



Visual statistical learning at basic and subordinate category levels in real-world images

Jihyang Jun^{1,2} · Sang Chul Chong³

Published online: 16 July 2018
© The Psychonomic Society, Inc. 2018

Abstract

Visual statistical learning (VSL) has been proposed as a powerful mechanism underlying the striking ability of human observers to handle complex visual environments. Previous studies have shown that VSL can occur when statistical information is embedded at multiple levels of abstraction, such as at semantically different category levels. In the present study, we further examined whether statistical regularities at a basic category level (e.g., a regular sequence of a bird, then a car, and then a dog) could influence the ability to extract statistical regularities at the subordinate level (e.g., a regular sequence of a parrot, then a sports car, and then an Eskimo dog). In the familiarization phase, participants were exposed to a stream of real-world images whose semantic categories had temporal regularities. Importantly, the temporal regularities existed at *both* the basic and subordinate levels, or the regularities existed at *only* the subordinate level, depending on the experimental condition. After completing the familiarization, participants performed a surprise two-alternative forced choice (2AFC) task for a familiarity judgment between two triplets in which the temporal regularities were either preserved or not preserved. Our results showed that the existence of statistical regularities at the basic level did not influence VSL at the subordinate level. The subsequent experiments showed these results consistently even when the basic-level categories had to be explicitly recognized and when the stimuli were not easily categorized at their subordinate level. Our results suggest that VSL is constrained to learn a particular level of patterns when patterns are presented across multiple levels.

Keywords Visual statistical learning · Basic and subordinate levels · Object categorization

In visual environments, we as human observers can recognize objects and scenes rapidly without much effort or scrutiny (Humphreys & Riddoch, 2001). This is one of the most striking aspects of visual perception, considering how complex and dynamic visual environments are (Biederman, 1987; DiCarlo & Cox, 2007). That is, visual environments consist of dynamic information, such as moving objects and changing light sources. Most objects of even the same category differ from one another in their physical properties, such as variation in color, size, or internal structure. Even visual inputs are

ambiguous, because observers mostly are given partial information, due to objects overlapping.

Our striking ability to process highly complex visual inputs is possible in various types of visual learning—one of which is visual statistical learning (VSL; Chun & Jiang, 1998; Fiser & Aslin, 2002; Orbán, Fiser, Aslin, & Lengyel, 2008). VSL refers to the ability of human observers to extract and learn statistical regularities, such as spatial configurations or temporal sequences, that repeatedly happen in our environments (Turk-Browne, 2012). For instance, in our daily commute, we may experience a repeated sequence of landmarks or encounter a set of objects that tend to co-occur. Every morning we may first drop by a café where we can encounter a familiar part-timer who mostly works at the counter, then we may go to a parking lot where we may park in a certain spot more frequently, and then we may enter a building where our office is located on a particular floor. The visual experience in this commute consists of various types of statistical regularities, such as temporal regularities (e.g., the regular sequence of the café, then the parking lot, and then the office building) and spatial regularities (e.g., the spatial association in which the

✉ Sang Chul Chong
scchong@yonsei.ac.kr

¹ Department of Psychology, University of Minnesota, Minneapolis, MN, USA

² Center for Cognitive Science, Yonsei University, Seoul, South Korea

³ Graduate Program in Cognitive Science, and Department of Psychology, Yonsei University, 50 Yonsei-ro, Sinchon-dong, Seodaemun-gu, Seoul 03722, South Korea

familiar part-timer appears mostly on a particular side of the café). So far, VSL has been proposed as a powerful learning mechanism to deal with the massive amounts of information that human observers receive from the world (Orban, Fiser, Aslin, & Lengyel, 2008).

One of challenges that human observers confront in this type of learning is that statistical regularities are embedded in a highly complex manner in our environments. Statistical information could exist at multiple levels of abstraction, such as various taxonomic classes at different hierarchical levels. Take the daily commute, for example—landmarks or objects could be labeled differently at the basic (e.g., coffee at the café), superordinate (e.g., drink at the shop), and subordinate (e.g., café latte from Starbucks) levels. Depending on which level is selected, different types of semantic categories could be used to define statistical regularities. At the basic level, statistical regularities could be defined as a sequence of a coffee from the café, then a car in the parking lot, and then an office in the building. At the subordinate level, the same regularities could be characterized as a sequence of café latte, then a Hyundai sedan car, and then my office.

Previous studies on VSL have examined this ability to extract statistical regularities that could be defined as multiple levels of semantic categories and tested the transferability of VSL across different category levels. Using real-world images with semantic information, Brady and Oliva (2008) found that participants extracted statistical regularities at the basic category level by abstracting the basic-level information from a sequence of exemplars that were always different at the subordinate level. In the familiarization phase of this study, participants were exposed to a stream of real-world scene images having temporal regularities among their basic-level categories (e.g., a regular sequence of a bathroom, then a bedroom, and then a kitchen), but every image per category was always different (e.g., every bathroom image was a novel exemplar). After this familiarization had been completed, participants performed a 2AFC familiarity judgment task between two triplets that had either old or new temporal regularities. Their results showed that the mean percentage of correct trials—that is, trials in which participants selected the old-sequence triplet as a more familiar sequence—was significantly above chance level. These results suggest that participants could abstract the basic-level category information and learn the temporal regularities at this category level.

Similarly, Otsuka, Nishiyama, Nakahara, and Kawaguchi (2013) showed that VSL could be successfully transferred from the subordinate to basic category levels. In their study (Experiment 5, in particular), participants were exposed to a stream of line drawings, which consisted of four triplets to which three exemplar images were assigned, and these images always appeared in the same order. Contrary to Brady and Oliva (2008), the exemplars were repeatedly presented in the familiarization. In the test phase, instead of the old exemplars

(e.g., a pug as an exemplar of dog during the familiarization), new exemplars (e.g., a shepherd in the test) were used for each image comprising both old and new triplets. The results again showed that the percentage of trials in which participants correctly reported the old triplets to be more familiar was higher than chance level, suggesting that participants learned the statistical regularities from the image sequence, and transferred the learning of the subordinate-level categories to the basic-level concepts. However, such flexibility of VSL between different category levels was not observed at the superordinate level. In Otsuka, Nishiyama, and Kawaguchi (2014), participants were exposed to the repeated sequence of images having the temporal regularities within triplets and performed the 2AFC familiarity judgment task. Similar to their previous study (Otsuka et al., 2013), new exemplars were used for both old and new triplets. However, in this study the new exemplars shared the same superordinate-level categories as in its preceding familiarization (e.g., when the superordinate category was a fruit, one image used in the familiarization was an apple, and the other used in the test was an orange). The results showed that the test performance was at chance level, suggesting that participants did not learn the temporal regularities at the superordinate level. Thus, these results suggest that when statistical regularities are structured at the multiple levels of semantic categories, participants can abstract the statistical regularities of the subordinate level to the basic level, but they cannot abstract the basic level to the superordinate level.

In the present study, we extended the study of VSL at semantically different levels by testing whether the statistical regularities at the basic and subordinate levels interact with each other. As far as we know, only one study (Emberson & Rubinstein, 2016) has investigated this question. In their study, when observers learned statistical information about everyday objects, their learning was biased toward the subordinate level and the basic-level concepts did not influence the degree of VSL. Emberson and Rubinstein employed a typicality effect in order to manipulate whether the basic-level concepts, rather than the subordinate exemplars, were initially represented and learned. In the familiarization, for half of the participants, the basic-level categories were easily extractable because the exemplars were typical. For example, participants viewed a sequence of typical exemplars, in which two exemplars always appeared in the same order (e.g., a rose, as a typical case of flower, was always followed by a goldfish, which is a typical type of fish). For the other half of participants, the basic-level categories were not easily accessible because the exemplars were atypical cases. For instance, whereas the former group was exposed to the pair association between the rose and the goldfish, the latter group was exposed to the association between a bird of paradise flower and a blowfish. After the familiarization, participants performed the 2AFC familiarity judgment task between two pairs: One was the same pair as in the familiarization, and the other was a

novel pair with new statistical regularities at the subordinate level (Experiment 2) or both at the basic and subordinate levels (Experiment 3). In both Experiments 2 and 3, the degree of VSL was not influenced by whether the basic-level categories were easily accessible. These results suggest that the basic category information (i.e., typicality) does not facilitate learning of statistical regularities at the subordinate level (i.e., VSL of exemplars). Considering that both their Experiments 2 and 3 were very well powered (.74 in Experiment 2 and .99 in Experiment 3), the lack of a typicality effect in both experiments could not be the failure to reject a false null hypothesis. These findings are surprising, considering that the basic-level category has been understood as receiving initial and predominant information processing (Murphy, 2002) and people could flexibly represent statistical regularities at this category level (Brady & Oliva, 2008; Otsuka et al., 2013). Moreover, people could flexibly use them at different hierarchical levels of novel objects (Jun & Chong, 2016). Participants in the VSL paradigm did not exploit the basic-level concepts, despite its robust representation, as an additional source of information.

