown in the test array, participants perform better if they appear in a *new* location, presumably because feature combinations are bound into objects (e.g., Luck & Vogel, 1997), which are bound to spatial locations in turn, and the memory for the entire objects is less likely to be accessed when the objects are shown at a new location.

To test whether the relative insensitivity of change detection tasks to PI might be due to their spatial organization, Makovski (2016) first replicated Endress and Potter's (2014a) experiments where all memory items were presented at the center of the screen (albeit with relatively small set-sizes of only 4 and 8 items) and observed sizable PI effects as in Endress and Potter's (2014a) experiments. In further experiments, spatial cues seemed to substantially reduce the strength of the PI effect. Specifically, when, in his Experiment 3, all sample items were presented simultaneously on an imaginary circle (instead of being presented sequentially at the center of the screen), he found the PI effect only for Set-Size 8, but not for Set-Size 4. Likewise, when, in his Experiment 4, items were presented sequentially but with unique spatial positions on an imaginary circle, the PI effect was completely abolished.

In contrast, and in line with the interpretation that binding memory items to spatial positions makes the position-item combinations more distinct and thus reduces PI, a PI effect reemerged if items were presented sequentially on a circle, but if the test item was presented at the center of the screen (Footnote 2), presumably because this manipulation abolished the usefulness of the spatial cues.

Makovski's (2016) results thus suggest that spatial cues can reduce PI among memory items because they make the items more distinct. However, before accepting this conclusion, it is worth pointing out another difference between Makovski's (2016) and Endress and Potter's (2014a) experiments. While Makovski (2016) used relatively small set-sizes of up to 8 items drawn from a total pool of 21 items, Endress and Potter (2014a) used set-sizes of up to 20 items in those experiments where they used a total pool of 21 items. Items were thus repeated much more frequently in Endress and Potter's (2014a) experiments. In the next section, I characterize this difference more precisely.

1.4 PI and waiting time

Reducing the set-size while keeping the total pool-size constant increases the mean waiting time between two occurrences of the same item. As an increasing delay between repeated occurrences of an item reduces PI (presumably by making items temporally more distinctive; e.g., Loess & Waugh, 1967; Kincaid & Wickens, 1970; Peterson & Gentile, 1965; Shipstead & Engle, 2013), this effect might have contributed to the relative insensitivity to PI of Makovski's (2016) experiments when items were spatially distributed. (While I use terms such as time and delay to describe these effects, they do not necessarily reflect absolute times measured in seconds, but might rather be relative times measured with respect to some experimental variables such as the presentation duration.)

Specifically, for a set-size of S pictures presented on each trial and a total pool-size of T pictures presented in the entire (block of an) experiment, the probability that a given picture appears as a sample (in a sample sequence or in a sample array) is $p = S/T$. The probability of a lag of N trials between successive occurrences of a sample picture is thus $P(\text{lag} = N) = p(1-p)^N$, the probability of waiting *at least* N trials is $P(\text{lag} \geq N) = (1-p)^N$, and the mean lag is $\frac{1-p}{p}$ trials.[3,4]

Interestingly, only waiting times between occurrences of an item as a sample depend on the set-size, while the probability of occurrence as a test item does not.[5] Critically, however, as observers can accumulate long-term memory traces from repeated exposure to even briefly presented items (e.g., Endress & Potter, 2014b; Melcher, 2001, 2006; Pertzov, Avidan, & Zohary, 2009), it seems plausible that a reduced waiting time between consecutive occurrences of an item might increase PI.

As shown in Figure 1a, the probability of lags greater than, say, 5 trials is small irrespective of the set-size. However, Figure 1b shows that small set-sizes allow for substantial probabilities of lags of *at least* 5. For example, and as mentioned above, in his Experiments 3 and 4, Makovski (2016) used the set-sizes 4 and 8, with a total Pool-Size for 21 items. This corresponds to probabilities of have at lags of at least 5 trials of 35% and 9%, respectively (and 19% if I assume an average set-size of 6). Figure 1c shows the mean lag between two occurrences of the same item as a sample item. The average lag with set-sizes 4, 6 and 8 is 4.3, 2.5 and 1.6 trials, respectively. At least for the smaller set-sizes, such waiting times are considered low-PI conditions in the recent probes task (e.g., Craig et al., 2013; D'Esposito et al., 1999; Jonides & Nee, 2006; Loosli et al., 2014). As a result, spatially distributed items might well show PI effects when PI is strengthened by reducing the waiting time between subsequent occurrences of an item.

1.5 Spatial information vs. waiting time

The discussion so far raises two questions. First, are the effects of PI determined by other factors over and above the waiting time between consecutive occurrences of the same memory items? Second, what are the memory items that interfere with each other: are they simple items (e.g., a pictures of dog)

[3]By noting that $P(\text{lag} \geq N) = 1 - P(\text{lag} < N)$, we can apply the geometric series formula and obtain $P(\text{lag} \geq N) = 1 - \sum_{k=0}^{N-1} p(1-p)^k = 1 - p\frac{1-(1-p)^N}{1-(1-p)} = (1-p)^N$.

[4]I first note that we can drop the first term in the series: $\langle \text{lag} \rangle = \sum_{N=0}^{\infty} NP(\text{lag} = N) = \sum_{N=1}^{\infty} NP(\text{lag} = N) = p\sum_{N=1}^{\infty} N(1-p)^N$, which is a polylogarithm of order -1. We thus obtain $\langle \text{lag} \rangle = p\frac{1-p}{(1-(1-p))^2} = \frac{1-p}{p} = \frac{1}{p} - 1$.

[5]The probability of an item occurring as a test item is the sum of the probability of it occurring as an *old* test item and as a *new* test item (i.e., a foil): $P(\text{test item}) = P(\text{test item}|\text{old trial})P(\text{old trial}) + P(\text{test item}|\text{new trial})P(\text{new trial})$. An item can be an *old* test item if (i) it is selected as a sample item, (ii) it is not in the first or the last position, and (iii) if it is selected as a test item: $P(\text{test item}|\text{old trial}) = \frac{S}{T}\frac{S-2}{S}\frac{1}{S-2} = \frac{1}{T}$. Likewise, an item can be a new test item if (i) it is not selected as a sample item, and (ii) it is selected as a test item: $P(\text{test item}|\text{new trial}) = \left(1 - \frac{S}{T}\right)\frac{1}{T-S} = \frac{1}{T}$. In sum, the probability of an item occurring as a test item is thus given by $P(\text{test item}|\text{old trial})P(\text{old trial}) + P(\text{test item}|\text{new trial})P(\text{new trial}) = 2\frac{1}{T}\frac{1}{2} = \frac{1}{T}$ and does not depend on the set-size.

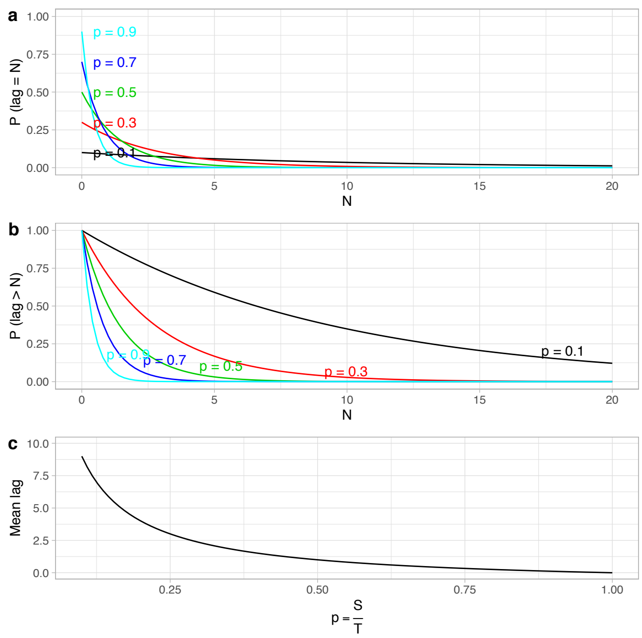


Figure 1. (a) Probability of a lag of N trials between two successive occurrence of an item as a sample. The different curves show different ratios between the set-size and the size of the total pool of items. The combinations of set-sizes and pool-sizes in Makovski's (2016) experiments correspond to $p = 4/21 \approx .19$ and $p = 8/21 \approx .38$, respectively. (b) Probability of waiting for *at least* N trials between two occurrences of an item as a sample item. Again, the different curves represent different ratios between the set-size and the total pool-size. (c) Mean waiting time (in trials) between two consecutive occurrences of an item as a sample item as a function of the ratio between the set-size and the total pool-size.

or rather item-location combinations (e.g., a picture of a dog presented in a specific spatial location)?

A comparison of Makovski's (2016) Experiments 1 and 4 strongly suggest that waiting times between occurrences of individual items do provide a sufficient explanation of the data. In both experiments, participants viewed sequences of objects and completed a recognition test after each sequence; the only difference between these experiments was that, in Experiment 1, all items were presented at the center of the screen while in Experiment 4, items were presented on an imaginary circle. As a result, the waiting times in both experiments were identical. Still, PI effects were observed only in Experiment 1 but not in Experiment 4.[6]

While these results suggest that the spatial distribution of items has *some* effects on PI, the conclusion that observers encode item-location combinations is not the only interpretation. In fact, if spatial information were the main source of information that eliminates PI among items, PI should not only be reduced with sequentially presented items, but also with simultaneously presented items, as long as they are spatially distributed. However, in his Experiment 3, Makovski (2016) found reliable PI effects for spatially distributed items when they were presented simultaneously rather than sequentially, though the waiting time between two occurrences of the same item was much reduced compared to Experiment 4 due to the fast presentation speed.

More generally, it is not clear if encoding item-location combinations would be a viable memory strategy, for two reasons. First, observers predominantly encode *relative* spatial positions of objects (e.g., Jiang et al., 2000; Treisman & Zhang, 2006; Udale et al., 2017) rather than the absolute ones required for item-location combinations. Second, in everyday cognition, such a strategy would be most useful for long-term retention of inanimate objects rather than for a general Working Memory system used in moment-to-moment reasoning, given that objects tend to change location, especially if they are animate. In line with this view, item-information is more stable over time than location-information in probabilistic foraging tasks (e.g., Téglás et al., 2011) and the two types of information have dissociable neural substrates (e.g., Goodale & Milner, 1992; Mishkin & Ungerleider, 1982).

[6]More specifically, performance in the Unique Condition was reduced between Experiments 1 and 4, while performance in the Repeated Condition was identical.

An alternative view is that the presence of spatial information changes the response strategies rather than the underlying representations. For example, participants might use the spatial information provided during test to initiate a memory search; if PI is weak enough so that a memory search is successful even in the Repeated Condition, one would expect performance in the Unique and the Repeated condition to become equivalent, which is just what Makovski's (2016) Experiment 4 shows.

To use an analogy with face recognition, similarly to the Unique Condition in the absence of spatial cues, it is relatively easy to discriminate faces of friends from faces of strangers, while, similarly to the Repeated Condition in the absence of spatial cues, it is harder to discriminate faces of good friends we have seen very recently from faces of good friends we have seen somewhat less recently.

In contrast, introducing additional retrieval cues changes the task demands. For example, we might try to discriminate faces of individuals who were at some party from individuals who were not. In this case, it is possible to initiate a memory search through our friends to decide who was at the party and who was not, though this search would become much more difficult when most of our friends are party animals who attend most parties. In contrast, discriminating previously unknown partygoers from complete strangers should still be possible, even though it might be harder to link a stranger to a party than a friend due to the lack of familiarity. In other words, changing the task demands by introducing retrieval cues might reduce the gap between the Unique and the Repeated Condition, unless interference in the Repeated Condition becomes too strong to make memory search unfeasible.

These views make different predictions, because they make different assumptions about the total pool of possible items. If participants encode item-location combinations, the effective total pool-size is the number of locations (i.e., the set-size) multiplied with the total number of items, $S \times T$. As a result, the mean waiting time between two occurrences of the same item-location combination is $T - 1$ and does not depend on the set-size.[7] While this result might appear counterintuitive at first, it can be made plausible through the following observation. If the set-size, and thus the number of positions, equals the total number of items T, once an item (e.g., a dog) has appeared in some position, we need to wait of the order of T trials for it to appear again *in that position*. In contrast, if we have twice as many items in total as we have positions, and if the dog happens to be selected on every trial, we need to wait of the order of $T/2$ trials for it to reappear in a position again (since we have $T/2$ positions). However, since the dog will be selected only on half of the trials, the waiting time for it to reappear in an earlier position is again T. Hence, if participants represent item-location combinations, the waiting time between two occurrences of am item-location combination, and thus the strength of their PI, only depends on the total number of items, but not on the set-size.

In contrast, if the participants use spatial information just as a retrieval cue, the critical total pool is still that of the available items, and the mean waiting time between subsequent occurrences of an item is $\frac{T}{S} - 1$.[8]

If participants encode item-location combinations, they should thus be relatively insensitive to the ratio between the set-size and the total pool-size; in contrast, if the predominant memory representations are item-based, the critical determinant of the strength of PI is the ratio between the set-size and the total pool-size.

[7] According to the formula in the previous section, the mean waiting time is giving by $\frac{1}{p} - 1 = \frac{1}{\frac{S}{S \times T}} - 1 = T - 1$

[8] This follows again from the expression for the mean waiting time derived in the previous section: $\frac{1}{p} - 1 = \frac{1}{\frac{S}{T}} - 1 = \frac{T}{S} - 1$.

