Statistical Regularities Across Trials Bias Attentional Selection

Ai-Su Li and Jan Theeuwes
Vrije Universiteit Amsterdam and Institute of Brain
and Behavior Amsterdam (iBBA), Amsterdam, the Netherlands

Previous studies have shown that attentional selection can be biased toward locations that are likely to contain a target and away from locations that are likely to contain a distractor. It is assumed that through statistical learning, participants are able to extract the regularities in the display, which in turn biases attentional selection. The present study employed the additional singleton task to examine the ability of participants to extract regularities that occurred across trials. In four experiments, we found that participants were capable of picking up statistical regularities concerning target positions across trials both in the absence and presence of distracting information. It is concluded that through statistical learning, participants are able to extract intertrial statistical associations regarding subsequent target location, which in turn biases attentional selection. We argue here that the weights within the spatial priority map can be dynamically adapted from trial to trial such that the selection of a target at a particular location increases the weights of the upcoming target location within the spatial priority map, giving rise to a more efficient target selection.

Public Significance Statement
Our visual system is sensitive to statistical regularities in the environment. When exposed to regularities, humans are able to extract them and use them to improve visual search efficiency in the form of target facilitation or/and distractor suppression. Even though many studies have focused on the extraction of statistical regularities regarding the location and features of the target and distractor, few studies so far have examined how attentional selection is affected when there are regularities across trials. In four experiments, we investigated the ability of participants to learn statistical regularities regarding the target location across trials. We found that attentional selection was affected by the across-trials regularities indicating that participants were capable of detecting these regularities not only in the absence but also in the presence of distracting information. Our results demonstrate that statistical learning is highly flexible and can bias attentional selection across trials.

Keywords: intertrial regularities, statistical learning, visual search, attentional bias

In everyday life, we need to focus our attention on objects and events that are relevant to us and ignore information that could distract us. The overload of visual input requires us to be able to effectively extract relevant and ignore irrelevant information. The world we live in is highly structured, and contains numerous regularities regarding objects and events occurring across time and space. Statistical learning (SL) refers to the ability to extract these regularities from the environment, which in turn allows a more effective selection and processing of information (e.g., Chun & Jiang, 1998; Saffran, Aslin, & Newport, 1996). For example, regularities facilitate speech segmentation learning (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996), improve classification learning (Aron et al., 2004; Poldrack et al., 2001), and expand visual short-term memory capacity (Brady, Konkle, & Alvarez, 2009; Umemoto, Scolari, Vogel, & Awh, 2010). SL has been described as an implicit learning process that allows the automatic extraction of structure from the world around us (Aslin, Saffran, & Newport, 1998).

Statistical Learning and Attention

Associative learning of statistical relationships between visual objects and events has been shown both for objects appearing in space (Chun & Jiang, 1998; Fiser & Aslin, 2001, 2002b) and in time (Fiser & Aslin, 2002a; Fiser, Scholl, & Aslin, 2007; Kirkham, Slemmer, & Johnson, 2002; Olson & Chun, 2001). For example, Fiser and Aslin (2002a) investigated SL of temporal relationships among sequentially presented shapes. Participants watched an

animation for 6 min in which a single object moved on a horizontal path across the display, cycling back and forth behind a central occluder. Each time it passed behind the occluder, the shape of the object changed. This change was not fully random but, unbeknown to the participant, could contain a structured sequence of shapes. Participants were confronted with a two-interval forced-choice task in which they had to indicate which of two consecutive shapes looked more familiar. Participants correctly identified 95% of the structured sequence of shapes as more familiar, indicating SL of visual temporal sequences.

Following this important study, many studies have focused on the role of attention in obtaining statistical learning. Several studies using auditory stimuli have shown that attention can improve SL while others have shown that without attention there is no learning. Toro, Sinnett, and Soto-Faraco (2005) showed that if there is less attention available, auditory SL is negatively impacted. Others have shown that instructions to attend to one auditory pattern (i.e., words) improved learning, possibly at the expense of learning other aspects of the word stream (such as the grammar; see Finn, Lee, Kraus, & Hudson Kam, 2014). Other studies have demonstrated that the role of attention in SL is limited. For example, Saffran, Newport, Aslin, Tunick, and Barrueco (1997) had children and adults listen to unsegmented artificial language while performing a cover task of creating computer illustrations. Participants were not told they listened to a language. Nonetheless, both adults and children learned this artificial language equally well, even when they performed the additional cover task. It was concluded that learning may occur in the absence of directed attention to the input. Also, Batterink and Paller (2019) showed that participants can learn statistical properties of language even if they do not focus their attention on the speech input.

For visual statistical learning, the evidence seems to indicate that attention is needed for learning to occur. In the study conducted by Turk-Browne, Jungé, and Scholl (2005), participants were presented with sequentially presented streams of nonsense shapes. Participants had to perform a demanding n-back task while attending two separate interleaved streams, one of which had to be attended (i.e., shapes in the color green) and one that needed to be ignored (i.e., shapes in the color red). The results showed that there was only learning of the attended stream and no learning of the unattended stream. Turk-Browne et al. (2005) concluded that selective attention determines the input for statistical learning. Note that a modified replication of this study did show learning of the unattended stream (Musz, Weber, & Thompson-Schill, 2015). However, in that study the task differed from the version of Turk-Browne et al. (2005), as participants needed to press for a particular stimulus when it appeared in the relevant color, but had to refrain from responding when it appeared in the irrelevant color. Because of this instruction, participants may not have fully ignored the stream that needed to be unattended (see Forest & Finn, 2018 for a similar argument).

Visual Statistical Learning (VSL) and Selective Attention

While the research review above has focused on whether attention is necessary for SL to occur, a different literature has focused on how VSL affects attentional selection. Only recently it was recognized that VSL may play a key role in attentional selection,

having an effect that cannot be explained by current goals (top-down selection) or by the stimulus-driven saliency of objects (bottom-up selection; Awh, Belopolsky, & Theeuwes, 2012; Theeuwes, 2018, 2019). This effect due to VSL has been described as "selection history," referring to the influence of lingering biases from past selection episodes (Failing & Theeuwes, 2018).

Several recent studies have provided evidence that VSL biases attentional selection. It has been shown that in visual search, participants learn that the target appears more often in one location than other locations, speeding up target detection (e.g., Geng & Behrmann, 2002, 2005; Jiang, Swallow, Rosenbaum, & Herzig, 2013). Also, participants can learn to avoid distractors that appear more often in one location than in all other locations (e.g., Ferrante et al., 2018; Wang & Theeuwes, 2018), suggesting that participants learn to suppress locations that may contain distracting information. There is also quite some research showing that participants can learn temporal sequences of objects (Howard, Howard, Dennis, & Kelly, 2008; Turk-Browne & Scholl, 2009; Yu & Zhao, 2015; Zhao, Al-Aidroos, & Turk-Browne, 2013). For example, it was shown that learning of temporal regularities could flexibly transfer from one sequence to the other sequence, reflecting sensitivity to a changing environment (Turk-Browne & Scholl, 2009; Yu & Zhao, 2015). Zhao et al. (2013) provided evidence that attention was biased to the location containing a structured stream relative to the location containing a random stream even though the regularity itself was irrelevant to the task at hand. In addition, it was shown that in a cue-cue-target paradigm (a revised serial reaction time [SRT] task), participants were able to learn nonadjacent regularities such that the first cue event predicted the third target event while the second cue was unrelated (Howard et al., 2008).

