Statistical learning of task relevance, rather than stimulus prevalence, improves visual working memory recall

Luzi Xu[1,A], Andre Sahakian[1], Stefan Van der Stigchel[1], Chris L.E. Paffen[1]
Surya Gayet[1]

1. Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 1, 3584 CS, Utrecht, The Netherlands

A. Corresponding author: l.xu2@uu.nl

Abstract

Our visual environment often overwhelms us with more information than we can process. We can, however, enhance our processing efficiency by leveraging inherent regularities in the environment, such as prioritizing likely target objects over unlikely ones. Although there is evidence that statistical learning improves visual working memory performance, it is unclear whether this improvement occurs through stimulus prevalence or task relevance. To distinguish between these hypotheses, we examined whether frequent appearance or frequent probing of stimuli (or both) contributes to improvements in visual working memory recall. Participants were asked to recall and replicate the orientation of one of two previously presented Gabors as precisely as possible. In two experiments, we manipulated (1) stimulus prevalence by presenting Gabors more frequently on one side (either left or right), and (2) task relevance by probing Gabors more frequently on one side (either left or right). We found comparable orientation recall performance for to-be-remembered stimuli appearing at probable versus improbable locations, which suggests that regularities in stimulus prevalence did not improve memory recall. Contrastingly, we found better recall for stimuli appearing at locations that were more versus less likely to be probed. Specifically, task relevance enhanced working memory recall, both by reducing the number of categorical errors (e.g., misremembering a vertical Gabor as horizontal) and by increasing fine-grained recall precision (i.e., the variance of the response errors, after excluding categorical errors). These findings demonstrate that statistical learning of task relevance, but not of stimulus prevalence, enhances visual working memory recall performance.

Keywords: statistical learning, working memory, attention, task relevance, stimulus prevalence

Introduction

Human beings are proficient in extracting patterns and structures embedded in visual environments. For example, when searching for a book, we typically prioritize probable locations like tabletops and shelves over less probable locations like a chair or the floor. This ability, known as statistical learning, is fundamental to a wide range of both basic and higher-level cognitive processes (for reviews see Bogaerts et al., 2020; Frost et al., 2019; Sherman et al., 2020). Over the last decade, numerous studies on statistical learning have primarily focused on its role in visual attention (e.g., Awh et al., 2012; Chun & Jiang, 1998; Wang & Theeuwes, 2018; Zhao & Luo, 2017). Several studies have suggested that visual working memory, a mechanism that is closely related to visual attention (Chun, 2011; Downing, 2000; Olivers, 2008) and is essential for maintaining and manipulating visual information, also benefits from statistical learning (Brady et al., 2009; Haines et al., 2023; Umemoto et al., 2010; Won & Leber, 2017). It is surprising, however, that far less is known about which aspects of statistical learning benefit visual working memory performance. It is conceivable that a better memory recall for a certain coffee packaging, for example, could result from the frequent use of it for tasks (e.g., searching for the product in the supermarket), and/or being visually exposed to it more often (e.g., on television). Previous studies on statistical learning often confound these two factors (e.g., Jiang et al., 2013; Umemoto et al., 2010; Van Moorselaar & Theeuwes, 2023; Xu et al., 2024), making a probable (versus improbable) stimulus *displayed* more frequently and at the same time *probed* more frequently. It is thus unclear whether stimulus prevalence, task relevance, or both, contribute to the effects of statistical learning on visual working memory.

The hypothesis that statistical learning of stimulus prevalence may enhance visual working memory recall (termed as *the stimulus prevalence hypothesis*) is supported by a prevalent theory is that statistical learning operates through mere visual exposure implicitly (Aslin, 2017; Fiser & Aslin, 2001), without the involvement of top-down processes driven by explicit goals (i.e., task relevance). According to this theory, frequent visual exposure to probable stimuli should be a critical factor in how statistical learning influences visual working memory. Furthermore, a large body of studies illustrates that memory and related processes (such as attention) benefit from frequent exposure to visual stimuli. Stimulus repetition has been shown to lead to quicker target recognition (Van Strien et al., 2005), enhanced attentional processing (Chun & Jiang, 1998; Kristjánsson & Campana, 2010), more efficient working memory encoding (Xu et al., 2024), and larger memory capacity (Brady et al., 2009). Also, visual repetition induces neural adaptation underlying learning and memory (Turk-Browne et al., 2007), including strengthened neural connections (Brewer et al., 1998; Lamprecht & LeDoux, 2004; Xue et al., 2010) and reduction in neuronal firing rates and hemodynamic responses (Miller & Desimone, 1994; Barron et al., 2006).

There is also evidence supporting the hypothesis that statistical learning facilitates visual working memory in a strategical manner – through task relevance (termed as *the task relevance hypothesis*). A study using a change-detection task found that the repetition of visual arrays alone, does not enhance memory performance unless it is linked to task relevance, such as associating a repeated visual array with a specific probe location where a change frequently occurs (Olson et al., 2005). Consistently, some studies found that frequent probing of an item can lead to better memory recall (Won & Leber, 2017) and faster retrieval of that item (Haines et al., 2023). Theories on visual attention also provide support for *the task relevance hypothesis*: Task relevance is considered as a key factor in determining attentional priority, with task-relevant stimuli typically gaining more attentional resources than irrelevant ones. Since attention gates visual working memory (Chun, 2011; Downing, 2000; Olivers, 200), task-relevant stimuli may gain priority in visual working memory (Chun, 2011; Downing, 2000; Olivers, 2008) because they receive more attentional resources (Luck et al., 2021).

The present study was set out to investigate whether statistical learning improves visual working memory performance through (1) stimulus prevalence (i.e., learning that stimuli *appear* more frequently at particular locations), (2) task relevance (i.e., learning that stimuli *are probed* more frequently at particular locations), or both. In Experiment 1, we manipulated the prevalence of stimuli at different locations, by presenting to-be-remembered stimuli four times more frequently on one side than the other. In this experiment, the probability of a stimulus being probed on either side was kept constant, thus keeping task relevance equal. In contrast, in Experiment 2, we varied task relevance, by probing stimuli appearing on one side four times more frequently than stimuli appearing on the other side. In this experiment, to-be-remembered stimuli were presented equally often on both sides, thus keeping stimulus prevalence equal. If statistical learning affects visual working memory recall performance—either through stimulus prevalence (Experiment 1) or through task relevance (Experiment 2)—we would observe differential visual working memory recall performance for stimuli in high-probability and low-probability conditions.

Notably, the effects of statistical learning on visual working memory could entail the reduction of *categorical* errors (i.e., failure to memorize items, such as misremembering a vertical-oriented Gabor as a horizontal one) and/or the increasement of *fine-grained* precision (i.e., reduced variances in recall of memorized items). In fact, previous studies have only found statistical learning effects in reducing *categorical* errors (Umemoto et al., 2010; Won & Leber, 2017) but not in increasing *fine-grained* precision (Umemoto et al., 2010). Specifically, Umemoto et al. (2010) found that higher detection accuracy for probable locations was present only for *categorical* changes (e.g., a circle changing into a square) but not for *fine-grained* changes (e.g., parallel lines changing into intersecting lines within a shape). The forced choice task that they used, however, only provided *categorical* options, and thus it may not have allowed for establishing differences in *fine-grained* recall performance. To address this issue, we employed a continuous report task where participants used a 360 degrees rotating response wheel to reproduce orientation of Gabor patches as precisely as possible. The continuous data we obtained from this task allowed us to distinguish between categorical errors (i.e., large errors indicating that stimuli are not memorized) and fine-grained precision (i.e., minor deviations from correct answers in the recall of memorized stimuli, after excluding categorical errors).