2 The current experiments

In the experiments below, I ask four questions. First, does PI affect memory for spatially distributed items? Second, are observers sensitive to the strength of PI? Third, is the strength of PI determined by PI between simple item representations or between representations of item-location combinations? Fourth, does spatial information *per se* affect memory performance and the susceptibility to PI?

In Experiment 1, I ask whether the strength of PI affects memory performance when memory items are spatially distributed. In each trial, participants viewed a sequence of 8 items, presented sequentially on an imaginary circle. Following this, they viewed another (test) item and had to indicate whether or not this latter item had been part of the sequence. Critically, in different blocks, I varied the size of the total pool from which items could be drawn: Items came from (1) a pool of 9 items in total, (2) a pool of 22 items in total or (3) were trial-unique (*Pool-Size* = ∞), respectively. Given that items are repeated more often when there are fewer items, I expected greater PI for smaller pool-sizes.

To preview the results, I found substantial PI for Pool-Size 9, but not for the larger Pool-Size 22, where performance was similar to the Unique Condition. Participants are thus sensitive to the strength of PI among spatially distributed items. However, these results are ambiguous as to whether PI occurs between the representations of simple items or of item-location combinations.

This ambiguity is addressed in Experiment 2 (and in a pilot experiment reported in Appendix A). In Experiment 2, I used a constant pool-size similar to the pool-sizes for which neither Makovski (2016) nor Experiment 1 above detected PI effects among spatially distributed items. Critically, however, I increased the set-size. As mentioned above, if people maintain representations of item-location combinations, the strength of PI should be unaffected by the set-size; if they maintain representations of simple items, the strength of PI should be determined by the ratio between the set-size and the total number of (simple) items.

Specifically, participants in Experiment 2 completed a block where pictures were trial-unique and one where they were sampled from a total pool of 21 pictures. Within each block, participants viewed sequences of 12 or 20 images that were presented on an imaginary circle, proceeding in a clock-wise direction. The set-size was thus increased compared to Experiment 1. (In the Pilot Experiment, the Set-Size was either 8 or 16, while the Pool-Size was 17.)

To preview the results, I found significant PI under these conditions, suggesting that participants were sensitive to the ratio between the set-size and the total number of items and thus that the strength of PI is determined by the waiting time between subsequent occurrences of simple items rather than item-location combinations.

In Experiment 3, I ask whether spatial information *per se* has an effect on memory performance and susceptibility to PI. Further, I asked if binding items to spatial location is automatic. Participants viewed sequences of 15 items, followed by a single item test image. In different blocks, the items appeared either (1) at the center of the screen, (2) on an imaginary circle with items proceeding in a clock-wise direction or (3) on an imaginary circle where locations were chosen in a random order. Within each block, there was a sub-block where items were *trial-unique* and a sub-block where items were repeatedly drawn from a total pool of 16 items. I asked if performance differed depending on whether items were spatially distributed and depending on whether participants can attentionally anticipate where an item will appear.

To preview the results, I found that spatially distributing items impaired memory but did not affect the strength of PI; in contrast, the predictability of item location affected neither memory nor the strength of PI.

The final demographic information in the sample is given in Table 1.

3.2 Apparatus

Stimuli were presented on a Dell P2213 22" (55.88 cm) LCD (resolution: 1650×1050 pixels at 60 Hz), using the Matlab psychophysics toolbox (Brainard, 1997; Pelli, 1997) on a Mac Mini computer (Apple Inc., Cupertino, CA). Responses were collected from pre-marked "Yes" and "No" keys on the keyboard.

3.3 Materials

Stimuli were color pictures of everyday objects taken from Brady, Konkle, Alvarez, and Oliva (2008). These were randomly selected for each participant from a set of 2,400. Stimuli were presented on an imaginary circle with a radius of 190 pixels (except in one condition in Experiment 3, as noted below), corresponding to a viewing angle of 5.1 dva at a typical viewing distance of 60 cm.

Images were scaled to the maximal size so that they would not overlap on the circle.[9] In Experiment 1, items were thus scaled to a size of 130×130 pixels (3.5 dva), while items were scaled to a size of 70×70 pixels (1.9 dva) in Experiments 2 and 3.

3.4 Procedure

As shown in Figure 2, trials started with a screen indicating that the stimuli were being loaded for as long as they were being loaded (i.e., in general the screen was invisible). Following this, participants had to press the space bar to continue (except in Experiment 2, where I added an extra 700 ms to the fixation instead), followed by a fixation cross presented for 300 ms, a blank screen presented for 200 ms and then the sample sequence. Sample images were presented for 250 ms each, either on an imaginary circle or at the center of the screen (*Center* Condition in Experiment 3).

Following the sample sequence, a blank screen was shown for 900 ms, followed by the test item. Test items were presented for 800 ms but participants had unlimited time to respond.

Test items had appeared in the sequence on half of the trials; on these trials, they appeared in the spatial position in which they had appeared in the sample sequence. "Old test items" were chosen equally often from two sequence-initial, two sequence-medial and two sequence-final positions, respectively.

A new trial started immediately after a participant response.

Verbal suppression was not administered because earlier research has shown that neither memory nor PI is affected by verbal suppression at the current presentation rates (Endress & Siddique, 2016).

3.5 Analysis strategy

The analyses below will be based on two types of measures. First, I will seek to analyze performance in terms of (1) memory per se and (2) susceptibility

[9] Items will not overlap if the circles inscribing them are more distant than their diameter. The distance between two points on a circle with radius r is given by $d = 2r \sin(\alpha/2)$ (where α is the angle between the two points on the circle); the diameter of the circle inscribing the objects is $\sqrt{2}x$ (where x is the edge length of the square). To make sure that the pictures do not overlap, the condition $d > \sqrt{2}x$ thus needs to be satisfied, or, equivalently with N pictures, $x < \frac{2}{\sqrt{2}} \sin(180/N) r$.

Table 1
Demographics of the final sample after exclusion criteria have been applied.

Experiment	N	Females	Males	Mean age	Age range
Exp. 1	30	16	14	27.7	18–43
Exp. 2	30	20	10	23.9	18–57
Exp. 3	60	38	22	23.5	17–45
Pilot Exp.	32	21	11	24.4	18–44

to PI. Second, I will analyze the data in terms of raw accuracy.

3.5.1 Memory and susceptibility to PI. Memory per se will be operationalized as the performance in the *Unique Condition*. This condition should be a relatively pure measure of memory performance, as participants just need to encode, store and retrieve items from memory in the absence of interference (other than the fact to have completed other trials with other stimuli).

Susceptibility to PI will be operationalized as the relative Cost of PI (see Endress & Siddique, 2016), defined as

$$\text{Cost of PI} = \frac{\text{Unique} - \text{Repeated}}{\text{Unique}}$$

This measure gives the relative performance decrement in the *Repeated Condition* compared to the *Unique Condition*, normalized by the performance in the *Unique Condition*. For example, if a participant has an accuracy of 80% in the *Unique Condition* and 60% in the *Repeated Condition*, the relative performance decrement due to PI is (80% – 60%) / 80% = 25%.

As shown in Table 2, some cells in some experiments do not meet the assumption of normality. As my predictions for the first set of analyses involve pairwise comparisons, I seek to apply the same statistical tests across experiments and thus use pairwise Wilcoxon tests in all experiments instead of using Gaussian-based statistics in those experiments where the assumptions are met. I note, however, that Gaussian-based statistics would give the similar results.

3.5.2 Raw accuracy. The second set of analyses involves performance in terms of accuracy in the different conditions, using generalized linear mixed models (GLMMs) treating the accuracy in individual trials as a binary random variable. The advantage with respect to the first set of analyses is that I can jointly analyze memory and susceptibility to PI; the disadvantage is that the measure of susceptibility to PI (i.e., performance in the repeated condition) is contaminated by contributions from memory per se.

4 Experiment 1: The role of PI strength

In Experiment 1, I asked if PI effects are more likely to be observed with stronger PI. I manipulated the strength of PI by manipulating the size of the total pool from which stimuli could be drawn. With smaller pools, items are repeated more frequently, which should lead to stronger PI in turn.

4.1 Materials and methods

In all trials, 8 items were presented on an imaginary circle as described above. Critically, across blocks, the size of the pool from which items could be drawn was set to infinite (i.e., in the unique condition), 22 or 9.

The order of blocks was counterbalanced across participants. A third of the participants completed

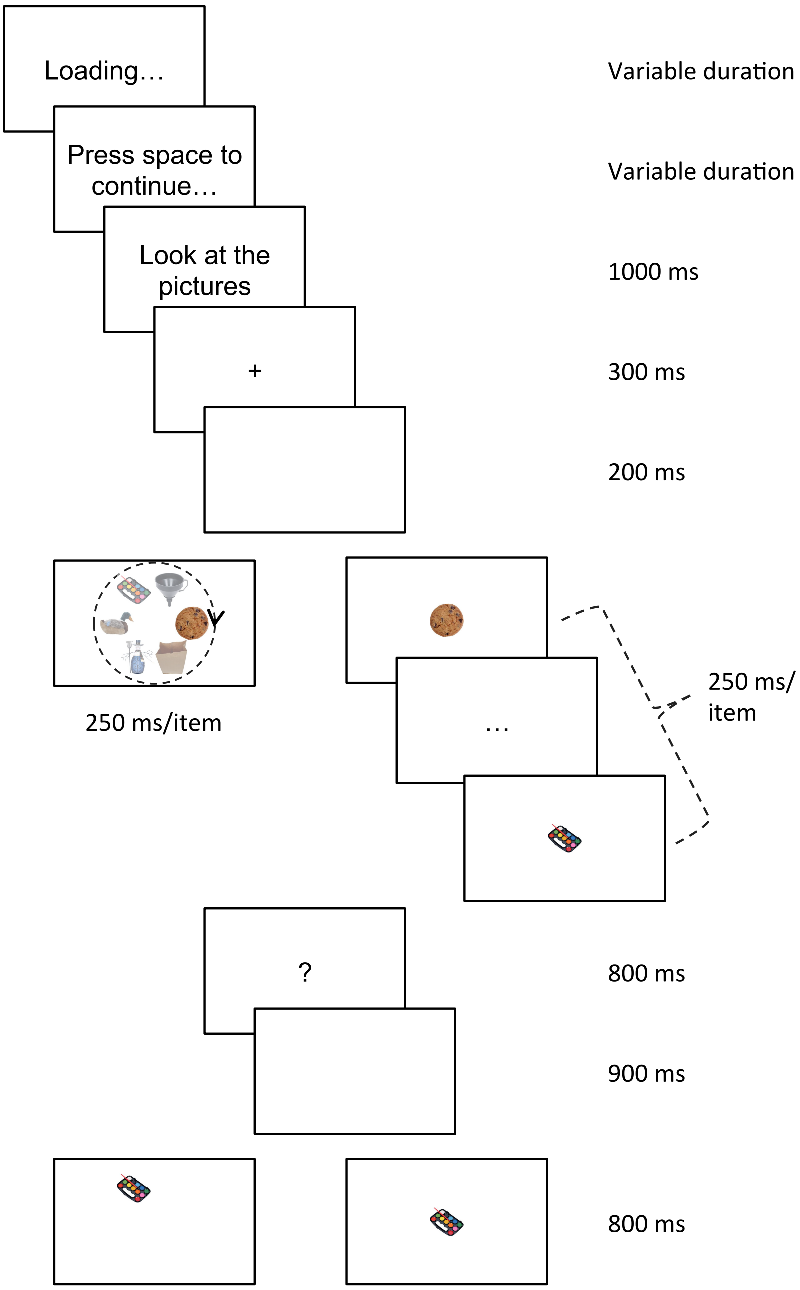


Figure 2. Trial schedule in the different experiments. Trials started with a (usually invisible) screen indicating that the stimuli were being loaded, followed by a button-press to start a trial, a fixation cross, a blank screen and finally the sample sequence. Samples were either presented on an imaginary circle (left) or at the center of the screen (right). After the sample sequence, participants saw a question mark, followed by a blank screen and finally the test item. If the test item had been part of the sample sequence, it appeared in its original position.

Table 2
Cells across experiments where a violation of normality was detected by a Shapiro-Wilk test when performance was measured in terms of accuracy and the Cost of PI, respectively.

Experiment	*Set-Size*	*Location Condition*	*PI Condition*	W	*p*
Accuracy					
Exp. 1	8	NA	Unique	0.88	0.002
Pilot Exp.	16	NA	Unique	0.93	0.036
Cost of PI					
Exp. 3	15	Circle - Random	NA	0.953	0.028
Pilot Exp.	16	NA	NA	0.919	0.017

the experiment in each of the Pool-Size orders ∞–*22–9*, *9–22–∞* and *22–∞–9*. Each block comprised 84 trials. Participants could take a break after each block. Before starting the experiment, participants were given two training trials. (There were no training trials in the other experiments.)

4.2 Results

I first analyze the results in terms of the performance in the *Unique Condition* and in terms of the Cost of PI and then in terms of the raw accuracy in the *Unique Condition* and the *Repeated Condition*, respectively.

4.2.1 Analyses of memory vs. susceptibility to PI. As shown in Table 3 and Figure 3, performance in the *Unique Condition* was well above chance.