A study by Baker, Olson, and Behrmann (2004) investigated how the distribution of attention across the visual field affected SL. In this study, participants were presented with two shapes that either were connected by a bar or were unconnected. If SL would occur, one expects faster and more accurate performance for frequent pairs than for infrequent shape pairs. It was shown that when participants only attended one location, no learning was found. However, when the two shapes were connected by a bar, learning occurred, suggesting that perceptual grouping forced learning of the connected pairs of shapes.

Learning Across Trials

The present study investigates the learning of pairwise associations between sequentially presented search displays. In essence, it combines the logic of VSL involving sequentially presented nonsense shapes (like Turk-Browne et al., 2005) with visual search involving the additional singleton paradigm (Theeuwes, 1991, 1992) in which participants have to search for a target singleton among seven other elements. In the current study, the target was equally likely to appear at any of the eight locations in the search array. However, unknown to participants, we built in particular regularities regarding specific target positions across trials. For example, if on one trial, the target happened to be presented at the rightmost position of the display (say the "3 o'clock" position of an analog clock), it would appear on the next trial at the leftmost position (say the "9 o'clock" position of an analog clock). The regularity pair was a two-element (e.g., T_1T_2) temporal sequence

concerning target location, which meant the position of T_1 on the current trial was 100% predictive of the position of T_2 on the following trial. The remaining trials whose target locations were not paired served as filler trials. Thus, trials in which the target location was predicted by the previous trial were regular (predicted) ones (T_2 trials), while trials whose target positions were not predicted by the previous trial were nonregular ones (nonpaired trials and T_1 trials). We hypothesized that if participants are able to extract and learn these regularities, the performance in regular (predicted) trials should be better than that in nonregular (unpredicted) trials.

Whether participants are able to extract these regularities across trials and subsequently bias their attention accordingly is an open question. Experiments 1 and 2 addressed this question. In Experiments 3 and 4, we introduced a color singleton distractor that is known to capture attention automatically (Theeuwes, 1991, 1992) and we determined whether the bottom-up capture would affect the extraction of the regularities in the visual field.

Experiment 1

Experiment 1 was designed to determine whether it was possible at all to learn statistical regularities across trials. If participants can extract trial-to-trial statistical regularities of target positions across trials, RTs for trials that are predicted by the previous trial should be faster than comparable trials that are not predicted by a previous trial.

Method

Participants. Using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), with $\alpha = .05$, $1-\beta = .8$, and a moderate effect size of 0.25, the minimum sample size of a repeated-measures analysis of variance (ANOVA) with two measurements (regular vs. nonregular) is 34 participants. In the current experiment, 34 naïve undergraduate students (26 females; mean age: 21.1 ± 3.2 years) participated for course credit or money compensation. Two participants were replaced because their overall error rates were > 2.5 standard deviations (SDs) below group average. All participants reported normal or corrected-to-normal visual acuity and gave written informed consent in accordance with the Declaration of Helsinki before the experiment. The study was approved by the Ethics Committee of the Department of Experimental and Applied Psychology of Vrije Universiteit Amsterdam.

Apparatus and stimuli. The experiment was presented using Matlab 2013a and PsychToolbox (Brainard, 1997) running on an HP Compaq 6300 SFF computer. Stimuli were presented against a black (RGB: 0/0/0, ~0 cd/m²) background on a 22-in. liquid crystal display monitor (Samsung SyncMaster 2233RZ; resolution: 1680 × 1050 pixels) at a refresh rate of 120 Hz. Participants were seated in a dimly lit room with their chin resting on a chin rest, viewing the monitor from a distance of approximately 73 cm. Behavioral responses were collected via a standard keyboard.

The stimuli used are illustrated in Figure 1. The search displays consisted of eight unfilled stimuli in a single color (red [RGB: 200/0/0, ~20 cd/m²] or green [RGB: 0/170/0, ~23 cd/m²]), one circle (2° diameter), and seven diamonds (2° × 2°), or vice versa. Each shape (0.15° line width) was centered 4° from a white (RGB: 255/255/255, ~103 cd/m²) central fixation square (0.3° × 0.3°)
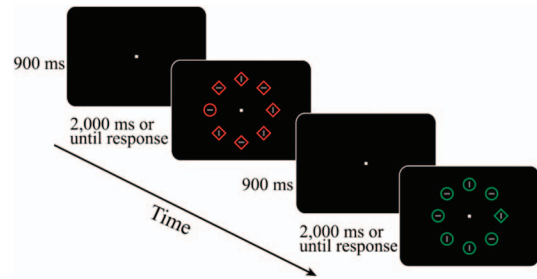

Figure 1. Stimulus display sequences. Participants were required to search for a unique shape and respond whether the orientation of the line segment inside was horizontal or vertical. In this example, the trial in which the target was presented at the leftmost location was followed by the trial where the target was presented at the rightmost location. See the online article for the color version of this figure.

and contained a horizontally or vertically oriented gray line (RGB: 128/128/128, ~22 cd/m²; 1.2° × 0.15°).

Design and procedure. In each block, the target singleton shape appeared in one of eight locations with equal probability. Also, the target appearing at each location was equally likely to be a red/green circle or diamond. All these factors (target location/shape/color) were randomized and counterbalanced across trials. In addition, particular regularities regarding target location across trials were built in. For example, for half of the participants, if on the previous trial the target was presented at the leftmost position of the display, it was always followed by the target presented at the rightmost position of the display on the following trial. Also for the same group of participants, if the target was presented at the top position in the display, it was always followed by the target at the bottom position on the following trial. For the other half of participants, the regularity pairs had opposite directions (rightmost (R) → leftmost (L), bottom → top). It is noteworthy that regularities only concerned the location of the target; its color and shape varied randomly across trials. The regularity pairs and nonpaired trials were intermixed randomly with the constraint that regularity pairs with the same direction could not repeat back to back (i.e., LRLR was not allowed).

The test consisted of one practice block and eight experimental blocks. During practice there were no regularities, implying that all trials were completely randomized. Each experimental block contained 64 trials, including 48 nonregular trials and 16 regular trials.