However, there are several possibilities that make it difficult to draw a clear conclusion, and more empirical data will be needed to test whether these results are not simply null findings. Although the study by Emberson and Rubinstein (2016) is a very well-powered study, the following possibilities, which mainly originated from the design of their study, make it difficult to rule out an opposite conclusion that participants might still utilize basic-level information when they acquired the subordinate-level regularities. The first possibility is simply between-subjects noise, such as significant variance between different groups that could negate any results. To test this possibility, our study adopted a within-subjects design in Experiment 1, whereby we attempted to control for any possible confounding effect of individual differences and increased the power.

Second, the memory load in the learning phase might be too low to detect the effect of the basic-level category representation. In Emberson and Rubinstein (2016), a very small number of basic-level concepts (four concepts: dog, flower, fish, and bird) were tested as a pair association (e.g., the association between birds and dogs). Each participant was exposed to a total of four pairs, each of which was repeated 112 times. In fact, more than half of participants (42 out of 78 participants across four experiments) reported explicit knowledge of the repeated patterns, suggesting that participants might have extracted and remembered the patterns so easily that the effect of the basic-level concept representation might not have an enough room to influence. To deal with this possibility, we increased the memory load by employing a larger number of basic concepts (six categories rather than four) with a larger size of tractable patterns (triplets rather than pairs) and each triplet was repeated for a fewer number of times (84

repetitions in Experiments 1A, 1B, and 2A, and 56 repetitions in Experiment 2B, rather than 112).

Third, the paired associations rather than triplets presented in Emberson and Rubinstein (2016) make it difficult to compare their results with those from previous studies of VSL, which have mostly used triplets (Brady & Oliva, 2008; Fiser & Aslin, 2002; Turk-Browne & Scholl, 2009). Because of this, we used triplets instead of pairs. Participants performed a one-back task, which has also been used as a cover task with which to keep participants alert in previous studies that have used real-world image stimuli (Brady & Oliva, 2008; Turk-Browne, Jungé, & Scholl, 2005). In this way, our study can be directly compared to previous findings of VSL.

Finally, a very small number of basic-level category sets (four concepts: dog, flower, fish, and bird) were repeatedly used across the four experiments, which makes it difficult to generalize their findings. To overcome this limitation, we increased the variability of the categories by employing 12 different categories that included different types of objects (six concepts in Exps. 1A and 1B: dog, flower, fish, bird, car, and building) and scenes (six concepts in Exps. 2A and 2B: bathroom, bedroom, kitchen, closet, staircase, and living room). In this way, we attempted to rule out the possibility that the results were simply confounded by the limited sets of basic-level concepts used in their experiments.

To provide an overview of the present study, in Experiment 1A our participants were exposed to a stream of real-world images, which consisted of six triplets that featured temporal regularities in the presentation order of three images. Half of the triplets had temporal regularities at both the basic and subordinate category levels, whereas the other three did not have such regularities at the basic level. Instead, in the latter triplets the temporal regularities were presented only at the subordinate level. Participants were not informed about such order structure but conducted a cover task (one-back repetition detection task). After this familiarization, participants performed a 2AFC task in which participants had to choose (or guess, if necessary) which one they felt to be more familiar, and two triplets including the same or different temporal regularities were presented sequentially for every trial. To preview our results, participants learned temporal regularities from this image sequence, but the additional existence of statistical regularities at the basic level did not influence the extent of VSL. The same pattern of results was consistently found in our subsequent experiments, even after controlling for potential contamination factors. In Experiment 1B we conducted the same task, except that participants were asked to explicitly focus on extracting the basic-level categories and to remember those categories. In Experiment 2A, we adopted a between-subjects design in order to avoid any possible memory interference effect, and we also used a different set of stimuli (indoor scene images) for the purpose of increasing the variability of the categories and reducing the saliency of

the subordinate-level category information. In Experiment 2B, we reduced the memory load and the length of the familiarization period, to avoid any possible effects due to the loss of motivation. All of these experiments consistently showed that when participants learned temporal regularities from a sequence of real-world images, the presence of extra regularities at the basic level neither improved nor reduced the extent of learning at the subordinate level.

Experiment 1A

Method

Participants

Twenty naïve students from Yonsei University participated in Experiment 1A in exchange for course credit or monetary compensation (5,000 won). The sample size was predetermined to be comparable to those of previous studies that had tested statistical-learning ability using real-world images (Brady & Oliva, 2008; Otsuka et al., 2013, 2014). All participants had normal or corrected-to-normal visual acuity. In this and the following experiments, the study protocol was approved by the Yonsei University Institutional Review Board, and participants provided written informed consent forms.

Apparatus and stimuli

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

We used a total of 18 grayscale, luminance-controlled square images in six different basic categories (bird, car, dog, flower, building, and fish). Each of the six basic categories contained three images, which were different at the subordinate level. For example, the bird category consisted of images of a parrot, an owl, and a magpie. The width and height of the images were 5° and 5° , respectively. The background of the image was gray. All the images used in this and the following experiments are presented in the Appendix.

Design and procedure

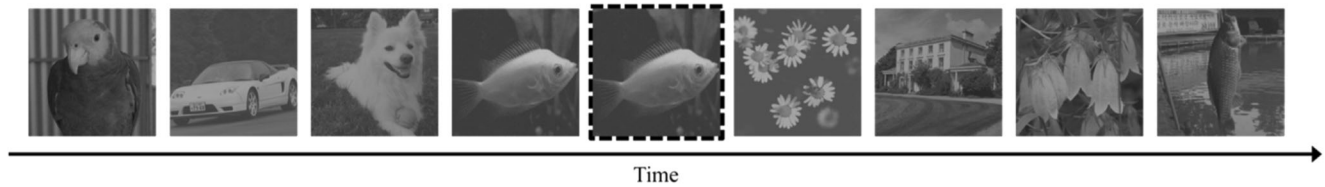
In Experiment 1A we comprised two phases: a familiarization phase and a test phase. Participants completed the familiarization phase and then performed the test phase.

Familiarization phase In the familiarization phase (see Fig. 1a), we presented a sequence of 1,680 images, in which the 18 images across the six different basic categories repeatedly appeared one after another. Unbeknownst to the participants, the 18 images were randomly distributed to one of the six triplets, and each triplet featured temporal regularity, in that the three images were always presented in the same order.

The sequence in the familiarization phase had two within-subjects conditions (Table 1). Specifically, depending on the conditions, the triplets had statistical regularities at both the basic and subordinate levels (the *basic- and subordinate-level* familiarization condition), or the statistical regularities existed at only the subordinate level (the *subordinate-level* familiarization condition). The former condition included three triplets that consisted of nine different images across three basic-level categories (e.g., the bird, car, and dog categories). In this condition, the three basic-level categories always appeared in the same order across the three triplets (e.g., a constant sequence of the bird, then the car, and then the dog), so that the temporal regularities existed at the basic level. At the same time, three different images in each triplet always appeared in the same order (e.g., a constant sequence of a parrot, then a sports car, and then an Eskimo dog in one triplet set), so that each triplet also presented the temporal regularities at the subordinate level. The latter condition included the other three triplets, which also consisted of nine different images across three basic-level categories (e.g., the fish, flower, and building categories). In this condition, the three basic-level categories did not appear in the same order across the three triplets (e.g., a sequence of the fish, then the flower, and then the building in one triplet set, but another sequence of the flower, then the fish, and then the building comprising a different triplet set), so that temporal regularities did not exist at the basic level across the three triplets. By contrast, in each triplet, three different images always appeared in the same order (e.g., a repeated sequence of a goldfish, then a daisy, and then a three-floor building in one triplet set), so that every triplet presented temporal regularities at the subordinate level.