As shown in Table 3 and Figure 3, the Cost of PI differed from zero only when items came from a total pool of 9 items in total, but not when items came from a pool of 22 items. A paired Wilcoxon Test showed that the Cost of PI for differed significantly between *Pool-Sizes* 9 for *Pool-Size* 22, V = 76, p = 0.001, $CI_{.95}$ = -0.113, -0.029.

Taken together, these results suggest that PI needs to be sufficiently strong to have a noticeable effect. When, as in Makovski's (2016) experiments, 8 items were picked from a total pool of 22 items, no PI effects were observed. In contrast, when the total Pool-Size was limited to 9 items, a sizeable PI effect emerged (Cohen's d = .618).

4.2.2 Analysis in terms of accuracy. As shown in Figure 4, performance in terms of raw accuracy was better in the *Unique Condition* than when items were drawn from a total pool of 9 items, while performance was similar in *Unique Condition* and for *Pool-Size* 22.

These results were confirmed in series of Generalized Linear Mixed Models, treating the trial-by-trial accuracy as a binary random variable. I first fit a model with the random factor *Participants* and the fixed factor *Pool-Size*, treating the *Unique Condition* as the reference level. As shown in Table 4, performance was better in the *Unique Condition* than for *Pool-Size* 9, while the *Unique Condition* did not differ statistically from the *Pool-Size* 22 condition.

Further, and as shown in Table 4 (bottom), a model fit to the data after removing the *Unique Condition* showed that performance was significantly worse in the *Pool-Size* 9 condition than in the *Pool-Size* 22 condition.

4.3 Discussion

The results of Experiment 1 show that PI effects are readily observed with spatially distributed items when PI is strong enough: When, on each trial, 8 items are shown on an imaginary circle, perfor-

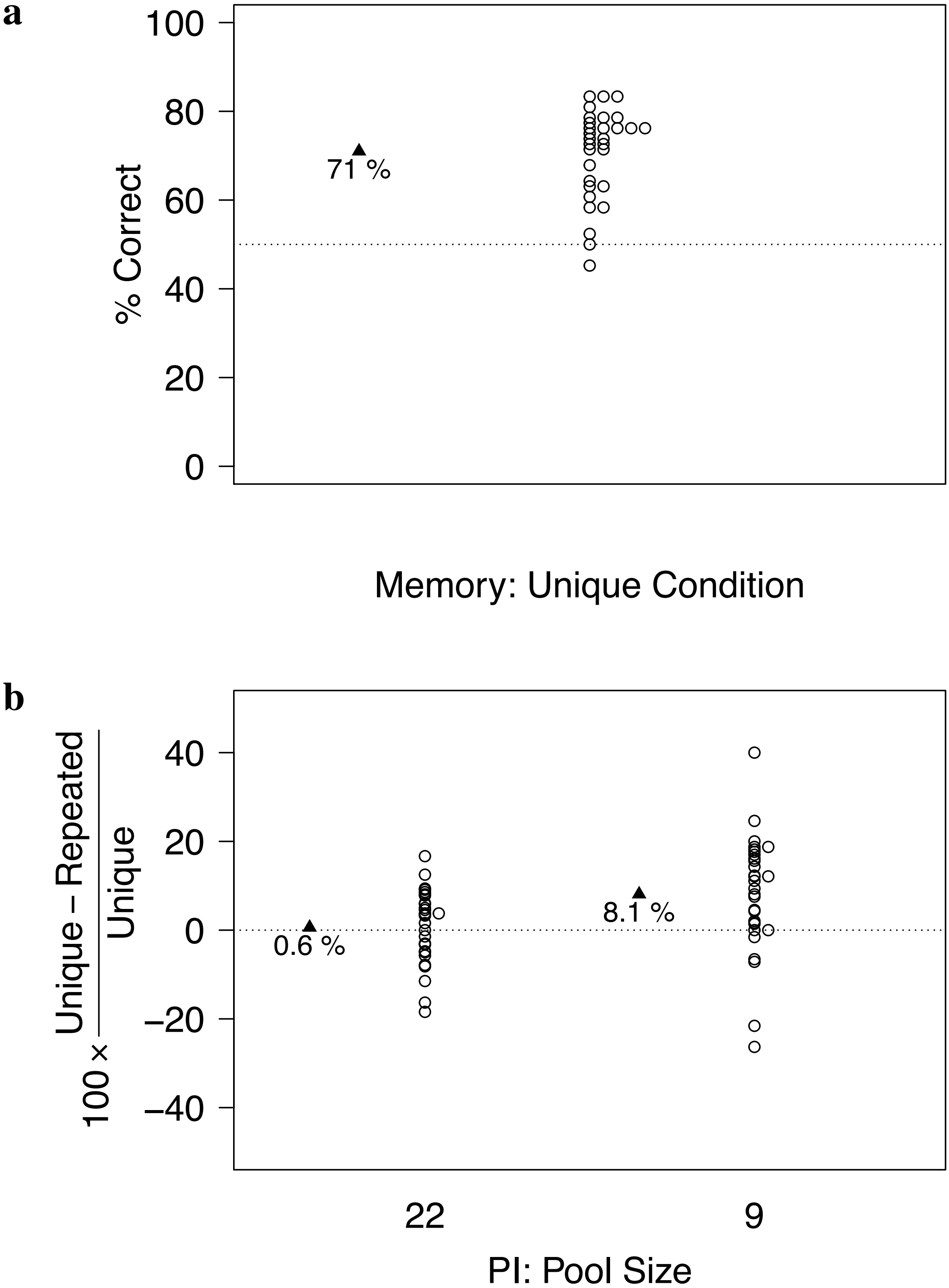


Figure 3. Results of Experiment 1 in terms of the performance in the *Unique Condition* (a) and the relative *Cost of PI* (b).

Table 3
Descriptive statistics in terms of raw accuracy and Cost of PI in Experiment 1. $p_{Wilcoxon}$ indicates the p *values of a Wilcoxon test against the chance levels of 50% (accuracy) and 0 (Cost of PI), respectively.*

Pool-Size	N	*M*	*SD*	*SE*	$p_{Wilcoxon}$	Cohen's *d*
Accuracy						
∞	30	0.706	0.102	0.019	< .001	2.03
22	30	0.699	0.099	0.018	< .001	2.02
9	30	0.641	0.087	0.016	< .001	1.62
Cost of PI						
22	30	0.006	0.082	0.015	0.430	0.073
9	30	0.081	0.132	0.024	0.002	0.618

Table 4
Results of a generalized linear model for Experiment 1, with the fixed factor Pool-Size*, trating the* Unique Condition *as a baseline. Performance was impaired for Pool-Size 9, but not for Pool-Size 22. The results show two models that include (top) or exclude (bottom) the* Unique Condition.

Effect	Estimate	Std. Error	*CI*	*t*	*p*
Model including the *Unique Condition*					
Pool-Size 22	-0.035	0.062	-0.158, 0.0873	-0.563	0.574
Pool-Size 9	-0.307	0.061	-0.427, -0.187	-5.025	< 0.001
Model excluding the *Unique Condition*					
Pool-Size 9	-0.271	0.061	-0.39, -0.152	-4.46	< 0.001

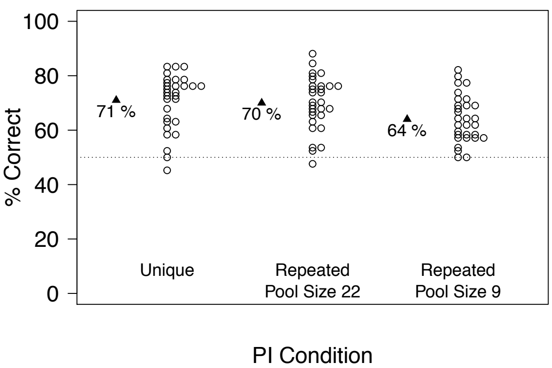


Figure 4. Results of Experiment 1 in terms of raw accuracy. When items were drawn from a total pool of 22 items, performance was undistinguishable from performance in the Unique Condition, where items were trial-unique. In contrast, when items were drawn from a total pool of 9 items, a sizeable PI effect emerged.

mance is impaired when items are taken from a limited pool of 9 items compared to when they are trial-unique. In contrast, when items come from a pool of 22 items in total, performance is equivalent to the Unique Condition.

Observers are thus sensitive to the strength of PI. However, the results of Experiment 1 are ambiguous as to whether PI occurs among the representations of simple items or rather among representations of item-location combinations. As discussed in the Introduction, these possibilities make different predictions with respect to the role of the set-size. If participants represent simple items, the strength of PI should be determined by the ratio between the set-size on each trial and the total number of possible items. After all, this ratio determines for how many trials we need to wait before see a given item again.

5.2.1 Analyses of memory vs. susceptibility to PI.

As shown in Figure 5 and Tables 5 and 6, both the performance in the *Unique Condition* and the Cost of PI were well above chance, but neither measure was affected by the set-size. Accordingly, a paired Wilcoxon test revealed no difference between the set-sizes, neither for accuracy in the *Unique Condition*, $V = 169$, $p = 0.871$, $CI_{.95}$ = -0.04, 0.03, nor for the Cost of PI, $V = 239$, $p = 0.903$, $CI_{.95}$ = -0.06, 0.06.

5.2.2 Analysis in terms of accuracy.

As shown in Figure 6 and Table 5, performance in terms of raw accuracy differed across the *PI Conditions*, but was unaffected by the *Set-Size*. This was confirmed by a generalized linear model with the within-subject predictors *PI Condition* and *Set-Size*, treating the trial-by-trial accuracy as binary random variables. Following Baayen, Davidson, and Bates (2008), I then removed the interaction term from the model as it did not contribute to the model likelihood. The results are shown in Table 7.

While performance was significantly worse for the Repeated Condition, the difference between the Set-Sizes was not significant.

Given the somewhat surprising absence of a difference between the set-sizes, I calculated the likelihood ratio in favor of the null hypothesis in the following way. First, for each participant and set-size, I averaged the accuracy, and, for each participant, subtracted the averages for the Set-Sizes from each other. A Shapiro-Wilk test did not detect any deviations from normality for these differences. Then, following Glover and Dixon (2004), I calculated the likelihood ratio of the hypotheses that (1) the differences were not different from zero and (2) that that they were. The null hypothesis was 3.148 more likely than the alternative hypothesis after correction with the Akaike information criterion, and 5.455 more likely after correction with the Bayesian Information criterion.

5.3 Discussion

The results of Experiment 2 revealed reliable PI effects with total pool-size of 21 items when the set-sizes were increased to 12 or 20 items, compared to 8 items in Experiment 1 and in Makovski's (2016) experiments. As the total pool-size was similar, one would not expect PI to be observed if participants had encoded item-location combinations, as they should not be sensitive to the set-size. The results of Experiment 2 thus clearly rule out that PI is driven exclusively by the total pool-size, and thus by interference between item-location combinations.

However, the results of Experiment 2 are also problematic for the most straightforward version of an account where participant remember simple items. After all, if PI strength is monotonically related to the waiting time between simple items, one would expect a further increase in PI between Set-Sizes 12 and 20.

In the Unique Condition, the lack of an effect of set-size is consistent with earlier results (e.g., Endress & Potter, 2014a). In the Repeated Condition, in contrast, this finding is unexpected, but has a relatively straightforward (post-hoc) interpretation in terms of dose-effect relations in PI.

As mentioned above, the role of spatial information in reducing PI might not be to allow participants to encode item-location combinations, but rather to use spatial locations as retrieval cues to initiate memory searches during test. In the party analogy above, when trying to decide who among our friends was present at a specific party, we might use the party as a retrieval cue to run a memory search through our friends.

However, such a memory search will not succeed if we see our friends too often; if we see them every day anyhow, it might be hard to decide if we have seen them at some party *as well*. Further, the level of interference in a memory search does not necessarily show a linear relationship with the frequency of meeting friends. For example, it is possible that

Table 5
Descriptive statistics in terms of raw accuracy for Experiment 2. $p_{Wilcoxon}$ *indicates the* p *values of a Wilcoxon test against the chance level of 50%.*

PI Condition	*Set-Size*	N	*M*	*SD*	*SE*	$p_{Wilcoxon}$	Cohen's *d*
Repeated	12	30	0.616	0.057	0.010	< .001	2.05
Repeated	20	30	0.616	0.068	0.012	< .001	1.70
Unique	12	30	0.656	0.066	0.012	< .001	2.36
Unique	20	30	0.658	0.061	0.011	< .001	2.60

Table 6
Descriptive statistics in terms of the Cost of PI for Experiment 2. $p_{Wilcoxon}$ *indicates the* p *values of a Wilcoxon test against the chance level of 0.*

Set-Size	N	*M*	*SD*	*SE*	$p_{Wilcoxon}$	Cohen's *d*
12	30	0.053	0.116	0.021	0.024	0.461
20	30	0.056	0.131	0.024	0.026	0.430

a certain meeting frequency is required to make a memory search hard, but that the search difficulty increases much more slowly afterwards.

Endress and Szabó (2017) made a similar point for more theoretical reasons. They noted that people need to search through some memory space and that this memory space might be more or less crowded. To the extent that memory spaces function similarly to other representational spaces, they pointed out that visual crowding occurs only if items are closer than some critical distance (e.g., Pelli, Palomares, & Majaj, 2004; van den Berg, Roerdink, & Cornelissen, 2007); if items are closer than that distance, recognition is impaired, but not if items are more separated.

