The role of Gestalt grouping principles in visual statistical learning

Arit Glicksohn · Asher Cohen

Published online: 13 January 2011
© Psychonomic Society, Inc. 2011

Abstract A major issue in visual scene recognition involves the extraction of recurring chunks from a sequence of complex scenes. Previous studies have suggested that this kind of learning is accomplished according to Bayesian principles that constrain the types of extracted chunks. Here we show that perceptual grouping cues are also incorporated in this Bayesian model, providing additional evidence for the possible span of chunks. **Experiment 1** replicates previous results showing that observers can learn three-element chunks without learning smaller, two-element chunks embedded within them. **Experiment 2** shows that the very same embedded chunks are learned if they are grouped by perceptual cues, suggesting that perceptual grouping cues play an important role in chunk extraction from complex scenes.

Keywords Perceptual organization · Statistical learning · Chunking · Scene perception

Typical visual scenes in daily life are immensely complex, leading to the challenging computational problem of identifying novel scenes (Barlow, 1989). A series of recent studies (e.g., Fiser & Aslin, 2005; Orbán, Fiser, Aslin, & Lengyel, 2008) suggested that one way to overcome this computational difficulty is by capturing regularities that exist between scenes. Specifically, observers capture chunks of scenes that repeat across multiple scenes and use these chunks to identify novel scenes. Chunks are defined as composed of objects that not only co-occur but

are also predictive of each other. Fiser, Aslin, and colleagues suggested that the extraction of chunks is performed primarily by processes that generally follow Bayesian rules (e.g., Orbán et al., 2008). We aim to demonstrate that perceptual grouping contributes to the nature of the chunks that are extracted during this learning.

The research on learning of recurring visual chunks has been done within a broader line of research concerning implicit learning. Numerous studies, using a variety of paradigms, have shown that people can learn regularities in the environment even without intention and without awareness of what was learned (e.g., Chun & Jiang, 1998; Cohen, Ivry, & Keele, 1990). One line of research has focused on statistical learning, which was first demonstrated in the auditory domain (Saffran, Aslin, & Newport, 1996). In typical studies, people passively encounter stimuli whose presentation follows statistical regularities over time and/or space, and then display learning of these regularities. Fiser and Aslin (2001) developed the observational learning (OL) paradigm, a visual version of statistical learning. In the original study, participants viewed a succession of visual scenes. In each scene, six unfamiliar shapes, consisting of three spatially invariant pairs of shapes sampled from an inventory of six such base pairs, were presented in six cells of a 5×5 grid. After this familiarization phase, participants were asked in a subsequent two-alternative forced choice (2AFC) familiarity test to choose between a base pair and two shapes randomly drawn from two different base pairs. Participants selected the base pairs more often than the random pairs, suggesting that learning consists at the very least of computing joint probabilities between elements. In a subsequent experiment, the researchers further showed that observers actually learn conditional probabilities between elements of a chunk.

A. Glicksohn (✉) · A. Cohen
Department of Psychology, The Hebrew University,
Mount Scopus,
Jerusalem 91905, Israel
e-mail: arit.glicksohn@mail.huji.ac.il

The nature of chunk extraction from visual scenes was later investigated by Fiser and Aslin (2005). They used an OL paradigm in which participants viewed six-element scenes composed of two spatially invariant triplets (three elements arranged in a fixed spatial configuration) out of four possible such triplets. Within each scene, the two triplets were spatially contiguous. Triplets have a hierarchical structure, because each triplet embeds within its structure two spatially invariant pairs whose co-occurrences are identical to those of the triplets. In the subsequent 2AFC test phase, participants reliably chose the triplets over identical spatial configurations of three shapes belonging to three different triplets. Surprisingly, though, the participants were not able to discriminate between the embedded pairs and the pairs composed of shapes from different triplets. This puzzling finding suggests that learning cannot be based on conditional probability per se, because such computation should also discriminate embedded pairs from random pairs. To account for these findings, Orban et al. (2008) suggested a Bayesian chunk-learner model, according to which the learner extracts possible chunks during the familiarization scene, choosing an optimal level of higher-order chunks without necessarily extracting their embedded chunks.