Given this design, we constructed six triplets with two constraints. First, images in the same category (e.g., a parrot and an owl, both of which belong to the bird category) were not assigned to the same triplet. In this way, the six basic-level categories were evenly distributed across the six triplets. Second, nine different images in three categories (e.g., the bird, car, and dog categories) comprised three triplets, and the remaining nine images in the other three categories comprised the other three triplets (e.g., the nine different images from the fish, flower, and building categories). In this way, the six basic-level categories were evenly distributed to two within-subjects conditions (i.e., the *basic- and subordinate-level* familiarization condition and the *subordinate-level* familiarization condition). Each triplet was repeated 84 times, and the same triplet set was never repeated successively, so

(a) Familiarization phase



(b) Test phase

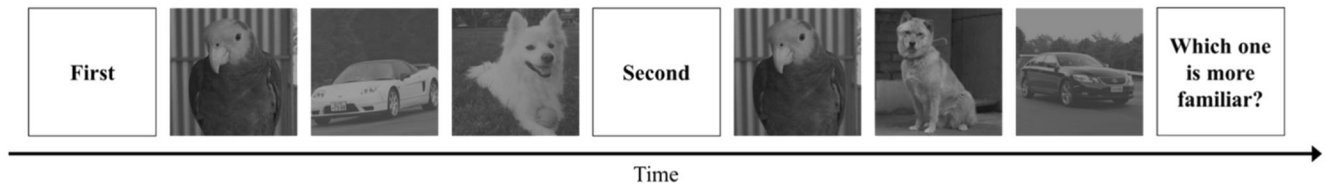


Fig. 1 An example trial from Experiments 1A and 1B. (a) In the familiarization phase, the participants in both experiments were asked to detect a repetition of the same image (e.g., the goldfish shown here in the dotted box), but the participants in Experiment 1B were further asked to categorize each image and remember a total of six categories they could later

find. (b) After completing the familiarization phase, participants performed the surprise two-alternative forced choice task of making familiarity judgments. In this test, participants were asked to choose the more familiar sequence after the base and nonbase triplets had been presented sequentially

that a sequence of 1,512 images was created. In addition, we inserted 168 repeat images into the sequence, so that either the first or the third image in a triplet was occasionally repeated. Repeating only the first or the third image within triplets allowed us to maintain the triplet structure and to prevent the repeat images from providing a cue to segment the triplets from one another (Brady & Oliva, 2008; Otsuka et al., 2013). There was no segmentation cue between the triplets. The duration of each image was 300 ms, and the duration of the blank interval was 700 ms. It took 29 min to complete the familiarization phase.

While this sequence of images was presented, the task was to detect a repetition of the same image. This cover task was intended to prevent participants from being explicitly aware of the temporal statistics in the stream and to help participants to view the stream more actively (Brady & Oliva, 2008; Otsuka et al., 2013; Turk-Browne, Jungé, & Scholl, 2005).

Test phase In Experiment 1A, we tested whether participants could extract temporal statistics from each triplet, and whether the temporal statistics at the basic level could influence the

extent of statistical learning. To investigate this, we tested participants' ability to discriminate a base triplet from a nonbase triplet as a measure of statistical learning (see Fig. 1b). The temporal order was maintained for the base triplets, whereas the temporal order was changed for the nonbase triplets.

There were four within-subjects variables. First, we presented two kinds of test conditions, in which temporal VSL was tested either at both the basic and subordinate levels (the *basic- and subordinate-level test condition*) or at only the subordinate level (the *subordinate-level test condition*). The base triplets in the basic- and subordinate-level test condition had temporal regularities at both the basic and subordinate levels, because we used the triplets that had appeared in the basic- and subordinate-level familiarization condition for the base triplets. The base triplets in the subordinate-level test condition also had temporal regularities, but only at the subordinate level, because the base triplets were triplets that had appeared in the subordinate-level familiarization condition. In both test conditions, the nonbase triplets did not have temporal regularities at any category level. The position in the original

Table 1 Sample triplets for the familiarization in Experiments 1A and 1B

Basic- and Subordinate-Level Familiarization	Subordinate-Level Familiarization
Triplet 1: Bird ₁ → Car ₁ → Dog ₁	Triplet 4: Fish ₁ → Flower ₁ → Building ₁
Triplet 2: Bird ₂ → Car ₂ → Dog ₂	Triplet 5: Building ₂ → Flower ₂ → Fish ₂
Triplet 3: Bird ₃ → Car ₃ → Dog ₃	Triplet 6: Flower ₃ → Fish ₃ → Building ₃

In the basic- and subordinate-level familiarization condition (left column), the three triplets involved statistical regularity at the basic level (i.e., Bird → Car → Dog, in this example), as well as at the subordinate level (e.g., Bird₁ → Car₁ → Dog₁, in Triplet 1). In the subordinate-level familiarization condition (right column), the other three triplets involved statistical regularities that existed at the subordinate level only. Each category (e.g., bird) included three exemplars that were different at the subordinate level (e.g., parrot, owl, and magpie)

base triplets (e.g., Bird₁ in the first position) was not always preserved in these nonbase triplets, for the following reason: If we maintained the original position (e.g., Bird₁ → Car₂ → Dog₃ as a possible foil triplet), the temporal regularities would be violated only at the subordinate level in the test phase. However, in our study, it was critical to test the ability to exploit the regularities at the basic level, because the main purpose of the study was to test whether this ability could facilitate the degree of VSL at the subordinate level. Thus, when we created the nonbase triplets, we violated the temporal regularities not only among the subordinate-level exemplars, but also at the basic level, by changing the image positions within their original triplets. Second and third, in each test condition there were three kinds of base triplets and three kinds of nonbase triplets. The base triplets included three different images that always appeared in the same order as in the familiarization phase. The joint probability of the three images in each of the base triplet was .056, as during the familiarization. Three different images in each nonbase triplet appeared in a different order than during the familiarization. The joint probability of the three images in each of the nonbase triplets thus became 0. The base and nonbase triplets had the same kinds of basic-level categories (e.g., the bird, car, and dog categories in the basic- and subordinate-level test condition), and the only difference between the base and nonbase triplets was the temporal order at the tested level. That is, we constructed the nonbase triplets by choosing only one image from each of the three different base triplets in the same test condition. Fourth, there were two different orders of presentation, in which either the base triplet was presented first or the nonbase triplet was presented first. Therefore, the test phase included 36 randomized trials. The frequency of all 18 images was equalized between the base and nonbase triplets.

Participants performed a surprise 2AFC task in which they were asked to make familiarity judgments. For each trial, the base and nonbase triplets were sequentially presented, with the word “First” or “Second” appearing on a blank screen for 1 s in order to segment the two triplets. The durations of the images and of the blank interval between the images were identical to those during familiarization. After participants had viewed two sequences (i.e., the base and nonbase triplets), they were asked to judge which sequence was more familiar. Participants were asked to base their judgment on every trial on their experience of the previous 29-min movie, by pressing “1” for the first sequence and “2” for the second sequence.

We additionally checked explicit awareness of the temporal regularities among the participants. After performing the 2AFC familiarity judgment task, participants were further asked to complete a binary confidence judgment task (Bertels, Franco, & Destrebecqz, 2012). In this task, participants chose either of two statements: The first statement indicated that participants had based their judgments during the 2AFC task on some kind of explicit knowledge (i.e., “I chose

the answers based on some kind of knowledge that I learned during the familiarization.”), whereas the second statement indicated that the test had been performed by guessing (i.e., “I guessed the answers based on my intuition.”).¹ If participants chose the first statement, we additionally asked those participants to describe what kind of specific knowledge they had used, and further checked whether they had been aware of the temporal order during the familiarization.