Experiment 1

In Experiment 1, we tested whether statistical regularities in stimulus prevalence alone influences visual working memory performance. To this aim, 80% of the stimuli appeared on one side of the screen (either left or right, varied between participants), with the remaining 20% of stimuli appearing on the opposite side. Any stimulus had a 50% chance of being probed upon presentation, irrespective of whether it was presented on the left or right of fixation. Thus, task relevance was equated between the left and right sides of the display, so that no statistical learning of task relevance could occur.

Method

Participants

Twenty-four participants (12 women and 12 men, mean age = 26.25, SD = 5.36) were recruited via Prolific (www.prolific.co) and received monetary reward (7.5 GBP per hour). As there were no prior comparable studies, we based the current sample size on prior experience and practical constraints (i.e., financial compensation for participants). Importantly, the use of Bayesian statistics allows us to distinguish between null-effects and experimental insensitivity. All participants self-reported (corrected to) normal vision, indicated to be fluent in English (as the instructions were provided in English), and had an approval rate of 95% or higher in Prolific. All experiments and procedures were approved by the Ethics Committee of the university.

Apparatus, procedure, and stimuli

The experiment was programmed in the code editor Visual Studio Code (version 1.75; https://code.visualstudio.com) using the JavaScript libraries jsPsych (version 7) (de Leeuw, 2015) and Jspsychophysics (Kuroki, 2021). The task was implemented on the web service Gorilla (https://app.gorilla.sc). We required participants to use a laptop or desktop computer along with a computer mouse for the experiment. Data analysis was performed in MATLAB (R2021a; The Mathworks, Natick, MA) and JASP (2022; Version 0.16.1).

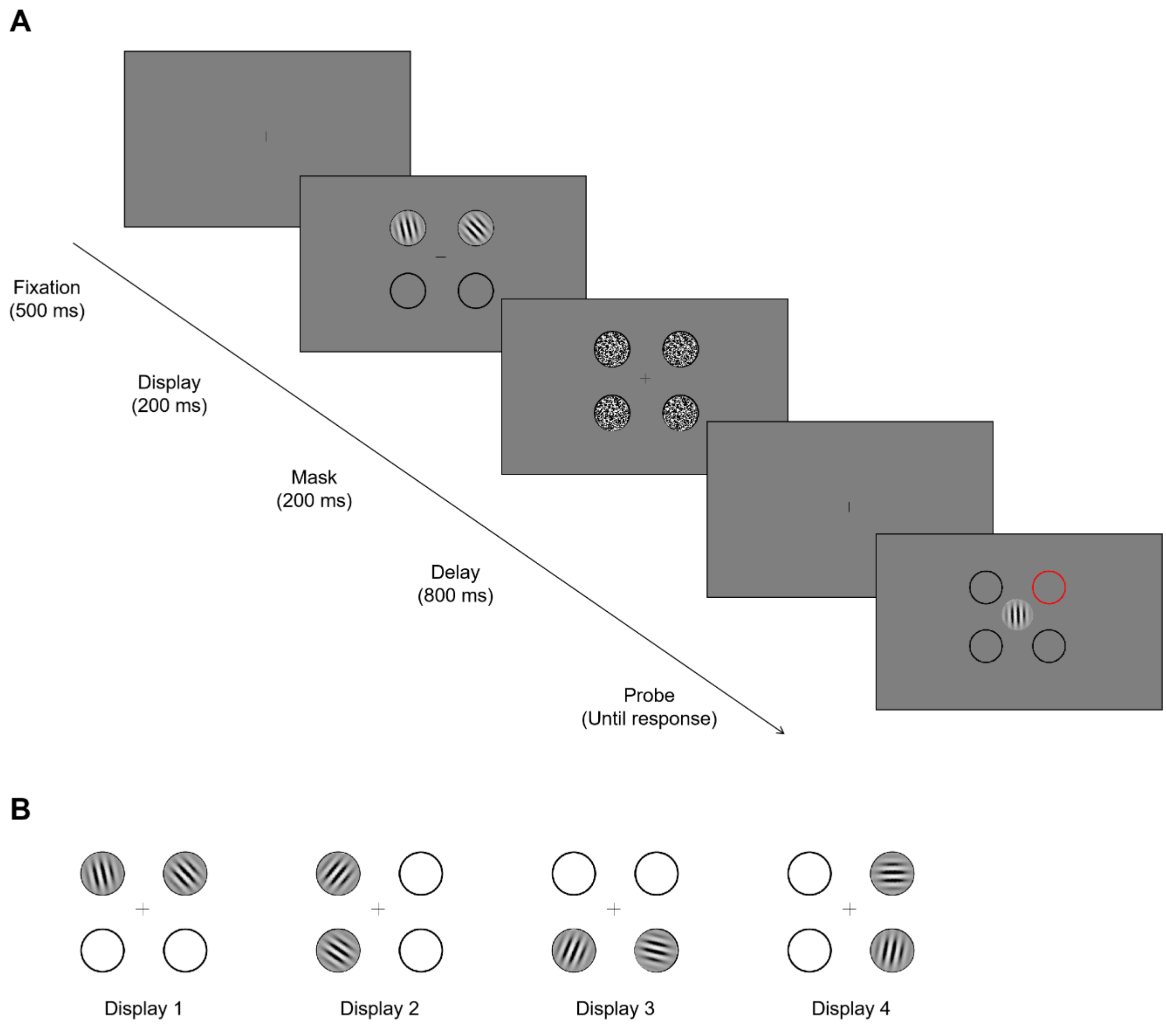


Fig. 1. (A) Schematic illustration of a trial from Experiments 1 and 2. Participants began each trial with a central fixation (500 ms), followed by a visual display of two Gabors (200 ms) with random orientations appearing in two of four placeholders. Subsequently, backward masks (200 ms) appeared at the positions of all four placeholders to reduce after images. After a delay of 800 ms, one of the placeholders was highlighted in red, instructing participants to reproduce the orientation of the Gabor that was previously presented at that location, by moving the mouse horizontally to rotate a centrally presented response Gabor (without any time constraints). In Experiment 1, stimuli appeared four times more frequently on one side (either left or right), while stimuli on either side were equally likely to be probed for reproduction. In Experiment 2, stimuli appeared equally frequently on the left and right sides, but stimuli on one side (either left or right) were four times more likely to be probed for reproduction. (B) The four different lay-outs used as memory displays in Experiments 1 and 2 (the two orientations were randomly determined on every trial).