Given the complexity of typical visual scenes and the known problem of extracting chunks in complex scenes (e.g., Bellman, 1961), it is likely that learners will use as much information as possible to constrain the extraction of chunks. It has long been noted by Gestalt psychologists that visual scenes are typically organized into perceptual groups defined by principles of continuity, proximity, common fate, and so forth (Palmer, 1999). We suggest that perceptual grouping principles may provide additional evidence for the possible span of chunks that may be extracted from visual scenes.

A number of studies are compatible with this suggestion. There is evidence that Gestalt grouping principles reflect the operation of attentional mechanisms, with grouping principles guiding attention to perceptual groups (Driver & Baylis, 1998). To the extent that attention is involved in chunk extraction, it implies that Gestalt principles affect this learning as well. More specifically, there is evidence that attention is a necessary prerequisite for statistical learning to occur. Turk-Browne, Jungé, and Scholl (2005) found that participants learn statistical regularities only for those objects to which their attention is directed. The interaction of attention and perceptual grouping processes has been further demonstrated in the form of statistical learning. Baker, Olson, and Behrmann (2004) presented on each trial a target and a distractor, and some target–distractor combinations appeared more often than others. In one condition, target and distractor shapes were connected by a line (thereby making them a single group), and in the second condition

the two shapes were not connected. When participants did not know in advance where the target and distractor would be (and presumably directed their attention to the locations of both target and distractor), they learned the frequency of the different target–distractor combinations, regardless of grouping. However, when participants did know in advance the location of the target—and attended the location of the target only—they learned the frequency of combinations only when the two shapes were connected, not when they were separated. This study clearly demonstrated the importance of both attention and perceptual grouping in learning statistical regularities. The generality of this study, however, is limited. This form of statistical learning is fairly simple, consisting of frequencies of co-occurrence of two shapes that may not involve chunking. A compatible pattern of results was found in modified versions of the Saffran et al. (1996) paradigm under dual-task conditions, with learning being compromised when participants were required to perform a secondary task (Toro, Sinnett, & Soto-Faraco, 2005; see also Fernandes, Kolinsky, & Ventura, 2010). Yet all of the above studies did not isolate the effect of perceptual learning on statistical learning.

The main purpose of our study is to show that perceptual grouping plays an important role in the extraction of chunks during statistical learning. To this aim, we employed the OL paradigm used by Fiser and Aslin (2005). In this paradigm, there are four triplets of spatially invariant shapes. In the familiarization phase, participants view a sequence of scenes. In each of these scenes, two of the four triplets are positioned in six cells of a 5×5 grid of cells. **Experiment 1** used the original design of Fiser and Aslin (2005), with black shapes positioned on a white background, and like those researchers, we found learning of the triplets without learning of their embedded pairs. **Experiment 2** used the same design, with one critical difference: Two of the elements within each triplet had one color, and the third element had a different color. Consequently, the shapes of one embedded pair of each triplet always appeared with the same color (creating a perceptual group), and the shapes of the other embedded pair of the same triplet always appeared with two different colors. If perceptual grouping principles help extract chunks, we expected to find learning of the embedded pair that was grouped by the same color.

Experiment 1

Method

Participants A group of 30 participants from the Hebrew University were given either course credit or payment for their participation.

Stimuli Twelve arbitrary shapes adapted from Fiser and Aslin (2005) were presented within a 5×5 grid on a white background. Viewed from a distance of 60 cm, the grid spanned approximately 16.1° ; the height and width of each shape was 2.86° . The 12 shapes were organized into four base triplets, each defined as three unique shapes that always appeared in a fixed spatial relation. A total of 112 unique scenes were created by placing two of the triplets within the grid. The rules of constructing the scenes were as follows: Two of the four triplets were randomly positioned on the grid, with at least two elements of each triplet positioned to either the right or left of two elements from the second triplet. Each triplet and each individual element appeared an equal number of times across the 112 scenes.