Results and discussion

In the familiarization phase, participants successfully completed the repetition detection cover task with a mean detection accuracy of 88.18%. Participants committed false alarms on 0.57% of the probe-absent trials (2.88 times) during the entire familiarization.

In the 2AFC familiarity test phase, we measured the extent of statistical learning by calculating the mean percentage of correct trials, in which participants selected the base triplets as being more familiar than the nonbase triplets. These results are shown in Fig. 2. Overall, the mean percentage of correct responses in the two test conditions was 57.78%, which was significantly higher than chance level (50%), as revealed by a one-sample *t* test, $t(19) = 2.38$, $p = .028$, $d = 0.53$. That is, participants had learned the temporal regularities across the six triplet sets. Specifically, in the basic- and subordinate-level test condition, the mean correct percentage was 57.22%, which was marginally significant, $t(19) = 2.02$, $p = .058$, $d = 0.45$. In the subordinate-level test condition, the mean correct percentage was 58.33%, which was significantly higher than chance level, $t(19) = 2.15$, $p = .045$, $d = 0.48$. Importantly, the degree of learning was not significantly different between the two test conditions, as assessed by a paired-samples *t* test, $t(19) = -0.31$, $p = .761$, $d = -0.07$. To quantify the evidence favoring the null hypothesis, we tested a Bayesian paired *t* test using the JASP software (Love et al., 2015). Using the inverted Bayes factor (BF₀₁) statistic, we computed the ratio of the Bayesian probability that the observed data support the null hypothesis to the probability that they support the alternative hypothesis, meaning that a BF₀₁ value greater than 1 indicated that the null hypothesis was favored. The Bayes factor moderately favored the null model, with the inverse BF₀₁ = 4.12, suggesting that our data were 4.12 times more likely to occur under the null hypothesis than under the alternative hypothesis. This is moderate evidence against the possibility that the additional existence of statistical regularities at the basic category level could have influenced the degree of VSL.

¹ The two statements in the binary confidence judgment task were adopted from Bertels et al. (2012). Bertels et al. ran this confidence task after each trial in order to measure the degree of conscious knowledge about statistical structure in a systematic way. However, because the role of explicit knowledge in the degree of VSL was outside the scope of our study, we ran this binary confidence judgment task only at the end of an experiment.

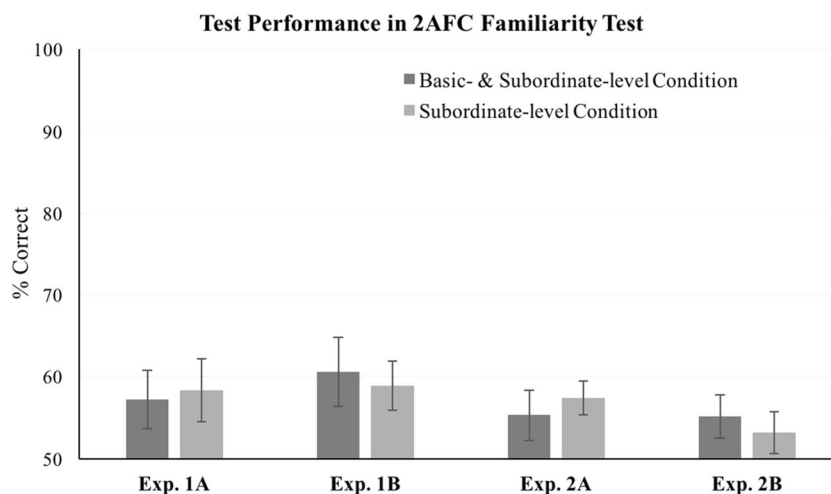


Fig. 2 Results of Experiments 1A, 1B, 2A, and 2B. The graph shows the mean percentages of correct discriminations between base and nonbase triplets in the 2AFC familiarity test. Each bar in dark or light gray

represents the basic- and subordinate-level condition or the subordinate-level condition, respectively. The error bars represent the standard errors of the means

The presentation order had no effect on participants' performance. The mean correct percentage on the half of trials in which the base triplet was presented first (test performance: 56.94%) did not significantly differ from the mean correct percentage on the other half of trials, in which the nonbase triplet was presented first (58.61%), as revealed by a paired-samples t test, $t(19) = -0.33$, $p = .743$, $d = 0.07$. In addition, statistical learning did not occur during the test phase, in that the mean correct percentages during the first half (56.56%) and the second half (60.56%) of trials were not significantly different, $t(19) = -0.89$, $p = .382$, $d = 0.20$.

In the binary confidence judgment, six out of the 20 participants reported having used some kind of knowledge (test performance: 70.37%), whereas the other 14 participants reported having relied on guesswork (test performance: 52.38%). Among these six participants, only three participants (test performance: 81.48%) specifically reported having been aware of the temporal regularities, whereas the other three participants (test performance: 59.26%) reported ineffective strategies (e.g., they responded to images that they thought had been more frequently repeated during the familiarization).

Experiment 1A showed that participants did extract temporal regularities across the six triplets. Importantly, our results showed that the degree of temporal VSL was not different in the test conditions in which temporal regularities existed at both the basic and subordinate levels (the basic- and subordinate-level test condition) or at only the subordinate level (the subordinate-level test condition). These results suggest that temporal regularities at the basic level did not influence the learning of temporal regularities at the subordinate level.

However, there is another possibility, that participants might not have categorized each image at the basic level, because such

categorization was not required in the repetition detection task during familiarization. To test this possibility, in Experiment 1B, in addition to detecting the repetition of the same image, participants were directly asked to categorize each image and were informed that they would recall a total of six categories after the image stream had ended. In this way, participants had to explicitly categorize each image at the basic level while they simultaneously detected repetition of the same image.

Experiment 1B

Method

Participants

Twenty-two new and naïve students from Yonsei University participated in Experiment 1B, in exchange for course credits or monetary compensation (5,000 won). All had normal or corrected-to-normal visual acuity. Two participants misunderstood the task instructions and did not perform the repetition detection task. The data from these participants were not analyzed. In this and the following experiments, the sample size of 20 was predetermined so as to be equal to that of Experiment 1A, because we were attempting a conceptual replication of Experiment 1A.

Apparatus and stimuli

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

Design and procedure

As in Experiment 1A, the familiarization phase was followed by a surprise test phase.

Familiarization phase The design and procedure of the familiarization phase were identical to those aspects of Experiment 1A, except that the participants were asked to categorize each image according to its basic-level concept and to remember the six concepts in total. Before the familiarization phase began, participants had been informed that there would be a total of six categories in the image sequence that they would see, and that they would be asked to verbally recall the six categories after the familiarization phase. That meant that we explicitly asked participants to categorize each image by a certain category name and to try to remember all six of the categories that they would observe. Therefore, in Experiment 1B the participants performed a dual task: They were asked to categorize each image and remember the categories, and simultaneously to detect a repetition of exactly the same image.

Test phase Participants were first asked to verbally recall the categories that they had observed during the familiarization phase. After this, they then performed the 2AFC familiarity judgment task, which had the same design and procedure as in Experiment 1A.

Results and discussion

The mean repetition detection accuracy in the familiarization phase was 85.15%, which was not statistically different from the mean accuracy in Experiment 1A, $t(37) = 0.75$, $p = .457$, $d = 0.24$. The false alarm rate was 0.53% (2.65 times) of the probe-absent trials during the entire familiarization phase, which did not statistically differ from the false alarm rate in Experiment 1A, $t(37) = 0.33$, $p = .741$, $d = 0.11$. Importantly, all participants recalled the six basic-level categories correctly. Thus, the participants in Experiment 1B did successfully categorize the images at the basic level, while at the same time they performed the repetition detection task comparably to the participants in Experiment 1A.