As illustrated in Figure 1, each trial began with a fixation dot presented at the center of the screen. After 900 ms, the search array with eight items was presented until the participant responded. Participants' task was to search for the shape singleton and discriminate the orientation of the line segment inside. They were instructed to maintain fixation on the central fixation dot throughout the whole block and to press the appointed key ("c" and "m" for horizontal and vertical line) as fast and as accurately as possible. If participants did not respond within 2000 ms or responded incorrectly, a text display "You missed it!" or "Your response was wrong!" appeared for 800 ms. At the same time, an 800-Hz tone sounded for 300 ms. At the end of each block, feedback of current block's accuracy and mean RTs (correct) was given to each participant. Breaks between blocks were controlled by participants themselves. After finishing the whole task, 22 participants were

required to recall whether they were aware of any regularities regarding target locations between trials and write down what kind of regularity they were aware of. The other 12 participants were asked to complete an eight-alternative forced-choice task to choose at which location the target was most likely to appear after the first item of the regularity pairs. Specifically, we showed participants a search array consisting of a diamond (target at the predicting location) among seven circles on the left side of the display and an array of eight circles representing the eight locations on the right side of the display. Participants were asked to choose the location in the array on the right side that they thought the target was most likely to appear following the search array that was displayed on the left side of the display. The same question was asked for the other regularity pair.

Results

Awareness of the regularities. None of the 22 participants indicated being aware of the regularities present in the display, and none of the 12 participants correctly chose both corresponding regular locations.

Analysis. RTs were limited to correct trials (91.62%) only. For the remaining trials of each block of each participant, RTs were submitted to a nonrecursive trimming procedure (Van Selst & Jolicoeur, 1994) that uses cell size to determine a criterion number of SDs from the mean beyond which an observation is considered as an outlier (2.77% of total trials). Then, trials with RTs < 200 ms (0%) were also excluded from analysis. Finally, mean RTs and accuracy were analyzed with repeated-measures ANOVAs.

Learning effect. Mean RTs and accuracy as a function of target regularity across blocks are shown in Figures 2A and 2B, separately. A two-way repeated-measures ANOVA on mean RTs revealed a significant main effect of target regularity, $F(1, 33) = 30.26$, $p < .001$, $\eta_p^2 = 0.48$, with faster responses in the regular condition (811 ms) than in the nonregular condition (842 ms). The Block × Target Regularity interaction did not reach statistical significance, $F < 1$. Two-tailed paired-samples t-test of the first block revealed that RTs were significantly slower for the nonregular condition than that for the regular condition, $t_{33} = 4.06$, $p < .001$, $d = 0.70$. This result suggests that participants learned the statistical regularities very quickly, as the effect was already present at the first block and stayed present throughout the experiment. The ANOVA on accuracy also revealed a significant main effect of target regularity, $F(1, 33) = 7.47$, $p = .01$, $\eta_p^2 = 0.19$, with higher accuracy in the regular condition (92.78%) than that in the nonregular condition (91.23%).

Intertrial target distance analysis. Although participants showed faster responses in the regular condition than in the non-regular condition, we cannot conclude with certainty that they actually learned the statistical regularities. Visual search efficiency is facilitated by discouraging attention from returning back to the previously attended locations (inhibition of return; see Klein, 2000). As our regularity pairs were always four items away (e.g., a target at the leftmost position was always followed by a target at the rightmost position), it was possible that the RTs to the regular target were fastest not because of learning the regularity but because the target was always four items away from the target on the previous trial. To test this, we restricted the distance to four items and performed an additional paired-samples t-test between orthogonal (regular) and oblique (nonregular) direction conditions. Critically, when comparing RTs for subsequent trials that were four items away, RTs in regular trials (810 ms) were significantly faster than in nonregular trials (825 ms), $t_{33} = 2.21$, $p = .03$, $d = 0.38$, suggesting that it was not the four-item distance that drove the effect but the learned regularity.

Discussion

In this experiment, we observed that participants were faster in responding to a target that appeared at a regular location than at a nonregular location. It implies that participants have extracted the statistical regularities concerning the trial-to-trial target locations, which in return improved visual search efficiency.

However, it is clear that participants learned the statistical regularities very fast as the effect was already present at the first block. To ensure that the effect was truly the result of learning and not due to some idiosyncratic strategy on part of the participants, we wanted to replicate the result in an experiment in which we could observe the transition process from searching freely to optimizing search efficiency via extracting statistical regularities.

Experiment 2

In Experiment 1, we observed the target facilitation effect of regular trials compared to nonregular trials, and found that participants could extract the regularities very quickly as there was already an effect of regularity during the first block. To rule out any idiosyncratic strategy that may have contributed to the effect and to ensure that participants indeed learned the regularities, we designed Experiment 2 in which all trials were totally randomized in the first four blocks while regularities were included in the last four blocks, allowing us to see the change in search behavior.
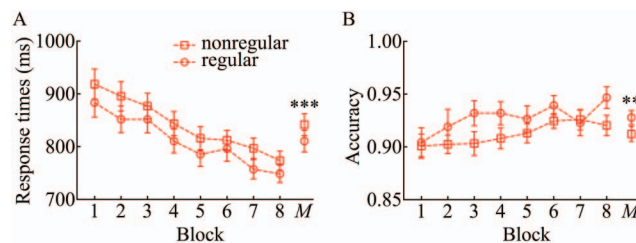

Figure 2. Mean RTs (A) and accuracy (B) for nonregular and regular conditions over blocks. Standard error bars are shown. ** $p < .01$. *** $p < .001$. See the online article for the color version of this figure.

Method

The method was identical to that of Experiment 1 except that all trials were random in the first four blocks. Participants completed eight consecutive blocks. At the end of the experiment, participants were required to indicate whether they were aware of any regularities regarding some target locations across trials in the last four blocks and write down what kind of regularity they were aware of. Thirty-four naïve undergraduate students (27 females; mean age: 20.1 ± 1.6 years) participated for course credit or money compensation. One participant was replaced because the overall error rate was >2.5 SDs below group average. All participants reported having normal or corrected-to-normal visual acuity.

Results

Awareness of the regularities. None of the 34 participants indicated being aware of the regularities present in the display.

Analysis. RTs were limited to correct trials (91.67%) only. For the remaining trials of each block of each participant, RTs were submitted to a nonrecursive trimming procedure (Van Selst & Jolicoeur, 1994) that uses cell size to determine a criterion number of SDs from the mean beyond which an observation is considered as an outlier (2.7% of total trials). Then, trials with RTs < 200 ms (0%) were also excluded from analysis. Finally, mean RTs and accuracy were submitted to a two-way repeated-measures ANOVA consisting of the factors "half" (1st and 2nd) and target regularity (nonregular and regular). It is noteworthy that target regularity conditions in the first half in which no regularity pairs were included referred to the corresponding target locations in the second half where statistical regularities were included.

