*Research Article*

# Statistical Learning Using Real-World Scenes

## Extracting Categorical Regularities Without Conscious Intent

**Timothy F. Brady and Aude Oliva**

*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*

ABSTRACT—*Recent work has shown that observers can parse streams of syllables, tones, or visual shapes and learn statistical regularities in them without conscious intent (e.g., learn that A is always followed by B). Here, we demonstrate that these statistical-learning mechanisms can operate at an abstract, conceptual level. In Experiments 1 and 2, observers incidentally learned which semantic categories of natural scenes covaried (e.g., kitchen scenes were always followed by forest scenes). In Experiments 3 and 4, category learning with images of scenes transferred to words that represented the categories. In each experiment, the category of the scenes was irrelevant to the task. Together, these results suggest that statistical-learning mechanisms can operate at a categorical level, enabling generalization of learned regularities using existing conceptual knowledge. Such mechanisms may guide learning in domains as disparate as the acquisition of causal knowledge and the development of cognitive maps from environmental exploration.*

One of the primary tasks of the brain is to extract regularities from the environment in order to make inferences and guide behavior in novel situations. Indeed, sensitivity to statistical regularities plays an important role in various domains of perception and cognition, from preparing motor actions (Nissen & Bullemer, 1987) to parsing language (Trueswell, 1996) and making higher cognitive judgments, such as predicting the amount of money a movie will earn (Griffiths & Tenenbaum, 2006).

Address correspondence to Timothy Brady, Department of Brain and Cognitive Sciences, 46-4078, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, e-mail: tfbrady@mit.edu.

Recent work on statistical learning has demonstrated that observers extract the covariance between syllables, tones, or shapes that appear in a predictable order or spatial arrangement (Behrmann, Geng, & Baker, 2005; Chun & Jiang, 1998; Fiser & Aslin, 2002; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Turk-Browne, Jungé, & Scholl, 2005). In a visual variant of such an experiment, observers are presented with a sequence of shapes and are not told that the sequence consists of arrangements of four temporal triplets (i.e., groups of three shapes that always appear in the same order, as in the sequence ABCGHIDEFABCJKLDEF . . . . After a brief exposure to this stream, observers are able to reliably identify the triplets (e.g., ABC) as more familiar than foil sequences (e.g., AEI), despite the fact that they have seen all the individual shapes an equal number of times. This indicates that they have learned which shapes appeared together, although their verbal reports indicate no awareness of the structure in the stream.

To date, these experiments have used simple novel shapes to investigate visual statistical learning. In the real world, however, the units over which learning must operate are considerably more information rich, and people often learn novel statistical regularities that pertain to objects and environments that have many prior associations. Furthermore, objects and scenes in the world have semantic knowledge associated with them, and important regularities occur at multiple levels of semantic abstraction. For instance, the co-occurrence of different places when you move through your home may allow you to infer that your kitchen will lead to your living room (i.e., you form a cognitive map; Tolman, 1948), but also allows you to construct a hierarchy of such associations, moving from specific instances ("my kitchen and living room appear together") to basic-level categories ("kitchens and living rooms tend to appear together") and more abstract representations ("indoor rooms appear together").

  

The purpose of the present study was to examine the extent to which statistical-learning mechanisms can operate automatically at an abstract, conceptual level. Specifically, we used a novel modification of the statistical-learning paradigm to test whether observers incidentally learn category-level relationships, even when the task does not require processing information at a categorical level.

We used real-world scenes as a case study for investigating categorical statistical learning and transfer of learning across levels of representation. In Experiment 1, we demonstrated statistical learning with complex real-world scenes. In Experiment 2, we demonstrated that statistical-learning mechanisms can extract regularities at a categorical level, by showing transfer of learning between a particular image and another exemplar of the same category. Finally, in Experiments 3 and 4, we showed that these categorical representations can be accessed lexically, which indicates that the regularities are present in an abstract format that is divorced from the visual details of the images.

## EXPERIMENT 1: STATISTICAL LEARNING USING VISUAL IMAGES

Previous experiments investigating the extraction of statistics of covariance from the visual world have used simple novel shapes. In Experiment 1, we verified that the visual statistical-learning mechanism computes relationships among complex stimuli with which subjects have prior associations: visual scenes.