Procedure The experiment consisted of two phases. During the familiarization phase, lasting approximately 11 min, participants viewed each of the 112 scenes twice, in random order. Each scene was presented for 2 s, with a 1-s pause between scenes. Participants were requested to pay attention to the sequence of the scenes.

The test phase consisted of a 2AFC test. In the triplet key trials, participants were presented with two successive triplets: a base triplet and a triplet composed of three shapes, each randomly selected from a different base triplet. The two triplets on each trial were presented at the center of a 5×5 grid for 2 s, with a 1-s pause between triplets. Participants then chose whether the first or second triplet seemed more familiar to them. In the pair key trials, two pairs were presented, an embedded pair from one of the base triplets and a second pair composed of two random shapes. A total of 12 key trials were presented (four triplets and eight embedded pairs). In the dummy trials, participants were presented with a sequence of two single elements, pairs, triplets, or quadruples, all composed of random elements. Obviously, in these trials there was no correct answer, and they were inserted to keep the appearance frequencies of individual shapes in each type of test trial equal. There were 29 dummy trials. Test trials were randomized for each participant, and the order of the two test items in a trial was counterbalanced.

At the end of the experiment, participants filled out a questionnaire assessing whether they had acquired explicit knowledge of what they had learned. Specifically, participants were asked whether they had noticed any regularity in the appearance of the stimuli, and to describe any regularity they had detected.

Results and discussion

The main results are presented in Fig. 1. Participants reliably selected the base triplets over the random triplets: The mean

Fig. 1 Results of Experiment 1: Mean percentages correct of triplets and embedded pairs

percentage of correct responses for base triplets was 62%, $t(29) = 2.92$, $p < .01$, Cohen's $d = 0.52$. In contrast, the mean percentage of correct responses for embedded pairs did not exceed chance ($M = 55\%$), $t(29) = 1.2$, $p = .1$. Because on the dummy trials there was no correct answer, participants were expected to select each single shape or group of shapes as often as the other member of the trial. Indeed, in these trials participants showed no significant deviation from chance performance ($p > .05$).

To assess whether participants were aware of the statistical regularities they had learned, we analyzed the self-report awareness questionnaires as follows: If the participant expressed no knowledge of the statistical regularities, s/he received 0 points (no awareness); if the participant answered that s/he had detected some form of regularity (e.g., that pairs of shapes reoccurred throughout the experiment), s/he received 1 point (partial awareness); and if the participant was able to report the full regularity (that four triplets of shapes reoccurred throughout the experiment), s/he received 2 points (awareness). Only 4 of the participants in this study were partly aware of the regularity, receiving a score of 1. We analyzed the experiment once again without these 4 participants and obtained similar results: Participants reliably selected the base triplets over the random triplets, with a mean percentage of correct responses of 58%, $t(25) = 1.873$, $p < .05$, Cohen's $d = 0.38$. Performance for the embedded pairs did not exceed chance, with a mean percentage of 50%, $t(26) = 0.1$.¹

¹ The 4 “aware” participants showed significant learning for both triplets and embedded pairs: for triplets, $M = 93.7\%$, $p < .01$, and for embedded pairs, $M = 87.5\%$, $p < .01$. Both performance rates were significantly higher than the corresponding ones for the unaware group: for triplets, $t(28) = 3.32$, $p < .01$; for embedded pairs, $t(28) = 3.24$, $p < .01$. Thus, being aware of the learned chunks may lead to learning of parts of the chunks as well. However, the sample size of the aware participants was too small for firmer conclusions.

Experiment 2

Method

Participants A group of 30 new participants from the Hebrew University were given either course credit or payment for their participation.