Learning effect. A repeated-measures ANOVA on mean RTs with half (1st and 2nd) and target regularity (nonregular and regular) as factors was conducted. There was no main effect of regularity, $F(1, 33) = 3.03$, $p = .09$, $\eta_p^2 = 0.08$. Importantly, as visualized in Figure 3, the Half × Target Regularity interaction was significant, $F(1, 33) = 4.9$, $p = .03$, $\eta_p^2 = 0.13$. The additional simple effects analysis showed that the difference was insignificant in the first half [$F(1, 33) = 0.01$, $p = .90$, $\eta_p^2 = 0$], but was

highly significant in the second half [$F(1, 33) = 7.94$, $p = .008$, $\eta_p^2 = 0.19$], with faster responses in the regular condition (763 ms) than in the nonregular condition (786 ms). ANOVA on mean accuracy did not show the significant main effect of target regularity ($p = .12$, $\eta_p^2 = 0.07$), nor was there a significant Half × Target Regularity interaction ($p = .27$, $\eta_p^2 = 0.04$).

Discussion

The current experiment involving 34 new participants confirmed the results of Experiment 1 and unequivocally demonstrated learning of the regularities introduced.

Now that we have shown that participants can learn to extract regularities across trials, the next goal is to determine whether this learning is affected by distraction. It is well-known that in the additional singleton task (Theeuwes, 1991, 1992), the presence of a salient color distractor results in attentional capture, and this capture may affect learning adversely.

Experiment 3

Experiment 3 was designed to examine whether participants could extract statistical regularities concerning the target location when a highly salient distractor singleton was present. If the salient distractor does not affect the ability to extract statistical regularities regarding target location, we still would observe the learning effect. However, if the automatic capture of attention disrupts learning and prevents the extraction of these regularities, then we would expect no difference between regular and nonregular trials.

Method

Participants. A new set of 35 naïve undergraduate students (30 females; mean age: 21.7 ± 4.6 years) participated for course credit or money compensation. Three participants were replaced because their overall error rates were >2.5 SDs below group average. All participants reported normal color vision and normal or corrected-to-normal visual acuity.

Stimuli. As illustrated in Figure 4, on 70% of the trials, a uniquely colored distractor was presented. The maximum response times extended to 3000 ms.

Design and procedure. As in previous experiments, the target of the search array had a unique shape irrespective of its color. The target appeared at one of the eight locations with equal probability. On 70% of trials, one of the seven identically shaped items was a color singleton, either red with all other objects green, or vice versa. The color distractor was equally likely to appear at the other seven locations except the target location, yielding 56 combinations of target-distractor position in total for each block. Consistently, the target was equally likely to be a red/green circle or diamond (accounting for 1/4 each in every block). All factors (target-distractor location and target shape/color) were randomized and counterbalanced within a block. The same regularities concerning the target location were built into the experimental blocks. Still, regularities only concerned the location of the target; its color and shape as well as the presence of the colored distractor varied randomly across trials. After finishing the task, 22 participants were required to report whether they were aware of regularities regarding some target locations across trials and write down what
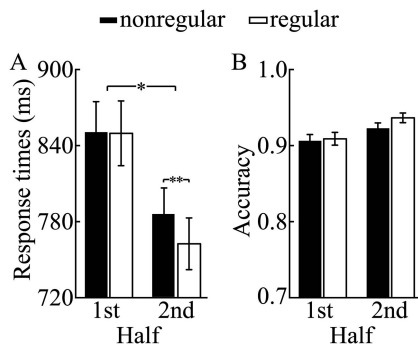
Figure 3. Mean RTs (A) and accuracy (B) for nonregular and regular conditions. Conditions in the first half, in which no regularity pairs were included, referred to the corresponding target locations in the second half, where statistical regularities were included. Standard error bars are shown. * $p < .05$. ** $p < .01$.
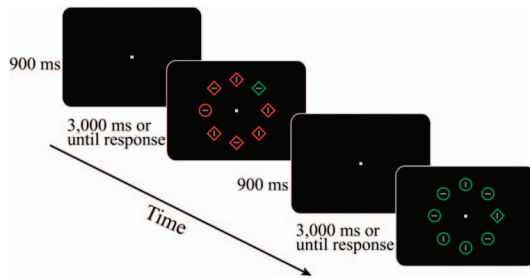
Figure 4. Stimulus display sequences in which a uniquely colored distractor was included. The target color changed unpredictably from trial to trial (either red or green). The regularity pairs concerning target locations were also included. See the online article for the color version of this figure.

kind of regularity they were aware of. The other 13 participants were asked to complete the same eight-alternative forced-choice task as we used in Experiment 1 in which they chose the most likely target location following a predicting trial.

The test consisted of one practice block and eight experimental blocks. Each experimental block contained 80 trials, including 56 distractor-present trials (42 nonregular trials and 14 regular trials) and 24 distractor-absent trials (18 nonregular trials and 6 regular trials). So, a 2 (distractor condition: present and absent) × 2 (target regularity: nonregular and regular) within-participant factorial manipulation was employed in a typical additional singleton paradigm.

Results

Awareness of the regularities. None of the 22 participants indicated being aware of the regularities present in the display. One of the 13 participants correctly chose both corresponding regular locations. If we removed this "aware" participant, the results remained the same.

Analysis. RTs were limited to correct trials (93.12%) only. For the remaining trials of each distractor condition (distractor-absent and -present) under each block of each participant, RTs were submitted to a nonrecursive trimming procedure that uses cell size to determine a criterion number of *SD*s from the mean beyond which an observation is considered as an outlier (2.59% of total trials). Then, trials with RTs < 200 ms (0%) were also excluded from analysis. Mean RTs and accuracy were submitted to a two-way (distractor condition and target regularity) repeated-measures ANOVA. To evaluate the strength of the evidence for the alternative hypothesis (H1) over the null hypothesis (H0) whenever a comparison using traditional null hypothesis testing was insignificant, we also quantified the Bayes factor (BF) using Bayesian hypothesis testing in JASP (Wagenmakers et al., 2018).

Learning effect. Mean RTs and accuracy under nonregular and regular trials as a function of distractor condition are shown in Figures 5A and 5B, separately. A two-way repeated-measures ANOVA on mean RTs with distractor condition and target regularity as factors revealed a significant main effect of distractor condition, $F(1, 34) = 143.39$, $p < .001$, $\eta_p^2 = 0.81$, with faster RTs under the distractor-absent (952 ms) condition than that in the distractor-present (1084 ms) condition. The main effect of target

regularity was also significant, $F(1, 34) = 4.39$, $p = .04$, $\eta_p^2 = 0.11$, with faster RTs in the regular condition (1000 ms) than in the nonregular condition (1016 ms). The Distractor Condition (absent/present) × Target Regularity (regular/nonregular) interaction was insignificant, $p = 0.22$, $\eta_p^2 = 0.04$, $BF_{10} = 0.30$, suggesting that there were no differences between nonregular and regular trials under distractor-absent and distractor-present conditions. The ANOVA on accuracy revealed that the main effect of distractor was significant, $F(1, 34) = 38.88$, $p < .001$, $\eta_p^2 = 0.53$, with higher accuracy in the distractor-absent (95.58%) condition than that in the distractor-present (92.36%) condition. The main effect of target regularity also reached significance, $F(1, 34) = 6.78$, $p = .01$, $\eta_p^2 = 0.17$, with higher accuracy in regular (94.37%) condition than that in nonregular (93.57%) condition.