### Method

#### *Observers*

Ten naive observers were recruited from the MIT participant pool (age range = 18–35) and received \$5 for their participation. All gave informed consent.

#### *Apparatus and Stimuli*

Stimuli were presented using MATLAB with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Twelve scene categories were used (see Fig. 1): bathroom, bedroom, bridge, building, coast, field, forest, kitchen, living room, mountain, street, and waterfall. Each category contained 120 different full-color images. For each observer, 1 picture was drawn from each of the 12 categories at random, resulting in a set of 12 different images. The images were centered and subtended $7.5° \times 7.5°$ of visual angle. Observers sat 60 cm from a 21-in. monitor.


**Fig. 1.** Examples from each of the categories of real-world scenes used in all four experiments. Overall, 120 different images of each category were selected.

Each of the 12 selected images was randomly assigned a position in one of four triplets (e.g., ABC)—sequences of three images that always appeared in the same order. Then a sequence of images was generated by randomly interleaving 75 repetitions of each triplet, with the constraints that the same triplet could never appear twice in a row and the same set of two triplets could never appear twice in a row (e.g., ABCGHIABCGHI was disallowed). In addition, 100 repeat images were inserted into the stream such that sometimes either the first or third image in a triplet repeated immediately (e.g., ABCCGHI or ABCGGHI). Allowing only the first or third image in a triplet to repeat served to keep the triplet structure intact, yet prevented the repeat images from being informative for delineating triplets from one another.

*Procedure*

Observers watched a 20-min sequence of 1,000 images, presented one at a time for 300 ms each with a 700-ms interstimulus interval (ISI). During this sequence, the task was to detect back-to-back repeats of the same image and to indicate repeats as quickly as possible by hitting the space bar. This cover task was intended to help prevent observers from becoming explicitly aware of the structure in the stream (Turk-Browne et al., 2005), and also avoided having observers simply view the stream passively (which would make it unclear what they were processing). Note that they were never informed that there was any structure in the stream of images.

Following this study period, observers were asked if they had recognized any structure in the stream and then were given a surprise forced-choice familiarity test. On each test trial, observers viewed two 3-image test sequences, presented sequentially at the center of the screen with the same ISI as during the study phase and segmented from each other by an additional 1,000-ms pause. One of these test sequences was always a triplet of images that had been seen in the stream (e.g., ABC), and another was a foil constructed from images from three different triplets (e.g., AEI). After the presentation of the two test sequences, observers were told to press either the "1" or the "2" key to indicate whether the first or second test sequence seemed more familiar from the initial study period. Each of the four triplets was tested eight times, paired twice with each of four different foil sequences (AEI, DHL, GKC, JBF), for a total of 32 test trials. Observers' ability to discriminate triplet sequences from foil sequences was used as a measure of statistical learning.

**Results and Discussion**

All 10 of the observers completed the repeat-detection task during the study period with few errors, detecting an average of 91% of the repetitions ($SD = 5\%$) and committing between one and five false alarms. These results demonstrate that observers were attending to the sequence of images. However, when asked, no observers reported explicitly noticing that the study stream had any structure.[1] Nonetheless, performance on the familiarity test indicated very robust statistical learning, with triplets being successfully discriminated from foils (86.6% of the test sequences chosen were triplets, and 13.4% were foils), $t(9) = 8.72$, $p = .00001$ (see Fig. 2).[2]

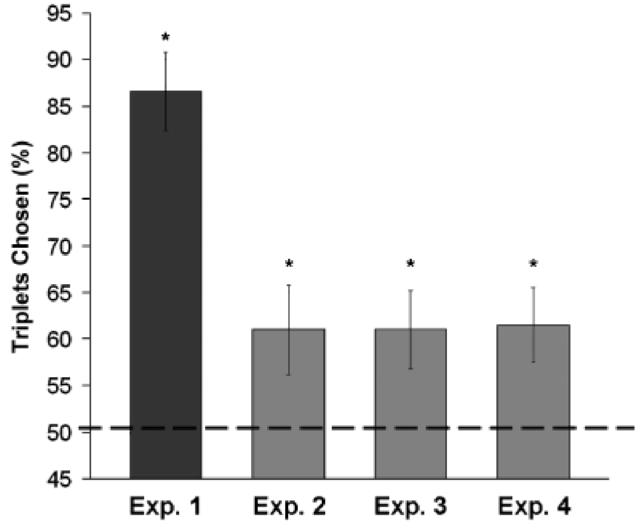


**Fig. 2.** Percentage of triplets chosen as familiar in each experiment. Error bars represent standard errors of the means. Chance level, indicated by the dashed line, is 50%. Asterisks indicate percentages significantly different from chance, $p < .05$.