The results for the 2AFC familiarity test are shown in Fig. 2. The mean percentage of correct responses was 59.72%, which was significantly higher than chance level, $t(19) = 3.16$, $p = .005$, $d = 0.71$, and an independent-samples t test revealed that this test performance was not significantly different from that in Experiment 1A, $t(38) = -0.43$, $p = .667$, $d = 0.14$. Again, as in Experiment 1A, we found that participants had learned the temporal regularities across the six triplet sets. Specifically, in the basic- and subordinate-level test condition, the mean correct percentage was 60.56%, which was significantly higher than chance level, $t(19) = 2.49$, $p = .022$, $d = 0.56$, and this test performance was not significantly different

from that in Experiment 1A, $t(38) = -0.60$, $p = .55$, $d = 0.19$. In the subordinate-level test condition, the mean correct percentage was 58.89%, which was again significantly higher than chance, $t(19) = 2.94$, $p = .008$, $d = 0.66$, and this test performance was not significantly different from that in Experiment 1A, $t(38) = -0.11$, $p = .91$, $d = 0.04$. Again, importantly, the degree of learning was not significantly different between the two test conditions, $t(19) = 0.41$, $p = .687$, $d = 0.09$. The Bayes factor ($BF_{01} = 3.99$) suggested that our data were 3.99 times more likely to occur under the null hypothesis than under the alternative hypothesis; that is, it moderately favored the null hypothesis. Combining the test performance in Experiments 1A and 1B, there was no significant main effect of the two experiments, $F(1, 38) = 0.19$, $p = .667$, $n_p^2 = .01$; no significant main effect of the two test conditions (the basic- and subordinate-level test condition or the subordinate-level test condition), $F(1, 38) = 0.01$, $p = .919$, $n_p^2 < .001$; and no significant interaction between the within- and between-subjects variables, $F(1, 38) = 0.26$, $p = .612$, $n_p^2 = .01$. Thus, in Experiment 1B, in which participants had to actively extract the basic-level categories and explicitly remember those categories, we replicated the findings that the presence of statistical regularities at the basic level did not influence temporal VSL at the subordinate level.

The presentation order did not influence the test performance. We observed no significant difference between the mean correct percentages in the half of trials in which the base triplets appeared first (test performance: 63.61%) and in the other half, in which the nonbase triplets appeared first (test performance: 55.83%), $t(19) = 1.20$, $p = .246$, $d = 0.28$. Additionally, participants did not learn the statistical regularities during the test phase, in that the mean correct percentages in the first half (58.89%) and the second half (60.56%) of trials were not significantly different, $t(19) = -0.47$, $p = .641$, $d = 0.11$.

In the binary confidence judgment, five out of the 20 participants reported having used some kind of knowledge (test performance: 70.56%), whereas the rest reported having guessed (test performance: 56.11%). Of the five participants, four specifically reported having been aware of the temporal regularities (test performance: 77.08%), whereas the other one reported another strategy that was not at all effective (test performance: 44.44%).

The main finding of Experiment 1A was replicated in Experiment 1B, in which participants not only had to extract the basic-level regularities actively and explicitly, but they also had to attend to the repetition in the image sequence that contained the temporal regularities. That is, in both Experiments 1A and 1B, we found that participants learned the temporal statistics from the six different triplets in real-world images. Importantly, the presence of temporal regularities at the basic category level did not help temporal VSL at the subordinate level.

However, there are two other possibilities for other factors that might have interfered with participants' ability to take advantage of the basic-level category information when learning the temporal regularities. One possibility is that in the stimuli in Experiments 1A and 1B, it might have been so easy to name a category at the subordinate level (e.g., participants might have categorized a bird image as a parrot at the subordinate level, instead of as a bird at the basic level) that the subordinate-level category information might have been more salient than the basic-level category information. That is, participants might not have detected the temporal regularities at the basic level because of low saliency, leading to the results in Experiments 1A and 1B, in which the temporal regularities at the basic category level did not influence the extent of temporal VSL. To rule out this possibility, in Experiments 2A and 2B, we used indoor scene photographs as our real-world images. Contrary to most of the object images used in Experiments 1A and 1B, which could easily be associated with certain words (e.g., for the bird category, *parrot*, *owl*, and *magpie*), the indoor scene photographs in this experiment were not associated with such names, but it would be more likely that the images could be differentiated by various descriptions (e.g., among the living rooms, there was a living room that featured a pretty large space between two sofas, another living room that had a unique tree, and a final living room that had a checkerboard pattern carpet). By using the indoor scene images, we intended to make it difficult to verbalize the subordinate-level information, and thus to lessen the saliency at this level. In addition, by using these different types of categories, we could test generalization of the findings from Experiments 1A and 1B.

Another possibility is that the triplets with statistical regularities at only the subordinate level might have impeded temporal VSL at the basic level, considering that we treated the basic- and subordinate-level versus subordinate-level familiarization conditions as a within-subjects variable and presented all of the six triplets in an interleaved manner. Although we initially chose the within-subjects design in order to reduce the confounding effects associated with the between-subjects design, we assumed that the high memory load (i.e., 18 exemplars across six triplets) might have caused a memory interference effect. To exclude this possibility, we adopted a between-subjects design in Experiments 2A and 2B, but used a sample size (40 participants in Experiment 2A) larger than those in previous experiments (20 participants on average; Emberson & Rubinstein, 2016), to increase the statistical power. Although participants were exposed to the same number of triplets as in Experiments 1A and 1B, for half of the participants all triplets had temporal regularities at both the basic and subordinate levels; for the other half, all triplets had temporal regularities at only the subordinate level.

Experiment 2A

Method

Participants

Forty new and naïve students from Yonsei University participated in Experiment 2A, in exchange for course credits or monetary compensation (5,000 won). All had normal or corrected-to-normal visual acuity.

Apparatus and stimuli

The apparatus was the same as in Experiments 1A and 1B. For the stimuli, we used a total of 18 grayscale square images from six indoor categories (a bathroom, a bedroom, a kitchen, a living room, a staircase, and a dressing room). As in Experiments 1A and 1B, each of the six basic categories contained three images that were different at the subordinate level. For instance, the bathroom category contained three scenes that were taken from three different bathrooms. The width and height of the images was the same as in Experiments 1A and 1B. The background of the image was gray.

Design and procedure

The familiarization phase was followed by a surprise test phase.

Familiarization phase As in Experiments 1A and 1B, we presented a sequence of 1,680 images, in which the 18 scene images in six different indoor scene categories repeated one after another (see Fig. 3a). The 18 images comprised six triplets (Table 2), in each of which three images presented temporal regularities by always appearing in the same order; except that the category condition (the basic- and subordinate-level and subordinate-level conditions) became a between-subjects condition, the sequence in the familiarization phase was presented in the same way as in Experiments 1A and 1B. The number of triplets presented to each group was six, as in Experiments 1A and 1B.

Test phase As in Experiments 1A and 1B, participants performed a surprise 2AFC task for familiarity judgments, to test their ability to learn temporal regularities at the basic and subordinate levels (see Fig. 3b), by performing the same binary confidence judgment task (Bertels et al., 2012). The only exception was that temporal VSL was tested either at both the basic and subordinate levels (the *basic- and subordinate-level*

(a) Familiarization phase



(b) Test phase

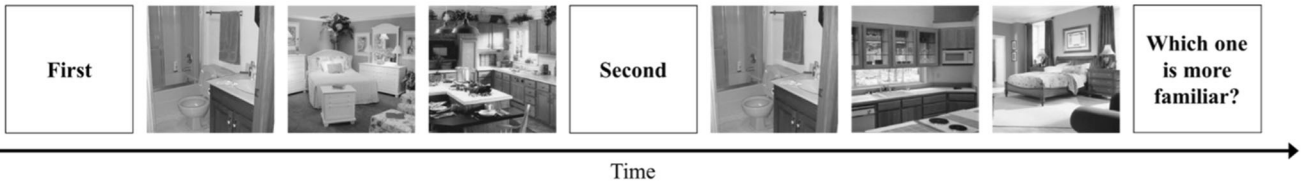


Fig. 3 An example trial from Experiments 2A and 2B. (a) In the familiarization phase, the participants performed the same dual task as in Experiment 1B. That is, they were asked to detect a repetition of the same image (e.g., the living room in shown here in a dotted box), as well as to categorize each image and remember a total of six categories that they

would observe. (b) After completing the familiarization phase, participants performed a surprise two-alternative forced choice task of making familiarity judgments. They were asked to choose the more familiar sequence after the base and nonbase triplets had been presented sequentially.

test condition) or at only the subordinate level (the subordinate-level test condition), as a between-subjects variable.

