

Visual statistical learning of temporal structures at different hierarchical levels

Jihyang Jun¹ · Sang Chul Chong^{1,2}

Published online: 11 April 2016
© The Psychonomic Society, Inc. 2016

Abstract Visual environments are complex. In order to process the complex information provided by visual environments, the visual system adopts strategies to reduce its complexity. One strategy, called *visual statistical learning*, or VSL, is to extract the statistical regularities from the environment. Another strategy is to use the hierarchical structure of a scene (e.g., the co-occurrence between local and global information). Through a series of experiments, this study investigated whether the utilization of the statistical regularities and the hierarchical structure could work together to reduce the complexity of a scene. In the familiarization phase, the participants were asked to passively view a stream of hierarchical scenes where the shapes were concurrently presented at the local and global levels. At each of the two levels there were temporal regularities among the three shapes, which always appeared in the same order. In the test phase, the participants judged the familiarity between 2 triplets, whose temporal regularities were either preserved or not. We found that the participants extracted the temporal regularities at each of the local and global levels (Experiment 1). The hierarchical structure influenced the ability to extract the temporal regularities (Experiment 2). Specifically, VSL was either enhanced or impaired depending on whether the hierarchical structure was informative or not. In summary, in order to process a complex scene, the visual system flexibly uses statistical regularities and the hierarchical structure of the scene.

Keywords Visual statistical learning · Hierarchical structure

The visual system has to cope with complex and ever-changing inputs in daily life. For example, most objects differ from one another in their surface features (e.g., color, brightness, orientation, texture), internal configurations, locations in space, and so on (Biederman, 1987). In addition, due to the continuous movements of objects and ourselves, our visual experience is dynamic rather than stationary (Freyd, 1987; Gibson, 1979; Pylyshyn & Storm, 1988). Despite the complexity of visual information, we can still perceive the world vividly (Brady, Konkle, Alvarez, & Oliva, 2008) and recognize visual environments without considerable effort (Humphreys & Riddoch, 2001). How can our visual system process complex information efficiently?

One possible way to examine this is to use statistical regularities in a scene (Chun, 2003; Turk-Browne, 2012). Our visual environments are not random but contain statistical regularities such as repeated contexts (Chun, 2000, 2003; Chun & Jiang, 1999; Oliva & Torralba, 2007), spatial configurations (Fiser & Aslin, 2001), or regular sequences of objects (Fiser & Aslin, 2002a). For example, sequences of objects (e.g., the parking lot, elevator, and office room during one's daily commute) constantly appear in the same order. Using a process called *visual statistical learning*, or VSL, the visual system can both learn and use these statistical regularities (Chun & Jiang, 1998; Fiser & Aslin, 2001; Orbán, Fiser, Aslin, & Lengyel, 2008). When statistical regularities are present in a scene or a sequence, observers can learn these regularities (Brady & Oliva, 2008; Chun & Jiang, 1998; Fiser, 2009; Fiser & Aslin, 2001, 2002a, 2002b, 2005; Fiser, Scholl, & Aslin, 2007; Kirkham, Slemmer, & Johnson, 2002; Otsuka, Nishiyama, Nakahara, & Kawaguchi, 2013; Turk-Browne, Scholl, Chun & Johnson, 2009).

✉ Sang Chul Chong
scchong@yonsei.ac.kr

¹ Graduate Program in Cognitive Science, Yonsei University, Seoul, South Korea

² Department of Psychology, Yonsei University, Seoul, South Korea

The learning of statistical regularities can compensate for the limited capacity of the visual system (Palmer, 1975). For instance, Brady, Konkle, and Alvarez (2009) showed that when a display contained statistical regularities within pairs of colored items, the participants could hold almost five colors at once, exceeding their capacity limit of three or four. In another study by Umemoto, Scolar, Vogel, and Awh (2010), the participants could detect changes more rapidly and accurately if these changes repeatedly took place in a given quadrant. Similarly, Zhao, Al-Aidroos, and Turk-Browne (2013) found that the participants searched for targets faster in a location that featured statistical regularities. In this location, three shapes always appeared in the same order, whereas in other locations every shape appeared in random order.

Another way to cope with the complexity of a scene is to use the hierarchical structure of our environments (Brady, Konkle, & Alvarez, 2011; Im & Chong, 2014). Specifically, the visual information of a scene can be organized into local- or global-level information (Navon, 1977). Whether the visual information is local or global is decided by the relative level where the properties exist within the hierarchical structure of a scene (Kimchi, 2014). For example, when we perceive a building, we estimate the average height and overall shape of the building (e.g., the height of 25 or 30 stories with the shape of a skyscraper). On the other hand, we may notice features of several components that comprise the building (e.g., small gray bricks and white balconies on each floor). In this example, the former (i.e., the overall properties of the building) is at the top of the hierarchical structure compared to the latter (i.e., the features of individual items). Therefore, extracting the overall information is at a relatively more global level than encoding the features of the individual elements that comprise the building. This hierarchical structure of visual information is present in our memory representations, and the visual memory at the two different hierarchical levels can interact with one another (Brady et al., 2011).

The encoding of hierarchical structures helps observers to compensate for their limited capacity and to acquire detailed information (Brady et al., 2011). When the number of items in a display exceeds the limited capacity of working memory, observers experience an information overload (Cowan, 2001). However, Brady and Alvarez (2011) found that observers could accurately memorize the size of individual items when their estimations were close to ensemble statistics (e.g., the mean size of all items in the display or the mean size of a set of items sharing the same color). They suggested that the encoding of hierarchical structures helped observers to base their judgments on global-level aspects of the display (i.e., ensemble statistics) instead of guessing. Im and Chong (2014) also showed that observers could better estimate the mean sizes of groups of items when the items that shared the same color were spatially grouped, thereby allowing extraction of the hierarchical relationship between the items to facilitate the process.

In summary, these two mechanisms (i.e., the utilization of the statistical regularities and the hierarchical structure) can explain how the visual system is able to handle complex scenes despite its limited capacity. For example, people who commute every day can learn the statistical regularities on their route: there is a parking lot on the right side of the main entrance (i.e., spatial regularities), and they have to pass the parking lot and take an elevator before they reach their office (i.e., temporal regularities). On the other hand, they can represent the hierarchical structure of the inputs: Their car, next to a friend's car (i.e., local information), is situated on the left side of the parking lot (i.e., global information). By taking advantage of these statistical regularities and hierarchical structure, they can efficiently represent the location of their cars, the entrance, the elevator, their desks, computers, and so on.

How are these two mechanisms related to one another? According to Fiser and Aslin (2005), when statistical regularities and hierarchical structure coexist in the same spatial domain, the two mechanisms seem to interfere with one another. Specifically, spatial statistical learning appears to be constrained by the hierarchical structure when spatial regularities are present at different hierarchical levels. In their experiments, four shapes always appeared in the same configuration and formed a quadruple, with the constraint that spatial regularities not only existed in the quadruple but also in a pair embedded in the quadruple. Therefore, the statistical regularities in the embedded pair could be considered local whereas those in the quadruple could be considered global. Spatial statistical learning only occurred at the global level of the quadruple but not for the local-level embedded pair. However, spatial statistical learning occurred for another type of pair that appeared adjacent to, but separate from, the quadruple. In other words, VSL occurred when a pair was not part of a hierarchical structure (e.g., the quadruple). These results indicate that the learning of statistics at a global level (i.e., in the quadruple) could constrain the extraction of regularities at the local level (i.e., in the embedded pair).

However, in our visual environments, hierarchical structures and statistical regularities do not always coexist in the same spatial domain. For instance, while the hierarchical structure of a scene may be represented in the spatial domain (e.g., a building consists of six floors), the statistical regularities of that scene can exist in the temporal domain (e.g., after passing the parking lot and the elevator, people can reach their office). When the hierarchical structure and statistical regularities do not occur in the same domain, how do the two mechanisms interact with one another? Based on several findings suggesting that spatial and temporal information can be processed independently (Bengtsson, Ehrsson, Forssberg, & Ullén, 2004; Karabanov & Ullén, 2008; Kornysheva, Sierk, & Diedrichsen, 2013; Ullén & Bengtsson, 2003), we assume that they can work together. In Ullén and Bengtsson's (2003) sequential learning task, the participants learned and reproduced three types of sequences in

different blocks: one temporal, one spatial, and one combined sequence. In the temporal sequence, each stimulus appeared for varying durations, but always in the same location, so that only temporal information was present in the sequence. In the spatial sequence, each stimulus appeared in certain locations, but always for the same duration, so that only spatial information was present in the sequence. The combined sequence was a combination of these temporal and spatial sequences so that both temporal and spatial information were present in the sequence. When the participants learned the combined sequence first, the ensuing learning of the temporal and spatial sequences was facilitated. In addition, when the participants separately learned the temporal and spatial sequences first, the learning of the combined sequence was also facilitated in both temporal and spatial domains. Based on these results, the authors suggested independent representations of temporal and spatial structures. Furthermore, a functional magnetic resonance imaging (fMRI) study (Bengtsson et al., 2004) showed a dissociation between the brain regions involved in the learning of temporal structures (i.e., the presupplementary motor area, right inferior frontal gyrus, precentral sulcus, and bilateral superior temporal gyri) and those involved in the learning of spatial structures (i.e., the lateral fronto-parietal areas, basal ganglia, and cerebellum).