In this experiment, there were 25 practice trials and 150 formal trials that were divided into 5 blocks of 30 trials each. Each trial (depicted in Fig. 1) started with the presentation of a central fixation cross for 500 ms. Participants were instructed to fixate the central fixation throughout the trial. Then, the memory display was presented, consisting of two Gabor patches presented for 200 ms within two out of four possible locations (upper left, upper right, lower left, and lower right of fixation). Immediately after the Gabor patches disappeared, Gaussian noise masks appeared for 200 ms within all four placeholder to interfere with iconic memory and reduce afterimages. Following a delay of 800 ms, the outline of one of the (now empty) placeholders was highlighted in red, and a new (horizontally oriented) Gabor patch was presented at fixation. Participants were instructed to reproduce the orientation of one of the two Gabors that were previously presented. Either of the two presented Gabors was equally likely to be probed (i.e., 50% chance) irrespective of the position in the memory display. The location where the probed Gabor was previously presented was highlighted by a red placeholder. Participants used the mouse to rotate the orientation of the central Gabor to replicate the orientation of the probed Gabor: moving the mouse left or right rotated the central Gabor counterclockwise or clockwise respectively. Participants could then fix the orientation of the Gabor by clicking the left mouse button, after which participants were presented with two options, "register" or "adjust". Clicking the "adjust" option enabled them to readjust the orientation of the Gabor, while choosing "register" finalized their response for that trial and initiated the next trial. Participants were instructed to replicate the probed Gabor as precisely as possible, with no time constraints. After each block, participants received feedback on their performance (average angular deviation of their registered responses relative to the true stimulus orientations) and got the opportunity to take a short break. At the end of the experiment, to test whether participants were aware of the statistical regularities, we asked participants to indicate on which side of fixation (left or right) they thought stimuli had appeared most frequently. They were asked to also rate their confidence regarding this question on a scale from 1 to 7.

In the memory display, two target stimuli appeared on two out of four possible locations, demarcated by circular placeholders: the upper left, upper right, lower left, and lower right relative to fixation. The two stimuli in a memory display were always presented in horizontally or vertically adjacent placeholders. As is shown in Fig. 1B, this resulted in four possible layouts of the memory display: (1) upper left and upper right (“Display 1”), (2) upper left and lower left (“Display 2”), (3) lower left and lower right (“Display 3”), and (4) upper right and lower right (“Display 4”). To manipulate stimulus prevalence, we varied the occurrence of these four displays across trials, thus creating a design in which the Gabors appeared more frequently on either the left side or the right side of the screen (80%), compared to the opposite side (20%). For example, when left was chosen as the probable side, Display 1 and 3 appeared on 16% of all trials each, Display 2 appeared on 64% of all trials, and Display 4 appeared on 4% of all trials. Whether the left or right side was more prevalent was counterbalanced across participants. Target stimuli were Gabor patches (sine-wave gratings with a Gaussian envelope) with a standard deviation of 12 pixels, a frequency of 0.1 cycles per pixel, and a fixed phase of 0. On each trial, the orientations of both Gabor patches were randomly selected from (range 0 to 360 degrees). Thus, there was no manipulation of statistical regularities of the orientations; all frequency manipulations pertained to the stimulus locations. The masks consisted of white noise masks with a Gaussian envelope. On each trial, four masks were randomly drawn without replacement from a set of 40 pre-generated masks.

In an attempt to maximally equate the stimulus sizes for different participants using different displays, we implemented a calibration procedure, before the main experiment. Participants were instructed to place a credit card or other standard-sized card (typically 8.56 cm wide) on the screen, within a designated rectangle. They then adjusted the size of the rectangle using the mouse, to match the dimensions of the card. After calibration, the radius of the Gabor patches and the inner radius of each place holder was approximately 1.1 cm, and the distance between the center of each place holder and fixation (at the center of the screen) was approximately 3.2 cm.

Data analysis

We used the circular standard deviation (CSD) of the response errors to assess the recall performance of visual working memory. CSD is often used for tasks involving the recall or reproduction of circular features (e.g., Bays et al., 2016), such as the replication of Gabor patch orientations in our study. It measures the dispersion of response errors, akin to the traditional standard deviation, while taking into account the circular nature of angular data. In the current study, the calculation of CSD involves computing angular differences between each response and the correct orientation, wrapping them within the [-90°, 90°] range, squaring them, averaging the result, and then taking the square root. As a result, lower CSD values indicate better performance in orientation reproduction.

A directional Bayesian paired-samples *t*-test was conducted to test the main hypothesis of Experiment 1 that statistical learning of stimulus prevalence leads to better working memory recall (i.e., lower CSD values) in the high-probability versus the low-probability condition. The *Bayes Factor* (directional test: $\boldsymbol{BF_{-0}}$; bi-directional test: $\boldsymbol{BF_{10}}$) obtained from these two analyses indicate the degree of evidence for *the alternative hypothesis* (i.e., the two conditions differ). Following convention, we interpret values between 1 and 3 (or between 1 and 1/3) representing inconclusive evidence, values between 3 and 10 (or between 1/3 and 1/10) indicating moderate evidence, values between 10 and 30 (or between 1/10 and 1/30) representing strong evidence, and values larger than 30 (or smaller than 1/30) representing very strong evidence (Kass & Raftery, 1995).

In the analysis described above, we included all data when calculating the circular standard deviation (CSD). This provides an overall measure of recall performance. Next, we set out to dissociate the potential effects of statistical learning on fine-grained recall precision and on coarse-scale categorical errors. Importantly, we wanted to avoid relying on an arbitrary criterion to distinguish between fine-grained recall errors and categorical recall errors (e.g., labeling response errors above 45 degrees as categorical errors). To do so, we tested for the effect of statistical learning on recall performance in two different sets of analyses, focusing either on fine-grained precision or categorical errors, while iterating through an exhaustive range of criteria for defining ‘categorical errors’.

First, to isolate fine-grained recall precision, we conducted directional Bayesian *t*-tests, testing whether the CSD of the response error was smaller in the high-probability condition than the low probability condition, excluding all trials above a certain criterion (i.e., ‘categorical errors’). We systematically iterated through multiple exclusion criteria, covering the full range of possible response errors ([0, 90] degrees) in steps of 1 degree. Accordingly, a criterion of 0° would exclude all trials from analysis, a threshold of 45° would include only trials with a response error below 45 degrees, and a criterion of 90° would include all trials from analysis (akin to the main analyses, described above). If statistical learning improves fine-grained recall precision, we would see evidence (e.g., $\boldsymbol{BF_{-0}} > 3$) that the CSD is smaller in the high-probability condition than in the low-probability condition, after excluding trials with ‘categorical errors’ (e.g., errors of 45° or more).

Second, to isolate categorical recall errors, we again iterated through all possible response error magnitudes ([0, 90] degrees, in steps of 1 degree). In this case, however, we computed the proportion of trials with a response error exceeding this criterion. In this case, a criterion of 0° labels all trials as categorical errors (so the proportion of ‘categorical’ errors is 1), and a criterion of 90° labels none of the trials as categorical errors (so the proportion of ‘categorical errors’ is 0). For most values in between (e.g., 45°), there is a proportion of trials with 'categorical errors' (in this example, errors of 45° or more). If statistical learning reduces the number of categorical errors, we will see evidence (e.g., $\boldsymbol{BF_{-0}}$ > 3) for a smaller proportion of 'categorical errors' in the high-probability condition, compared to the low-probability (e.g., when labeling trials as 'categorical errors' when they exceed 45°). Note that categorical/coarse-scale recall errors are measured as a proportion of trials (rather than using the CSD), because the exact magnitude of a categorical error is, by definition, irrelevant.