Discussion

In the current experiment in which a salient color distractor was present, we still observed the learning effect. This suggests that participants still learned the regularities even when attention was captured by the color singleton distractor. The difference in RTs between the distractor-present and -absent conditions indicates that the capture was large, yet this did not affect the learning of the regularities.

It is noteworthy that in the previous two experiments, we varied the stimulus colors randomly from trial to trial, which basically had no effect on performance. In the current experiment, we also switched the color randomly. Yet in this experiment the target and distractor colors also swapped unpredictably from trial to trial, which implies that the color of the target on one trial could be the color of the distractor on the next trial. This will result in intertrial priming effects between targets and distractors (Pinto, Olivers, &
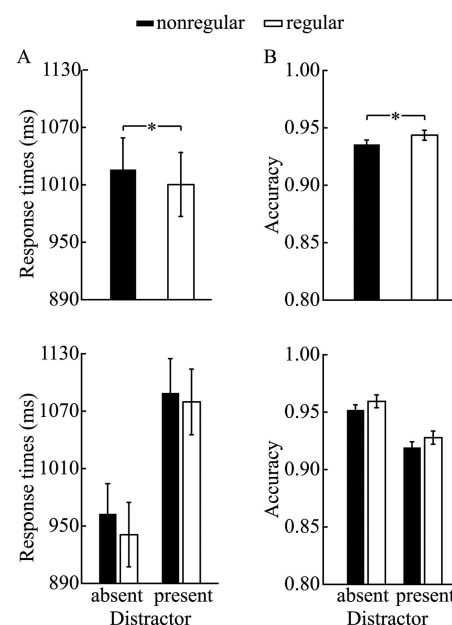

Figure 5. Mean RTs (A) and accuracy (B) for nonregular and regular conditions (top), separate for distractor-absent and distractor-present conditions (bottom). Standard error bars are shown. * $p < .05$.

Theeuwes, 2005), which increases attentional capture and may adversely affect the magnitude of learning effect. In Experiment 4, we wanted to determine the amount of learning taking place when this intertrial effect was removed. To do so, the color of the target and that of the distractor never changed roles. Experiment 4 also introduced either a high- or a low-salience color distractor to determine whether the amount of capture would modulate the learning effect. It is assumed that a highly salient color distractor would cause more disruption than a color distractor that has a low saliency.

Experiment 4

Experiment 4 was designed to substantiate the findings of Experiment 3 when intertrial priming was ruled out. In addition, to determine whether the saliency of the distractor modulated the magnitude of learning effect, we introduced either a highly salient color distractor (in red against a gray search array background) or a low salient color distractor (in brown against a gray search array background).

Method

The method was identical to that of Experiment 3, with the following changes. First, a new set of 34 naïve undergraduate students (25 females; mean age: 20.9 ± 2.7 years) completed 12 blocks for course credit or money compensation. More blocks than previous experiments were used to ensure enough trials under each condition. Specifically, there were 288 distractor-absent trials (216 nonregular trials and 72 regular trials), 336 low-saliency distractor trials (252 nonregular trials and 84 regular trials), and 336 high-saliency distractor trials (252 nonregular trials and 84 regular trials). Two participants were replaced: one because overall error rates were > 2.5 SDs below group average and one because overall RTs were > 2.5 SDs above group average. All participants reported normal color vision as well as normal or corrected-to-normal visual acuity. Second, as illustrated in Figure 6, all the stimuli were gray (RGB: 100/100/100, ~ 14 cd/m^2) except for the uniquely colored distractor, which had high salience (red, [RGB: 200/0/0, ~20 cd/m^2]) or low salience (brown, [RGB: 120/90/45, ~12 cd/m^2]). Third, all factors (target-distractor location and target shape as well as distractor color) were randomized and counterbalanced across trials. Finally, after finishing the whole task, 23
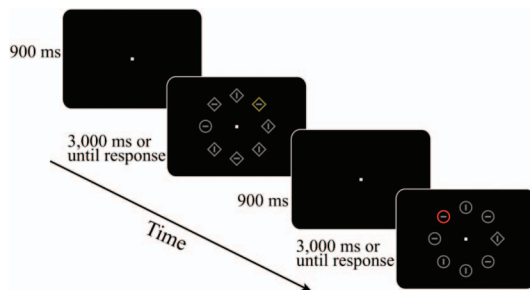

Figure 6. The search array was gray with a high-salience (red) or a low-salience (brown) distractor. The regularity pairs concerning target location were also included. See the online article for the color version of this figure.

participants were required to recall whether they were aware of any regularities regarding target locations between trials and write down what kind of regularity they were aware of. The other 11 participants were asked to complete an eight-alternative forced-choice task, the same as in Experiment 1, to choose at which location the target was most likely to appear after the first item of the regularity pairs.

Results

Awareness of the regularities. None of the 23 participants reported being aware of any regularities throughout the whole experiment. Only 1 of the 11 participants correctly chose both regular locations. Exclusion of this participant did not influence the results.

Analysis. RTs were limited to correct trials (93.5%) only. For the remaining trials of each distractor condition (no distractor, low-salience distractor, high-salience distractor) under each block of each participant, RTs were submitted to a nonrecursive trimming procedure that uses cell size to determine a criterion number of SDs from the mean beyond which an observation is considered as an outlier (3.08% of total trials). Then, trials with RTs < 200 ms (0%) were also excluded from analysis. Finally, mean RTs and accuracy were submitted to a two-way (distractor condition and target regularity) repeated-measures ANOVA.