In this study, we used a hierarchical structure and statistical regularities occurring in different domains. While the hierarchical structure was constructed in the spatial domain, the statistical regularities occurred in the temporal domain. We investigated whether the participants could extract the temporal regularities represented at different hierarchical levels simultaneously (Experiment 1A). We additionally tested whether the participants used the hierarchical structure, not two different features of a hierarchical stimulus (i.e., smaller and larger shapes) to extract the temporal regularities (Experiments 1B and 1C). Next, we examined whether the processing of the hierarchical structure could influence the degree of statistical learning (Experiment 2). In each experiment, a familiarization phase was followed by a test phase. In the familiarization phase, the participants passively viewed a sequence of Navon-like objects (see Fig. 1), where two different shapes formed a hierarchical structure at the local and global levels. Unbeknownst to the participants, the sequence featured temporal regularities among the three shapes (i.e., the triplet) at both the local and global levels, as these three shapes always appeared in the same order. The familiarization phase of the two experiments was identical. In the test phase, the participants performed a surprise two-alternative forced-choice (2AFC) familiarity judgment task. In each trial, the participants were asked to choose the more familiar sequence of the two triplets; that is, a base triplet and a nonbase triplet. In the base triplet, the temporal regularities were preserved compared to the familiarization phase whereas in the nonbase triplet the temporal regularities were changed.

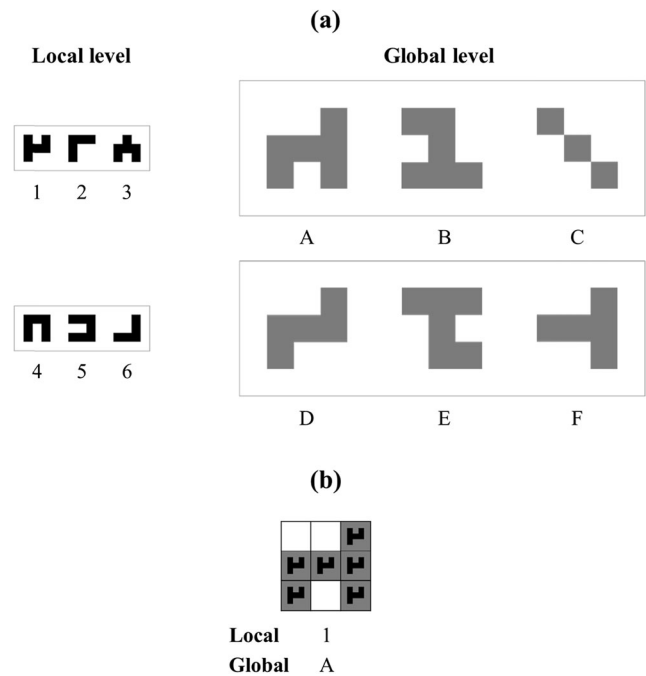


Fig. 1 Stimuli for the 12 novel shapes and an example of a display actually viewed by the participants. (a) Half of the shapes were randomly chosen to form one of the two triplets at the local level, and the other half were allocated to the two triplets at the global level. The three shapes at the local level were identified with numbers whereas those at the global level were identified with capital letters below gray boxes. The numbers, letters, and gray boxes were only for demonstration and were not part of the actual display. (b) The hierarchically structured display was composed of two different local (e.g., 1) and global (e.g., A) shapes. The backgrounds of the cells where the local shapes appeared were displayed in gray color to make the global shape evident

In Experiment 1A, the temporal statistical learning was tested at each local and global level. Depending on the tested level (e.g., local level), each display contained only one shape (e.g., a shape at the local level) that had appeared in that level during the familiarization phase, so that the two levels were tested separately. To rule out an alternative possibility that the participants might have used two different shapes of the hierarchical stimulus, not the hierarchical structure per se, in Experiment 1B, we again tested the temporal VSL at the global level by eliminating the shape information and only providing the spatial hierarchy. In Experiment 1C, we conducted the same test conditions as in Experiment 1B, after equating the number of test trials at the local level with that at the global level.

To further examine the relationship between the two mechanisms, in Experiment 2 we tested whether the hierarchical structure could influence the degree of statistical learning. We preserved the hierarchical structure during the test phase; that is, the two shapes at both levels simultaneously appeared the same as in the familiarization phase. The temporal statistical learning was tested at both levels simultaneously, as well as at each of the two levels. In other words, depending on the tested level (e.g., local level), only the shapes that had occurred at that level (e.g., shape at the local level) during the

familiarization phase changed the temporal regularities in the nonbase triplets. However, the shapes at the other level (e.g., shape at the global level) featured the same temporal regularities as in the base triplets or during the familiarization phase.

To provide a preview of our results, temporal statistical learning was possible at each of the different hierarchical levels (Experiment 1A). In addition, the degree of statistical learning did not differ between the two levels. We also found that the participants used the hierarchical structure when they extracted the temporal regularities (Experiments 1B and 1C). In Experiment 2, statistical learning was influenced by the hierarchical structure. When the hierarchical structure was helpful for familiarity judgments, it enabled VSL; conversely, it impaired VSL when it was not informative.

Experiment 1A

Visual statistical learning at different hierarchical levels

In Experiment 1A, we investigated whether temporal statistical learning could occur at different hierarchical levels (i.e., local and global). Previously, Fiser and Aslin (2005) showed that VSL did not occur when statistical regularities and hierarchical structures coexisted in the same spatial domain. However, this study used the temporal domain to present the statistical regularities and the spatial domain to present the hierarchical structure. Experiment 1A tested whether the participants could extract temporal regularities at the local and global levels.

Method

Participants

Twenty naïve students from Yonsei University participated in Experiment 1A. They were paid 5,000 Won for their participation. All of the participants reported normal or corrected-to-normal vision. The protocol of the experiment was approved by the Yonsei University Institutional Review Board, and the

participants were informed of their rights before signing written consent forms.

Apparatus and stimuli

We presented the stimuli using MATLAB and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The display was a linearized Samsung 21-in. monitor with a resolution of 1,600 × 1,200 pixels and a refresh rate of 85 Hz. The experiment was conducted in a dark room. The participants’ heads were fixed on a chin-and-forehead rest at a viewing distance of 90 cm; one pixel subtended 0.016° at this distance.

We created a total of 12 novel shapes (see Fig. 1a). Half of the shapes were randomly chosen to be used as shapes at the local level and the other half at the global level. As in Navon (1977), the shape of the global-level structure was composed of the local-level shapes (see Fig. 1b). The two shapes at the two different levels were always different and presented in a 3 × 3 grid. Each local-level shape occupied a single cell of the grid. To form the global-level shape, this local shape repeatedly appeared in more than three (but less than seven) cells. The background of the cells was filled with gray color in order to make the global shape evident. The color of the local shapes was black (0.10 cd/m²) and that of the global shapes was gray (18.8 cd/m²). The line color of the grid was black and the color of the background was white (91.4 cd/m²). The maximum width and height of the shapes was 2.29° and 12° at the local and global level, respectively. The side of a cell in the grid was 4°.

Design and procedure

Experiment 1A comprised two phases: a familiarization phase and a test phase. The participants completed the familiarization phase (see Fig. 2) and then performed the test phase (see Fig. 3a and b).