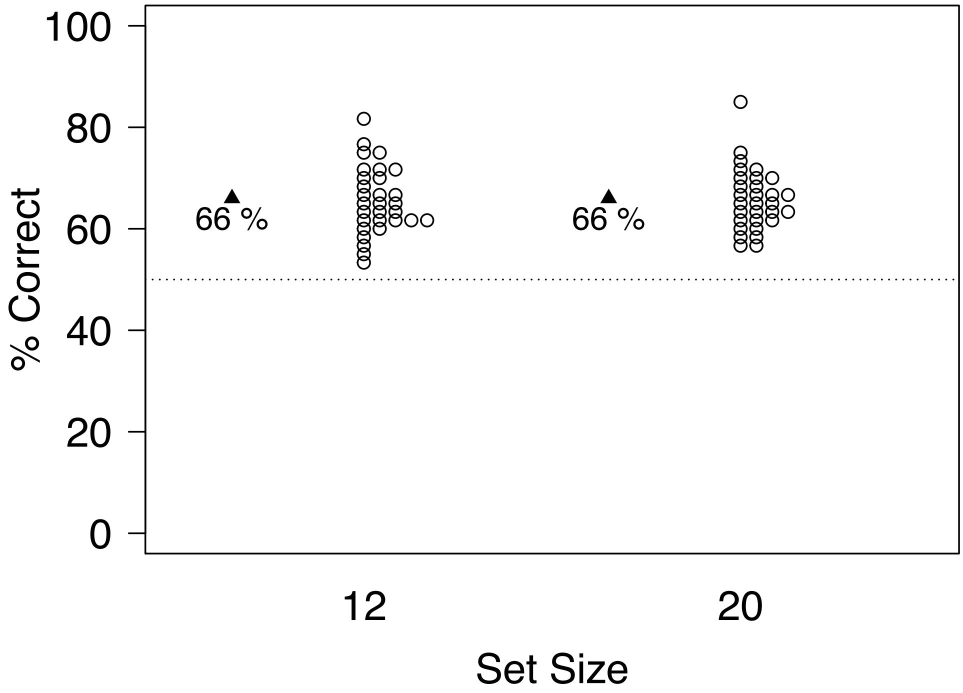
If a similar situation holds for memory, the strength of PI (as measured by the waiting time between consecutive occurrences of the same item) would need to be greater than some critical value for interference effects to be observable, but, beyond this critical value, the strength of PI might increase much more slowly. As a result, with pool-size of 21 items, a set-size of 12 items might be sufficient to establish PI even for spatially distributed items; once this critical value is crossed, however, changes in the strength of PI might be much slower.

However, before accepting this (post-hoc) explanation, one needs to directly and parametrically investigate the Cost of PI as a function of the set-size and the pool-size. Be that as it might, the combined results of Experiments 1 and 2 rule out that participants rely on item-location combinations, and are consistent with the view that spatial cues might act as retrieval cues as long as PI is not too strong. However, more parametric research is needed to find out what exactly determines the strength of PI.

6 Experiment 3: The role of spatial distribution and predictability

In Experiment 3, I asked if PI effects are affected by the spatial distribution of the memory items has an effect of on memory performance and susceptibility to PI. I kept the set-size and the pool-size constant (with a high ratio between the set-size and the pool-size) and presented items either centrally on the screen, on an imaginary circle in a predictable sequence, or on an imaginary circle in random order.

a



b

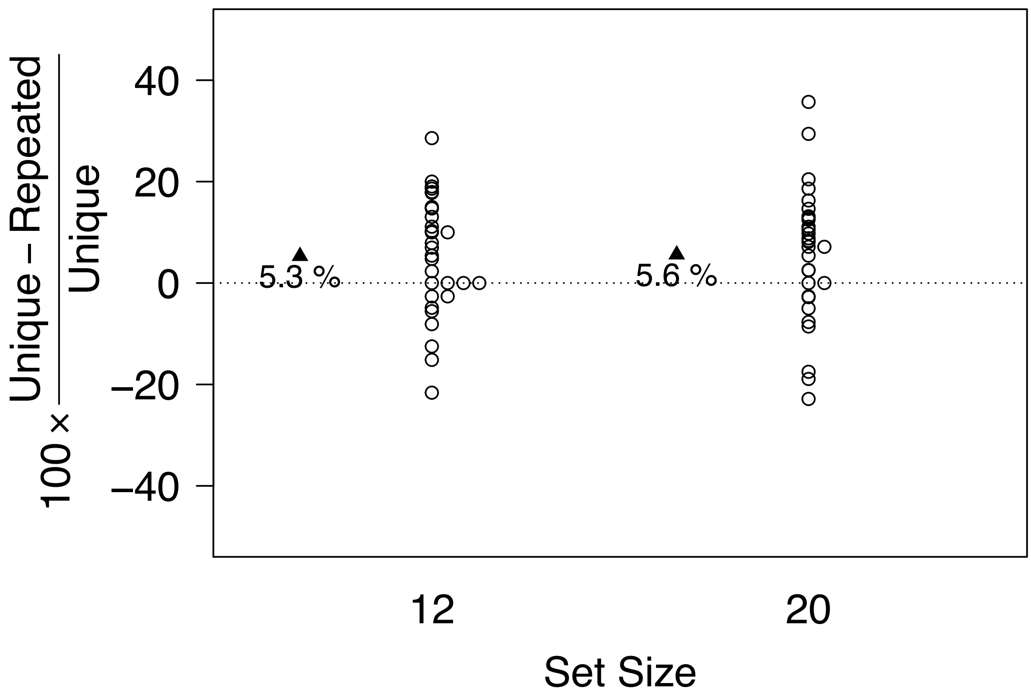


Figure 5. Results of Experiment 2 in terms of the accuracy in the *Unique Condition* (a) and of the relative *Cost of PI* (b). Neither measure appears to be affected by the Set-Size.

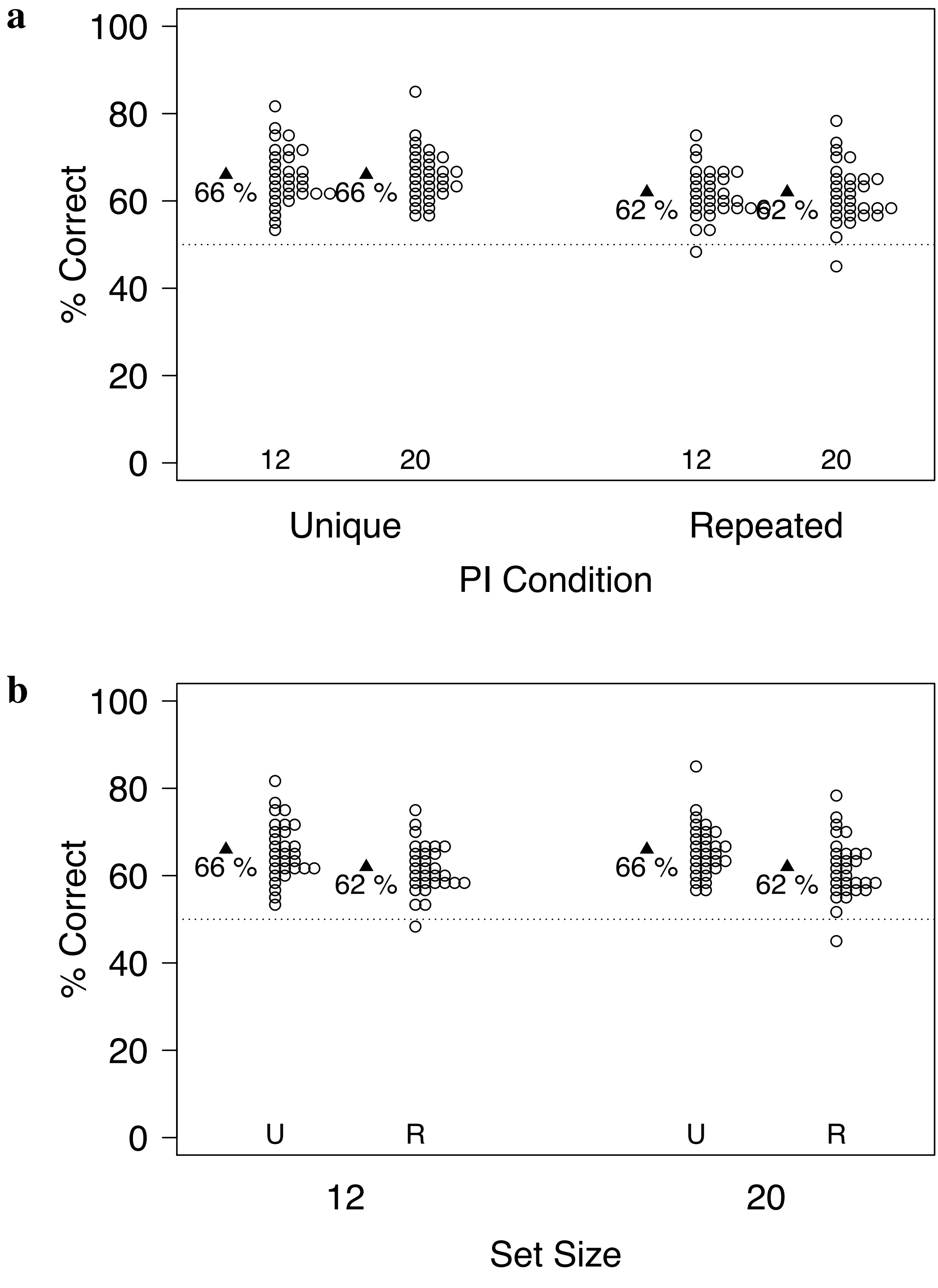


Figure 6. Results of Experiment 2 in terms of raw accuracy. Both figures show the same data, grouped by the *PI Conditions* (a) or the *Set-Sizes* (b). While PI impairs memory performance, memory performance is largely unaffected by the Set-Size.

Table 7
Results of a generalized linear model for Experiment 2. I treated the Unique Condition *and a* Set-Size *of 12 as the reference levels for the two predictors.*

Effect	Estimate	Std. Error	*CI*	*t*	*p*
PI Condition: Repeated	-0.177	0.049	-0.273, -0.0808	-3.604	0.000
Set-Size: 20	0.004	0.049	-0.0926, 0.0999	0.074	0.941

6.1 Materials and methods

In all trials, participants viewed a sequence of 15 objects. Critically, across blocks, items were presented either at the center of the screen (in the *Center Condition*), on an imaginary circle where items proceeded in a clockwise direction (in the *Circle-Ordered Condition*) or on an imaginary circle where items positions on the circle were chosen at random (in the *Circle-Random Condition*). In both circle conditions, the position of the first sample item was randomly chosen.

Each block comprised a sub-block with trial-unique items and a sub-block where items were drawn from a pool of 16 items in total.

The order of blocks and sub-blocks was counterbalanced across participants. Specifically, I used all 6 possible orders of the three blocks; further, for half of the participants, the *Repeated Condition* (within each block) preceded the *Unique Condition*, while the order was reversed for the remaining participants, leading to 12 counterbalancing conditions in toal.

Each of the 6 blocks (3 location conditions $\times$ 2 PI conditions) comprised 48 trials; participants were offered the opportunity to take a break every 24 trials. Experiment 3 comprised 288 trials in total.

6.2 Results

I first analyze the results in terms of the performance in the *Unique Condition* and in terms of the Cost of PI and then in terms of the raw accuracy in the *Unique Condition* and the *Repeated Condition*, respectively.

6.2.1 Analyses of memory vs. susceptibility to PI. As shown in Figure 7 and Table 8, memory performance in the *Unique Condition* was better when items were presented at the center of the screen compared to the two conditions where they were spatially distributed, with no difference between the latter two conditions. In contrast, the Cost of PI was not affected by the *Location Condition* manipulation all.

This impression was confirmed by pairwise Wilcoxon tests among the conditions. As shown in Table 10, performance in the *Unique Condition* was better in the *Center* Condition than in the *Circle Random* Condition and than in the *Circle Ordered* Condition. In contrast, performance did not differ significantly among the two circle conditions, nor were there any statistically significant differences in terms of the Cost of PI.

In sum, memory (as operationalized in the Unique Condition) was impaired when items were spatially distributed, with little difference between predictable and random positions. In contrast, the strength of PI was largely unaffected by these manipulations.