Third, we also examined potential inter-trial priming effects. Because stimuli in high-probability locations are inherently more likely to be probed on two consecutive trials than stimuli in low probability locations, we wanted to ensure that a potential effect of stimulus prevalence on memory performance was not caused by mere repetition of the probe on two consecutive trials. To examine whether such priming (or: "N-1") effects indeed affect working memory performance, we conducted a Bayesian paired-samples *t*-test (repetition: repeat vs. change) to compare participants' performance when the probed item was at exactly *the same location* (i.e., 'repeat') as in the previous trial, to performance when the probed item was at *the other location* (i.e., 'change') of the same side compared to the previous trial (irrespective of whether it is a high-probability or low-probability side). The comparison of change versus repeat trials within the same probability-side (rather than between probability-sides) enables us to examine such priming effects independently of a potential effect of stimulus prevalence (in Experiment 1) or task relevance (Experiment 2). To test for these priming effects, a Bayesian paired-samples *t*-test was conducted to test the effects of Repetition (repeat vs. change).

Finally, we analyzed if participants have awareness of statistical regularities. If their guess on the high-probability (displayed or probed) side was correct, their confidence score was recorded as a positive number; if incorrect, it was recorded as a negative number. This value was used as a proxy for their subjective awareness of the statistical regularities and was compared to zero (representing no awareness) in a Bayesian one sample *t*-test. Then, we further examined whether a benefit in memory recall for high probability relative to low probability stimuli is specific to participants that were particularly aware of the statistical regularities. To this end, we correlated their recall performance with their awareness level of statistical regularities. Since the awareness ratings are categorical, we conducted a Kendall's tau Bayesian correlation test.

Results

Statistical learning effect. Bayesian directional *t*-tests provide moderate evidence ($\boldsymbol{BF_{-0}}$ = 0.15) against an effect of stimulus prevalence on recall performance (high probability: CSD = 0.93, SD = 0.24; low probability: CSD = 0.90, SD = 0.29), as shown in Fig. 2A (left). This result indicates that statistical regularities in stimulus prevalence alone do not influence visual working memory recall performance; stimuli that appear at more likely locations are not memorized better.

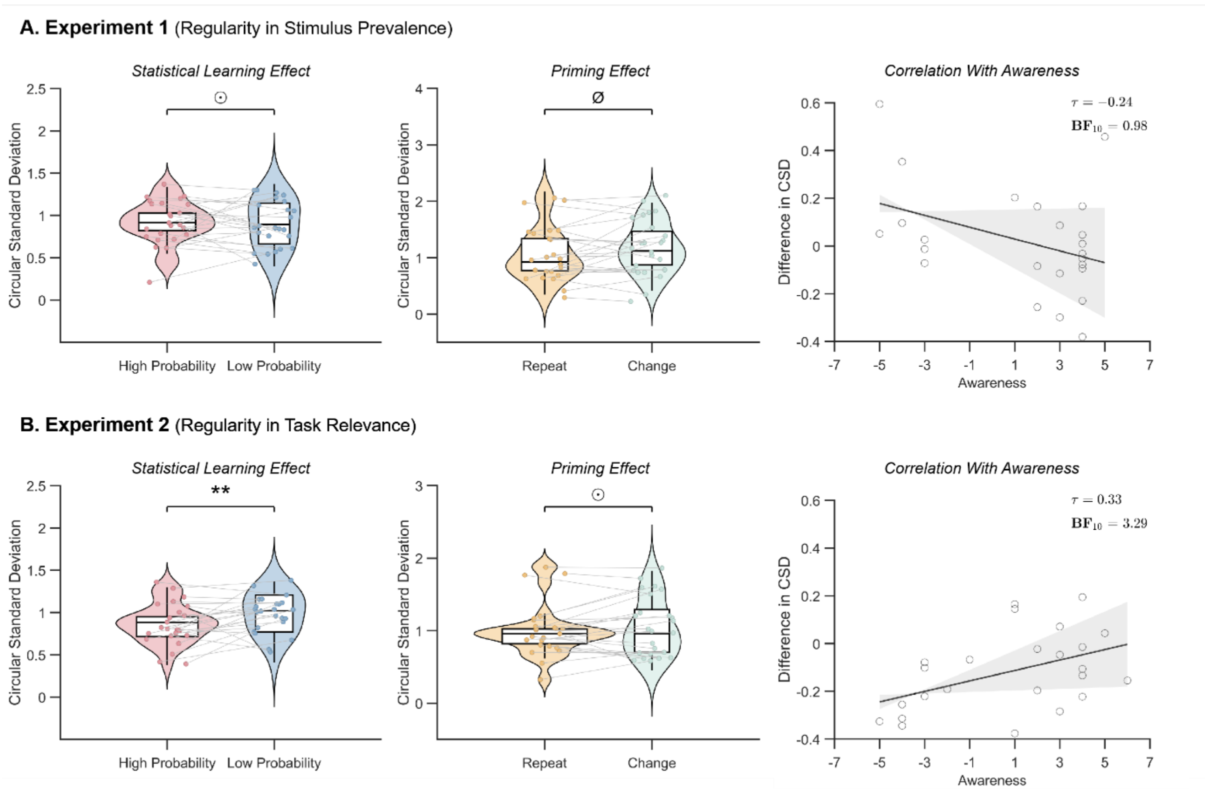


Fig. 2. Results of Experiments 1 and 2. The left panels display comparisons of circular standard deviations (CSD) between high-probability and low-probability locations. The middle panels display comparisons of CSD between repeat trials (i.e., items appearing at the same location as in the previous trial) and change trials (i.e., items appearing at a different location from the previous trial). The right panels display the correlation between participants' awareness of regularities (on the x-axis) with the difference in CSD between high- and low-probability locations (on the y-axis). In violin plots, the extent of the box plots shows the upper/lower quartile, and the whiskers extend to the most extreme data point within 1.5 inter-quartile range above/below the upper/lower quartile; individual dots show individual participant means; color contours outside the box plots show the probability density function of these participant means. The scatter plots show individual participant means and linear fit. **: Bayes factor > 10, Ø: 1/3 < Bayes factor < 3, ⊙: Bayes factor < 1/3

Next, we specifically tested for differences in fine-grained recall precision, by excluding categorical errors (regardless of the definition of a 'categorical error'). As shown in Fig. 3A (left), CSD at high-probability locations was not lower than at low-probability locations ($\boldsymbol{BF_{-0}}$ < 1/3).

When specifically investigating categorical errors, we found that the proportion of categorical errors was not lower at high-probability locations than at low-probability locations ($\boldsymbol{BF_{-0}}$ < 1/3), regardless of the definition of a ‘categorical error’. Thus, statistical learning of stimulus prevalence does not improve visual working memory recall, neither by reducing categorical errors nor by increasing fine-grained precision.