Familiarization phase The familiarization phase included a movie with a 2 × 2 design, that is, two levels of display (local and global) and two kinds of triplets in each level (Triplets 1 and 2). Unbeknownst to the participants, the 12 shapes were randomly allocated to one of the four triplets: two triplets were

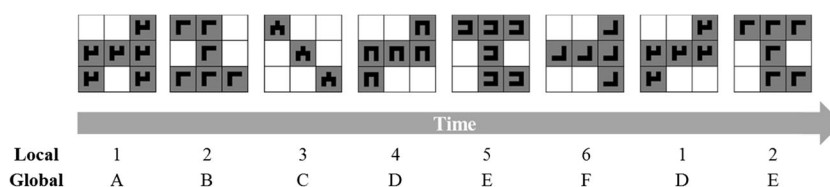


Fig. 2 In the familiarization phases of Experiments 1 and 2, the participants passively viewed a sequence of hierarchically structured displays. In these displays, each of the local and global levels contained temporal regularities where three shapes always appeared in the same order. For example, triplets across the local and global levels appeared

in the following manner: 1-2-3 or 4-5-6 at the local level, and A-B-C or D-E-F at the global level. There was no segmentation cue between the triplets, whereas each display sequentially appeared for 1 s with a blank interval of 750 ms

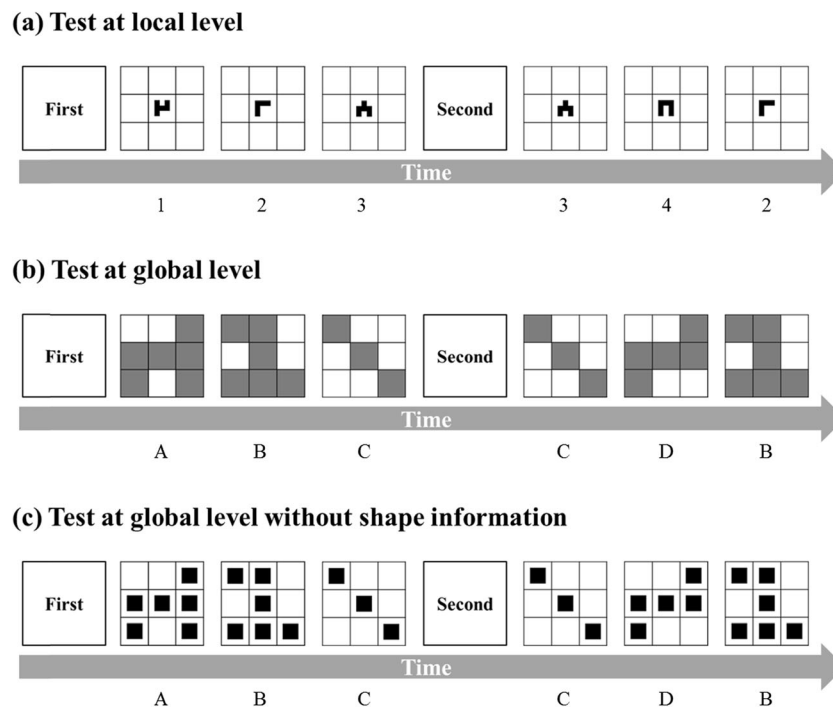


Fig. 3 In the test phase of Experiment 1A, the participants were asked to choose the more familiar sequence after base and nonbase triplets were sequentially presented. While the temporal regularities in the base triplets were preserved from the familiarization to the test phase (e.g., 1-2-3 or A-B-C), those in the nonbase triplets were changed (e.g., 3-4-2 or C-D-B).

The temporal visual statistical learning (VSL) was tested at both (a) the local and (b) global levels. In addition to the two conditions, in Experiments 1B and 1C, we added (c) the test condition at the global level without providing shape information (i.e., the gray shading and the black shapes

displayed at the local level and the other two were shown at the global level. The three shapes in each triplet presented temporal regularities, as they always appeared in the same order. For instance, if naming the six shapes of the local level as numbers (i.e., 1–6) and those of the global level as capital letters (i.e., A–F), the 12 shapes across the four triplets could appear in the following manner: 1-2-3 or 4-5-6 at the local level and A-B-C or D-E-F at the global level. Every display contained a hierarchical structure composed of the local and global shapes. When two different shapes occurred simultaneously at the local and global levels, each shape had its own temporal position (i.e., first, second, or third) within its triplet. In this way, the combination of the two triplets at the local level with the other two triplets at the global level created four possible kinds of triplet displays that were hierarchically structured (i.e., 1A-2B-3C, 1D-2E-3 F, 4A-5B-6C, and 4D-5E-6 F). As the four kinds of triplet displays were repeated 35 times, the familiarization movie could be segmented into 140 triplet displays or 140 triplets for each level; however, there was no segmentation cue between the triplet displays or the triplets. The same triplet display was not presented successively. As each of the 140 triplet displays consisted of three displays presented sequentially, 420 displays were presented to the participants one after another. The frequency of the 12 individual shapes was the same. At either the local or the

global level, the joint probability of the three shapes in each triplet was 0.17.

The experimental procedure adopted an observational learning paradigm (Fiser & Aslin, 2001). The participants passively viewed a 12-min movie without any instructions. The movie consisted of a series of 420 hierarchically structured displays, with blank intervals between each display. The duration of each display was 1 s, and the duration of the blank interval was 750 ms.

Test phase In Experiment 1A, we tested whether the participants could extract temporal statistics at the local (see Fig. 3a) and global level (see Fig. 3b), and whether the extent of the statistical learning was different between the two levels. To investigate this, we tested whether the participants could discriminate a base triplet from a nonbase triplet. The temporal order was preserved for the base triplets but changed for the nonbase triplets.

There were four within-subject variables. First, the extent of the temporal VSL was tested at each of the two different levels: local and global. Depending on the tested level, only individual shapes that had been present at the corresponding level during the familiarization phase were sequentially presented, whereas shapes that had been present at the other level during the familiarization phase were not presented. Second,

there were two kinds of base triplets for each of the two levels: 1-2-3, 4-5-6 for the local level, and A-B-C, D-E-F for the global level. Third, there were two kinds of nonbase triplets for each of the two levels: 2-4-1 and 6-3-5 or 3-4-2 and 6-1-5 for the local level or B-D-A and F-C-E or C-D-B and F-A-E for the global level. We constructed the nonbase triplets by choosing three shapes from two different base triplets such that the joint probability of the three shapes in the nonbase triplet became 0. In contrast, the joint probability of the three shapes in the base triplets was 0.17, as in the familiarization. Fourth, there were two different orders of presentation: the base triplet first or the nonbase triplet first. The participants therefore performed 16 randomized trials. The frequency of the 12 individual shapes was equal for the base and nonbase triplets.

The procedure adopted a two-alternative forced-choice (2AFC) task for familiarity judgments. Before either the base or nonbase triplets were presented, the words *First* and *Second* were shown on a blank screen for 1 s in order to distinguish the two triplets. The duration of each display and of the blank intervals between the displays were identical to those of the familiarization. After viewing the two sequences (i.e., the base and nonbase triplets), the participants were asked to decide what sequence was more familiar, based on their experience of the previous 12-min movie, by pressing 1 for the first sequence or 2 for the second sequence.

After the familiarity test, we also examined whether explicit awareness of temporal regularities influenced visual statistical learning, by using a binary confidence judgment task including two different statements (Bertels, Franco, & Destrebecqz, 2012). The first statement indicated that the test had been performed based on some kind of explicit knowledge (i.e., “I chose the answers based on some kind of knowledge that I learned during the familiarization phase.”). The second statement indicated that the test had been performed in an implicit manner (i.e., “I guessed the answers based on my intuition.”). Only the participants who had chosen the first statement were further asked what kind of knowledge they had by reporting whether they were aware of the existence of the temporal regularities and how many shapes regularly appeared in a sequence.

Results and discussion

In this and the following experiment, the extent of the temporal VSL was measured with the mean percentage of correct trials, that is, trials where the participants had selected the base triplets as being more familiar than the nonbase triplets were. The results of Experiment 1A are shown in Fig. 4.

Overall, the mean percentage of correct responses was 66.25 %, which was significantly higher than 50 % (chance level), as assessed by a one-sample *t* test, $t(19) = 3.73$, $p = .001$. Specifically, the mean percentage for the local level was

67.50 %, which was significantly higher than chance, $t(19) = 3.50$, $p = .002$, while the mean percentage for the global level was 65.00 %, which also significantly differed from chance, $t(19) = 3.21$, $p = .005$. Therefore, the participants had learned the temporal regularities at both the local and global levels. However, the degree of learning did not differ significantly between the two levels, as revealed by a paired-samples *t* test, $t(19) = .59$, $p = .560$. In this and subsequent experiments, the participants could not rely on the frequency of the individual shapes, as every shape in the base and nonbase triplets had appeared equally frequently in the familiarization phase. The change of temporal regularities in the nonbase triplets for either of the two levels was the only source of information that the participants could base their judgments.

The presentation order did not influence the participants' performance. A paired-samples *t* test showed no significant difference between the trials when the base triplets were presented first (test performance: 68.75 %) and those when they were presented second (test performance: 63.75 %), $t(19) = .85$, $p = .408$. We also tested the possibility that the participants might have used the test phase itself to learn parts of the base triplets. Comparing the performance of the first half (67.50 %) with that of the second half (65.00 %), we found no significant difference, $t(19) = .44$, $p = .666$, indicating that learning had not occurred during the test phase.

The awareness of the temporal regularities improved statistical learning. In the binary confidence judgment, 11 out of the 20 participants reported having used some kind of knowledge (test performance: 77.85 %). Among these 11 participants, eight specifically reported being aware of the temporal regularities (test performance: 85.15 %), whereas the other nine participants reported that they had responded by guessing (test performance: 52.08 %). The eight participants who were aware of the temporal regularities performed better than the other 12 who were not aware of the temporal regularities, as assessed with an independent-samples *t* test, $t(18) = 5.91$, $p < .001$.