I further investigate these results in two ways. First, I split the *Location Condition* predictor into *two* predictors, indicating whether the spatial location of items was predictable (i.e., in the Center and the Ordered conditions) or not (in the Random condition), and whether the items were spatially distributed or not. I fit models to the data (separately for the

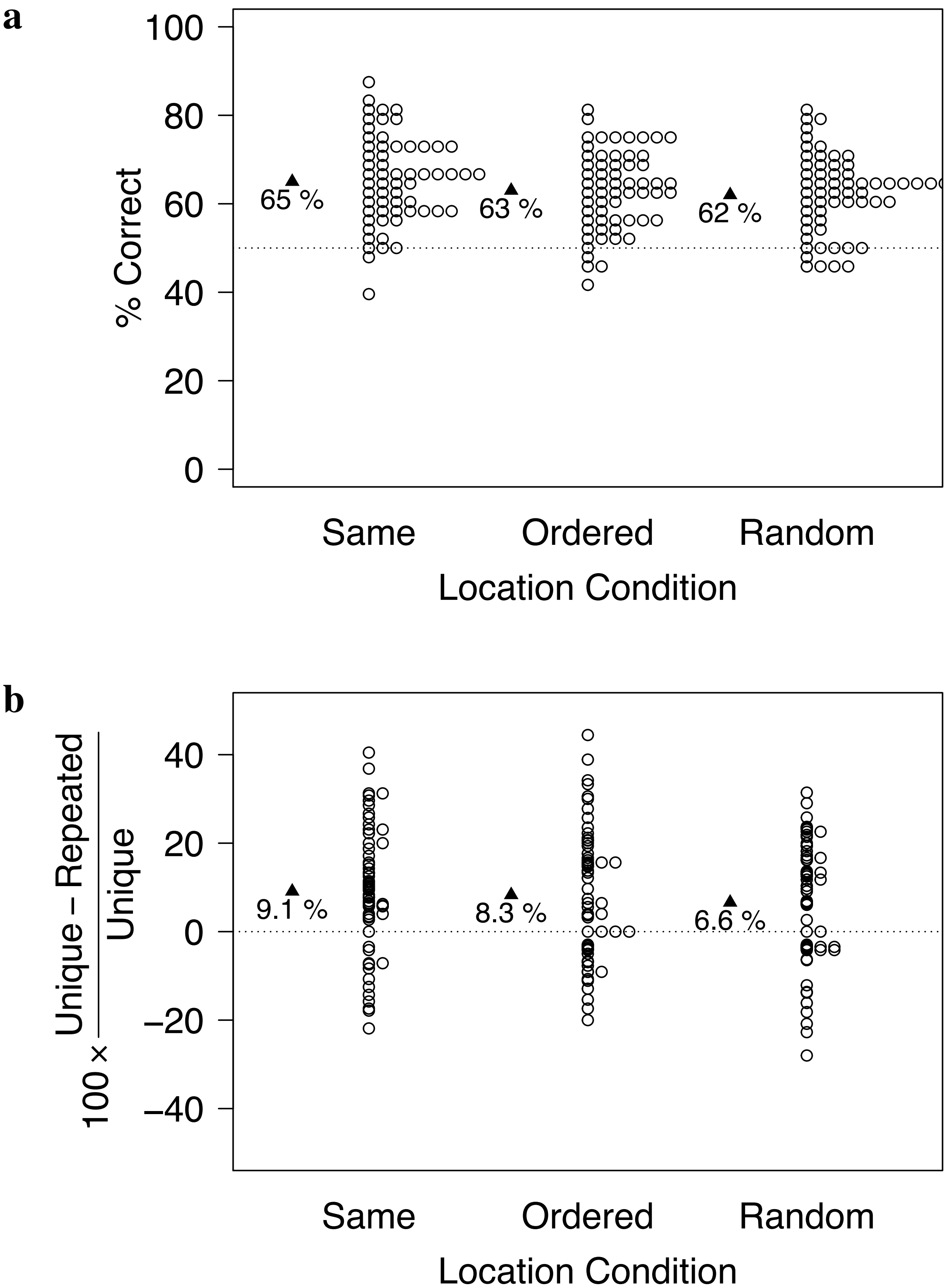


Figure 7. Results of Experiment 3 in term of the raw performance in the *Unique Condition* (a) and the relative Cost of PI (b). While performance in the Unique Condition was better when items were presented at the center of the screen than when they were spatially distributed, the Cost of PI was relatively unaffected by these spatial manipulations.

Table 8
Descriptive statistics in terms of raw accuracy for Experiment 3. $p_{Wilcoxon}$ indicates the p *values of a Wilcoxon test against the chance level of 50%.*

PI Condition	*Location Condition*	N	*M*	*SD*	*SE*	$p_{Wilcoxon}$	Cohen's *d*
Unique	Center	60	0.655	0.099	0.013	< .001	1.557
Unique	Circle - Random	60	0.624	0.088	0.011	< .001	1.413
Unique	Circle - Ordered	60	0.631	0.089	0.012	< .001	1.459
Repeated	Center	60	0.586	0.080	0.010	< .001	1.072
Repeated	Circle - Random	60	0.576	0.083	0.011	< .001	0.922
Repeated	Circle - Ordered	60	0.570	0.079	0.010	< .001	0.879

Table 9
Descriptive statistics in terms of the Cost of PI for Experiment 3. $p_{Wilcoxon}$ indicates the p *values of a Wilcoxon test against the chance level of 0.*

Location Condition	N	*M*	*SD*	*SE*	$p_{Wilcoxon}$	Cohen's *d*
Center	60	0.091	0.143	0.019	< .001	0.638
Circle - Ordered	60	0.083	0.150	0.019	< .001	0.554
Circle - Random	60	0.066	0.141	0.018	0.001	0.468

unique condition and the Cost of PI) that included both predictors and models that included only a single one. (The models for the accuracy data treated data as binomial, while the models for the Cost of PI treated it as Gaussian, though similar results are obtained when both models assume Gaussian distributions.) Following this, I removed one of the predictors to establish whether it significantly contributed to the model likelihood.

As shown in Table 11, removing *Spatial Distribution* from the model significantly impaired the model fit for the Unique Condition, but not for the Cost of PI. In contrast, *Location Predictability* did not contribute to the model likelihood for either measure. These analyses thus confirm that memory is affected by spatially distributing memory items irrespective of whether their positions are predictable or not, while PI seems invariant under such manipulations.

To provide direct evidence for the null hypothesis, I investigated these results using likelihood ratios (Glover & Dixon, 2004) corresponding to these contrasts. Specifically, for each participant, I calculated the difference (1) between performance in the predictable conditions and performance in the unpredictable ones, and (2) between performance in the spatially distributed conditions and performance with central presentation of the items. I then asked if a model fitting a non-zero value to these differences would fit the data better than a model where the differences were fixed to zero, assuming the data were normally distributed and correcting for the different numbers of parameters (i.e., whether an intercept was fit) using Bayesian Information Criterion and the Akaike Information Criterion. (This approach is similar to an analysis using Bayes factors, except that it is frequentist and does not make arbitrary assumptions about the prior distribution of the effect sizes.)

The results are shown in Table 12. For the Unique Condition, the likelihood ratios strongly favored the alternative hypothesis in the case of Distributivity, while they were ambiguous as to whether *Location Predictability* affected memory performance.

For the Cost of PI, there was evidence favoring the null hypothesis: For both the alternative hypothesis that the Cost of PI is affected by (1) the presence of spatial information and (2) its predictability, the null hypothesis is about 5 times (BIC) or twice (AIC) as likely.

In other words, memory performance was clearly

Table 10
Pairwise Wilcoxon test for the Location Conditions in Experiment 3.

Measure	*Location Condition* 1	*Location Condition* 2	V	*p*	*CI*
Unique	Center	Circle - Ordered	1074	0.049	4.76e-05, 0.0625
Unique	Center	Circle - Random	994	0.030	3.41e-05, 0.0625
Unique	Circle - Ordered	Circle - Random	805	0.772	-0.0209, 0.0313
Cost of PI	Center	Circle - Ordered	978	0.645	-0.039, 0.0624
Cost of PI	Center	Circle - Random	1092	0.194	-0.0127, 0.0798
Cost of PI	Circle - Ordered	Circle - Random	949	0.632	-0.041, 0.0669

Table 11
Contributions of the predictability and the spatial distribution of the items. While the spatial distribution contributed to the model likelihood in terms of the accuracy in the Unique Condition*, predictability of item locations did not. In contrast, the* Cost of PI *was unaffected by either manipulation.*

Measure	Removed Predictor	χ^2	Df	*p*
Accuracy (Unique Condition)	Spatial Distribution	3.763	1	0.052
Accuracy (Unique Condition)	Location Predictability	0.272	1	0.602
Cost of PI	Spatial Distribution	0.128	1	0.721
Cost of PI	Location Predictability	0.494	1	0.482

impaired when items were spatially distributed, replicating Makovski's (2016) results. However, memory performance was fairly unaffected by the predictability of the item locations.

In contrast, the Cost of PI was unaffected by either factor. The latter result is somewhat unexpected, though it is consistent with previous finding that the strength of PI is surprisingly unaffected by temporal manipulations as well (Endress & Siddique, 2016).

6.2.2 Analysis in terms of accuracy. As shown in Figure 8, performance was better in the *Unique Condition* than in the *Repeated Condition*, and when the items were not spatially distributed than when they were, with little difference between the spatially distributed conditions.

I confirmed this impression using a generalized linear model with the within-subject predictors *PI Condition* and *Set-Size*, treating the trial-by-trial accuracy as a binary random variable. Following Baayen et al. (2008), I then removed the interaction term from the model as it did not contribute to the model likelihood. As shown in Table 13, performance was better in the *Unique Condition* than in the *Repeated Condition* and better in the *Center Condition* than in either of the *Circle Conditions*.

A model fit to the data after removing the *Center Condition* did not reveal a statistically significant difference among the *Circle Conditions* (see Table 13, bottom).

Taken together, these results provide no evidence for the hypothesis that spatially distributing items affects the strength of PI. They reveal indistinguishable PI effects across the *Location Conditions*. In contrast, these results do suggest that spatially distributing items impairs memory, irrespective of whether the spatial locations are predictable or not, though the performance cost is rather modest.

6.3 Discussion: Are there limits to the number of item-location bindings that can be encoded?

The results of Experiment 3 suggest that the susceptibility to PI is not affected by spatial information. This is consistent with the interpretation of Experiments 1 and 2 that spatial information provides retrieval cues when PI is not too strong, but that it does not lead to memory for item-location

Table 12
Likelihood ratios for the null *hypothesis in Experiment 3 after correction with the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Likelihood ratios below 1 favor the* alternative *hypothesis.*

Measure	Predictor	AIC	BIC
Accuracy (Unique Condition)	Predictability	0.658	1.746
Accuracy (Unique Condition)	Distributivity	0.128	0.341
Cost of PI	Predictability	1.731	4.597
Cost of PI	Distributivity	2.032	5.394

Table 13
Results of a generalized linear model for Experiment 3. For the PI Condition*, the reference level was the* Unique Condition*. For the* Location Condition*, the reference level was the* Center Condition *when it was present (top) and the* Circle-Ordered Condition *when the* Center Condition *was removed (bottom).*

Effect	Estimate	Std. Error	*CI*	*t*	*p*
Model including the *Center Condition*					
PI Condition: Repeated	-0.251	0.031	-0.312, -0.189	-7.98	0.000
Location Condition: Circle-Ordered	-0.087	0.038	-0.162, -0.0112	-2.25	0.024
Location Condition: Circle-Random	-0.087	0.038	-0.162, -0.0112	-2.25	0.024
Model excluding the *Center Condition*					
PI Condition: Repeated	-0.228	0.038	-0.303, -0.153	-5.96	0
Location Condition: Circle-Random	0.000	0.038	-0.075, 0.075	0.00	> .999

combinations.

However, an alternative interpretation of these results is that people face difficulties when encoding more than a few item-location bindings. This possibility might take four different (but not mutually exclusive) forms, but, as I will argue below, all of them are either refuted by the data or point to a limited role of space in the resolution of PI.

First, item-location bindings might automatic, but when there are too many locations, each location becomes less distinct and thus less useful. For example, when eight objects are presented on a circle, a resolution of 45° is sufficient, while one would need a resolution of 24° with 15 objects. Critically, however, if item-location bindings are automatic, the items should still be bound to some *approximate* location, even if the resolution is not sufficient to encode the location precisely. As a result, the effect of PI should still be reduced relative to when items are presented at the same central location, which was clearly not the case. The results of Experiment 3 thus rule out that the lack of an effect of spatial distribution is due to a limited precision of the location encoding system.

Second, item-location bindings might be automatic, but observers might only be able to maintain a limited number of bindings in memory, similarly to how some authors propose that we can maintain only a limited number of items in memory (e.g., Cowan, 2001; Fukuda et al., 2010; Hartshorne, 2008; Luck & Vogel, 1997; Rouder et al., 2008; Zhang & Luck, 2008). If so, one would predict that observers retain the *last* bindings. If binding is automatic, they keep binding items to locations as they experience more item-location combinations; but as they cannot retain all of them, earlier bindings are overwritten by later item-location combinations.

I tested the predictions of this account in two ways. (The detailed results are presented in Appendix B). In the first analysis, I consider the *Cost of PI* in the *Circle-Ordered Condition*. One would expect a main effect of the sequential position: the *Cost of PI* should be reduced in later positions compared to earlier positions. However, there was no trace of

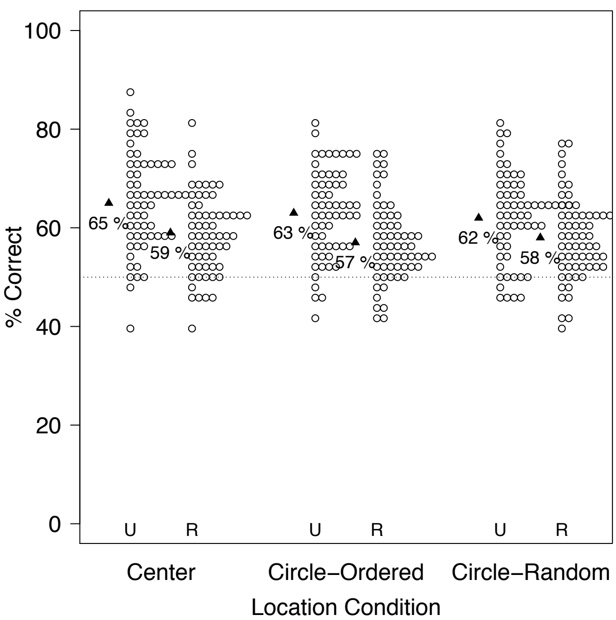


Figure 8. Results of Experiment 3 in terms of raw accuracy. Performance was better in the Unique Conditions than in the Repeated Conditions and when items were presented at the center for the screen than when they were spatially distributed. In contrast, the strength of the PI effect seemed unaffected by the spatial distribution of the items.

such an effect, and, in fact, likelihood ratio analysis provided evidence in favor of the null hypothesis.