Stimuli and procedure The stimuli were identical to those in [Experiment 1](#), with one notable difference: Six elements were colored red, and six were colored green. For each triplet, two adjacent elements were colored the same color, while the third element was colored the second color (e.g., red–red–green, red–green–green; see [Fig. 2](#)). The assignment of colors to triplets was randomized across participants, but remained constant for each participant throughout the experiment. This procedure resulted in two embedded pair types within each triplet: same-color and different-color pairs. The test phase was also identical to that of [Experiment 1](#), except that the elements were presented in their original red and green colors.

Results and discussion

[Figure 3](#) depicts the results of the main conditions. Participants reliably selected both the base triplets and embedded pairs over random elements: The mean percentage of correct responses for base triplets was 60%, $t(29) = 2.18$, $p < .05$, Cohen's $d = 0.4$, which was identical to the result for the embedded pairs, $t(29) = 3.17$, $p < 0.001$, Cohen's $d = 0.4$. To assess whether the grouping cues were responsible for the learning of the embedded pairs, we compared performance for the same-color and the different-color pairs. When an embedded pair consisted of two elements with the same color, participants reliably chose them over random pairs ($M = 67.5\%$), $t(29) = 3.88$, $p < .001$, Cohen's $d = 0.7$. In sharp contrast, when the embedded pairs consisted of two elements with different colors, performance did not exceed chance ($M = 53\%$), $t(29) = 0.8$, $p = .2$.² Direct comparison

² Each trial of the base pairs in the test phase included a correct pair and an incorrect, random pair. The two elements of the incorrect pair thus had equal probabilities of having the same color or different colors. To rule out the possibility that the results were due to a bias of participants to prefer a pair with identical colors (which would lead to the same pattern of results that we obtained), we did the following two comparisons: First, for the same-color correct pair, we only analyzed trials in which the elements of the alternative, incorrect pair also had identical colors. Second, for the different-color correct pair, we only analyzed trials in which the alternative, incorrect pair included elements with different colors. The pattern of results was qualitatively the same: Participants preferred the correct same-color pair on 69% of the trials, $t(29) = 3.03$, $p < .001$, Cohen's $d = 0.56$; participants preferred the correct different-color pair on 57% of the trials, which did not exceed chance, $t(29) = 1.05$, $p = .15$.

of the two types of pairs also showed better performance for the same-color pairs, $t(29) = -2.48$, $p < .01$. Performance on the dummy trials, as expected, did not deviate from chance.

In this experiment, 6 of the participants revealed some form of awareness of the regularities underlying the experiment, with 5 of them receiving a score of 1 (partial awareness), and 1 receiving a score of 2 (awareness). Analyzing the results without these participants, a similar pattern emerged: Participants reliably selected the embedded pairs, with a mean percentage of 59%, $t(23) = 2.48$, $p = .01$, Cohen's $d = 0.3$. The selection of the triplets approached statistical significance ($M = 57\%$), $t(23) = 1.49$, $p = .07$. Most importantly for our present purposes, the pattern of results for the same-color and different-color pairs remained the same: Participants reliably chose the same-color pairs over random pairs ($M = 65\%$), $t(23) = 2.9$, $p < .01$, Cohen's $d = 0.57$, but could not reliably choose the different-color pairs over random pairs ($M = 54\%$), $t(23) = 0.8$, $p = .2$.³

Our main claim is that pairs are not learned in [Experiment 1](#) whereas pairs within a perceptual group in [Experiment 2](#) are learned. To further verify this claim we contrasted the performance of pairs in [Experiment 2](#) with the performance of pairs in [Experiment 1](#). In accord with our hypothesis, the performance for the same-color pairs in [Experiment 2](#) ($M = 67.5\%$) was significantly higher than the performance for pairs in [Experiment 1](#) ($M = 55\%$), $t(58) = -1.9$, $p < .05$.