Experiment 1B

Visual statistical learning at the global level without shape information

Experiment 1A found that temporal statistical learning took place both at the local and global levels when the statistical learning and hierarchical structure were presented in separate spatiotemporal domains. In addition, there was no difference between the extent of statistical learning at the local and global level. However, one might argue that observers might not represent the hierarchical structure but, rather, they might have used the two different shapes to perform the familiarity judgment task, such as the gray (global) and the black (local)

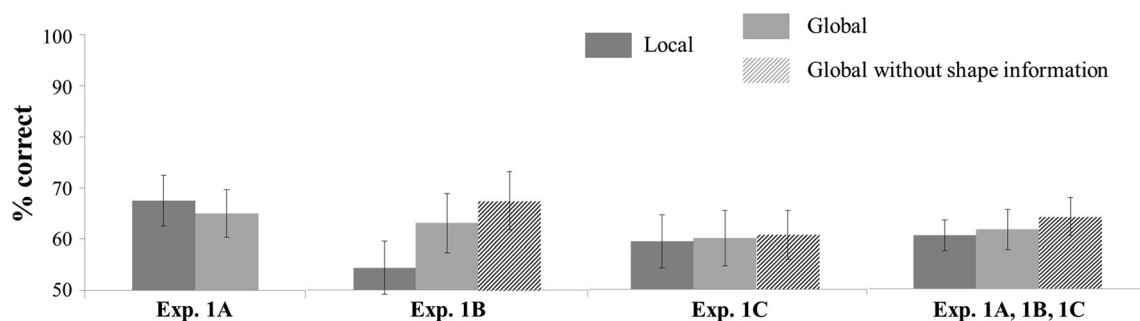


Fig. 4 Results of Experiment 1. The mean percentage of correct discriminations between base and nonbase triplets in Experiments 1A, 1B, and 1C (the first three panels), and the mean percentage collapsed across the three experiments (the fourth panel) are displayed. Each bar in

dark gray, gray, and a tilted pattern represents the local level, the global level, and the global level without shape information, respectively. The error bar represents the standard error of the mean

shapes. To rule out this possibility, in Experiment 1B we again tested the temporal VSL at the global level without providing strong shape cues (i.e., the gray shading and the black shapes). In the global level of Experiment 1B, new shapes replaced the local shapes that had appeared in the familiarization phase, thus preserving the hierarchy of the stimuli without the two shape cues. However, we hypothesized that the participants could still judge the familiarity correctly if they could use the spatial hierarchy of the local and global shapes that they had represented during the familiarization phase.

Method

Participants

Twenty new and naïve students from Yonsei University participated in Experiment 1B in exchange for course credit. All had normal or corrected-to-normal visual acuity. The study protocol was approved by the Yonsei University Institutional Review Board, and the participants provided informed consent forms.

Apparatus and stimuli

The apparatus and stimuli were the same as in Experiment 1A, except for stimuli used in a new test condition. In this condition, a square with the same color (black, 0.10 cd/m^2), width, and height (2.29°) of the local shape was used as a new shape.

Design and procedure

As in Experiment 1A, the familiarization phase (see Fig. 2) was followed by the test phase (see Fig. 3).

Familiarization phase The familiarization phase was the same as in Experiment 1A.

Test phase The new test condition at the global level without providing the shape information (see Fig. 3c) was added to the local- (Fig. 3a) and global-level test conditions (Fig. 3b), which were the same as in Experiment 1A. This new condition intended to test whether the participants could use the spatial hierarchy of the local shapes, not the gray shapes, when extracting the temporal regularities at the global level. For this purpose we eliminated the gray shading and the local shapes, and we displayed the global structure by replacing the local shapes that had appeared in the familiarization phase with the new shapes (i.e., the black squares, which did not appear in the familiarization phase).

We employed a $3 \times 2 \times 2$ within-subjects design. The first variable was whether the temporal VSL was tested at the global level where the shape information was either provided or not, and at the local level. The other three variables were the same as in Experiment 1A.

The procedure applied the same 2AFC familiarization judgment task as in Experiment 1A. After the familiarity test, the participants responded to the same binary confidence task to check the explicit awareness of the temporal regularities.

Results and discussion

The results of Experiment 1B are shown in Fig. 4. Overall, the mean correct percentage for the three conditions was 61.67 %, which was significantly higher than chance, $t(19) = 2.51$, $p = .021$. Specifically, for the global level without the shape information, the mean percentage of correct responses was 67.50 %, which was significantly higher than chance, $t(19) = 3.07$, $p = .006$, and did not differ from the test performance at the global level in Experiment 1A, $t(38) = .251$, $p = .803$. That is, participants could judge the global-level regularities correctly when they could rely only on the spatial hierarchy of the local shapes. For the global level with the gray shapes, the mean correct percentage was 63.13 %, which was also significantly higher than chance, $t(19) = 2.25$, $p = .037$, and did not

differ from the global performance in Experiment 1A, $t(38) = .339$, $p = .736$. In addition, the test performances at the two global levels (i.e., with or without the shape information) were not different from one another, $t(19) = .941$, $p = .358$. Therefore, the results showed that the extent of learning the temporal regularities was reliable at the global level, regardless of the presence of the shape cues.

On the other hand, the test performance at the local level was 54.38 %, which was not significantly different from chance, $t(19) = .84$, $p = .413$. That is, in Experiment 1B, the participants could not extract the temporal regularities at the local level, failing to replicate the results at the local level in Experiment 1A. One possibility that explains these results is that the discrepancy in the number of test trials at the local level (eight trials) and at the two global levels (16 trials, with or without the shape cues) might have induced the participants to focus mainly on the global level in the test phase. To test this hypothesis, in Experiment 1C, we equated the number of test trials at the local level (16 trials) with that of the two global levels (16 trials) and tested the temporal VSL again.

The presentation order did not influence the test performance as it did in Experiment 1A. There was no significant difference between the trials when the base triplets appeared first (test performance: 62.92 %) and those when the base triplets appeared second (test performance: 60.42 %), $t(19) = .49$, $p = .627$. In addition, the participants did not learn the temporal regularities during the test phase, because the performance did not differ between the first (61.67 %) and the second (61.67 %) halves of the test, $t(19) < .001$, $p > .999$.

The awareness of the temporal regularities tended to influence temporal statistical learning. The results in the binary confidence judgments showed that among eight participants who reported having some kind of knowledge (test performance: 69.27 %), seven reported being aware of the temporal regularities (test performance: 72.62 %), while the other 12 reported having responded by guessing (test performance: 56.60 %). The difference in the test performance between the seven participants who were aware of the temporal regularities and the other 13 who were not aware approached statistical significance, $t(18) = 1.83$, $p = .083$.

Experiment 1C

The equated number of test trials at the local and global levels

In Experiment 1B, the participants could still extract the temporal regularities at the global level when they could rely only on the spatial hierarchy of the local shapes. Therefore, the participants used the hierarchical structure when they learned the temporal regularities at the global level. In addition, the participants showed a reliable ability to extract the global-

level temporal regularities, regardless of the presence of shape information. Contrary to the participants in Experiment 1A, who could also extract the temporal regularities at the local level, those in Experiment 1B could not extract these local regularities. We hypothesize that the discrepancy in the number of test trials between the local and global levels might influence the familiarity judgments at the local and global levels in a different way. To test this hypothesis, we conducted the same experiment as in Experiment 1B, except that we equated the number of test trials at the local level with that at the global levels.

Method

Participants

Twenty new naïve students from Yonsei University who did not take part in Experiments 1A or 1B participated in Experiment 1C for course credit. All reported normal or correct-to-normal visual acuity, and signed informed consent forms. The Yonsei University Institutional Review Board approved the study protocol.

Apparatus and stimuli

The apparatus and stimuli were the same as in Experiment 1B.

Design and procedure

As in Experiments 1A and 1B, the familiarization phase (see Fig. 2) was followed by the test phase (see Fig. 3).

Familiarization phase The familiarization phase was the same as in Experiments 1A and 1B.

Test phase The procedure of the test phase was the same as in Experiment 1B, except that we increased the number of test trials at the local level to a total of 16 trials. For this purpose, the same eight trials as in Experiment 1B were repeated twice. The participants performed 32 trials that were randomly interleaved. After this familiarity test, the same binary confidence task as in Experiments 1A and 1B was conducted again.

Results and discussion

The results of Experiment 1C are shown in Fig. 4. The overall mean percentage for correct responses was 59.84 %, which was significantly higher than chance, $t(19) = 2.43$, $p = .025$. Specifically, for the local level, the mean correct percentage was 59.38 %, which approached statistical significance, $t(19) = 1.80$, $p = .088$. Despite this marginal significance, the local-level performance in Experiment 1A, which was significantly higher than chance, did not differ from that in Experiment 1C,

$t(38) = 1.12, p = .268$. Furthermore, the collapsed performance (63.44 %) from the local conditions in the two experiments was significantly higher than chance, $t(39) = 3.71, p = .001$. These results suggest that the participants learned the temporal regularities at the local level.