These results extend previous demonstrations of visual statistical learning in two ways. First, they demonstrate visual statistical learning for scene stimuli, which are more complicated and information rich than the stimuli for which statistical learning has been demonstrated previously. Second, choosing the correct triplets at test in this experiment required not just forming episodic associations between the correct pictures, but also overcoming prior knowledge about how the scenes represented are associated in the world (e.g., bridges are rarely associated with living rooms).

In this experiment, learning likely occurred at the image level, because identical stimuli were repeated throughout the learning and test phases (and statistical learning has been previously demonstrated for shape and color: e.g., Turk-Browne, Isola, Scholl, & Treat, 2008). Therefore, to examine the role of category-level semantics in statistical learning, we performed a second experiment, in which the same string of images was never presented twice, but a pattern occurred at the categorical level.

[1]Observers were asked, "Did you notice any patterns in the stream of images?" and then "So, for example, if I asked you what images generally followed mountains, would you be able to tell me?" All observers responded negatively to both questions.

[2]For all experiments, a Lilliefors (1967) test failed to reject the null hypothesis that the data were normally distributed, $p > .10$. In addition, Monte Carlo methods that take into account the actual distribution of the stimuli (the average over subjects of a sum of 32 Bernoulli trials) indicated that our data were unlikely given the null hypothesis (all $ps < .01$).

## EXPERIMENT 2: STATISTICAL LEARNING OF CATEGORICAL REGULARITIES

Recent work has suggested that statistical learning abstracts away at least some of the perceptual information in the input: Learning can be preserved despite changes in color and even the order of the stimuli at test (Turk-Browne et al., 2008; Turk-Browne & Scholl, in press). In the real world, however, it may be useful to learn at an even more abstract level: the conceptual, rather than perceptual, level. Categorization allows one to compress information, because rather than storing the same information about each possible object or scene, one can store the information only one time and apply it to many different stimuli (Mervis & Rosch, 1981). This suggests that for statistical learning to be most useful (e.g., for statistical learning to maximize the number of stimuli to which a given regularity will apply), observers should learn statistical relationships between semantic categories, as well as between individual objects.

In order to learn information at the categorical level during everyday visual experience, people would need to extract the category of a given stimulus even though such information is task irrelevant in most everyday interactions. Therefore, in Experiment 2, we sought to determine (a) whether observers automatically extract the basic-level category of real-world scenes even when doing so is task irrelevant, and (b) whether observers learn the statistics of covariation at this categorical level. We used a design much like that of Experiment 1, except that the same 12 pictures were not repeated throughout the study and test periods; instead, new pictures from the same categories were drawn each time a particular triplet was shown (see Fig. 3). Thus, with the exception of the repeats required for the repeat-detection task, no images were ever repeated, and the statistical structure of the stream was defined only by the basic-level categories of the images. Observers still performed an image-level repeat-detection task during the study period, and were unaware that the category of the images was relevant or that there was a statistical structure in the stream of images.

### Method

Eleven naive observers participated in this experiment. The apparatus and stimuli were identical to those in Experiment 1, with one exception: Every time a triplet appeared in the stream, new images from the same categories were used. Thus, the triplet ABC would be a particular mountain image, bathroom image, and street image the first time it was presented, and a different mountain image, bathroom image, and street image the next time (e.g., $A_1B_1C_1G_1H_1H_1I_1D_1E_1F_1A_2B_2C_2$. . .). Repeats (for the