Results and discussion

Overall, the mean repetition detection accuracy in the familiarization was 83.79%, and the false alarm rate was 0.88% (5.21 times). Specifically, in the basic- and subordinate-level group, the mean detection accuracy was 86.22%, and the false alarm rate was 0.63% (3.15 times). In the subordinate-level group, the mean detection accuracy was 81.37%, and the false alarm rate was 1.13% (5.7 times). There was no significant difference between the two groups in either mean detection accuracy, $t(38) = 1.03, p = .310, d = 0.03$, or false alarm rate, $t(38) = -1.52, p = .136, d = 0.48$. Participants successfully recalled the six basic-level categories. In the basic- and subordinate-level group, 18 participants

recalled all of the six categories correctly, and two participants reported only five categories. In the subordinate-level group, 17 participants recalled all categories, and three participants recalled five.

The results for the 2AFC familiarity test are shown in Fig. 2. Overall, the mean percentage of correct responses in the 2AFC familiarity test was 56.32%, which was significantly different from chance level, $t(39) = 3.42, p = .001, d = 0.54$. Specifically, in the basic- and subordinate-level group, the mean accuracy was 55.28%, which showed a marginally significant difference from chance level, $t(19) = 1.71, p = .10, d = 0.38$. In the subordinate-level group, the mean accuracy was 57.36%, which was significantly different from chance level, $t(19) = 3.54, p = .002, d = 0.79$. Again, importantly, the mean accuracy was not statistically different between the two groups, $t(38) = -0.56, p = .580, d = -0.18$. The strength of

Table 2 Sample triplets for the familiarization in Experiment 2A

Basic- and Subordinate-Level Familiarization	
Triplet 1: Bathroom ₁ → Bedroom ₁ → Kitchen ₁	Triplet 4: Closet ₁ → Staircase ₁ → Livingroom ₁
Triplet 2: Bathroom ₂ → Bedroom ₂ → Kitchen ₂	Triplet 5: Closet ₂ → Staircase ₂ → Livingroom ₂
Triplet 3: Bathroom ₃ → Bedroom ₃ → Kitchen ₃	Triplet 6: Closet ₃ → Staircase ₃ → Livingroom ₃
Subordinate-Level Familiarization	
Triplet 1: Kitchen ₁ → Bedroom ₁ → Bathroom ₁	Triplet 4: Livingroom ₁ → Staircase ₁ → Closet ₁
Triplet 2: Bedroom ₂ → Bathroom ₂ → Kitchen ₂	Triplet 5: Closet ₂ → Staircase ₂ → Livingroom ₂
Triplet 3: Bathroom ₃ → Kitchen ₃ → Bedroom ₃	Triplet 6: Staircase ₃ → Livingroom ₃ → Closet ₃

Half of the participants, who were in the basic- and subordinate-level familiarization condition (first four rows), were presented with six triplets, in which statistical regularity existed at both the basic (i.e., Bathroom → Bedroom → Kitchen or Closet → Staircase → Living Room, in this example) and the subordinate (e.g., Bathroom₁ → Bedroom₁ → Kitchen₁, in Triplet 1) levels. The other half of the participants, who were in the subordinate-level familiarization condition (last four rows), were presented with six triplets that involved statistical regularity at the subordinate level only

evidence favoring the null model was weak or anecdotal, $BF_{01} = 2.86$, indicating that our data were 2.86 times more likely to occur under the null hypothesis than under the alternative hypothesis.

The presentation order did not influence test performance. The mean correct percentages were not significantly different between the trials in which the base triplets were presented first (test performance: 53.61%) and those in which the nonbase triplets appeared first (test performance: 59.03%), $t(39) = -1.60$, $p = .119$, $d = 0.05$. Additionally, learning did not occur during the test phase, because the performance means in the first half (57.50%) and the second half (55.14%) were not significantly different, $t(39) = 0.85$, $p = .400$, $d = 0.27$.

In the binary confidence judgment, only two out of 40 participants reported having used knowledge in the test (one in the basic- and subordinate-level group and one in the subordinate-level group; test performance: 73.61%), and the rest reported having guessed (test performance: 55.41%). Of the two participants, only the one participant in the basic- and subordinate-level group reported having detected the existence of the temporal regularities (test performance: 86.11%), whereas the other participant reported an ineffective strategy (test performance: 61.11%).

The results in Experiments 1A and 1B were again replicated in Experiment 2A, in that participants learned the temporal regularities from six different triplets, and the presence of temporal regularities at the basic level did not help VSL at the subordinate level. These results ruled out the possibility that our finding was due to memory interference, either among the triplets for which the temporal regularities existed at different hierarchical levels or due to the characteristics of the images themselves, where the basic-level information was more salient than the subordinate-level categories.

However, there could still be other confounding possibilities. One possibility was that memory load for the six triplets was so high that participants could not take advantage of the temporal regularities at the basic level. To test this, in Experiment 2B participants were provided with fewer triplets (four triplets). In addition, by using the same number of triplets that had been used in previous studies of VSL in real-world images (Brady & Oliva, 2008; Otsuka et al., 2013, 2014), we attempted replication of a previous finding that VSL could be based on object information in real-world images. Another possibility was that the time spent in the familiarization phase (29 min) was so long that participants in the dark room could have lost motivation to sustain their attention in order to track the statistical information at different levels. Considering that the ability to sustain attention is highly associated with the length of a task (Kindlon, 1998), in Experiment 2B we reduced the length of the familiarization phase by more than half (13 min).

Experiment 2B

Method

Participants

Forty new and naïve students from Yonsei University participated in Experiment 2B in exchange for course credits or monetary compensation (5,000 won). All had normal or corrected-to-normal visual acuity.

Apparatus and stimuli

The apparatus and stimuli were the same as in Experiment 2A, except that a total of 12 indoor scene images were used. That is, we selected only 12 indoor scene images from among the 18 images used in Experiment 2A, whereas all six basic-level indoor scene categories were still used.

Design and procedure

The familiarization phase was followed by a surprise test phase.

Familiarization phase The familiarization phase was the same as in Experiment 2A, except with two changes. First, we used only four triplets per group, which were constructed from the 12 indoor scene images. In this way, we could reduce the total number of triplets that participants had to remember. This number was identical to the total number of triplets used in previous studies (Brady & Oliva, 2008; Otsuka et al., 2013, 2014). Second, each triplet was repeated 56 times, which was less than the number of repetitions (84 times) in Experiment 2A. In this way, the length of the familiarization phase was reduced by more than half (i.e., from 29 to 13 min), such that participants were prevented from being bored or losing the motivation to learn the temporal regularities. As compared to the length of the familiarization in previous studies that had tested temporal VSL, the 13-min duration was located in the middle—it was still longer than those in a group of studies that had used nonsemantic shapes (6 min in Fiser & Aslin, 2002; 5 min in Turk-Browne & Scholl, 2009), but it was shorter than in another group (20 min in general) that had used semantic images (Brady & Oliva, 2008; Otsuka et al., 2013). We inserted 48 images into the image sequence so that either the first or the third image in the triplets was occasionally repeated. Participants performed the same dual task as in Experiments 2A and 1B, which was to remember the six basic-level categories and to detect the repetition of the same image.

Test phase The test phase was the same as in Experiment 2A, except with two changes. First, learning of four (not six) triplets was tested in the 2AFC familiarity task. Second, the total number of trials in this task was 32. The joint probability of the three images in each of the base triplets was .083, which was identical to the levels of probability used in previous studies (Brady & Oliva, 2008; Otsuka et al., 2013, 2014), whereas the joint probability in the nonbase triplets was 0. The 32 trials were constructed by combining four base triplets, four nonbase triplets, and two presentation orders, in which the base triplet appeared either first or second.