In the second analysis, I consider the raw accuracy in the *Unique* and the *Repeated* Conditions of the *Center Condition* and the *Circle-Ordered Condition*, again as a function of the sequential position of the items. The critical prediction is a triple interaction between these factors. If spatial information reduces PI, one would expect an interaction between the *Location Condition* and the *PI Condition*. However, if observers cannot encode more than a few item-location bindings, this double interaction should be strongest for the most recent sequential positions, resulting in a triple interaction between the *Location Condition*, the *PI Condition* and the *Sequential Position*. Again, these predictions were unsupported.

The third way in which observers might face difficulties when encoding an excessive number of item-location bindings assumes that they automatically stop binding items to locations when there are more than a handful of locations. While the current experiments do not rule out this possibility, it does raise the question of why observers would not use a more approximate representation of the locations instead of forgoing this source of information altogether, and if an automatic item-location binding system really estimates the expected number of bindings even before the start of each trial to decide whether the bindings should be encoded.

Finally, and relatedly, it is possible that item-location binding is *not* automatic; rather, observers might strategically avoid encoding such bindings under some conditions. While this possibility is not ruled out by the data, it would suggest that location information does not necessarily reduce the effects of PI.

Taken together then, the possibility that people might face difficulties when encoding more than a few item-location bindings is either contradicted by the data or seems to suggest a limited role of spatial information for PI resolution.

7 General discussion

To what extent can PI be observed with spatially distributed items? Previous research has found only limited evidence for PI when items were spatially distributed and presented sequentially (though PI effects were reliable in very similar experiments where items were presented simultaneously; Makovski, 2016). This leads to two hypotheses about why PI is reduced when items were spatially distributed. On the one hand, people might store item-location combinations; as there are many more item-location combinations than simple items, this should reduce PI among the relevant memory representations, especially in experiments were PI was relatively weak to begin with. On the other hand, the introduction of spatial cues might change the task demands, in that they introduce retrieval cues that allow participants to perform memory searches. The critical predictions of these accounts are that (1) participants should be sensitive to both the set-size of items in a trial and the total pool-size if they store simple items, but that (2) they should be sensitive only to the total pool-size if they store item-location combinations.

I first show that PI is readily observed with spatially distributed items when it is sufficiently strong. Experiment 1 revealed that, compared to a condition with trial-unique items, performance is impaired when sequences of 8 items are drawn from a total pool of 9 items. In contrast, with a total Pool-Size of 21 items, performance is undistinguishable from performance with trial-unique items.

While Experiment 1 shows that PI exists with spatially distributed items, it leaves open the specific role of spatial cues. Experiment 2 attempted to address these issues by manipulating the set-size for spatially distributed items without changing the pool sizes. It revealed two critical results. First, when, for a pool-size of 21 or 22 items, the set-size is increased from 8 to 12, the Cost of PI increases from .6% (Cohen's d = .073) to 5.3% (Cohen's d = .461). Participants are thus sensitive to the set-size of the items, which rules out that they predominantly encode item-location combinations.

Second, when the set-size was further increases from 12 to 20, no further changes were observed, neither for memory performance nor for the Cost of PI. In terms of memory performance, this result is consistent with Endress and Potter's (2014a) finding that performance in the Unique Condition is relatively unaffected by the set-size, at least for larger set-sizes.[10] These results thus confirm that, in the absence of PI, we have a relatively large memory capacity for temporarily retaining information.

However, based on previous results, it is unclear why the Cost of PI did not further increase between Set-Sizes 12 and 20. As mentioned in the discussion of Experiment 2, one possibility is that the relationship between the waiting time between two occurrences of an item and the observable strength of PI is non-linear. PI effects might become observable once the waiting times between two occurrences of an item become too short, but there is little additional cost when they become even shorter.

Endress and Szabó (2017) made a similar suggestion based on crowding phenomena in visual perception (e.g., Pelli et al., 2004; van den Berg et al., 2007; Whitney & Levi, 2011). Just as visual items become hard to identify (but not to detect) when they are closer to surrounding items than a critical distance (e.g., Pelli et al., 2004; van den Berg et al., 2007; Whitney & Levi, 2011), there might be a critical distance in a memory space under which memory items are hard to identify. If the distance between items in that space is greater than that critical distance, PI might be hard to observe; but if the memory items become even more crowded, there might only be a limited cost in terms of *additional* PI. However, while such an account would unify representational principles of perception and memory, it is currently a post-hoc conjecture, and requires parametric studies of the strength of PI to be substantiated.

Finally, Experiment 3 investigated the role of spatial information of PI more directly. While, in line with previous experiments (Makovski, 2016), memory performance suffered when items were spatially distributed, spatially distributing item had no effect at all on the strength of PI. Interestingly, spatially distributing items impaired memory irrespective of whether the item locations were predictable or not, though it is unclear why participants did *not* perform better when the locations were predictable.

Taken together, the present results show that, just like other forms of memory, visual memory is susceptible to PI. This, in turn, raises the question of why visual Working Memory as implemented in

[10]In the pilot experiment presented in Appendix A, we did find an effect of Set-Size on performance in the Unique Condition, but this might be because the smaller set-size might have been relatively small.

change detection experiments is relatively *insensitive* to PI. The current results rule out that the role of spatial information is to make items more distinct by allowing observers to encode item-location combinations. However, spatial information might still have other effects.

One such possibility is that spatial cues provide participants with retrieval cues that allow them to perform memory searches. However, especially for the smaller set-sizes typically used in change detection experiments, spatial information might also allow observers to encode entire displays as configurations of items by binding together the items in the displays. If so, visual Working Memory limitations might reflect encoding rather than memory limitations (Tsubomi et al., 2013; see also Fukuda & Vogel, 2009, 2011). The memory representations are still likely to be susceptible to PI, but the PI effects might be relatively minor if the limitations due to encoding are much more stringent. This hypothesis would be in line with previous suggestions that an important function of Working Memory is to create temporary bindings (e.g., Bateman & Birney, 2019; Oberauer, Süß, Wilhelm, & Wittmann, 2008), though the bindings in the case of change detection experiments would be much more perceptual than those in earlier non-visual Working Memory experiments and would lead to a radically new view on the nature of visual Working Memory limitations in terms of encoding processes and strategies rather than memory per se.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. doi: 10.1037/0033-2909.131.1.30

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and iq in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20 - 29. doi: DOI:10.1016/j.jecp.2009.11.003

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390 - 412. doi: 10.1016/j.jml.2007.12.005

Baddeley, A. D. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13468-72.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173.

Bateman, J. E., & Birney, D. P. (2019). The link between working memory and fluid intelligence is dependent on flexible bindings, not systematic access or passive retention. *Acta psychologica*, *199*, 102893. doi: 10.1016/j.actpsy.2019.102893

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7.1–711. doi: 10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854. doi: 10.1126/science.1158023

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(2), 317-33. doi: 10.1037/a0014873

Bouffard, N., Stokes, J., Kramer, H. J., & Ekstrom, A. D. (2018). Temporal encoding strategies result in boosts to final free recall performance comparable to spatial ones. *Memory & cognition*, *46*, 17–31. doi: 10.3758/s13421-017-0742-z

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329. doi: 10.1073/pnas.0803390105

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Carroll, L. M., Jalbert, A., Penney, A. M., Neath, I., Surprenant, A. M., & Tehan, G. (2010). Evidence for proactive interference in the focus of attention of working memory. *Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale*, *64*, 208–214. doi: 10.1037/a0021011

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547 - 552. doi: DOI:10.1016/j.tics.2003.10.005

Cortis Mack, C., Dent, K., & Ward, G. (2018). Near-independent capacities and highly constrained output orders in the simultaneous free recall of auditory-verbal and visuo-spatial stimuli. *Journal of experimental psychology. Learning, memory, and cognition*, *44*, 107–134. doi: 10.1037/xlm0000439

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

Cowan, N. (2005). *Working memory capacity*. Hove, UK: Psychology Press.

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology. General*, *143*(5), 1806–1836. doi: 10.1037/a0036814

Craig, K. S., Berman, M. G., Jonides, J., & Lustig, C. (2013). Escaping the recent past: which stimulus dimensions influence proactive interference? *Memory & cognition*, *41*, 650–670. doi: 10.3758/s13421-012-0287-0

D'Esposito, M., Postle, B. R., Jonides, J., & Smith, E. E. (1999). The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional mri. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(13), 7514–7519.

Endress, A. D., & Potter, M. C. (2012). Early conceptual and linguistic processes operate in independent channels. *Psychological Science*, *23*(3), 235–245. doi: 10.1177/0956797611421485

Endress, A. D., & Potter, M. C. (2014a). Large capacity temporary visual memory. *Journal of Experimental Psychology. General*, *143*(2), 548–65. doi: 10.1037/a0033934

Endress, A. D., & Potter, M. C. (2014b). Something from (almost) nothing: Buildup of object memory from forgettable single fixations. *Attention, Perception and Psychophysics*, *76*(8), 2413–2423. doi: 10.3758/s13414-014-0706-3

Endress, A. D., & Siddique, A. (2016). The cost of proactive interference is constant across presentation conditions. *Acta Psychologica*, *170*, 186 - 194. doi: dx.doi.org/10.1016/j.actpsy.2016.08.001

Endress, A. D., & Szabó, S. (2017). Interference and memory capacity limitations. *Psychological Review*, *124*(5), 551–571. doi: 10.1037/rev0000071

Engel de Abreu, P. M., Conway, A. R., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, *38*(6), 552 - 561. doi: 10.1016/j.intell.2010.07.003

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19-23. doi: 10.1111/1467-8721.00160

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin and Review*, *17*(5), 673–679. doi: 10.3758/17.5.673

Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*, 8726–8733. doi: 10.1523/JNEUROSCI.2145-09.2009

Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological science*, *22*, 361–368. doi: 10.1177/0956797611398493

Gardiner, J. M., Craik, F. I., & Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 778 - 783. doi: https://doi.org/10.1016/S0022-5371(72)80012-4

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16. doi: 10.1002/acp.934

Glover, S., & Dixon, P. (2004). Likelihood ratios: a simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin and Review*, *11*(5), 791–806.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*, 20–25. doi: 10.1016/0166-2236(92)90344-8

Hartshorne, J. K. (2008). Visual working memory capacity and proactive interference. *PLoS One*, *3*(7), e2716. doi: 10.1371/journal.pone.0002716

Hopkins, R. H., Edwards, R. E., & Cook, C. L. (1972). The dissipation and release of proactive interference in a short-term memory task. *Psychonomic Science*, *27*(2), 65–67. doi: 10.3758/BF03328890

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of experimental psychology. Learning, memory, and cognition*, *26*, 683–702.

Jitsumori, M., Wright, A. A., & Shyan, M. R. (1989). Buildup and release from proactive interference in a rhesus monkey. *Journal of experimental psychology. Animal behavior processes*, *15*, 329–337.

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, *139*(1), 181–193. doi: 10.1016/j.neuroscience.2005.06.042

Kam, C. L. H., & Newport, E. L. (2009). Getting it right by getting it wrong: when learners change languages. *Cognitive Psychology*, *59*(1), 30-66. doi: 10.1016/j.cogpsych.2009.01.001

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(2), 336-358.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, *1*(3), 153 - 161. doi: 10.1016/S0022-5371(62)80023-1

Kincaid, J. P., & Wickens, D. D. (1970). Temporal gradient of release from proactive inhibition. *Journal of Experimental Psychology*, *86*(2), 313 - 316.

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language*, *59*(2), 200 - 222. doi: https://doi.org/10.1016/j.jml.2008.04.004

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, *13*(3), 120–126. doi: 10.1016/j.tics.2008.12.003

Lin, P.-H., & Luck, S. J. (2012). Proactive interference does not meaningfully distort visual working memory capacity estimates in the canonical change detection task. *Frontiers in Psychology*, *3*, 42. doi: 10.3389/fpsyg.2012.00042

Loess, H., & Waugh, N. C. (1967). Short-term memory and intertrial interval. *Journal of Verbal Learning and Verbal Behavior*, *6*(4), 455 - 460. doi: DOI:10.1016/S0022-5371(67)80001-X

Loosli, S. V., Rahm, B., Unterrainer, J. M., Weiller, C., & Kaller, C. P. (2014). Developmental change in proactive interference across the life span: evidence from two working memory tasks. *Developmental psychology*, *50*, 1060–1072. doi: 10.1037/a0035231

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. doi: 10.1038/36846

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology. General*, *130*(2), 199–207.