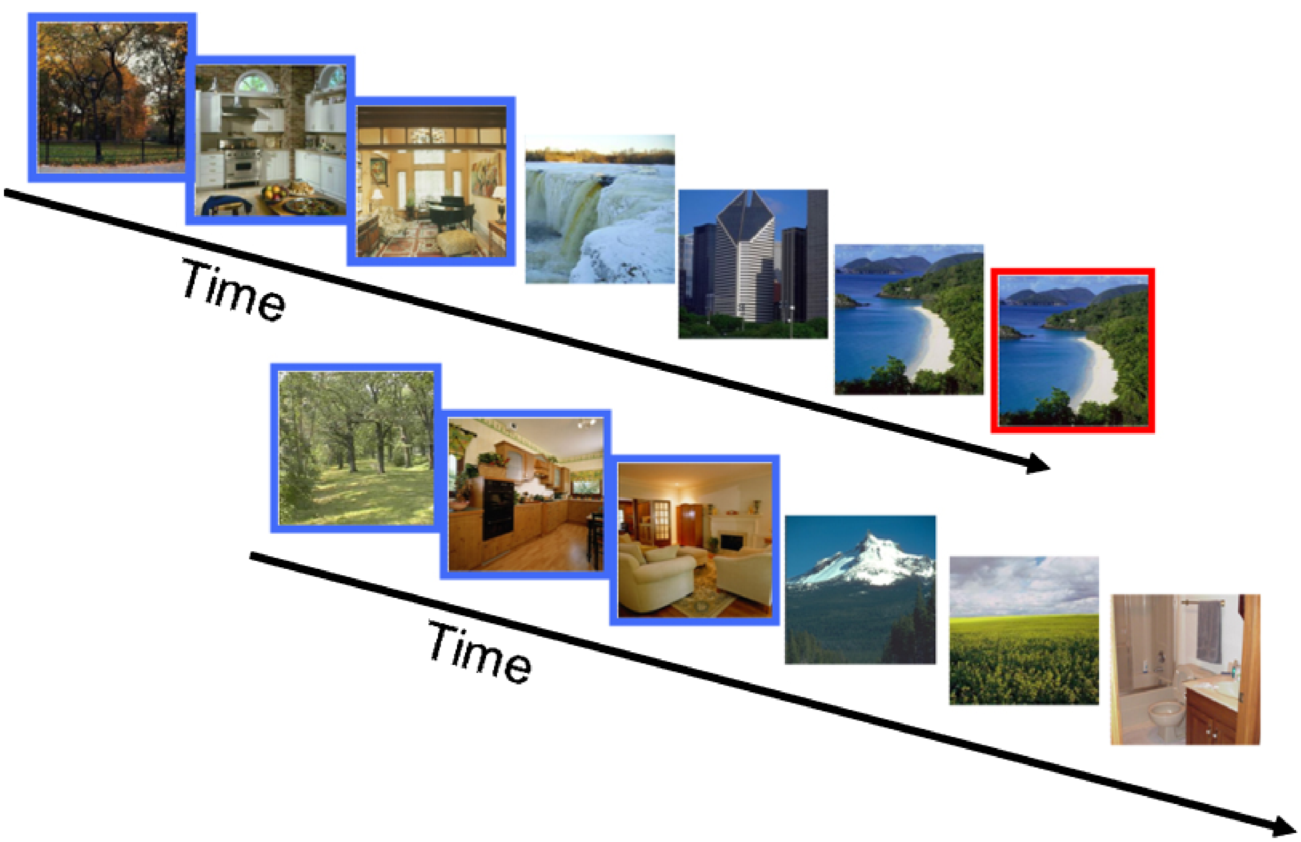


**Fig. 3.** Samples from a stream in Experiment 2. The order in which images were presented was predictable, but only at the categorical level. Thus, the first time a triplet was presented, one set of images was used; the next time, another set of images was used; and so forth. The blue outlines highlight images from the same triplet. Repeated images (e.g., the image outlined in red) were still physically identical.

repeat-detection task) were still of the exact same image, and observers were not informed that the images were drawn from a particular set of categories or that there was a structure in the stream. The test sequences consisted of entirely new images drawn from the same categories as the training images.

### Results and Discussion

Ten of the observers completed the repeat-detection task with few errors, detecting an average of 94% of the repetitions (*SD* = 3%) and committing between one and eight false alarms. The remaining observer was excluded from further analysis because of a high error rate on this task (having missed 22% of the repetitions). No observers reported noticing that the study stream had any structure.