Results and discussion

The mean repetition detection accuracy in the familiarization was 83.07%, and this performance was not statistically different between the basic- and subordinate-level (80.63%) and subordinate-level (85.52%) groups, $t(38) = -0.94$, $p = .352$, $d = 0.30$. The false alarm rate was 1.17% (2.63 times), and this performance was also not statistically different between the two groups (1.05%, or 2.35 times, for the former, and 1.29%, or 2.9 times, for the latter), $t(38) = -0.46$, $p = .65$, $d = 0.15$. Participants successfully recalled the basic-level categories. In the basic- and subordinate-level group, 15 participants recalled all six of the categories correctly, and five participants recalled five, whereas in the subordinate-level group, 18 participants recalled all six categories, and two participants recalled five of them.

The results from the 2AFC familiarity test are shown in Fig. 2. The mean percentage of correct responses in the 2AFC familiarity test was 54.14%, which was significantly higher than chance level, $t(39) = 2.29$, $p = .028$, $d = 0.36$. Specifically, the test performance in the basic- and subordinate-level group was 55.16%, which was marginally significantly higher than chance level, $t(19) = 1.95$, $p = .066$, $d = 0.44$. However, in the subordinate-level group, the test performance was 53.13%, which was not significantly different from chance level, $t(19) = 1.24$, $p = .231$, $d = 0.28$. Again, the test performance was not significantly different between the two groups, $t(38) = 0.56$, $p = .582$, $d = 0.18$. The strength of evidence favoring the null model was weak or anecdotal ($BF_{01} = 2.86$), meaning that our data were 2.86 times more likely to occur under the null hypothesis than under the alternative hypothesis. When we combined the test performance in Experiments 2A and 2B, there was no main effect of the two experiments, $F(1, 76) = 0.70$, $p = .407$, $n_p^2 = .01$; no main effect of the different learning groups (the basic- and subordinate-level group and the subordinate-level group),

$F(1, 76) < 0.01$, $p = .992$, $n_p^2 < .001$; and no interaction between the experiments and learning conditions, $F(1, 76) = 0.62$, $p = .433$, $n_p^2 = .01$.

The test performance was not influenced by the presentation order. The mean correct percentages did not significantly differ, depending on whether the base triplets were presented first (test performance: 53.59%) or second (test performance: 54.69%), $t(39) = -0.41$, $p = .687$, $d = 0.06$. In addition, learning did not happen during the test, in that the performance levels in the first half (55.63%) and the second half (52.66%) of trials did not differ significantly, $t(39) = 1.03$, $p = .308$, $d = 0.16$.

In the binary confidence judgment, six among the 40 participants reported having relied on knowledge (test performance: 53.65%), and more specifically, all six of these participants reported ineffective strategies (three in the basic- and subordinate-level group and three in the subordinate-level group). However, the other participants reported having guessed (test performance: 54.23%).

In Experiment 2B, in which participants were provided with fewer triplets to learn and with a shortened familiarization time, we again found that temporal regularities at the basic level did not enhance VSL at the subordinate level. These results suggest that this is not simply a null finding confounded by the memory overload or by an excessively long learning phase.

General discussion

In the present study, we examined whether temporal regularities at the basic category level could influence VSL at the subordinate category level. In Experiment 1A, participants were exposed to a stream of real-world images, which consisted of six triplets with temporal regularities. Half of the triplets included statistical regularities at only the subordinate level, and the other half had regularities at both the subordinate and basic levels. We found that participants learned the temporal regularities across the six triplets, but the additional presence of statistical regularities at the basic category level had no effect on the degree of statistical learning. This finding was consistently observed in subsequent experiments, suggesting that our results could not be explained by specific features of our design. In Experiment 1B, we attempted to rule out the possibility that participants might not have categorized each image at the basic level. We explicitly asked participants to extract the six basic-level categories from the image stream and to remember all of the categories, in order to encourage them to categorize each image at the basic level. Although nearly all of the participants suc-

successfully recalled the six basic-level categories, we again found that the additional presence of statistical regularities at the basic level did not influence the degree of learning. In Experiment 2A, by using indoor scenes, we attempted to make it more difficult to categorize images at the subordinate level. In this experiment, we also intended to avoid the memory interference effect by adopting a between-subjects design, though we initially had adopted a within-subjects design in Experiments 1A and 1B to reduce between-subjects noise. In Experiment 2B, we reduced the memory load by presenting four triplets (rather than six) and attempted to reduce motivation loss in maintaining attention throughout the learning phase by reducing its length by half. In these experiments, we again observed that the degree of VSL was not influenced by the extra regularities at the basic level.

Our results are consistent with the previous finding (Emberson & Rubinstein, 2016) that VSL of semantic categories was constrained to individual objects. That is, in the present study we found that being exposed to statistical regularities at both the basic and subordinate levels produced no benefit of boosting the extent of learning, as compared to being exposed to the regularities at the subordinate level only. Contrary to Emberson and Rubinstein, in which a typicality effect was employed to represent basic-level concepts without any explicit instruction, in our study we directly asked participants to extract the semantic concepts at the basic level, and almost all of the participants successfully recalled these concepts after the learning phase. Although participants had explicit knowledge of the basic-level information, the statistical regularities at this level still had no influence on the degree of VSL at the subordinate level. By ruling out several possibilities that had not been tested by Emberson and Rubinstein, our results strongly suggest that people do not utilize temporal regularities at both the basic and subordinate levels. In particular, in our study we sought to minimize the group variance by adopting a within-subjects design. We increased the memory load, in case the load was too low to detect the difference in the degree of VSL, and we also extended the number of the tested basic-level concepts (12 categories), to increase the generalizability of our findings. As was shown by Emberson and Rubinstein, our results supported the idea that VSL is constrained to learn a particular level of patterns when the patterns are presented across multiple levels. These findings are surprising, considering the prevalent view that basic-level category information is prioritized during perception (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). The basic-level category has been considered to consist of the most inclusive information, where

conceptually similar names possess many commonalities as well as differences (Roach & Lloyd, 1978). Basic-level categories are identified faster than those at the subordinate or superordinate identities (Murphy & Brownell, 1985). Despite the robust effect of basic-level prioritization, our data did not show any influence of basic-level regularities in the degree of VSL. The effect of the extra regularities among the basic-level concepts was neither facilitative nor disruptive.

One possibility for why our participants did not utilize the extra regularities is that VSL at multiple levels might be limited when both perceptual and semantic processes engage in the learning at the same time. In our study, regularity at the basic level was primarily defined in terms of semantic categories that participants had to abstract, whereas the regularity at the subordinate level was based on visual features as well as semantic names. Unlike in the present study, in a previous study (Jun & Chong, 2016) we had presented a series of nonsemantic shapes, for which perceptual features were the only available information from which to learn the statistical regularities. Each display contained two novel shapes in which the global-level shape was defined by the location of the local-level shapes. In both the present and previous studies, the statistical regularities were defined at multiple levels—that is, at basic and subordinate levels in the present study, and at global and local levels in the previous study. However, the results were not consistent. Contrary to the present results, the previous study showed an increase in VSL in the presence of statistical regularities at the other level. That is, the presence of regularities at both the global and local levels (BOTH & HS condition in Experiment 2) boosted the extent of VSL, as compared to the test condition, in which regularities were presented at only the local (or only the global) level (LOCAL & HS condition or GLOBAL & HS condition). Importantly, one of the major differences between the two studies was that different processes were engaged in the learning process. In the previous study, only the perceptual features were available, in that all the stimuli were novel and nonverbalized shapes. However, in the present study, both the perceptual and semantic features were available at the different levels. Our results imply that multiple features of stimuli could pose a constraint on utilization of the regularities at multiple levels. Specifically, we think that the semantic feature itself may not be sufficient to learn statistical regularities when VSL is associated with both perceptual and semantic features. For instance, a contextual cueing effect was not observed when semantic information was repeated without the repetition of visual features (Makovski, 2016, 2018), suggesting that perceptual

repetition is more critical than semantic categories in order for learning to happen.