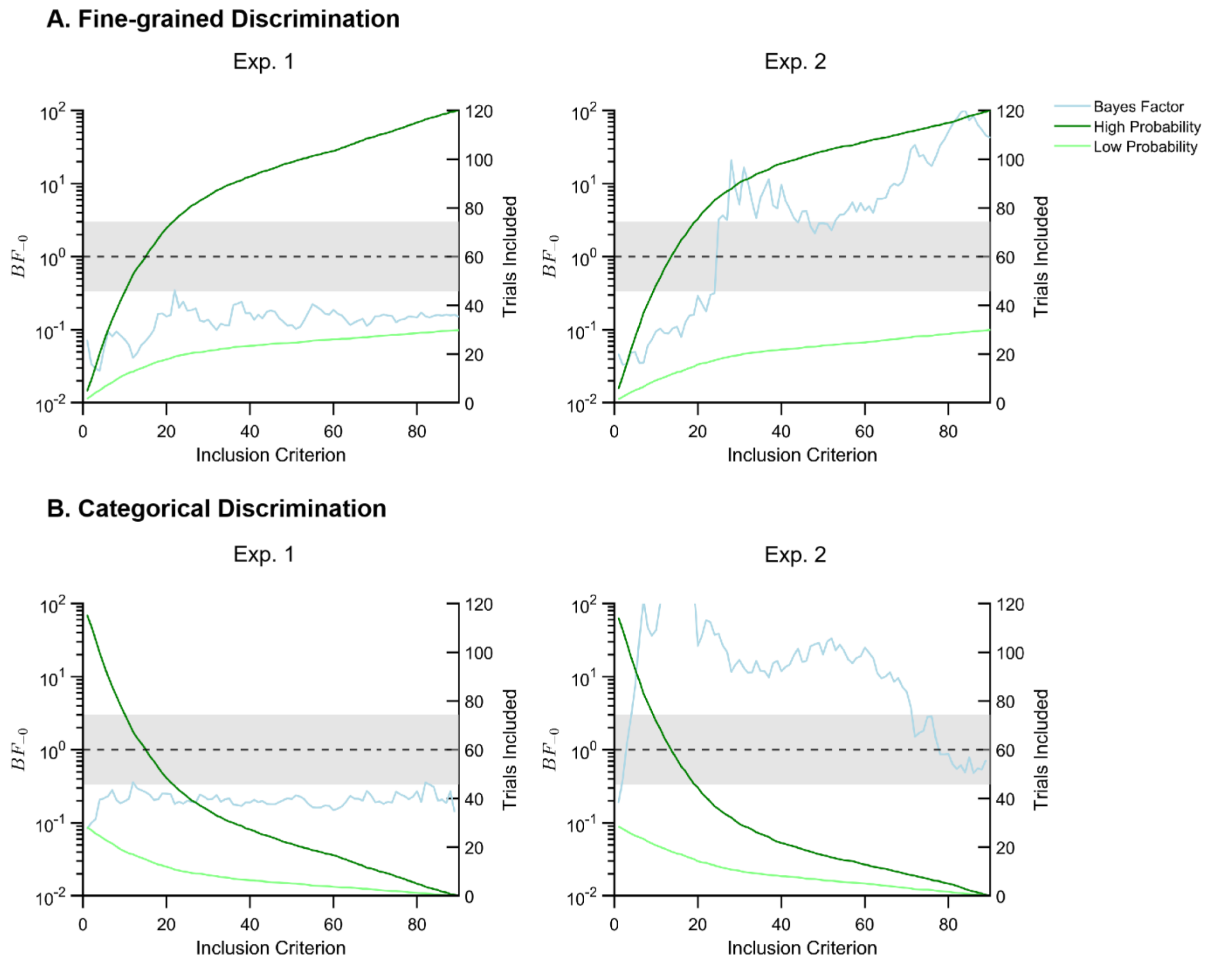


Fig. 3. The Bayes Factors (left y-axis) for the difference in memory recall performance between high- and low-probability locations, for an exhaustive range of inclusion criteria (on the x-axis), and for two outcome metrics (on Panel A and B). Panel A depicts the difference in CSD between probability conditions, for all trials up to the response error depicted on the x-axis (e.g., at a value of 45, the CSD is computed for all trials with an absolute error below 45). Panel B depicts the difference in error rate between probability conditions, given the threshold value depicted on the x-axis (e.g., at a value of 45, the error rate reflects the proportion of trials with an error above 45). The blue line shows the Bayes Factors for the directional t test, testing whether recall performance is better (lower CSD or error rate) for high-probability locations than for low-probability locations. The dark green curves depict the number of included trials in the high-probability condition, and the light green curves depict the number of trials in the low-probability condition (right y-axis). The shaded areas cover [$\frac{1}{3} \geq \boldsymbol{BF_{-0}} \leq 3$], so that values

above this area provide evidence for better recall performance in the high-probability condition compared to the low-probability condition, and values below this area provide evidence against such a difference. Values within the shaded area reflect insufficient evidence for either hypothesis.

Priming effect. There was inconclusive evidence ($\boldsymbol{BF_{10}}$ = 0.88) for the effects of probe repetition (repeat: CSD = 1.06, SD = 0.48; change: CSD = 1.16, SD = 0.42), Fig. 2A (middle), which indicates the absence of inter-trial priming effects.

Awareness of statistical regularities. There was inconclusive evidence for the difference between the subjective awareness (mean = 1.25, SD = 3.49) and zero ($\boldsymbol{BF_{10}}$ = 0.81) as well as the correlation between awareness of statistical regularities and the statistical learning effects (i.e., the CSD difference between high- and low-probability conditions), Kendall's tau = -0.24, $\boldsymbol{BF_{10}}$ = 0.98 (Fig. 2A right). These results suggest that participants were not aware of the statistical regularities at the group level, and that participants' awareness of regularities in stimulus prevalence did not predict their visual working memory precision.

Interim Discussion

In Experiment 1, we showed that presenting items more frequently at a certain location than at another location (given equal probe probability) did not improve memory recall, neither through a reduction in categorical errors nor through an increase in recall precision for stimuli appearing at high-probability locations. The absence of an effect was confirmed by Bayesian evidence for the null. We considered the possibility that the presentation duration of the memory display (i.e., 200 ms) might have led to near-floor performance, thus dampening a potential effect of statistical learning on recall performance. Therefore, in a supplementary experiment, we prolonged the presentation time of the memory display to 450 ms and confirmed that statistical regularity in stimulus prevalence does not improve visual working memory recall (see Supplementary Materials S1).

Experiment 2

Next, we set out to test whether statistical learning of task relevance (i.e., the likelihood of a stimulus being probed) does affect visual working memory performance. To manipulate task relevance, we probed stimuli appearing on one side of the screen (either left or right) on 80% of trials, thus probing stimuli on the other side of the screen only on 20% of trials. In this experiment, memory items appeared on both sides of the display with equal probability. To be able to achieve this ratio, we set Displays 1 and 3 to appear in 38% of all trials each, while Displays 2 and 4 appeared in 12% of all trials each (see Fig. 1B).

Methods

Participants

A new group of twenty-four participants (17 women and 7 men, mean age = 28.54, SD = 4.91) were recruited via Prolific (www.prolific.co) as in Experiment 1.

Apparatus, stimuli & procedure

The apparatus, stimuli, and experimental procedure in Experiment 2 were identical to Experiment 1, except that (1) stimuli appeared equally frequently on left and right sides of the screen (50% each) and that (2) stimuli were probed more frequently when they appeared on either the left side or the right side of the screen (80%) compared to the opposite side (20%). Which side was probed more often was varied (and counterbalanced) between participants.