Nonetheless, performance on the familiarity test indicated robust categorical statistical learning, with triplets successfully discriminated from foils (61.3% of the test sequences chosen were triplets, and 38.8% were foils), $t(9) = 2.61$, $p = .028$ (see Fig. 2). This indicates that observers were able to learn that images from particular categories co-occurred with images from other categories, even when the categorization was entirely unrelated to the task being performed.

The results also demonstrate that people extract category information automatically when confronted with real-world images (e.g., Grill-Spector & Kanwisher, 2005). Although this corroborates intuition in the literature on scene gist (Oliva, 2005; Potter, 1976) and fits well with the literature on semantic priming (pictures presented outside of conscious awareness result in priming that transfers to similar words, e.g., Dell'Acqua & Grainger, 1999), observers' ability to incidentally categorize visual scenes at the basic level while performing an unrelated task has not been previously reported, to our knowledge.

## EXPERIMENT 3: LEXICAL RECALL OF VISUALLY LEARNED REGULARITIES

Experiment 2 demonstrated that statistical learning operates at a categorical level, forming associations based on the basic-level category of scenes. However, because the images used were real-world scenes, many of the images within the same category shared low-level properties that they did not share with images from other categories (e.g., Oliva & Schyns, 2000; Oliva & Torralba, 2001). Thus, the statistical learning observed in Experiment 2 could have been operating over low-level regularities (color, orientation) that happened to covary strongly with image category, rather than directly over more conceptual representations. This distinction is tied to debates about the representation of conceptual knowledge and whether this representation is modality dependent (Barsalou, 1999) or more abstract (Pylyshyn, 1984), and furthermore is related to the degree to which categories are defined by visual similarity in the first place (Mervis & Rosch, 1981). However, we suggest that if the representations formed by statistical learning can be accessed in the absence of the low-level image regularities, they should be thought of as conceptual rather than simply low level (although low-level regularities are likely learned as well; this probably accounts for why learning was significantly stronger in Experiment 1 than in Experiment 2).

To probe the degree of abstraction present in the representations learned in our task, we changed the test sequences to words representing the categories, rather than images from the categories (e.g., in the triplet ABC, if A represented mountain images throughout training, A was the word "mountain" at test). Expression of learning in this context would indicate that the regularities learned in this task result in the association of more than just low-level perceptual features of the images.

### Method

Ten naive observers participated in this experiment. The apparatus and stimuli were identical to those in Experiment 2, with one exception: At test, words, rather than images, were presented. For example, during training, the triplet ABC might be a particular mountain image, bathroom image, and street image the first time it was presented, and a different mountain image, bathroom image, and street image the next time. At test, this triplet was represented by the words "mountain," "bathroom," and "street."

### Results and Discussion

All 10 of the observers completed the repeat-detection task with few errors, detecting an average of 92% of the repetitions (*SD* = 4%) and committing between one and four false alarms. No observers reported noticing that the study stream had any structure.

Nonetheless, performance on the familiarity test indicated robust statistical learning, with triplets of words being successfully discriminated from foils (61.1% of the test sequences chosen were triplets, and 38.9% were foils), $t(9) = 2.51$, $p = .036$ (see Fig. 2). This indicates that observers were able to incidentally learn that images from particular categories co-occurred with images from other categories and, surprisingly, were able to transfer this knowledge to a lexical task at test.

The results from Experiment 3 suggest that the associations formed in this task are not solely based on low-level image regularities, or at least can be expressed in the absence of these regularities. Thus, the learned information is automatically abstracted from the perceptual input into a high-level conceptual representation.

## EXPERIMENT 4: AUTOMATIC ABSTRACTION OF CATEGORICAL REGULARITIES

Experiments 2 and 3 demonstrated that statistical learning operates at a categorical level, forming associations based on the basic-level category of scenes. However, in both of these experiments, individual images were never repeated, meaning that the only regularity present in the stream was at the categorical level. Does categorical learning happen even when other regularities are present in the stream? Experiment 4 combined elements of Experiments 1 and 3 as follows: During the study stream, 12 identical images were repeated over and over again, but at test, we examined whether observers could transfer their learning to lexical items. Successful transfer would indicate that observers extracted the regularities at the categorical level even when regularities existed at the individual level.