Our results suggest that the formation of chunks at the basic level (i.e., a group of three different categories without a specified arrangement) is not enough to influence the ability to compute statistical regularities (i.e., the representation of specific temporal orders). In our study, at the basic level, the nonbase triplets were manipulated to have changes only in the specific temporal orders. That is, the basic-level concepts (e.g., ABC) of the base triplets were pitted against the nonbase triplets having the same three basic-level concepts but in a different temporal order (e.g., ACB, BAC, CBA). However, at the subordinate level, the nonbase triplets were manipulated in order to have changes in chunk information as well as in the temporal regularities. That means that the subordinate-level concepts (e.g., $A_1B_1C_1$, $A_2B_2C_2$, $A_3B_3C_3$) were pitted against the nonbase triplets, in which chunk information was violated (e.g., $A_1C_2B_3$, $B_2A_3C_1$, $C_3B_1A_2$) in new temporal orders. We intentionally designed the nonbase triplets in this way in order to avoid the possibility that triplets learned with different basic-level category members might be mixed up in the test phase. Although what people can acquire during VSL between chunk knowledge and temporal regularities was not the aim of our study, our results suggest that people might not utilize the specific representation of temporal orders at the basic level, but they might instead rely more on the subordinate-level information, for which both the chunk information and temporal order were available. Regarding this question, the relationship between temporal order and chunk information in statistical learning, Perruchet and Pacton (2006) posited three possibilities: (1) these processes are independent, (2) they are successive steps at which the chunking is inferred from the temporal orders, (3) or the chunking is a primitive process. Our results likely support the first possibility, in that the temporal orders at the basic level had no influence on the processing of chunk information at the subordinate level. To clarify this relationship in future research, one could attempt to tease apart the chunking and the computation of statistical information when statistical regularities exist at multiple levels.

The most important implication of our findings is that statistical learning can be limited to a particular level when regularities are simultaneously presented at multiple levels. Despite the previous finding that people could abstract statistical regularities at semantically different levels (Brady & Oliva, 2008), our study revealed that there was no enhancement in the degree of VSL when statistical information was available at multiple levels. This constraint of VSL might reflect the general limited capacity of our information processing. Human observers are “cognitive misers,” because they do

not utilize all possible information but flexibly adopt the most accessible and relevant information (Kahneman, 2011). Due to the limited capacity of mental resources and the effort involved, people do not maximize the utility of all relevant information that is available to them (Tversky & Kahneman, 1974). In the real world—for instance, in our commute—we may utilize only the portion of regularities that are currently relevant information (e.g., which café I will go to today, which space in the parking lot I should use this morning, and which part of the building I should head for), instead of representing all possible kinds of regularities that might not be immediately useful or might consist of too much abstract information (e.g., the sequence of the café, then the parking lot, and then the office building). Consistent with this idea, Higuchi, Ueda, Ogawa, and Saiki (2016) found that when statistical regularities were available in multiple dimensions (location and object identity), only the regularities that were most relevant to one’s task goal were prioritized, suggesting that observers prioritized the most relevant information for statistical learning, instead of learning all kinds of statistical regularities.

To conclude, in the present study we found that people do not utilize extra statistical regularities at the basic level when they learn them among a sequence of real-world images. Our natural environments feature statistical regularities at semantically multiple category levels, and our evidence suggests that statistical regularities at the basic level do not influence the acquisition of regularities at the subordinate level. This finding supports the idea that statistical learning may be constrained to a particular level when observers are confronted with multiple levels of regularities.

Author note This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science and ICT (Grant NRF-2016R1A2B4016171), and by the Yonsei University Future-Leading Research Initiative of 2017 (Grant 2017-22-0145).

Appendix: Visual stimuli used in our experiments

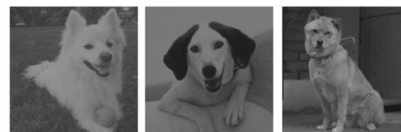
The stimuli were adopted from those of Kim (2008). (a) In Experiments 1A and 1B, there were a total of 18 images, in which each image was classified under one of six basic-level categories (bird, car, dog, fish, flower, and house). (b) Experiment 2A also featured a total of 18 images, all of which were indoor scene images. In Experiment 2B, the same visual images were used as in Experiment 2A, except that the last image was removed for each basic-level category, and thus a total of 12 images were included in this experiment.

(a) Bird

Fish

**Car**

Flower

**Dog**

House

**(b) Bathroom**

Kitchen

**Bedroom**

Livingroom

**Closet**

Staircase

**References**

- Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1425–1431. doi:<https://doi.org/10.1037/a0027210>
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147. doi:<https://doi.org/10.1037/0033-295X.94.2.115>
- Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, 19, 678–685. doi:<https://doi.org/10.1111/j.1467-9280.2008.02142.x>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436. doi:<https://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. doi:<https://doi.org/10.1006/cogp.1998.0681>
- DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11, 333–341. doi:<https://doi.org/10.1016/j.tics.2007.06.010>
- Emberson, L. L., & Rubinsten, D. Y. (2016). Statistical learning is constrained to less abstract patterns in complex sensory input (but not the least). *Cognition*, 153, 63–78. doi:<https://doi.org/10.1016/j.cognition.2016.04.010>
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822–15826. doi:<https://doi.org/10.1073/pnas.232472899>
- Higuchi, Y., Ueda, Y., Ogawa, H., & Saiki, J. (2016). Task-relevant information is prioritized in spatiotemporal contextual cueing. *Attention, Perception, & Psychophysics*, 78, 2397–2410. doi:<https://doi.org/10.3758/s13414-016-1198-0>
- 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.
- Jun, J., & Chong, S. C. (2016). Visual statistical learning of temporal structures at different hierarchical levels. *Attention, Perception, & Psychophysics*, 78, 1308–1323. doi:<https://doi.org/10.3758/s13414-016-1104-9>
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Strauss, & Giroux.
- Kim, J. (2008). *Temporal and spatial advantages of object detection over basic-level categorization*. (Master's thesis). Yonsei University, Seoul, South Korea.
- Kindlon, D. (1998). The measurement of attention. *Child Psychology and Psychiatry Review*, 3, 72–78.
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., & Wagenmakers, E.-J. (2015). JASP (Version 0.7) [Computer software]. Amsterdam: JASP Project.
- Makovski, T. (2016). What is the context of contextual cueing? *Psychonomic Bulletin & Review*, 23, 1982–1988. doi:<https://doi.org/10.3758/s13423-016-1058-x>
- Makovski, T. (2018). Meaning in learning: Contextual cueing relies on objects' visual features and not on objects' meaning. *Memory & Cognition*, 46, 58–67. doi:<https://doi.org/10.3758/s13421-017-0745-9>
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge: MIT Press.
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 70–84. doi:<https://doi.org/10.1037/0278-7393.11.1.70>
- 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. doi:<https://doi.org/10.1073/pnas.0708424105>
- Otsuka, S., Nishiyama, M., & Kawaguchi, J. (2014). Constraint on the semantic flexibility in visual statistical learning. *Visual Cognition*, 22, 865–880. doi:<https://doi.org/10.1080/13506285.2014.923548>
- Otsuka, S., Nishiyama, M., Nakahara, F., & Kawaguchi, J. (2013). Visual statistical learning on the basis of the perceptual and semantic information of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 196–207. doi:<https://doi.org/10.1037/a0028645>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. doi:<https://doi.org/10.1163/156856897X00366>
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233–238. doi:<https://doi.org/10.1016/j.tics.2006.03.006>
- Rosch, E., & Lloyd, B. B. (Eds.). (1978). *Cognition and categorization* (Vol. 1, p. 978). Hillsdale, NJ: Erlbaum.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439. doi:[https://doi.org/10.1016/0010-0285\(76\)90013-X](https://doi.org/10.1016/0010-0285(76)90013-X)
- 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., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, 134, 552–564. doi:<https://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. doi:<https://doi.org/10.1037/0096-1523.35.1.195>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>