Results

Statistical learning effect. A Bayesian directional *t*-test provided strong evidence ($\boldsymbol{BF_{-0}}$ = 42.36) that recall performance was better for items that were more likely to be probed (high probability: CSD = 0.87, SD = 0.23; low probability: CSD = 0.99, SD = 0.25), see Fig. 2B (left).

Next, we tested for differences in recall precision. As shown in Fig. 3A (right), the CSD was lower in the high-probability than the low-probability condition, when excluding categorical errors ($\boldsymbol{BF_{-0}}$ > 3). This was the case for a wide range of sensible criteria for defining errors as 'categorical', ranging from very strict (25˚) to very loose definitions of a categorical error (45˚ and higher). This finding demonstrates that statistical learning of task relevance influences the fine-grained recall precision of visual working memory.

Moreover, the proportion of categorical errors was lower in the high-probability condition than in the low-probability condition ($\boldsymbol{BF_{-0}}$ > 3), as shown in Fig. 3B (right). This was the case for a wide range of criteria for defining errors as 'categorical', ranging from very strict (5˚) to very loose definitions of a categorical error (up to 70˚). This shows that statistical learning of task relevance reduces the amount of categorical errors.

Priming effect. There was moderate evidence against ($\boldsymbol{BF_{10}}$ = 0.28) an effect of inter-trial priming on recall precision (repeat: CSD = 1.00, SD = 0.37; change: CSD = 1.03, SD = 0.39), see Fig. 2B middle.

Awareness of statistical regularities. We found inconclusive evidence for the difference between the subjective awareness (mean = 0.75, SD = 3.43) and zero, $\boldsymbol{BF_{10}}$ = 0.36, which suggests that participants were not aware of the statistical regularities at the group level. Besides, there was moderate evidence for a positive correlation between awareness of statistical regularities and the statistical learning effect on recall precision (i.e., the CSD difference between high and low probability conditions), Kendall's tau = 0.34, $\boldsymbol{BF_{10}}$ = 3.29, Fig. 2B (right). That is, the extent to which participants were aware of the regularities in task relevance predicts the extent to which recall precision was better for high probability (i.e., task relevant) relative to low probability stimuli.

Comparison between Experiments 1 and 2. To confirm whether stimulus prevalence (Experiment 1) and task relevance (Experiment 2) yield substantially different effects on recall performance, we compared the statistical learning effects (i.e., the difference in recall precision between high and low probability sides) between the two experiments. A Bayesian independent-samples *t*-test provided moderate evidence for the difference between the statistical learning effects in the two experiments (Experiment 1: ΔCSD = 0.02, SD = 0.23; Experiment 2: ΔCSD = -0.12, SD = 0.16), $\boldsymbol{BF_{10}}$ = 3.17. Results of a Bayesian Mann-Whitney (i.e., independent-samples non-parametric) *t*-test showed moderate evidence *against* a difference in awareness of statistical regularities between experiments (Experiment 1: CSD = 1.25, SD = 3.50; Experiment 2: CSD = 0.75, SD = 3.43), $\boldsymbol{BF_{10}}$ = 0.31. Thus, statistical learning of task relevance influences recall precision more than statistical learning of stimulus prevalence does, and this is not caused by a difference in awareness of the statistical regularities.

Interim Discussion

In Experiment 2, we found that statistical learning of task relevance affects visual working memory recall performance, even when stimulus prevalence is equated. This effect persisted when excluding categorical errors (even with definitions of categorical errors as strict as 25 degrees), demonstrating that statistical regularities affect the fine-grained recall precision of VWM representations. The effect of statistical regularities on recall performance was overall more pronounced for participants who were more aware of the statistical regularities. Furthermore, we observed no evidence for inter-trial priming effects; probing a stimulus at the exact same location on two consecutive trials (which inevitably occurs more often in the high-probability condition than the low-probability condition) does not improve recall performance. Thus, the effects of task relevance identified here cannot be attributed to inter-trial priming effects, and should therefore be attributed to statistical learning, that is, the learning of the overall distributional probability of task-relevant items.

General Discussion

In this study, we investigated whether statistical learning enhances visual working memory recall through learning of *stimulus prevalence* (e.g., stimuli appeared more frequently on certain locations), *task relevance* (e.g., stimuli were probed more frequently on certain locations), or both. In addition, we distinguished between two aspects of memory recall performance that could be influenced by statistical learning: (1) the proportion of *categorical* errors and (2) the *fine-grained* recall precision. Experiment 1 (and the Supplementary Experiment) showed that recall performance for stimuli appearing at more probable locations was not better than those appearing at less probable locations, neither reducing *categorical* errors nor increasing *fine-grained* precision. In contrast, Experiment 2 showed that regularities in *task relevance* did improve visual working memory recall, resulting in a smaller proportion of categorical errors and an increase in recall precision for stimuli that were more likely to be probed compared to those that were less likely to be probed. Importantly, Experiments 1 and 2 were directly comparable, because participants performed the same task on the same stimuli, and the statistical regularities were equal in objective strength (a 4:1 ratio between high and low-probability conditions) as well as subjective strength (i.e., no difference in the awareness measures between experiments). Together, these findings demonstrate that statistical learning improves visual working memory recall through task relevance, but not stimulus prevalence.

Stimuli appearing at high-probability locations were not recalled better (Experiment 1), and perhaps even yielded worse recall performance (Supplementary Experiment) than stimuli appearing at low probability locations. This may seem counterintuitive, since the high-probability location was more likely to contain stimuli, and was therefore also more likely to contain the task-relevant target stimulus. The lack of recall benefit is consistent, however, with previous work showing that simply repeating the same multi-item display throughout an experiment does not increase change-detection accuracy; change-detection accuracy only improved for repeated displays when the same change occurred (i.e., when the same location was probed; Olson et al., 2005). Thus, visual repetition of a display enhances recall performance only when the repeated display was associated with a task-relevant location (i.e., a repeated probe location). In Experiment 1 and the Supplementary Experiment, we purposefully did not associate stimulus prevalence with task relevance; that is, stimuli that appeared more frequently were not probed more frequently. Although the high-probability location inevitably contained the target stimuli more often, it also contained more distractor stimuli, which participants may have learned to ignore or suppress (Wang & Theeuwes, 2018). Therefore, observers may learn to ignore locations that convey more task-relevant information, when these locations also convey more task-irrelevant information.

In contrast, when manipulating how often items in a particular location would be probed (i.e., task relevance) while keeping stimulus prevalence constant, memory recall performance was substantially higher for stimuli that were probed at high-probability locations. This finding aligns with that of two previous studies showing that repeated probing of the same location or feature facilitates memory recall (Umemoto et al., 2010; Won & Leber, 2017). Interestingly, these studies showed that statistical learning only affects categorical recall errors (Umemoto et al., 2010; Won & Leber, 2017); that is, the chance of completely misremembering an item. One study even specifically tested and showed that statistical learning does not influence the *fine-grained* precision of visual working memory recall (Umemoto et al., 2010). Specifically, in a change-detection (i.e., two-alternative force choice) task, Umemoto et al. (2010) found that the learning of statistical regularities in probe location leads to improvements in detection accuracy for changes in categorical information (i.e., overall shape of an item) but not for changes in fine-grained information (i.e., small elements within that shape). Contrary to their results, our study showed that statistical learning not only affected the proportion of categorical errors (i.e., the likelihood of remembering an item), but also affected fine-grained recall precision (after excluding categorical errors). Therefore, our findings not only align with earlier work that statistical learning increases the number of items stored in working memory (Brady et al., 2009; Umemoto et al., 2010), but additionally demonstrates that statistical learning can improve the quality (or resolution) of visual working memory content.