Makovski, T. (2016). Does proactive interference play a significant role in visual working memory tasks? *Journal of experimental psychology. Learning, memory, and cognition*, *42*, 1664–1672. doi: 10.1037/xlm0000262

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory and Cognition*, *36*(1), 43–52.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory and Cognition*, *27*(5), 759–767.

Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, *412*(6845), 401. doi: 10.1038/35086646

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, *6*(1), 8–17. doi: 10.1167/6.1.2

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research*, *6*, 57–77. doi: 10.1016/0166-4328(82)90081-x

Nee, D. E., Jonides, J., & Berman, M. G. (2007). Neural mechanisms of proactive interference-resolution. *Neuroimage*, *38*(4), 740–751. doi: 10.1016/j.neuroimage.2007.07.066

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11–28.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: an interference model of complex span. *Psychonomic Bulletin and Review*, *19*(5), 779–819. doi: 10.3758/s13423-012-0272-4

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological review*, *124*, 21–59. doi: 10.1037/rev0000044

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. (2000). Working memory capacity — facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017 - 1045. doi: 10.1016/S0191-8869(99)00251-2

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*(6), 641– 652. doi: https://doi.org/10.1016/j.intell.2008.01.007

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437-42.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: distinguishing feature integration from detection. *Journal of Vision*, *4*(12), 1136–1169. doi: 10:1167/4.12.12

Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision*, *9*(10), 2.1–212. doi: 10.1167/9.10.2

Pertzov, Y., & Husain, M. (2014). The privileged role of location in visual working memory. *Attention, perception & psychophysics*, *76*, 1914–1924. doi: 10.3758/s13414-013-0541-y

Peterson, L. R., & Gentile, A. (1965). Proactive interference as a function of time between tests. *Journal of Experimental Psychology*, *70*(5), 473–478.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & cognition*, *1*, 19–40. doi: 10.3758/BF03198064

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509-22.

Potter, M. C. (2010). Conceptual short term memory. *Scholarpedia*, *5*(2), 3334. (revision #122765) doi: 10.4249/scholarpedia.3334

Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(3), 478-89. doi: 10.1037/0096-1523.30.3.478

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression,. *Journal of Memory and Language*, *39*(3), 418 - 436. doi: 10.1006/jmla.1998.2590

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 5975–5979. doi: 10.1073/pnas.0711295105

Rowe, G., Hasher, L., & Turcotte, J. (2010). Interference, aging, and visuospatial working memory: the role of similarity. *Neuropsychology*, *24*(6), 804–807. doi: 10.1037/a0020244

Schneegans, S., & Bays, P. M. (2016). No fixed item limit in visuospatial working memory. *Cortex*, *83*, 181–193. doi: 10.1016/j.cortex.2016.07.021

Shipstead, Z., & Engle, R. W. (2013). Interference within the focus of attention: working memory tasks reflect more than temporary maintenance. *Journal of experimental psychology. Learning, memory, and cognition*, *39*, 277–289. doi: 10.1037/a0028467

Shoval, R., Luria, R., & Makovski, T. (2019). Bridging the gap between visual temporary memory and working memory: The role of stimuli distinctiveness. *Journal of experimental psychology. Learning, memory, and cognition*. doi: 10.1037/xlm0000778

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability – and a little bit more. *Intelligence*, *30*(3), 261 - 288. doi: 10.1016/S0160-2896(01)00100-3

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*(6033), 1054–1059. doi: 10.1126/science.1196404

Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, *34*(8), 1704–1719. doi: 10.3758/BF03195932

Tsubomi, H., Fukuda, K., Watanabe, K., & Vogel, E. K. (2013). Neural limits to representing objects still within view. *The Journal of neuroscience*, *33*, 8257–8263. doi: 10.1523/JNEUROSCI.5348-12.2013

Udale, R., Farrell, S., & Kent, C. (2017). No evidence for binding of items to task-irrelevant backgrounds in visual working memory. *Memory & Cognition*, *45*(7), 1144–1159. doi: 10.3758/s13421-017-0727-y

Underwood, B. J., & Postman, L. (1960). Extraexperimental sources of interference in forgeting. *Psychological review*, *67*, 73–95. doi: 10.1037/h0041865

van den Berg, R., Roerdink, J. B. T. M., & Cornelissen, F. W. (2007). On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. *Journal of Vision*, *7*(2). doi: 10.1167/7.2.14

Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2011.02.005

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

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *2*(5–6), 440–445. doi: 10.1016/S0022-5371(63)80045-6

Yates, F. A. (1966). *The art of memory*. London: Routledge & Kegan Paul.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. doi: 10.1038/nature06860

Appendix A
Pilot Experiment: The role of the set-size (2)

The pilot experiment was similar to Experiment 2 (testing the role of the set-size), except that the set-sizes were 8 or 16 and that items were drawn from a total pool of 17 items.

A.1 Results and discussion

I first analyze the results in terms of the performance in the *Unique Condition* and in terms of the Cost of PI and then in terms of the raw accuracy in the *Unique Condition* and the *Repeated Condition*, respectively.

A.1.1 Analyses of memory vs. susceptibility to PI. As shown in Figure A1 and Tables A1 and A2, performance in the *Unique Condition* was well above chance for both Set-Sizes, while the Cost of PI was not significantly different from zero for either Set-Size, probably because, in the *Repeated Condition*, the set-sizes were too small relative to the size of the pool from which items were drawn.

Table A1
Descriptive statistics in terms of raw accuracy for the Pilot Experiment. $p_{Wilcoxon}$ *indicates the* p *values of a Wilcoxon test against the chance level of 50%.*

PI Condition	*Set-Size*	N	*M*	*SD*	*SE*	$p_{Wilcoxon}$	Cohen's *d*
Repeated	8	32	0.655	0.083	0.015	$< .001$	1.88
Repeated	16	32	0.620	0.065	0.011	$< .001$	1.85
Unique	8	32	0.678	0.080	0.014	$< .001$	2.21
Unique	16	32	0.635	0.093	0.017	$< .001$	1.44

Table A2
Descriptive statistics in terms of the Cost of PI for the Pilot Experiment. $p_{Wilcoxon}$ *indicates the* p *values of a Wilcoxon test against the chance level of 0.*

Set-Size	N	*M*	*SD*	*SE*	$p_{Wilcoxon}$	Cohen's *d*
8	32	0.029	0.116	0.020	0.217	0.247
16	32	0.005	0.164	0.029	0.439	0.029

Further, a paired Wilcoxon test revealed that performance in the *Unique Condition* was significantly better for the smaller set-size, $V = 308.5$, $p = 0.05$, $CI_{.95}$ = -5.3e-05, 0.0834, maybe mirroring Endress and Potter's (2014a) finding that performance in the absence of PI drops from small set-sizes to larger set-sizes, with a limited change in performance for larger set-sizes. In contrast, there was no difference between the set-sizes in terms of the Cost of PI, $V = 259$, $p = 0.837$, $CI_{.95} = -0.0624, 0.09$.

A.1.2 Analysis in terms of accuracy. As shown in Figure A2 and Table A1, performance in Terms of raw accuracy differed somewhat across the *Set-Size*, but the effect of *PI Condition* seemed small. This was confirmed by a generalized linear model with the within-subject predictors *PI Condition* and *Set-Size*, treating the trial-by-trial accuracy as a binary random variable. Following Baayen et al. (2008), I

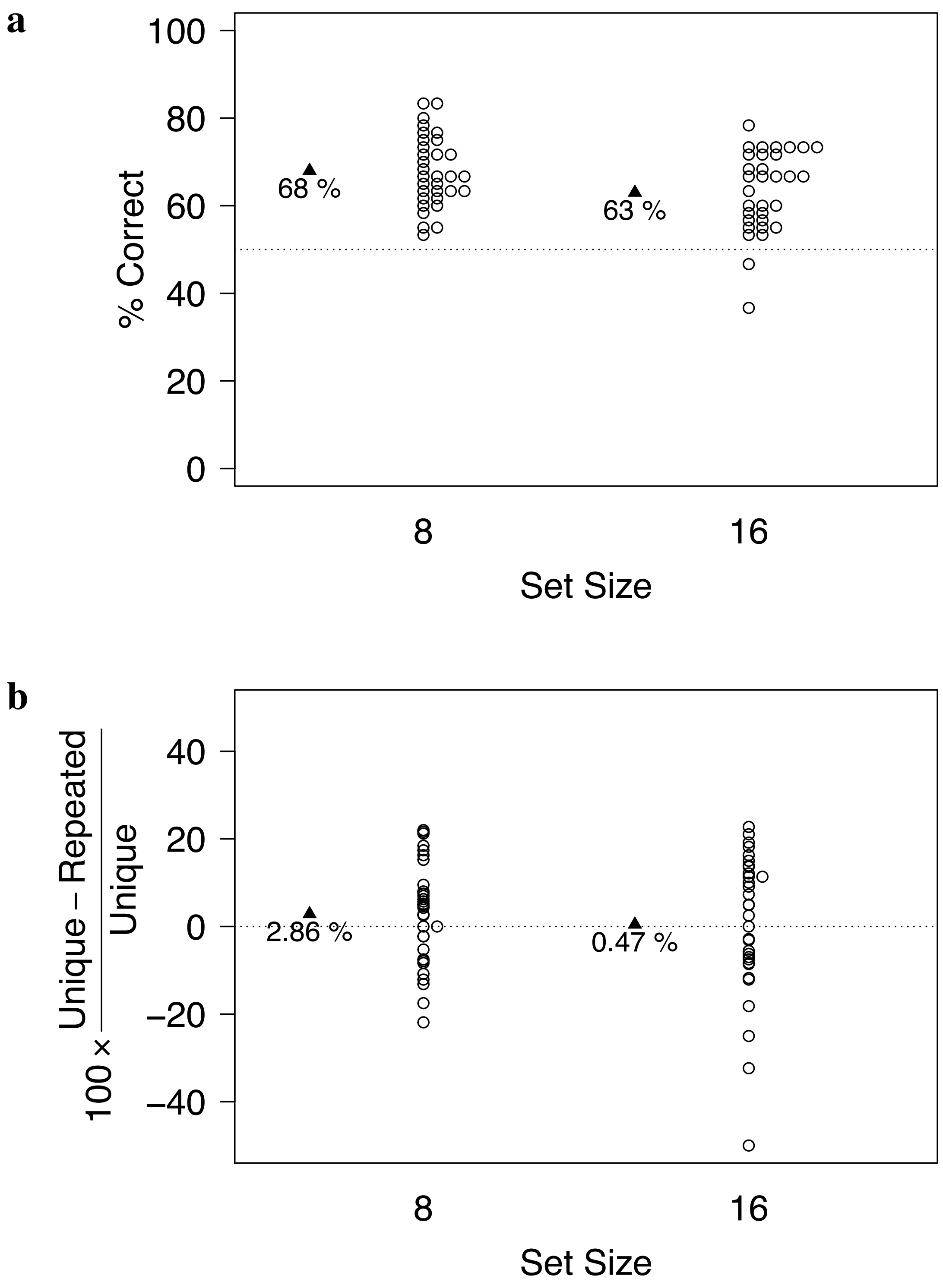


Figure A1. Results of the Pilot Experiment in terms of the accuracy in the *Unique Condition* (a) and the relative *Cost of PI* (b).

then removed the interaction term from the model as it did not contribute to the model likelihood. The results are shown in Table A3

Table A3
Results of a generalized linear model for the Pilot Experiment. I used the Unique Condition *and the* Set-Size *of 8 as the reference levels of the two predictors.*

Effect	Estimate	Std. Error	*CI*	*t*	*p*
PI Condition: Repeated	-0.083	0.048	-0.177, 0.011	-1.73	0.084
Set-Size: 16	-0.173	0.048	-0.267, -0.0789	-3.60	0.000

While performance was significantly worse for the *Set-Size* 16, the difference between the *Unique* and the *Repeated* condition was only marginal.