### Method

Fifteen naive observers participated in this experiment. The apparatus and stimuli were identical to those in Experiment 1, with one exception: During the test trials, we presented words, rather than images.

### Results and Discussion

All 15 of the observers completed the repeat-detection task during the study period with few errors, detecting an average of 93% of the repetitions ($SD = 5\%$) and committing between one and three false alarms. No observers reported noticing that the study stream had any structure.

Nonetheless, performance on the familiarity test indicated robust statistical learning, with triplets of words being successfully discriminated from foils (61.5% of the test sequences chosen were triplets, and 38.5% were foils), $t(14) = 2.14$, $p = .050$ (see Fig. 2). This indicates that observers were able to incidentally learn at a categorical level.

Our results suggest that categorical information is abstracted from the perceptual input into a high-level categorical representation even when regularities exist at both the perceptual and the categorical levels. This provides further evidence for the automaticity of learning at the categorical level. The difference between the learning in this experiment and in Experiment 1 was significant (86.6% triplets vs. 61.5% triplets), $p = .003$, and indicates that observers were better able to identify which triplets they saw when the exact same images were used at test than when words were used at test. We suggest this is because observers were simultaneously learning at both the perceptual and the categorical levels; in Experiment 1, both of those sources of knowledge helped participants choose the correct triplet at test, whereas in Experiment 3, only the conceptual regularities informed observers' choices at test.

## GENERAL DISCUSSION

We examined for the first time whether statistical learning can operate at an abstract conceptual level, to discover the co-occurrence of semantic categories. In Experiment 1, we showed that statistical learning operates over complex real-world scenes, extending the domain of visual statistical learning to semantically meaningful stimuli. In Experiment 2, we demonstrated that statistical learning can successfully extract categorical regularities in a stream of pictures even when the semantic category of the scenes is task irrelevant. In Experiment 3, we demonstrated that these learned regularities can be accessed lexically, which suggests that they are accessible in an abstract format. Finally, in Experiment 4, we showed that observers learned the categorical structure of the stream even when they could have learned regularities at the perceptual level only. In all four experiments, observers were unaware of the presence of a structure in the stream and were unaware that category was relevant.

The present results extend understanding of statistical-learning mechanisms, demonstrating that they are powerful mechanisms for organizing the input people receive from the world, capable of extracting its underlying structure without conscious intent or awareness. Statistical-learning mechanisms are capable of extracting many different regularities with only minutes of exposure (joint probabilities—Fiser & Aslin, 2002; nonadjacent dependencies—Newport & Aslin, 2004; etc.) and appear to be relatively ubiquitous, occurring in the auditory, tactile, and visual domains, and in infants, adults, and monkeys (Conway & Christiansen, 2005; Hauser, Newport, & Aslin, 2001; Kirkham, Slemmer, & Johnson, 2002). They have been used to explain processes as diverse as how listeners segment a speech stream and how observers create visual objects from low-level perceptual information (Saffran et al., 1996; Turk-Browne et al., 2008).

The present study demonstrates that statistical-learning mechanisms operate at multiple levels of abstraction, including the level of semantic categories. Making use of categories should reduce the amount of information that must be extracted and stored by statistical-learning mechanisms, which face particularly acute computational limitations because of the large number of possible statistics and units over which such statistics could be computed (Brady & Chun, 2007; Turk-Browne et al., 2005).[3] In other words, in the real world, where people need to track relationships among huge numbers of objects and locations, it would make sense to learn these statistics at a level where there is not a huge amount of redundancy; the level of the semantic category is one such level.

The present results also demonstrate a surprising degree of abstraction in statistical learning, as regularities learned with images were transferred to a lexical test. There has been some debate over how closely statistical-learning mechanisms are tied to the perceptual characteristics of the input, particularly in the literature on artificial-grammar learning (e.g., Conway & Christiansen, 2006; Marcus, Vijayan, Bandi Rao, & Vishton,

[3]Even the earliest statistical-learning studies used regularities that presumably operated over categories, although ones with no semantic meaning associated with them. For example, the results of Saffran et al. (1996) are interesting only if infants can learn not only that particular sound waveforms (e.g., the waveform used to represent "pa" and the waveform used for "bi") tend to occur in sequence, but also that the linguistic units these waveforms belong to (the syllable "pa" and the syllable "bi") tend to occur in that sequence.