Learning effect. Mean RTs and accuracy of nonregular and regular trials as a function of distractor condition are shown in Figures 7A and 7B, separately. A two-way repeated-measures ANOVA on mean RTs with distractor condition and target regularity as factors revealed the significant main effect of distractor condition, $F(2, 66) = 76.67$, $p < .001$, $\eta_p^2 = 0.70$. Post hoc tests indicated that RTs in the low-salience distractor condition (912 ms) were significantly slower than no-distractor condition (850 ms, $p < .001$) but were significantly faster than the high-salience distractor condition (941 ms, $p < .001$), suggesting that our manipulation of the distractor condition was valid. More importantly, the main effect of target regularity was significant as well, $F(1, 33) = 13.06$, $p < .001$, $\eta_p^2 = 0.28$, with faster RTs in the regular condition (887 ms) than that in the nonregular condition (915 ms). The current findings basically replicated the results of Experiment 3, showing that participants were able to learn the regularities even when very distracting information was present. The Distractor Condition (no distractor, low-salience distractor and high-salience distractor) × Target Regularity (regular and nonregular) interaction did not reach significance, $F(2, 66) = 1.54$, $p = .22$, $BF_{10} = 0.16$, indicating that there were no significant differences between nonregular and regular trials under three distractor conditions. ANOVA on accuracy revealed similar results. The main effect of distractor condition was significant, $F(2, 66) = 9.24$, $p < .001$, $\eta_p^2 = 0.22$, with higher accuracy in the no-distractor condition (94.83%) than in the low-salience distractor (93.08%, $p < .001$) and the high-salience distractor (93.22%, $p = .003$) conditions. The main effect of target regularity also reached significance, $F(1, 33) = 4.42$, $p = .04$, $\eta_p^2 = 0.12$, with higher accuracy in the regular condition (94.02%) than in the nonregular condition (93.41%).

Comparisons across experiments. To further determine whether the presence of a distracting color singleton affected the magnitude of learning effect, we calculated the overall learning effect ($RT_{nonregular} - RT_{regular}$) in Experiments 1, 3 and 4, respec-
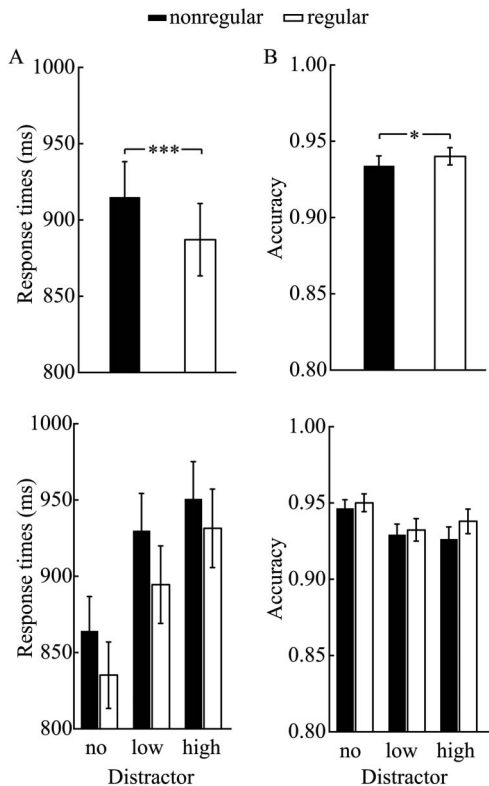
Figure 7. Mean RTs (A) and accuracy (B) for nonregular and regular trials (top), separate for each distractor condition (bottom). Standard error bars are shown. Low and high refer to low-salience and high-salience distractor condition, respectively. * $p < .05$. *** $p < .001$.

tively and conducted independent-samples *t*-tests across experiments. We also provided the BF values corresponding to the *t*-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009). The results showed that the magnitude of learning found in Experiment 1 was neither different from the learning effect in Experiment 3 [$t_{67} = 1.65$, $p = .10$, $BF_{10} = 0.78$] nor different from the learning effect in Experiment 4 [$t_{66} = 0.34$, $p = .74$, $BF_{10} = 0.26$]. There was also no difference in learning between Experiments 3 and 4, $t_{67} = 1.14$, $p = .26$, $BF_{10} = 0.43$. These results indicate that when a distractor singleton was present (Experiments 3 and 4), learning was just as effective as when no distractor was present (Experiment 1).

Discussion

In this experiment, we kept the color of search array constant throughout the task to prevent intertrial priming between targets and distractors. In other words, targets and distractors no longer shared the colors so there were no carry-over effects across trials. The results indicate that learning was just as robust as when these intertrial carry-over effects were removed. Experiment 4 also showed that the amount of distraction did not modulate the magnitude of learning: Learning was just as robust when a low salient distractor was present as when a highly salient distractor was present. Note that the condition in which a highly salient distractor was present was about 91 ms slower than when no distractor

was present, indicating strong attentional capture. Yet even in this condition learning was just as effective as when no distractor was present. These findings indicate that even after attention was erroneously captured to a completely random location (containing the salient distractor), attention was efficiently reallocated to the prioritized (learned) location.

Given the observation that (strong or weak) capture did not modulate learning, one can conclude that participants did not learn a sequence of attentional orienting (e.g., shifting from the center to the left was followed by a shift of attention from the center to the right) but instead learned to prioritize locations. For example, following the selection of the leftmost location, participants learned to prioritize the rightmost location for the upcoming next trial. The disruption by the color singleton distractor causing an erroneous shift of attention to a random other location did not modulate this learned prioritization process.

General Discussion

Across a series of four experiments, we found faster responses (and higher accuracy) to targets presented at regular locations than targets at nonregular locations. This indicates that people were able to learn statistical regularities regarding target locations across trials and made use of regularities in the environment to bias attention toward the future target location. Experiments 3 and 4 showed that the presence of a singleton distractor did not negatively affect learning of these regularities. Specifically, Experiment 4 showed that whether there was strong capture (high-saliency distractor) or weaker capture (low-saliency distractor), the learning effect remained basically the same.

The results of Experiments 3 and 4 showing that learning was not affected by strong or weak capture to the location of the color singleton distractor also suggest that people learned to predict the location of the predicted (upcoming) trial instead of learning a sequence of attentional orienting. Indeed, if participants would have learned a covert orienting sequence (e.g., orienting to the left was always followed by orienting to the right), then one would have expected that the capture by the distractor would have disrupted this sequence, and consequently no benefit on these trials would have been expected. We showed the same benefits for trials in which a salient distractor was present as when there was no distractor. On the basis of these findings, we conclude that the effect found here was not due to implicit spatial sequence learning analogous to motoric temporal sequence learning (e.g., SRT task in Nissen & Bullemer, 1987).

As a mechanism, we assume that the selection of the predicting location generates, for the upcoming trial, activation of the predicted location. Critically, this activation is not disrupted by erroneous attentional shifts to the location of the distractor. This suggests that the learning that we observed here likely affects the weights within the spatial priority map such that the selection of the predicting location up-regulates the weight of the location of the predicted location. Within the map, the weights are combined into a single topographic representation of the environment (Fecteau & Munoz, 2006; Itti & Koch, 2001), which determines the selection priority. Even though, in our Experiments 3 and 4, the strong sensory bottom-up input from the distractor singleton briefly calls attention to the distractor location, the increased weight within the priority map for the predicted location remains

intact and this location receives immediate attention following attentional capture.