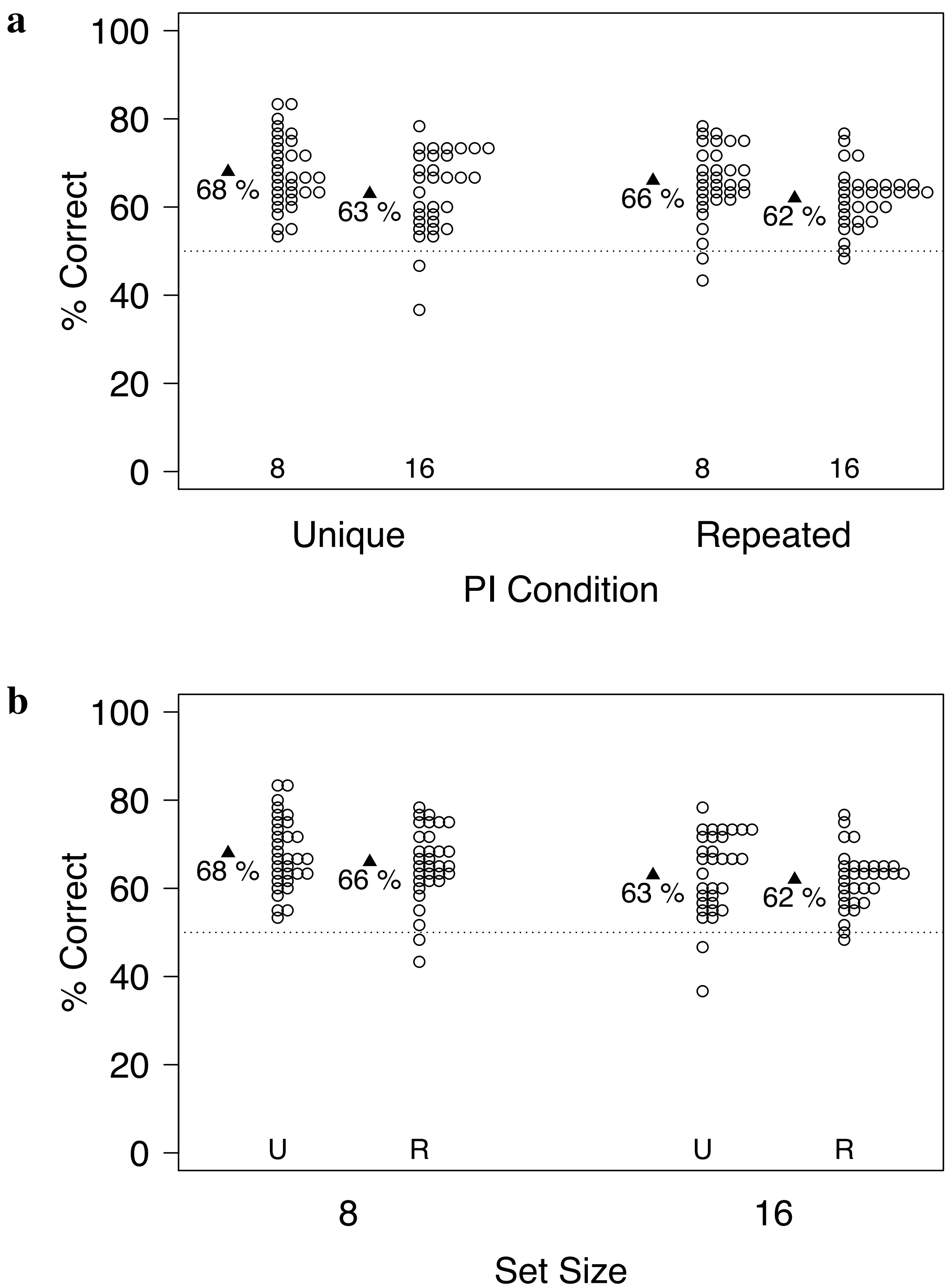


Figure A2. Results of the Pilot Experiment in terms of raw accuracy. Both figures show the same data, grouped by the *PI Conditions* (a) or the *Set-Sizes* (b).

Appendix B
The Cost of PI across serial positions

A possible limitation of Experiment 3 is that observers might face difficulties when encoding more than a few item-location bindings because later bindings overwrite earlier bindings. I analyze the predictions of this account in two ways, one involving the *Cost of PI* and one the raw accuracy.

In the first analysis, I consider the *Cost of PI* in the *Circle-Ordered Condition*. One would expect a main effect of the sequential position: the *Cost of PI* should be reduced in later positions. (Due to violations of normality, I will actually calculate pairwise differences rather than an ANOVA). As described in the Methods section above, the serial positions are two sequence-initial positions, two sequence-medial positions and two sequence-final positions.

In the second analysis, I consider the raw accuracy (on a trial-by-trial basis) in the *Unique* and the *Repeated* Conditions of the *Center Condition* and the *Circle-Ordered Condition*, again as a function of the sequential position of the items. The critical prediction is a triple interaction between these factors. If spatial information reduces PI, one would expect an interaction between the *Location Condition* and the *PI Condition*. However, if observers cannot encode more than a few item-location bindings, this double interaction should be strongest for the most recent sequential positions, resulting in a triple interaction between the *Location Condition*, the *PI Condition* and the *Sequential Position*.

B.1 Analyses in terms of the susceptibility to PI

Table B1
Cells in the Serial Position analysis for Experiment 3 where a violation of normality was detected by a Shapiro-Wilk test when performance was measured in terms of the Cost of PI.

Sequential Position	*Location Condition*	W	*p*	$p \leq .05$
First	Center	0.903	0.000	***
Middle	Circle - Ordered	0.941	0.006	**
First	Circle - Ordered	0.947	0.012	*

In the analyses below, I excluded one participant whose *Cost of PI* for the initial positions differed by more than 4.5 standard deviations from the mean. Further, as shown in Table B1, some cells showed deviation from normality. The analyses below are thus based on pairwise differences assessed by Wilcoxon tests.

As shown in Table B2, the present data do not support the prediction that the *Cost of PI* is most pronounced towards the beginning of the sequences. In fact, numerically at least, it is more pronounced for later positions. However, as shown in Table B3 and Figure B1, these numeric trends are not statistically reliable, as no pairwise difference between sequential positions reaches significance.

To provide evidence for the null hypothesis, I computed, for each participant, the pair-wise differences between the *Cost of PI* in the different sequential positions (e.g., the difference between the *Cost of PI* for sequence-initial positions and sequence-medial positions). For each difference, I then compared two linear models with Gaussian noise with a mean of zero and a standard deviation estimated from the

Table B2
Descriptive statistics in terms of Cost of PI *of the different sequential positions (top 3 rows), and in terms of differences of the* Cost of PI *across sequential positions (bottom 3 rows).*

	N	*M*	*SD*	*SE*
Cost of PI				
First	59	0.054	0.305	0.040
Middle	59	0.046	0.283	0.037
Last	59	0.098	0.210	0.027
Differences				
First-Middle	59	0.007	0.425	0.055
Last-First	59	0.044	0.375	0.049
Last-Middle	59	0.051	0.322	0.042

Table B3
Pairwise Wilcoxon test for the differences in the Cost of PI *across Sequential Positions for the* Circle-Ordered Condition *from Experiment 3. The last two columns show the likelihood ratios for the* null *hypothesis for these differences after correction with the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).*

Difference	*V*	*p*	*CI*	*AIC* in favor	*BIC* of null
Last–Middle	984	0.459	-0.0531, 0.135	1.37	3.61
Last–First	943	0.357	-0.0657, 0.148	1.95	5.12
First–Middle	859	0.622	-0.0972, 0.145	2.90	7.61

data. The null model set the intercept term to zero (i.e., there was no data fitting on top of estimating the variance), while the alternative model fitted the intercept term on top of estimating the variance. I then compared these models using likelihood ratios, adjust for the different number of parameters (i.e., the intercept term) using the Akaike Information Criterion and the Bayesian Information Criterion (Glover & Dixon, 2004).

As shown in Table B3, this analysis favored the null hypothesis. For example, the null hypothesis was about 5 times more likely than the alternative hypothesis the *Cost of PI* differed between the initial and the final position when correcting with the Bayesian Information Criterion, and about twice as likely when correcting with the Akaike Information Criterion. As a result, we can exclude that people automatically encode item-position bindings, and that earlier bindings get overwritten by later bindings.

B.2 Analysis in terms of accuracy

In the second serial position analysis, I used a generalized linear model with the within-subject predictors *PI Condition*, *Location Condition* and *Sequential Position* (two initial, medial, and final position, respectively). Sequential positions for "new" test items were assigned pro-forma, but, of course, new items has no real sequential position. In this analysis, I treated the data as binary. Following Baayen et al. (2008),

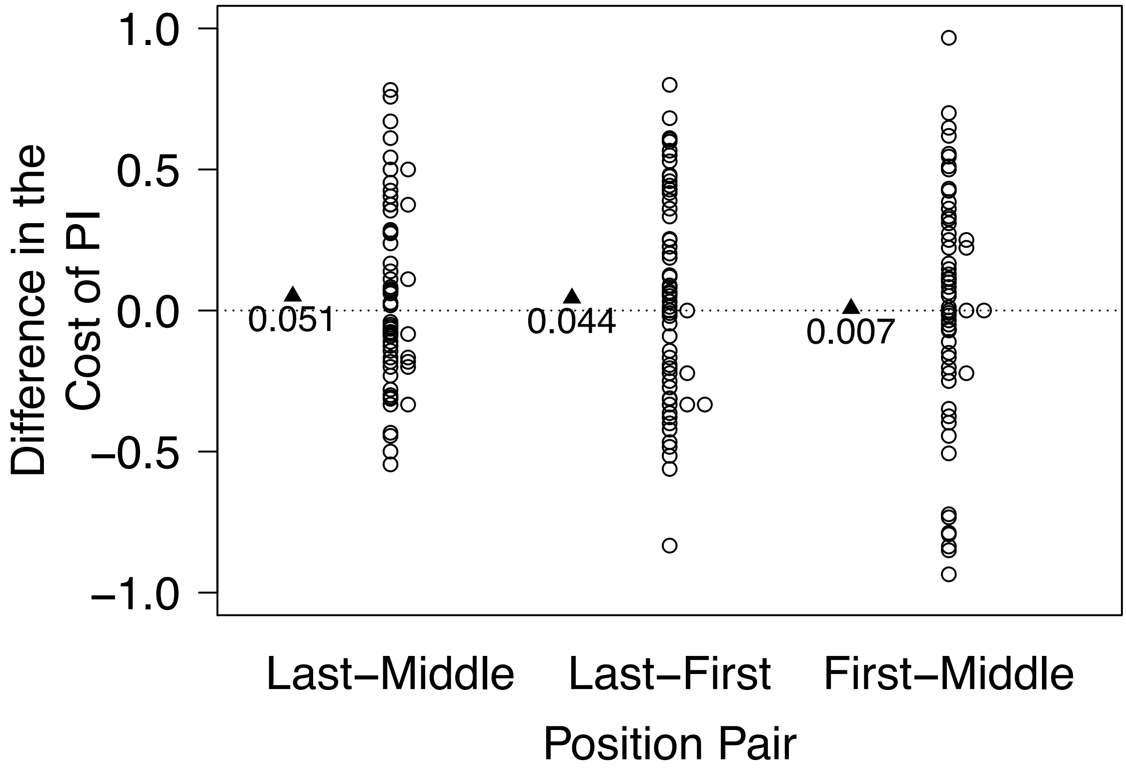


Figure B1. Differences between Sequential Positons in the *Circle-Ordered* Condition of Experiment 3.

I then removed all interaction terms from the model as they did not contribute to the model likelihood. The critical three-way interaction did not contribute to the model likelihood. In fact, the Akaike Information Criterion favored the reduced model by a factor of 55, while the Bayesian Information Criterion favored the reduced model by a factor of 7×10^7, though these values are not necessarily interpretable as likelihood ratios in a generalized linear model.

As shown in Table B4, performance was better in the *Unique Condition* than in the *Repeated Condition*, confirming the effect of PI; it was better in the *Center Condition* and in the *Circle-Random Conditions*, confirming that spatially distributing items has a cost for memory; and it was better for the final positions than for either initial or medial positions. In line with previous results, I thus observed a recency effect on overall accuracy, but recency did not affect the strength of PI.

For completeness, Table B5 shows the results of the full model. Critically, the triple interaction is not significant.

Table B4

Results of a generalized linear model for Experiment 3. For the PI Condition*, the reference level was the* Unique Condition*, for the* Location Condition*, the reference level was the* Center Condition*, and for the* Sequential Position*, the reference level were the final positions. The table shows only those effects that contributed to the model likelihood.*

Effect	*Estimate*	*Std. Error*	*CI*	*t*	*p*
PI Condition: Repeated	-0.277	0.039	-0.352, -0.201	-7.17	0.000
Location Condition: same	0.087	0.039	0.0114, 0.163	2.26	0.024
Sequential Position: middle	-0.269	0.048	-0.362, -0.176	-5.66	0.000
Sequential Position: first	-0.266	0.048	-0.359, -0.172	-5.59	0.000

Table B5

Results of a generalized linear model for Experiment 3. For the PI Condition*, the reference level was the* Unique Condition*, for the* Location Condition*, the reference level was the* Center Condition*, and for the* Sequential Position*, the reference level were the final positions. The table shows the results of the full model.*

Effect	*Estimate*	*Std. Error*	*CI*	*t*	*p*
PI Condition: Repeated	-0.372	0.097	-0.562, -0.183	-3.846	0.000
Location Condition: same	-0.029	0.099	-0.224, 0.165	-0.297	0.766
Sequential Position: middle	-0.381	0.097	-0.571, -0.192	-3.940	0.000
Sequential Position: first	-0.460	0.096	-0.649, -0.271	-4.772	0.000
PI Condition: Repeated × *Location Condition*: same	0.047	0.137	-0.221, 0.315	0.345	0.730
PI Condition: Repeated × *Sequential Position*: middle	0.156	0.134	-0.108, 0.419	1.159	0.247
PI Condition: Repeated × *Sequential Position*: first	0.179	0.134	-0.0835, 0.442	1.337	0.181
Location Condition: same × *Sequential Position*: middle	0.141	0.137	-0.127, 0.41	1.031	0.303
Location Condition: same × *Sequential Position*: first	0.257	0.137	-0.0117, 0.525	1.874	0.061
PI Condition: Repeated × *Location Condition*: same × *Sequential Position*: middle	-0.150	0.190	-0.523, 0.222	-0.791	0.429
PI Condition: Repeated × *Location Condition*: same × *Sequential Position*: first	-0.099	0.190	-0.472, 0.274	-0.519	0.604