1999). The transfer observed in the present study was between highly related stimuli (pictures from a category and the word representing that category), and therefore does not bear directly on the question of whether people learn amodal "algebraic rules" that can be applied to any new stimuli (as has been argued by Marcus et al., 1999). However, the abstraction we observed does indicate that statistical learning need not be tied to the perceptual features of the input, contrary to a previous suggestion (Conway & Christiansen, 2006).

Overall, our results suggest that statistical learning is about more than just organizing low-level perceptual inputs: It might be a useful tool for organizing conceptual knowledge as well. The learning demonstrated here may be a useful mechanism for encoding new relationships between visual episodes that are tied together in a particular context. For example, learning new cognitive maps can be thought of as extracting sequences of places that often occur together from a continuous stream of inputs (Tolman, 1948). In general, when observers are confronted with a new environment with a novel distribution of stimuli, statistical learning may provide them with the ability to implicitly learn the new distribution of regularities (or lack of regularities) in a context-dependent fashion, informed by, but not constrained by, the regularities found in other similar environments they have encountered.

More broadly, inferences about causality are often based on the observed covariance of causes (e.g., a block touches an apparatus) and effects (e.g., a noise is produced), and statistical-learning mechanisms are often cited as responsible for learning of such covariance information, particularly in infants (Gopnik & Schultz, 2004). If infants possess categorical statistical-learning mechanisms like those revealed by the present study of adults, they could have available the covariance information necessary to learn not just that a particular object creates a particular effect, but also that all objects of a known category would create the same effect (e.g., intuitive theories: Carey, 1985). At this point, however, the utility of statistical-learning mechanisms for building such high-level knowledge remains an open question. Assessing the scope and depth of knowledge that can be acquired via statistical-learning mechanisms presents a challenge for future work.

***Acknowledgments***—For helpful conversation and comments on earlier drafts, we thank George Alvarez, Michelle Greene, Talia Konkle, Mary Potter, Adena Schachner, and Todd Thompson. T.F.B. was supported by a National Institutes of Health training grant (T32 MH020007). A.O. was supported by a National Science Foundation CAREER award (0546262).

## REFERENCES


Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–609.

Behrmann, M., Geng, J.J., & Baker, C.I. (2005). Acquisition of long-term visual representations: Psychological and neural mechanisms. In B. Uttl, N. Ohta, & C. Macleod (Eds.), *Dynamic cognitive processes: The Fifth Tsukuba International Conference* (pp. 1–26). Tokyo: Springer Verlag.

Brady, T.F., & Chun, M.M. (2007). Spatial constraints on learning in visual search: Modeling contextual cueing. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 798–815.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Chun, M.M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.

Conway, C.M., & Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24–39.

Conway, C.M., & Christiansen, M.H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, *17*, 905–912.

Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*, *73*, B1–B15.

Fiser, J.Z., & Aslin, R.N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 458–467.

Gopnik, A., & Schultz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences*, *8*, 371–377.

Griffiths, T.L., & Tenenbaum, J.B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*, 767–773.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, *16*, 152–160.

Hauser, M.D., Newport, E.L., & Aslin, R.N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*, B53–B64.

Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, 35–42.

Lilliefors, H. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, *62*, 399–402.

Marcus, G.F., Vijayan, S., Bandi Rao, S., & Vishton, P.M. (1999). Rule-learning in seven-month-old infants. *Science*, *283*, 77–80.

Mervis, C.B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89–115.

Newport, E.L., & Aslin, R.N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127–162.

Nissen, M.J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J.K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.

Oliva, A., & Schyns, P.G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*, 176–210.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175.

Pelli, D.G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 509–522.

Pylyshyn, Z.W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.

Tolman, E.C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208.

Trueswell, J.C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*, 566–585.

Turk-Browne, N.B., Isola, P.J., Scholl, B.J., & Treat, T.A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 399–407.

Turk-Browne, N.B., Jungé, J., & Scholl, B.J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*, 552–564.

Turk-Browne, N.B., & Scholl, B.J. (in press). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance*.


(RECEIVED 12/13/07; REVISION ACCEPTED 12/14/07)