We consider two main reasons for why we did find that statistical learning of task relevance influences recall precision, where Umemoto and colleagues (2010) did not. First, we used a continuous report task, which may be more sensitive to capture subtle differences in recall precision than a two-alternative force choice task. Second, in the study of Umemoto et al. (2010) the definitions of categorical errors and of fine-grained precision were fixed, and determined by the experimenter, leaving up the possibility that an effect on fine-grained precision would have been observed with a (slightly) different definition. In contrast, we tested our hypotheses for an exhaustive range of definitions, circumventing the need for relying on an arbitrary definition. This method uniquely allowed us to distinguish between categorical recall errors and fine-grained recall precision and is applicable to any experiment with continuous report data.

What could be the mechanism underlying the influence of statistical learning on recall performance? One likely mechanism is that statistical learning of task relevance influences the allocation of (spatial) attention during stimulus encoding, thereby affecting visual working memory performance (e.g., Gaspelin et al., 2015). Given the brief presentation duration in our study (i.e., 200 ms in Experiments 1 and 2), participants were unlikely to attend both items (or both sides of fixation) consecutively. Thus, based on the statistical regularities, participants likely determined which location to attend prior to stimulus presentation. We found that participants did not have better (and perhaps even had worse) recall performance for stimuli *appearing* more frequently at particular locations. This indicates that participants allocated the same amount of attentional resources (or less) to probable and improbable stimulus locations. Conversely, the better memory recall for stimuli appearing at frequently *probed* locations (observed in Experiment 2) indicates that participants allocated more attentional resources to the probable locations compared to the improbable locations. Our findings therefore not only speak to the role of statistical learning on visual working memory, but also to a plethora of other processes that depend on (spatial) attention. In this light, our findings prompt a reconsideration of the idea that statistical learning – at large – operates solely as an implicit mechanism through mere exposure (Aslin, 2017), independent of top-down processes driven by explicit goals or task relevance (for a review, see Luck et al., 2021).- Instead, although distributional probabilities may be learned implicitly, their influence on perceptual, mnemonic, and attentional processes may be contingent on the behavioral goals of the observer.

Open practice statement

All data have been made publicly available at Open Science Framework (https://osf.io/q6sbu/?view_only=0d955063aa82477eab24e55d0a9185ec).

References

Aslin, R. N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*(1-2), e1373. https://doi.org/10.1002/wcs.1373

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437-443. https://doi.org/10.1016/j.tics.2012.06.010

Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific reports*, *6*(1), 19203. https://doi.org/10.1167/9.10.7

Boettcher, S. E., Shalev, N., Wolfe, J. M., & Nobre, A. C. (2022). Right place, right time: Spatiotemporal predictions guide attention in dynamic visual search. *Journal of Experimental Psychology: General*, *151*(2), 348–362. https://doi.org/10.1037/xge0000901

Bogaerts, L., Frost, R., & Christiansen, M. H. (2020). Integrating statistical learning into cognitive science. *Journal of Memory and Language*, *115*, 104167. https://doi.org/10.1016/j.jml.2020.104167

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*(4), 487–502. https://doi.org/10.1037/a0016797

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015). Direct evidence for active suppression of salient-but-irrelevant sensory inputs. *Psychological Science*, *26*(11), 1740–1750. https://doi.org/10.1177/0956797615597913

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129-135. https://doi.org/10.1016/j.tics.2011.11.014

Geng, J. J., & Behrmann, M. (2002). Probability cuing of target location facilitates visual search implicitly in normal participants and patients with hemispatial neglect. *Psychological Science*, *13*(6), 520-525.

Grimshaw, G. M., Kranz, L. S., Carmel, D., Moody, R. E., & Devue, C. (2018). Contrasting reactive and proactive control of emotional distraction. *Emotion*, *18*(1), 26-38. https://doi.org/10.1037/emo000033726.

Haines, O., Owens, K., Nobre, A. C., Draschkow, D., & Boettcher, S. (2023, June 25). Statistical learning biases working memory maintenance. https://doi.org/10.31234/osf.io/vrq2c

Jiang, Y. V., Swallow, K. M., Rosenbaum, G. M., & Herzig, C. (2013). Rapid acquisition but slow extinction of an attentional bias in space. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 87–99. https://doi.org/10.1037/a0027611

Klyszejko, Z., Rahmati, M., & Curtis, C. E. (2014). Attentional priority determines working memory precision. *Vision Research*, *105*, 70-76. https://doi.org/10.1016/j.visres.2014.09.002

Klyszejko, Z., Rahmati, M., & Curtis, C. E. (2014). Attentional priority determines working memory precision. *Vision Research*, *105*, 70-76. https://doi.org/10.1016/j.visres.2014.09.002

Kuroki, D. (2021). A new jsPsych plugin for psychophysics, providing accurate display duration and stimulus onset asynchrony. *Behavior Research Methods*, *53*, 301-310. https://doi.org/10.3758/s13428-020-01445-w

Luck, S. J., Gaspelin, N., Folk, C. L., Remington, R. W., & Theeuwes, J. (2021). Progress toward resolving the attentional capture debate. *Visual Cognition*, *29*(1), 1-21. https://doi.org/10.1080/13506285.2020.1848949

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347-356. https://doi.org/10.1038/nn.3655

Olson, I. R., Jiang, Y., & Moore, K. S. (2005). Associative learning improves visual working memory performance. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 889-900. https://doi.org/10.1037/0096-1523.31.5.889

Rutman, A. M., Clapp, W. C., Chadick, J. Z., & Gazzaley, A. (2010). Early top–down control of visual processing predicts working memory performance. *Journal of Cognitive Neuroscience*, *22*(6), 1224-1234. https://doi.org/10.1162/jocn.2009.21257

Sherman, B. E., Graves, K. N., & Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Current Opinion in Behavioral Sciences*, 32, 15–20. https://doi.org/10.1016/j.cobeha.2020.01.015

Souza, A. S., & Oberauer, K. (2017). The contributions of visual and central attention to visual working memory. *Attention, Perception, & Psychophysics*, *79*, 1897-1916. https://doi.org/10.3758/s13414-017-1357-y

Souza, A. S., Rerko, L., & Oberauer, K. (2016). Getting more from visual working memory: Retro-cues enhance retrieval and protect from visual interference. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(6), 890–910. https://doi.org/10.1037/xhp0000192

Umemoto, A., Scolari, M., Vogel, E. K., & Awh, E. (2010). Statistical learning induces discrete shifts in the allocation of working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1419–1429. https://doi.org/10.1037/a0019324

Van Moorselaar, D., & Theeuwes, J. (2023). Statistical learning within objects. *Psychological Science*, *34*(4), 501-511. ttps://doi.org/10.1177/095679762311548

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120-1135. https://doi.org/10.1167/4.12.11

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11. https://doi.org/10.1167/4.12.11

Won, B. Y., & Leber, A. B. (2017). Spatial constraints on probability learning in visual working memory. *Visual Cognition*, *25*(1-3), 34-50. https://doi.org/10.1080/13506285.2017.1346738