For the global level without the shape information, the test performance was 60.63 %, which was significantly higher than chance, $t(19) = 2.20, p = .040$. Therefore, our results again showed that the participants could use the hierarchical structure to extract the temporal regularities at the global level. For the global level with the shape information, the test performance was 60.00 %, showing a marginally significant learning effect, $t(19) = 1.85, p = .080$. As in Experiment 1B, the presence of the shape cues did not influence the degree of temporal VSL at the global level, $t(19) = -.113, p = .912$. The collapsed performance (60.1 %) of the two global conditions was significantly higher than chance, $t(19) = 2.39, p = .027$. The comparison of the test performance between Experiments 1A and 1C also supported the evidence of learning the temporal regularities at the global level. The performance at the same global test conditions with the shape information was not significantly different between the two experiments, $t(38) = .699, p = .489$. The collapsed data from the same global conditions in the two experiments again showed that the participants learned the temporal regularities at the global level (test performance: 62.66 %, $t(39) = 4.00, p < .001$).

The test performance was not significantly different between the local level and the global level with the shape cues, $t(19) = .091, p = .928$. Again, there was no significant difference between the local and the global level without the shape cues, $t(19) = .275, p = .786$. Therefore, regardless of the shape cues, the degree of temporal VSL was similar between the local and global levels.

The presentation order did not influence the test performance as in Experiments 1A and 1B because the test performances when the base triplets appeared first (59.69 %) and second (60.00 %) were not significantly different, $t(19) = .061, p = .952$. In addition, the learning did not occur during the test phase because the performance for the first half of the test phase (58.13 %) and that for the second half (61.56 %) was not significantly different, $t(19) = .825, p = .420$.

The awareness of the temporal regularities improved statistical learning. Eight out of 20 participants reported using some kind of knowledge (test performance: 72.66 %), and seven of the eight reported being aware of the temporal regularities (test performance: 75.00 %), whereas the other 12 participants reported guess responses (test performance: 51.30 %). The seven participants who were aware of the temporal regularities performed better than the other 13 who were not aware of them, $t(19) = 3.45, p = .003$.

In summary, our Experiment 1 results revealed that the participants extracted the temporal regularities at the different hierarchical levels. The fourth panel of Fig. 4

shows the results of the collapsed data across Experiments 1A, 1B, and 1C. The participants judged the temporal regularities correctly above chance with 62.59 % accuracy on the whole, $t(59) = 5.04, p < .001$, with 60.42 % accuracy at the local level, $t(59) = 3.46, p = .001$, and with 63.54 % accuracy at the global level, $t(59) = 4.97, p < .001$. The degree of learning was not significantly different between the local and global levels, $t(59) = 1.09, p = .280$. Regardless of the presence of the shape cues, the participants showed reliable performance to judge the temporal regularities at the global level (Experiments 1B and 1C). Therefore, the participants used the hierarchical structure, not the two different shapes of the hierarchical stimulus, when they extracted the temporal regularities. In addition, the overall results in Experiment 1 showed that the explicit awareness of the temporal regularities improved the statistical learning. The participants who reported being aware of the temporal regularities (22 participants, test performance: 77.94 %) performed the test of the temporal VSL better than those who did not reported this kind of awareness (38 participants, test performance: 53.70 %) $t(58) = 5.843, p < .001$.

Experiment 2

Influence of the hierarchical structure on the degree of visual statistical learning

Experiment 1 revealed that the temporal VSL occurred both at the local and global levels without difference in the extent of learning between the two levels. To extract the temporal regularities, the participants used the spatial hierarchy of the local and global shapes. However, it is still unclear how the ability to utilize the statistical regularities and the hierarchical structure are related to one another. In Experiment 2, we examined whether the utilization of the hierarchical structure influenced the degree of learning the statistical regularities. Specifically, unlike in Experiment 1, the test phase in Experiment 2 always maintained the same hierarchical structure as in the familiarization phase. The temporal VSL was tested at both levels: at the local level only, and at the global level only. Consequently, this experiment tested whether and how the hierarchical structure could influence the degree of temporal statistical learning. We also tested whether the participants used the hierarchical structure. Specifically, we tested whether the participants utilized the hierarchical co-occurrence between the local and global shapes, according to whether each of the shapes at one level (e.g., the first shape of the local triplets) co-occurred with particular shapes at the other hierarchical level (e.g., the first shape of the global triplets).

Methods

Participants

Eighty new Yonsei University students who were unaware of the purpose of the experiment participated in Experiment 2. All of the participants were paid and had normal or corrected-to-normal visual acuity. The Yonsei University Institutional Review Board approved the experimental protocol, and all the participants signed informed consent forms.

Apparatus and stimuli

The apparatus and stimuli were the same as in Experiment 1A.

Design and procedure

As in Experiment 1A, the participants first completed a familiarization phase (see Fig. 2) before performing the test phase (see Fig. 5).

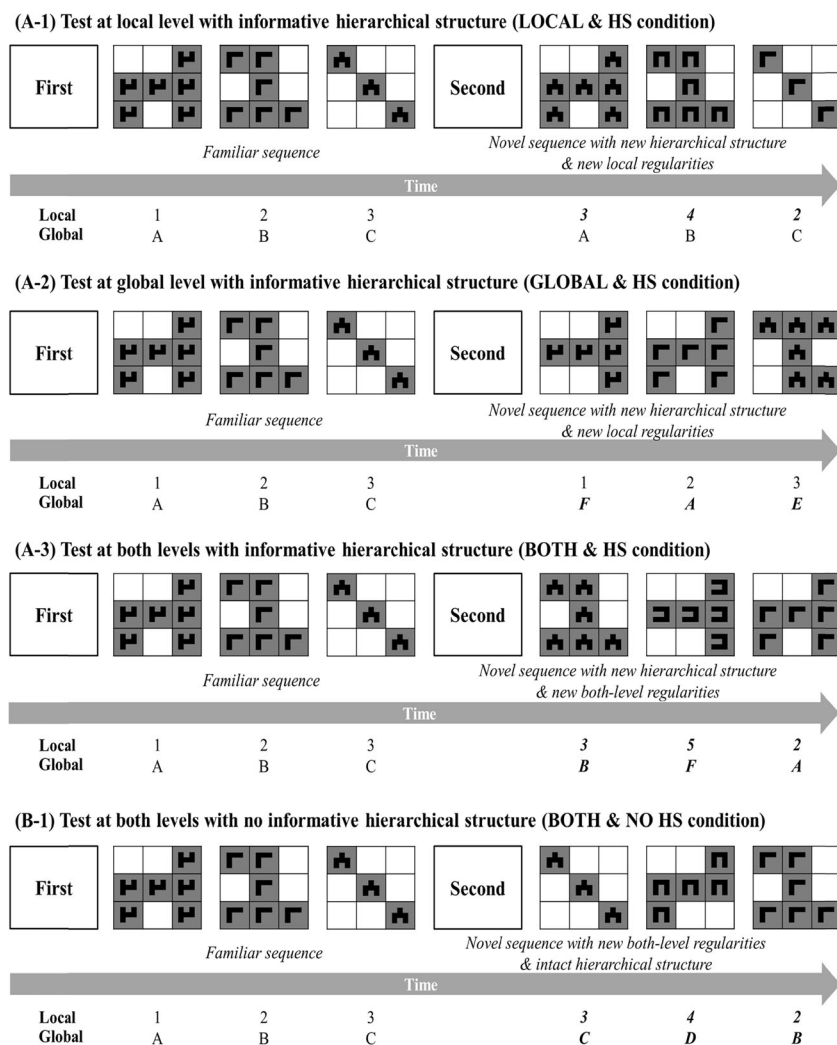


Fig. 5 In the test phase of Experiment 2, the participants in each of the four conditions were asked to choose the more familiar sequence after base and nonbase triplets were sequentially presented. In the base triplets, the hierarchical structure was composed of local and global shapes (e.g., 1A) and the temporal regularities at each of the two levels (e.g., 1-2-3 at the local level and A-B-C at the global level) were preserved from the familiarization. However, in the nonbase triplets, the hierarchical structure and temporal regularities could be either preserved or changed depending on the four experimental conditions. (A-1) In the LOCAL & HS condition, the temporal regularities at the local level (e.g., 3-4-2) and the hierarchical structure (e.g., 3A) were changed. (A-2) In the GLOBAL & HS

condition, the temporal regularities at the global level (e.g., F-A-E) and the hierarchical structure (e.g., 1F) were changed. (A-3) In the BOTH & HS condition, the temporal regularities at both levels (e.g., 3-5-2 at the local level and B-F-A at the global level) and the hierarchical structure (e.g., 3B) were changed. (B-1) In the BOTH & NO HS condition, the temporal regularities at both levels (e.g., 3-4-2 at the local level and C-D-B at the global level) were changed, whereas the hierarchical structure (e.g., 3C) was preserved. Therefore, depending on the experimental condition, the temporal regularities at the two levels and the hierarchical structure either did or did not provide useful information for the familiarity task

Familiarization phase The familiarization phase was the same as in Experiment 1A.