The effects observed here are comparable to cueing effect as originally described by Posner (Posner, 1980). Indeed, it is well-known that if participants have knowledge about the upcoming target location, attention is biased accordingly (see also Shaw & Shaw, 1977). For example, in the classic endogenous cueing paradigm (Posner's spatial cueing task), participants are told that the centrally presented cue (usually an arrow) is highly predictive (i.e., 80% validity) of the upcoming target location. The typical result is that RTs to the target presented at the cued (valid) location are faster than those to the target appearing at the uncued (invalid) location (Posner, 1980). Critically, this effect has not only been shown with centrally presented arrow cues (which may direct attention in an automatic way) but also with centrally presented numbers that represented the hand of a conventional clock (e.g., the number 9 represented a location to the left of fixation, whereas the number 12 represented the top location; see Theeuwes & Van der Burg, 2007). In these Posner-like cueing tasks, the location indicated by the cue usually varies from trial to trial and participants typically are perfectly able to anticipate the upcoming target location. If we take this perspective, one may assume that what participants may have learned here was that attending to one location generated expectations about where to attend on the next location, similar to Posner endogenous cueing. Even though feasible, this conjecture is unlikely, as our awareness measures showed that participants had little, if any, awareness regarding the regularities in the display. If anything, participants had no explicit top-down knowledge regarding the regularities of target locations. This is in sharp contrast with Posner endogenous cueing, which is considered a prime example of knowledge-based, top-down, effortful, trial-to-trial attentional control (see Theeuwes, 2018 for a discussion).

Consistent with the notion that the current findings are due to SL (instead of top-down guidance) is the finding that learning in our study was extremely fast: Already at the first block, the learning effect was established and did not change much in subsequent blocks. Fast learning was also found in previous studies that investigated SL of target probabilities (Ferrante et al., 2018; Jiang et al., 2013). It is likely that the medial temporal lobe (MTL), and in particular the hippocampus, is critical for the rapid extraction of regularities from the environment (Chun & Phelps, 1999; Turk-Browne, Scholl, Chun, & Johnson, 2009; Turk-Browne, Scholl, Johnson, & Chun, 2010). The MTL plays a critical role in the representation of space, as was demonstrated by the discovery of "place cells" in both rodents (O'Keefe & Dostrovsky, 1971) and humans (Ekstrom et al., 2003).

Instead of assuming top-down knowledge that drives selection, we argue that through statistical learning the weights within the spatial priority map are dynamically adapted from trial to trial. Upon selecting the predicting location, the weight within the spatial priority map of the predicted location is enhanced such that selection of that location is boosted. The function of the spatial priority map is basically priority control. It serves as a filter, passing on particular bits of information to downstream brain areas. It is assumed that the spatial priority map is associated with a brain network consisting of the frontal eye field (Thompson, Hanes, Bichot, & Schall, 1996), lateral parietal area (Bisley & Goldberg, 2010), superior colliculus (Fecteau & Munoz, 2006;

White, Kan, Levy, Itti, & Munoz, 2017), and the caudate nucleus in the basal ganglia (Hikosaka, Kim, Yasuda, & Yamamoto, 2014; Hikosaka, Takikawa, & Kawagoe, 2000). We argue here that the weights, within the spatial priority map, are dynamically adapted such that following the selection of one location (the predicting location), on the next trial, the weight of another location (the predicted location) is up-regulated.

The current study has some resemblances to a study conducted by Ono, Jiang, and Kawahara (2005). That study investigated the boundary conditions of intertrial temporal contextual cueing (a subclass of statistical learning, see Goujon, Didierjean, & Thorpe, 2015 for a review). Specifically, in their task, participants had to search for a rotated "T" target among 11 rotated "L" distractors. Unknown to participants, the repeated spatial layout of trial $N-1$ was predictive of the target location on the following trial N in the training phase. Even though the target location was predicted by the previous trial, distractors in these displays were presented at random locations. Ono et al. (2005) tested various display configurations and concluded that only when the complete repeated spatial layout of the previous trial was predictive of the target location on the next trial, intertrial SL occurred (see also Thomas, Didierjean, Maquestiaux, & Goujon, 2018). Ono et al. (2005) showed that any random variation in the spatial layout of the displays disrupted learning. Specifically, they argued that even when there were consistent target-to-target associations between the successive displays, the random variation of the nontarget elements in the display introduced so much noise that SL did not occur. Unlike the study from Ono et al. (2005), the current study did show intertrial SL regarding the target location even when there was a lot of random variation across and within any given display. Yet there are several differences between our study and theirs.

First, we had eight fixed locations presented on an imaginary circle around the fixation dot, and our array contained one target and seven nontargets. In Ono et al.'s study, 12 items were presented on an 8×6 grid, resulting in many more combinations of nontarget layouts. Second, the tasks used in these two studies were quite different. Our study utilized the well-known additional singleton task (Theeuwes, 1991, 1992), in which there was parallel search for the target singleton. In the study of Ono et al. (2005), search was serial (rotated "T" among rotated "L"s), where participants had to search items one-by-one or in a clump-wise fashion. It is possible that even though some target-to-target learning may have taken place, it is likely that serial search involving the selection of many nontarget items may have disrupted this learning (e.g., Jiang & Kumar, 2004). In our study, however, attention can be directed quickly toward the target position so that its position may remain activated and possibly memorized. It is likely that in our task, a strong association between two target selection episodes across trials can be formed, which is less likely the case when participants search serially through the display. We assume that participants learned the sequential relationship between these trials, which was possibly strengthened after each encounter. Once the predicting target position was selected, the target position on the next trial was boosted, giving rise to the faster and more accurate responses for the predicted location.

In sum, the present study shows that participants can learn pairwise associations between sequentially presented search displays. Specifically, we conclude that participants are able to ex-

tract intertrial statistics regarding subsequent target locations, which in turn biases attentional selection. We conclude that through SL, the weights within the spatial priority map can be dynamically adapted from trial to trial so that the selection of a target at a particular location increases the weights within the map of the upcoming target location, giving rise to faster and more efficient target selection.

References

Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology, 92,* 1144–1152. http://dx.doi.org/10.1152/jn.01209.2003

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321–324. http://dx.doi.org/10.1111/1467-9280.00063

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

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science, 15,* 460–466. http://dx.doi.org/10.1111/j.0956-7976.2004.00702.x

Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex, 115,* 56–71. http://dx.doi.org/10.1016/j.cortex.2019.01.013

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience, 33,* 1–21. http://dx.doi.org/10.1146/annurev-neuro-060909-152823

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,* 487–502. http://dx.doi.org/10.1037/a0016797

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

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

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience, 2,* 844–847. http://dx.doi.org/10.1038/12222

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature, 425,* 184–188. http://dx.doi.org/10.1038/nature01964

Failing, M., & Theeuwes, J. (2018). Selection history: How reward modulates selectivity of visual attention. *Psychonomic Bulletin & Review, 25,* 514–538. http://dx.doi.org/10.3758/s13423-017-1380-y

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences, 10,* 382–390. http://dx.doi.org/10.1016/j.tics.2006.06.011