Xu, L., Paffen, C., Van der Stigchel, S., & Gayet, S. (2023, November 1). Statistical Learning Facilitates Access to Awareness. https://doi.org/10.31219/osf.io/yb7cn

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience*, *14*(5), 656-661. https://doi.org/10.1038/nn.2773

Zhao, J., & Luo, Y. (2017). Statistical regularities guide the spatial scale of attention. *Attention, Perception, & Psychophysics*, *79*, 24-30. https://doi.org/10.3758/s13414-016-1233-1

Supplementary Materials for:

Statistical learning of task relevance, but not stimulus prevalence, improves working memory recall

S1. Supplementary Experiment

In Experiment 1, we observed no significant difference in orientation recall performance between high-probability locations (where stimuli were presented four times more often) and low-probability locations. To ensure the robustness of this finding, we conducted a supplementary experiment with the same stimuli and procedures, but with a longer presentation duration of the memory display. This would provide observers with more time to encode the stimuli, thus reducing the task-difficulty compared to Experiment 1, and thereby possibly increasing the experimental sensitivity to detect a difference in performance between probability conditions.

Method

Participants

A new group of twenty-four participants (12 women and 12 men, mean age = 29.17, SD = 4.68) were recruited via Prolific (www.prolific.co) as in Experiment 1.

Apparatus, stimuli &procedure

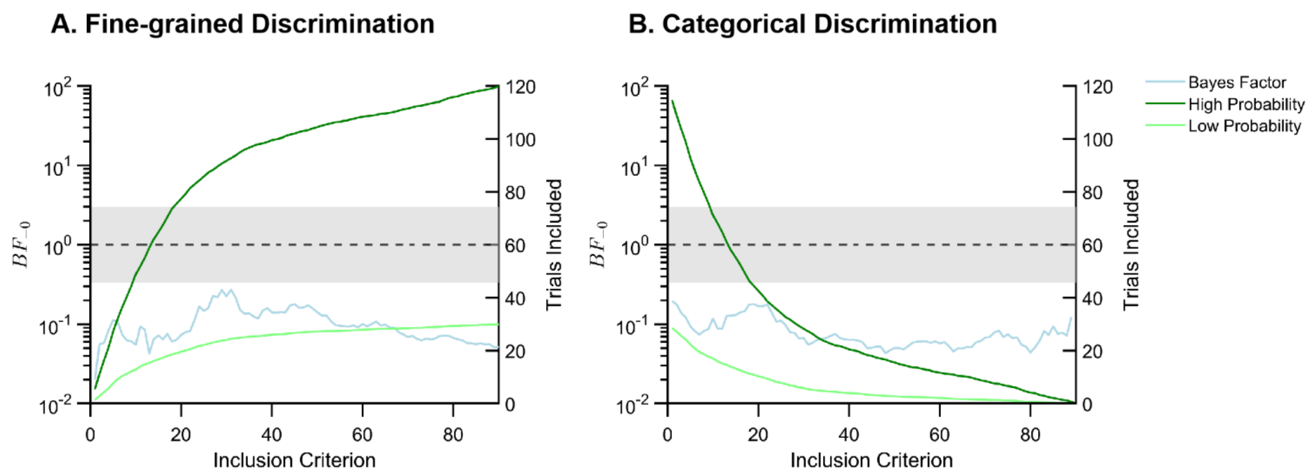
The experimental settings were identical to those in Experiment 1, except that the presentation time of the memory display was set to 450 ms (instead of 200 ms in Experiment 1) and that there were fewer practice trials (five practice trials).

Results

Statistical learning effect. Results of a directional Bayesian *t*-test provided strong evidence ($\boldsymbol{BF_{-0}}$ = 0.07) against the hypothesis that the circular standard deviation (CSD) of the response error is smaller at high-probability locations (0.86, SD = 0.18) than at low-probability locations (0.76, SD = 0.24). This replicates the key finding of

Experiment 1. To explore whether recall performance was actually worse at high-probability locations than at low-probability locations, we further conducted an exploratory bidirectional Bayesian t test and found moderate evidence for a higher CSD for high-probability compared to low-probability locations ($\boldsymbol{BF_{-0}}$ = 3.46). While it is unexpected that memory performance was worse for high-probability locations compared to low probability locations, the present results do confirm the findings of Experiment 1 that increasing stimulus prevalence (while keeping task relevance constant) does not *enhance* visual working memory performance. Together, these results indicate that statistical regularities in stimulus prevalence do not improve visual working memory recall performance.

As in Experiment 1, we also conducted directional Bayesian t tests to examine the CSD of responses to high- versus low-probability locations with different data inclusion thresholds (see Method of Experiment 1 in the main document for more details). Consistent with Experiment 1, results of Bayesian directional t tests consistently showed that, irrespective of inclusion threshold, both CSD (for isolating fine-grained discrimination, see Suppl. Fig.1A) and error rate (for isolating categorical discrimination, see Suppl. Fig.1A) were never lower at high-probability locations than at low-probability locations ($\boldsymbol{BF_{-0}} < 1$). These analyses provide further confirmation that statistical learning of stimulus prevalence does not improve visual working memory recall, neither at the level of categorical discrimination nor at the level of fine-grained discrimination.



Suppl. Fig. 1. The Bayes Factors (left y-axis) for the difference in memory recall performance between high-probability and low-probability locations, for an exhaustive range of inclusion criteria (on the x-axis), and for two outcome metrics (on Panel A and B). Panel A depicts the difference in CSD between probability conditions, for all trials up to the response error depicted on the x-axis (e.g., at a value of 45, the CSD is computed for all trials with an absolute error below 45). Panel B depicts the difference in error rate between probability conditions, given the threshold value depicted on the x-axis (e.g., at a value of 45, the error rate reflects the proportion of trials with an error above 45). The blue line shows the Bayes Factors for the directional t test, testing whether recall performance is better (lower CSD or error rate) for high-probability locations than for low-probability locations. The dark green curves depict the number of included trials in the high-probability condition, and the light green curves depict the number of trials in the low-probability condition (right y-axis). The shaded areas cover $[\frac{1}{3} \geq \boldsymbol{BF_{-0}} \leq 3]$, so that values above this area provide evidence for better recall performance in the high-probability condition compared to the low-probability condition, and values below this area provide evidence against such a difference. Values within the shaded area reflect insufficient evidence for either hypothesis.

Priming effect. Consistent with the results of Experiment 1, we found moderate evidence ($\boldsymbol{BF_{10}}$ = 0.27) against the presence of priming effects (repeat: 1.06, SD = 0.30; change: 1.00, SD = 0.35).

Awareness of statistical regularities. As in Experiment 1, results of a Bayesian one sample *t*-test provided moderate evidence against a difference between the subjective awareness scores (mean = 0.04, SD = 3.61) and zero, $\boldsymbol{BF_{10}}$ = 0.22. Therefore, participants were largely unaware of the statistical regularities at the group level. Besides, results provided inconclusive evidence for the correlation between awareness of statistical regularities and the statistical learning effects (i.e., the CSD difference between high- and low-probability conditions), Kendall's tau = 0.16, $\boldsymbol{BF_{10}}$ = 0.46. This result suggests, at least, that the awareness of regularities in stimulus prevalence does not strongly predict the effects of statistical learning on visual working memory recall.