Test phase As in Experiment 1A, the ability to learn the temporal regularities was tested again, although there were three differences to Experiment 2. First, the two shapes hierarchically structured at the local and global levels appeared at the same time in each display. Therefore, in comparison with Experiment 1A, Experiment 2 displayed the shapes in a more similar way to the familiarization phase.

Second, the hierarchical structure played a different role depending on the experimental condition that the participants had been randomly assigned. In three out of the four conditions (see Fig. 5A-1, A-2, and A-3), the hierarchical structure was helpful for familiarity judgments. In these conditions, the hierarchical co-occurrence between the local and global shapes was changed in the nonbase triplet displays, whereas this co-occurrence was preserved in the base triplet displays. To illustrate, let us take an example of one of these condition (see Fig. 5A-1) where the hierarchical structure was informative for the familiarity judgments and the temporal VSL was tested at the local level. In this condition in an abbreviated form of a LOCAL & HS condition, the first shape in the local triplets (e.g., 1) and the first shape in the global triplets (e.g., A) co-occurred in the base triplet displays as in the familiarization phase. However, in the nonbase triplet displays, the first shapes co-occurred with different local shapes (e.g., 3, in this figure). If the participants represented the hierarchical structure, they could utilize the change in the hierarchical co-occurrence to identify the familiar sequence correctly. On the other hand, in the last condition (Fig. 5B-1), the hierarchical structure did not help with familiarity judgments. The hierarchical co-occurrence between the local and global shapes was preserved in both the nonbase and base triplet displays.

Third, the hierarchical levels where the temporal VSL was tested were manipulated over three conditions. When the VSL was tested at a single level, the temporal regularities provided useful information for that single level only, that is, the local (Fig. 5A-1) or global (Fig. 5A-2) level. However, when the

learning was tested at both levels, the temporal regularities were helpful for both levels (Fig. 5A-3, B-1).

Therefore, there were four different experimental conditions in Experiment 2 (see Table 1): LOCAL & HS, GLOBAL & HS, BOTH & HS, and BOTH & NO HS conditions. The capitalized names of the four conditions represented the hierarchical level where the temporal VSL was tested (e.g., BOTH, if both levels were tested) and the informativeness of the hierarchical structure (e.g., HS, if the hierarchical structure was informative; NO HS, if not informative). The 80 participants were randomly assigned to the four conditions, so that each condition comprised 20 participants. The participants judged the degree of familiarity between a base and nonbase triplet display. The base triplet display was one of the four kinds of triplet displays used in the familiarization phase (i.e., 1A-2B-3C, 1D-2E-3 F, 4A-5B-6C, or 4D-5E-6 F). The nonbase triplet display consisted of the same individual shapes, although the shapes appeared in a different hierarchical co-occurrence or with different temporal regularities at the local and global levels.

In the LOCAL & HS condition (see Fig. 5A-1), the participants could rely on the temporal regularities at the local level as well as on the hierarchical structure, as both kinds of information were changed in the nonbase triplet displays. Specifically, by combining two nonbase triplets that had never appeared sequentially at the local level during the familiarization phase (i.e., either 2-4-1 and 6-3-5 or 3-4-2 and 6-1-5, selected randomly) and two base triplets at the global level (i.e., A-B-C and D-E-F), four kinds of nonbase triplet displays were generated (i.e., either 2A-4B-1C, 2D-4E-1 F, 6A-3B-5C, and 6D-3E-5 F or 3A-4B-2C, 3D-4E-2 F, 6A-1B-5C, and 6D-1E-5 F).

In the GLOBAL & HS condition (see Fig. 5A-2), the participants could base their judgments on the temporal regularities at the global level and on the hierarchical structure. In the same way as in the LOCAL & HS condition, the two kinds of information were altered in the nonbase triplet displays. Specifically, after combining the two nonbase triplets that had never occurred sequentially at the global level during the familiarization phase (i.e., either B-D-A and F-C-E or C-

Table 1 The four experimental conditions of Experiment 2. The capitalized names of the four conditions represent the hierarchical level where the temporal visual statistical learning (VSL) was tested (e.g.,

BOTH, if both levels were tested) and the informativeness of the hierarchical structure (e.g., HS, if the hierarchical structure was informative; NO HS, if not informative)

		Temporal regularities	
		Single	Both
Hierarchical structure	Informative	LOCAL & HS GLOBAL & HS	BOTH & HS
	Uninformative		BOTH & NO HS

D-B and F-A-E, selected randomly) and two base triplets at the local level (i.e., 1-2-3 and 4-5-6), four kinds of nonbase triplet displays were constructed (i.e., 1B-2D-3A, 1 F-2C-3E, 4B-5D-6A, and 4 F-5C-6E or 1C-2D-3B, 1 F-2A-3E, 4C-5D-6B, and 4 F-5A-6E).

In the BOTH & HS condition (see Fig. 5A-3), the participants could make use of the temporal regularities at both levels and of the hierarchical structure, as the two kinds of information were changed in the nonbase triplet displays. The combination of the two nonbase triplets at the local level (i.e., 3-5-2 and 4-1-6) with those at the global level (i.e., B-F-A and E-C-D) generated four kinds of nonbase triplets (i.e., 3B-5F-2A, 3E-5C-2D, 4B-1F-6A, and 4E-1C-6D).

In the BOTH & NO HS condition (see Fig. 5B-1), the participants could rely on the temporal regularities at both levels, just as in the BOTH & HS condition. However, they could not use the hierarchical structure. In the nonbase triplet displays, the temporal regularities were altered at both the local and global levels, while the hierarchical co-occurrence of two shapes was identical to that in the base triplet displays. Specifically, after combining two nonbase triplets that had never been presented sequentially at the local level during the familiarization phase (i.e., 3-4-2 and 6-1-5) with those at the global level (i.e., C-D-B and F-A-E), four kinds of nonbase triplet displays were generated (i.e., 3C-4D-2B, 3F-4A-2E, 6C-1D-5B, and 6F-1A-5E). Nevertheless, each display maintained the hierarchical co-occurrence of the two shapes used in the familiarization phase (e.g., 3C).

In each of the four conditions, there were three within-subjects variables. First, four kinds of base triplet displays were the same as the four kinds of triplet displays (e.g., 1A-2B-3C) presented in the familiarization. Second, there were four kinds of nonbase triplet displays, depending on the condition. Third, there were two kinds of presentation orders (i.e., the base triplet coming either first or second). Therefore, every participant in the four conditions performed 32 randomized trials. As in Experiment 1, the frequency of the 12 individual shapes was equal in the base and nonbase triplet displays. The joint probability of the three shapes in the base triplets at the local or the global level was 0.17, that is, the same as in the familiarization. However, the joint probability of the three shapes in the nonbase triplets was 0.

The procedure followed the same 2AFC familiarity judgment task as in Experiment 1A. After the familiarity test, the participants conducted a binary confidence task designed to investigate whether the explicit awareness of the temporal regularities influenced the temporal VSL.

Results and discussion

The results of Experiment 2 are shown in Fig. 6. The mean correct percentage for the four conditions was 71.60 %, which was significantly higher than chance, $t(79) = 10.38, p < .001$.

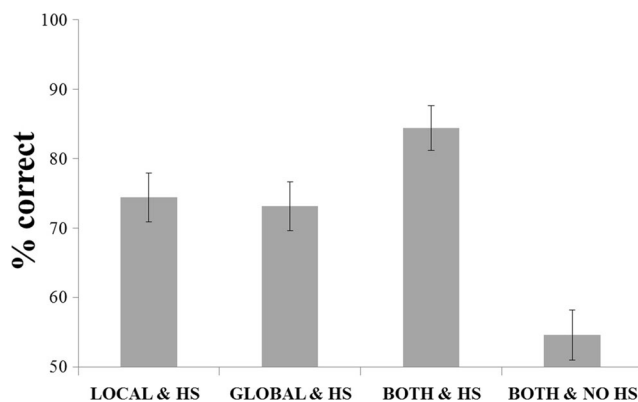


Fig. 6 Results of Experiment 2. The mean percentage of correct discriminations between the base and nonbase triplet displays containing hierarchical structures for the LOCAL & HS, GLOBAL & HS, BOTH & HS, and BOTH & NO HS conditions. The error bars represent the standard error of the mean

However, a one-way ANOVA revealed a significant difference between the four conditions, $F(3, 76) = 13.01, p < .001$.