Ferrante, O., Patacca, A., Di Caro, V., Della Libera, C., Santandrea, E., & Chelazzi, L. (2018). Altering spatial priority maps via statistical learning of target selection and distractor filtering. *Cortex, 102,* 67–95. http://dx.doi.org/10.1016/j.cortex.2017.09.027

Finn, A. S., Lee, T., Kraus, A., & Hudson Kam, C. L. (2014). When it hurts (and helps) to try: The role of effort in language learning. *PLoS ONE, 9,* e101806. http://dx.doi.org/10.1371/journal.pone.0101806

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

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 458–467. http://dx.doi.org/10.1037/0278-7393.28.3.458

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 15822–15826. http://dx.doi.org/10.1073/pnas.232472899

Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychonomic Bulletin & Review, 14,* 173–178. http://dx.doi.org/10.3758/BF03194046

Forest, T. A., & Finn, A. S. (2018). Attention selectively boosts learning of statistical structure. *Proceedings of the Annual Meeting of the Cognitive Science Society, 40,* 1674–1679. Retrieved from https://cogsci.mindmodeling.org/2018/papers/0322/0322.pdf

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,* 520–525. http://dx.doi.org/10.1111/1467-9280.00491

Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. *Perception & Psychophysics, 67,* 1252–1268. http://dx.doi.org/10.3758/BF03193557

Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences, 19,* 524–533. http://dx.doi.org/10.1016/j.tics.2015.07.009

Hikosaka, O., Kim, H. F., Yasuda, M., & Yamamoto, S. (2014). Basal ganglia circuits for reward value-guided behavior. *Annual Review of Neuroscience, 37,* 289–306. http://dx.doi.org/10.1146/annurev-neuro-071013-013924

Hikosaka, O., Takikawa, Y., & Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiological Reviews, 80,* 953–978. http://dx.doi.org/10.1152/physrev.2000.80.3.953

Howard, J. H., Howard, D. V., Dennis, N. A., & Kelly, A. J. (2008). Implicit learning of predictive relationships in three-element visual sequences by young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1139–1157. http://dx.doi.org/10.1037/a0012797

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203. http://dx.doi.org/10.1038/35058500

Jiang, Y., & Kumar, A. (2004). Visual short-term memory for two sequential arrays: One integrated representation or two separate representations? *Psychonomic Bulletin & Review, 11,* 495–500. http://dx.doi.org/10.3758/BF03196601

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,* 87–99. http://dx.doi.org/10.1037/a0027611

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83,* B35–B42. http://dx.doi.org/10.1016/S0010-0277(02)00004-5

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences, 4,* 138–147. http://dx.doi.org/10.1016/S1364-6613(00)01452-2

Musz, E., Weber, M. J., & Thompson-Schill, S. L. (2015). Visual statistical learning is not reliably modulated by selective attention to isolated events. *Attention, Perception, & Psychophysics, 77,* 78–96. http://dx.doi.org/10.3758/s13414-014-0757-5

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1–32. http://dx.doi.org/10.1016/0010-0285(87)90002-8

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research, 34,* 171–175. http://dx.doi.org/10.1016/0006-8993(71)90358-1

Olson, I. R., & Chun, M. M. (2001). Temporal contextual cuing of visual attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1299–1313. http://dx.doi.org/10.1037/0278-7393.27.5.1299

Ono, F., Jiang, Y., & Kawahara, J. I. (2005). Intertrial temporal contextual cuing: Association across successive visual search trials guides spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 703–712. http://dx.doi.org/10.1037/0096-1523.31.4.703

Pinto, Y., Olivers, C. N., & Theeuwes, J. (2005). Target uncertainty does not lead to more distraction by singletons: Intertrial priming does. *Perception & Psychophysics, 67,* 1354–1361. http://dx.doi.org/10.3758/BF03193640

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414,* 546–550. http://dx.doi.org/10.1038/35107080

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32,* 3–25. http://dx.doi.org/10.1080/00335558008248231

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928. http://dx.doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606–621. http://dx.doi.org/10.1006/jmla.1996.0032

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8,* 101–105. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00690.x

Shaw, M. L., & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 201–211. http://dx.doi.org/10.1037/0096-1523.3.2.201

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics, 50,* 184–193. http://dx.doi.org/10.3758/BF03212219

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51,* 599–606. http://dx.doi.org/10.3758/BF03211656

Theeuwes, J. (2018). Visual selection: Usually fast and automatic; seldom slow and volitional. *Journal of Cognition, 1,* 29.

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology, 29,* 97–101. http://dx.doi.org/10.1016/j.copsyc.2018.12.024

Theeuwes, J., & Van der Burg, E. (2007). The role of spatial and nonspatial information in visual selection. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1335–1351. http://dx.doi.org/10.1037/0096-1523.33.6.1335

Thomas, C., Didierjean, A., Maquestiaux, F., & Goujon, A. (2018). On the limits of statistical learning: Intertrial contextual cueing is confined to temporally close contingencies. *Attention, Perception, & Psychophysics, 80,* 1420–1435. http://dx.doi.org/10.3758/s13414-018-1519-6

Thompson, K. G., Hanes, D. P., Bichot, N. P., & Schall, J. D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology, 76,* 4040–4055. http://dx.doi.org/10.1152/jn.1996.76.6.4040

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97,* B25–B34. http://dx.doi.org/10.1016/j.cognition.2005.01.006

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564. http://dx.doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 195–202. http://dx.doi.org/10.1037/0096-1523.35.1.195

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21,* 1934–1945. http://dx.doi.org/10.1162/jocn.2009.21131

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *The Journal of Neuroscience, 30,* 11177–11187. http://dx.doi.org/10.1523/JNEUROSCI.0858-10.2010

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,* 1419–1429. http://dx.doi.org/10.1037/a0019324

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample-size on outlier elimination. *The Quarterly Journal of Experimental Psychology Section A, 47,* 631–650. http://dx.doi.org/10.1080/14640749408401131

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Morey, R. D. (2018). Bayesian inference for psychology: Part I. Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25,* 35–57. http://dx.doi.org/10.3758/s13423-017-1343-3

Wang, B., & Theeuwes, J. (2018). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 44,* 13–17. http://dx.doi.org/10.1037/xhp0000472

White, B. J., Kan, J. Y., Levy, R., Itti, L., & Munoz, D. P. (2017). Superior colliculus encodes visual saliency before the primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 114,* 9451–9456. http://dx.doi.org/10.1073/pnas.1701003114

Yu, R. Q., & Zhao, J. (2015). The persistence of the attentional bias to regularities in a changing environment. *Attention, Perception, & Psychophysics, 77,* 2217–2228. http://dx.doi.org/10.3758/s13414-015-0930-5

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science, 24,* 667–677. http://dx.doi.org/10.1177/0956797612460407