When the participants could utilize the hierarchical structure during the test phase, they could also extract the temporal statistical regularities. The mean percentage of correct trials was 74.38 % in the LOCAL & HS condition, $t(19) = 7.01, p < .001$, 73.13 % in the GLOBAL & HS condition, $t(19) = 6.65, p < .001$, and 84.38 % in the BOTH & HS conditions, $t(19) = 10.62, p < .001$, all of which were significantly higher than chance. In addition, the degree of VSL differed between the three conditions. Specifically, the test performance in the BOTH & HS condition, where the temporal regularities could be extracted at both levels, was significantly higher than that in the LOCAL & HS, $t(76) = 2.05, p = .044$, and GLOBAL & HS, $t(76) = 2.31, p = .024$, conditions. Moreover, in both conditions the temporal statistics could only be extracted at a single level. However, consistent with the results of Experiments 1A and 1C, the performance was not significantly different between the LOCAL & HS and GLOBAL & HS conditions, $t(76) = .26, p = .799$. As the hierarchical structure provided equally useful information in the three conditions, the difference between the three conditions likely arose from the amount of temporal regularities that the participants could extract at different hierarchical levels. In other words, this difference rules out the alternative possibility that the participants' performance above chance level can only be attributed to learning the hierarchical relationship between the local and global shapes and not to statistical learning.

On the other hand, no temporal VSL was observed in the BOTH & NO HS condition where the hierarchical structure was not informative. In the BOTH & NO HS condition, the hierarchical co-occurrence of the two shapes used in the familiarization phase was maintained in both the base and nonbase triplets. Therefore, in order to judge the familiarity correctly in the test phase, the participants had to ignore the

hierarchical structure and focus on the temporal regularities only. The mean correct percentage in this condition was 54.53 %, which did not differ significantly from chance, $t(19) = 1.26, p = .222$. In addition, this performance was significantly lower than it was for the BOTH & HS condition, where the participants could rely on the hierarchical structure as well as on the temporal regularities at both levels, $t(76) = 6.12, p < .001$. These results suggest that the participants utilized the hierarchical structure to judge familiarity. Therefore, if participants had to ignore the hierarchical structure in the BOTH & NO HS condition, the hierarchical structure information interfered with the extraction of the temporal regularities. In line with this assumption, the mean performance of the participants in the BOTH & NO HS condition was significantly lower than that in Experiment 1, $t(38) = 2.07, p = .045$. In Experiment 1, the hierarchical structure could not interfere with the participants' familiarity judgments, as we only presented a sequence of single shapes in the test phase. One might argue that although the hierarchical structure in our natural environment is stable as in the BOTH & NO HS condition, observers can still reliably extract the temporal regularities. We suggest that the difference in performance between the experimental condition (i.e., breakdown in the BOTH & NO HS condition) and the real-life environments (i.e., reliable performance) comes from the difference in the degree of learning. We assume that the temporal regularities can be more firmly established in our mental representation from real-life experience than from the experimental situation. For example, observers learn the temporal regularities long term and thereby with a high frequency, whereas the temporal regularities in the experiment appear for a short term (i.e., 12 min) and with a relatively low frequency. Therefore, we expect the degree of learning the temporal regularities to be much higher in real-life situations than in the experimental condition. Consequently, the temporal VSL acquired from a real-life experience could be less vulnerable to processing another type of information, such as the hierarchical structure.

Unlike in Experiment 1, the main effect of the presentation order was significant in that the participants' familiarity judgments were more correct when the base triplets were presented first (73.91 %) rather than second (69.30 %), as assessed with a paired-samples t test, $t(79) = 2.43, p = .017$. However, this tendency to choose the first order more frequently than the second one happened only in the LOCAL & HS condition, $t(19) = 2.81, p = .011$, while the performances in the other three conditions were not influenced by presentation order: GLOBAL & HS: $t(19) = .22, p = .832$; BOTH & HS: $t(19) = .30, p = .772$; BOTH & NO HS: $t(19) = .66, p = .517$. We speculate this was by accident and that the participants in this condition showed such tendency biased toward the first order. If a participant chose one particular order for the two-thirds or more trials out of the entire test (i.e., 21 trials or more out of the 32 trials in total), we defined this participant as the one

who made biased responses toward the specific order. Only three out of 20 participants in the LOCAL & HS condition showed this tendency toward the first order, whereas no participant in the other three conditions showed such a tendency. In addition, learning did not occur during the test, as the performance did not significantly differ between the first (72.50 %) and second (70.70 %) halves of the test, $t(79) = .98, p = .332$.

The awareness of the temporal regularities did not have any influence on temporal statistical learning, although this awareness enhanced the learning in Experiment 1. Of the 60 participants in the three conditions where temporal VSL occurred (LOCAL & HS, GLOBAL & HS, and BOTH & HS), the 36 participants (test performance: 84.72 %) who reported having used some kind of knowledge, either of the temporal regularities or the hierarchical structure, performed significantly better than the other 24 participants, who had responded by guessing (test performance: 66.15 %), $t(58) = 5.44, p < .001$. However, among these 36 participants, the 15 participants who had reported being aware of the temporal regularities did not perform significantly better than the other 21, who had reported using the hierarchical structure only without noticing the temporal regularities, $t(57) = .19, p = .851$. Therefore, their explicit awareness of the temporal regularities did not help the participants to use them. In addition, of the 20 participants in the BOTH & NO HS condition performing at chance, the 12 (test performance: 57.56 %) who reported having relied on some kind of knowledge did not perform significantly better than the other 8 participants who had responded by guessing (test performance: 50.00 %), $t(18) = 1.03, p = .316$. In this condition, only two participants reported being aware of the temporal regularities, and their test performance was 50.00 %. To explain why the explicit knowledge of the temporal regularities only helped the participants in Experiment 1, we think that the participants during the test phase might have been better focused on the temporal regularities in Experiment 1 than in Experiment 2. In other words, when the temporal regularities at both levels and the hierarchical structure were presented at the same time in Experiment 2, it was more difficult to utilize the temporal regularities explicitly.

To analyze the effect of the hierarchical structure on VSL, we compared the results of Experiments 1 (collapsed across Experiments 1A, 1B, and 1C) and 2. The test performance varied significantly between the different test conditions across the two experiments, $F(6, 133) = 6.98, p < .001$. This difference likely resulted from the test conditions, as the familiarization phase was identical in the two experiments. Therefore, the participants in the five conditions (i.e., Experiment 1 and LOCAL & HS, GLOBAL & HS, BOTH & HS and BOTH & NO HS conditions of Experiment 2) should have shown a similar amount of VSL and hierarchically structured representation right after the familiarization.

However, the test conditions differed in their familiarity source (i.e., the statistical regularities or hierarchical structure) and the amount of information that the participants could rely. Therefore, the different degrees of familiarity judgment suggest that the participants used the source of familiarity flexibly. In Experiment 1, where the statistical regularities were the only source of familiarity in the test, only temporal statistical learning was utilized for familiarity judgments. On the other hand, in Experiment 2, when both statistical regularities and a hierarchical structure were presented (i.e., LOCAL & HS, GLOBAL & HS, and BOTH & HS conditions), both mechanisms were utilized. In these three conditions, the post hoc contrast analysis showed that participants performed significantly better than in Experiment 1 when only VSL operated, $t(133) = 4.66, p < .001$.

It seems that participants utilized the familiarity sources that existed in the test display additively. When VSL was only helpful at a single level, the participants in Experiment 2 (i.e., in the LOCAL & HS and GLOBAL & HS conditions) possessed more information to judge the familiarity due to the hierarchical structure present in the test display than did the participants in Experiment 1. The participants' performance tended to be better in Experiment 2 than in Experiment 1 (local-level performance in Experiment 1 (60.42 %) vs. LOCAL & HS condition (74.38 %), $t(78) = 2.50, p = .015$; global-level test performance in Experiment 1 (63.54 %) vs. GLOBAL & HS condition in Experiment 2 (73.13 %), $t(78) = 1.87, p = .066$, as assessed by post hoc contrast analyses. Furthermore, the participants used the statistical regularities at each level additively. When the hierarchical structure was equally useful in Experiment 2, the familiarity judgments were enhanced more when temporal regularities could be relied on at both levels (i.e., in the BOTH & HS condition) than when they were only provided at a single level (i.e., in the LOCAL & HS and GLOBAL & HS conditions), $t(133) = 2.24, p = .026$, as shown by the post hoc contrast analysis.

General discussion

This study showed that the utilization of statistical regularities and hierarchical structure could work together in different spatiotemporal domains. In the two experiments, statistical regularities existed in the temporal domain while the stimuli were hierarchically structured in the spatial domain. In Experiment 1, we found that the participants extracted the temporal regularities that occurred at different hierarchical levels simultaneously, and the extent of learning was similar at the local and global levels. In addition, when extracting these regularities, the participants used the spatial hierarchy of the local and global shapes. In Experiment 2, the ability to extract the temporal regularities was influenced by the informativeness of the hierarchical structure. Specifically, the

hierarchical structure prevented the participants from extracting the temporal regularities when the hierarchical information was meant to be ignored. However, the participants extracted the temporal regularities when they could utilize the hierarchical structure.

Our findings seem to be inconsistent with Fiser and Aslin's (2005) study showing that the participants could not learn the statistical regularities if these regularities were embedded in a higher order hierarchical structure. To explain this discrepancy, we need to discuss two differences in the experimental paradigm between our study and the previous one. First, in both studies, shape features were hierarchically structured, but in different ways. In Fiser and Aslin, the smaller pair features were embedded in the larger quadruple structures; therefore, the features at the local pairs were necessarily tied to the larger quadruples at the global level. The authors argued that this embeddedness could constrain the statistical learning of the local features, which were simply a part of a larger structure. However, for the hierarchical structure used in this study, there was no such characteristic of embeddedness. Because the global structure was only defined by the location of the local shapes, and the two shapes at the different levels had different identities, it is not likely that the global structure constrained the statistical learning of the local shapes. Nevertheless, our results are still consistent with those of Fiser and Aslin, in that both studies showed that the hierarchical structure could influence the ability to extract statistical regularities. Depending on whether the hierarchical structure provided useful information, the VSL was possible (i.e., the LOCAL & HS, GLOBAL & HS, and BOTH & HS conditions) or not (i.e., the BOTH & NO HS condition), suggesting that the hierarchical structure could influence the extent of learning statistical regularities.

Second, in Fiser and Aslin's (2005) study, the statistical regularities and the hierarchical structure always appeared in the same spatial domain, whereas we presented the two kinds of information in different spatiotemporal domains. When VSL and the utilization of the hierarchical structure occur in separate domains, as in our study, there should be less competition between the two mechanisms. For example, the two mechanisms may ask for different sources of attention (i.e., temporal and spatial; Chun, Golomb, & Turk-Browne, 2011; Correa & Nobre, 2008). Furthermore, it has been reported that sequential learning of temporal and spatial structures could occur independently (Bengtsson et al., 2004; Karabanov & Ullén, 2008; Kornysheva et al., 2013; Ullén & Bengtsson, 2003). Therefore, this study was able to demonstrate cooperation between VSL and the utilization of the hierarchical structure, as each mechanism relied on different resources (i.e., temporal and spatial attention).

Despite these different attentional sources, in our study, attention to the hierarchical structure might precede attention to the temporal regularities. This is because hierarchical structures in our experiments included higher joint probabilities

than temporal regularities. Previous studies have shown that attention could be prioritized toward information with higher joint probabilities (Zhao et al., 2013). During the familiarization phase, the joint probability was higher for the hierarchical co-occurrence (i.e., .5) than for the temporal regularities (i.e., .17). Moreover, the hierarchical structure encoding required global attention to the two levels, whereas the computing of statistical regularities required local attention to the relations between the three shapes at the same hierarchical level. Global attention to the hierarchical structure can precede local attention to the temporal regularities (Love, Rouders, & Wisniewski, 1999; Navon, 1977) and can occur efficiently (Chong & Treisman 2003, 2005). Indeed, in Experiment 2, more participants reported being only aware of the hierarchical structure (28 participants) than only aware of the temporal regularities (three participants). Therefore, the lower degree of attention to the temporal regularities could explain why the participants in the BOTH & NO HS condition could not extract the temporal regularities. In order to understand the effects of VSL and hierarchical representation on familiarity judgments, future research will need to equate the level of attention to hierarchical structure and statistical regularities.

Hierarchical structure may present advantages for the handling of complex inputs such as different regularities happening at the same time. One possible advantage of the hierarchical structure may be its ability to reduce the burden of attention to multiple regularities. When different shapes in a scene can be organized into a local and global level, attention can be more efficiently deployed in order to extract the multiple regularities occurring simultaneously. In our study, it is not likely that the participants selectively attended to any particular shape in each scene, given that there was no task or instruction to this effect during the familiarization phase. Nevertheless, the participants could extract different regularities in each of the two levels, and the extent of learning was not different between the two levels. Indeed, we could not find any participant who showed strong preference for a certain level, when we further asked some of the participants who reported relying on explicit knowledge about the characteristics of this knowledge. On the other hand, we also admit that it is difficult to draw a clear conclusion from our data on whether the statistical learning at the different levels could vary according to the preference for a certain level, given that we did not measure the level preference, as this was not the aim of our study. We suggest that this issue can be better delivered by future studies, with a more precise method to measure the level of preference and its effect on statistical learning.

Another possible advantage of hierarchical structures is that they may be used as predictive information, where information of one level can help observers predict that of another level. In our experiment, based on the prediction that one particular local shape typically occurred with another particular global shape, the association between the local and global

shapes allowed participants to assess familiarity correctly. Indeed, when the hierarchical structure was removed in Experiment 1, the participants tended to perform worse than did those in the LOCAL & HS and GLOBAL & HS conditions of Experiment 2.

Our results revealed that VSL could simultaneously occur at different levels. Observers can flexibly extract statistical regularities not only at different hierarchical levels but also through different sensory modalities (Glicksohn & Cohen, 2013; Seitz, Kim, van Wassenhove, & Shams, 2007; Shams & Seitz, 2008). Seitz and colleagues (2007) showed that statistical learning could occur independent of sensory modality (i.e., vision and audition). Future studies should examine whether multiple levels of statistical regularities can also be extracted through other sensory modalities. For instance, when listening to choir music, the tones in the four different pitches (soprano, alto, tenor, and base) may also present statistical regularities at each pitch level. It would be interesting to investigate whether listeners are able to extract these regularities appearing simultaneously for different pitches.

In conclusion, the visual system can utilize both VSL and hierarchical encoding to handle complex scenes when they occur in different spatiotemporal domains. While observers can represent the location of their cars at the local (e.g., next to a friend's car) and global level (e.g., the left side of the parking lot), they also learn the statistical regularities between cars and buildings. We suggest that the cooperation of the two mechanisms may explain how the visual system is able to handle complex information in natural environments.

Acknowledgments This work was supported by the National Research Foundation (NRF) of Korea grant funded by the Korean government (MEST, No. 2011-0025005). We thank Min-Suk Kang for helpful comments and discussion on this manuscript.

References

- Bengtsson, S. L., Ehrsson, H. H., Forssberg, H., & Ullén, F. (2004). Dissociating brain regions controlling the temporal and ordinal structure of learned movement sequences. *The European Journal of Neuroscience*, *19*, 2591–2602.
- Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1425–1431.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory ensemble statistics bias memory for individual items. *Psychological Science*, *22*, 384–392.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*, 14325–14329.
- 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.

- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 1–34.
- Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*, 678–685.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*, 393–404.
- Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*, 891–900.
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*, 170–178.
- Chun, M. M. (2003). Scene perception and memory. *Psychology of Learning and Motivation*, *42*, 79–108.
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*, 360–365.
- Correa, Á., & Nobre, A. C. (2008). Spatial and temporal acuity of visual perception can be enhanced selectively by attentional set. *Experimental Brain Research*, *189*, 339–344.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*, 85–87.
- Fiser, J. (2009). Perceptual learning and representational learning in humans and animals. *Learning & Behavior*, *37*, 141–153.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499–504.
- Fiser, J., & Aslin, R. N. (2002a). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*, 15822–15826.
- Fiser, J., & Aslin, R. N. (2002b). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 458–467.
- Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, *134*, 521–537.
- Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychonomic Bulletin & Review*, *14*, 173–178.
- Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, *94*, 427–438.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review*, *20*, 1161–1169.
- Humphreys, G. W., & Riddoch, M. J. (2001). The neuropsychology of visual object and space perception. In E. B. Goldstein (Ed.), *Blackwell handbook of sensation and perception* (pp. 204–236). Malden: Blackwell.
- Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception*, *43*, 663–676.
- Karabanov, A., & Ullén, F. (2008). Implicit and explicit learning of temporal sequences studied with the process dissociation procedure. *Journal of Neurophysiology*, *100*, 733–739.
- Kimchi, R. (2014). The perception of hierarchical structure. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization*. Oxford: Oxford University Press.
- 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.
- Kornysheva, K., Sierk, A., & Diedrichsen, J. (2013). Interaction of temporal and ordinal representations in movement sequences. *Journal of Neurophysiology*, *109*, 1416–1424.
- Love, B. C., Roudier, J. N., & Wisniewski, E. J. (1999). A structural account of global and local processing. *Cognitive Psychology*, *38*, 291–316.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383.
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*, 520–527.
- Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*, 2745–2750.
- Otsuka, S., Nishiyama, M., Nakahara, F., & Kawaguchi, J. (2013). Visual statistical learning based on the perceptual and semantic information of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 196–207.
- Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519–526.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.
- Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, *36*, 1445–1453.
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*, 411–417.
- Turk-Browne, N. B. (2012). Statistical learning and its consequences. In M. D. Dodd & J. H. Flowers (Eds.), *The influence of attention, learning, and motivation on visual search* (pp. 117–146). New York: Springer.
- 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.
- Ullén, F., & Bengtsson, S. L. (2003). Independent processing of the temporal and ordinal structure of movement sequences. *Journal of Neurophysiology*, *90*, 3725–3735.
- 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.
- Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*, 667–677.