General discussion

In this study, we hypothesized that perceptual grouping cues constrain the span of repeating chunks that are extracted from complex scenes. Our participants watched a sequence of different scenes, each including two triplets from an inventory of four. The first experiment replicated previous findings ([Fiser & Aslin, 2005](#)) showing that participants learned the triplets without learning their embedded pairs. [Experiment 2](#) showed that when some of these embedded pairs were grouped by color, these pairs

³ Analyzing the data of the 6 “aware” participants, we found the following results: For the triplets, performance only approached significance, $M = 70.8\%$, $t(5) = 1.74$, $p = .07$, probably due to lack of statistical power. For the embedded pairs, participants performed significantly above chance: $M = 62.5\%$, $t(5) = 2.23$, $p < .05$. Importantly, the dissociation between same-color and different-color pairs remained: Participants reliably selected the same-color pairs over random ones, $M = 75\%$, $t(5) = 3.87$, $p < .01$, while performing at chance level for the different-color pairs, $M = 50\%$. The difference between the two conditions was significant, $t(5) = 3.87$, $p < .01$.

Fig. 2 The base triplets of Experiment 2, and a typical scene

were learned. Importantly, the probabilistic properties of the embedded pairs were identical in Experiments 1 and 2. These findings demonstrate that perceptual grouping cues affect the type of chunks that are learned while watching scenes, with participants tending to learn chunks that belong to the same perceptual group.

Interestingly, participants in Experiment 2 were also able to distinguish between the triplets, which contained both same-color and different-color pairs. It is not clear whether participants were able to learn the triplets, as such, in Experiment 2 or were able to use the same-color pair embedded in it as a cue for recognition of the larger triplet. Further research is required to uncover the reason for the ability of participants to recognize triplets in the setup of Experiment 2.

Parallel findings for the role of perceptual grouping cues on statistical learning have also been observed in the temporal domain. Fiser, Scholl, and Aslin (2007) demonstrated the importance of visual grouping cues in statistical learning of transitional probabilities in the temporal domain. A stream of pictures consisting of a single shape were presented and at some point—via an occlusion manipulation—split into two simultaneous streams. One of the two emerging streams was linked to the original stream by perceptual grouping cues, and the other was

unconnected. Participants showed better learning of the grouped stream over the ungrouped stream.

Perceptual grouping principles have also been found to govern auditory statistical learning. While it has been demonstrated that participants are able to learn statistical regularities occurring between adjacent auditory elements, they do not readily learn statistical regularities occurring between nonadjacent elements (Newport & Aslin, 2004). However, learning did occur when the nonadjacent elements were either two consonants with an intervening random vowel or two vowels with as intervening random consonant. Creel, Newport, and Aslin (2004) argued that this learning selectivity of nonadjacent dependencies is an instance of perceptual similarity—with speech elements from the same category, such as consonants and vowels, being similar to one another. Using tones instead of speech stimuli, the researchers found that participants learned statistical regularities occurring between nonadjacent tones if they were perceptually similar to each other—that is, sharing the same frequency or timbre—and perceptually dissimilar from the intervening tone. For example, learning occurred when the two nonadjacent tones shared the same pitch (e.g., high) and the intervening tone had a different pitch (e.g., low).

Our experiments demonstrate that participants use perceptual grouping cues even when they are unaware of the presence of chunks. However, we must emphasize two caveats in conjunction with this observation. First, the questionnaire that we used to assess awareness was quite simple and explored participants' awareness informally. Thus, we cannot rule out the possibility that participants had some sort of awareness of this learning that was not reflected in the self-report questionnaire. The results of the questionnaire do suggest that they did not employ a conscious strategy of learning chunks within groups, since most of the participants were not even able to report on the existence of such chunks. Second, and more importantly, our primary aim has been to understand the constraints that guide chunk formation from complex scenes, and it is of secondary importance whether this learning occurs with or without awareness.

Fig. 3 Results of Experiment 2. Left: Mean percentages correct of triplets and embedded pairs. Right: Mean percentages correct of same-color and different-color pairs. * $p < 0.01$

Author Note This work was funded by a grant from the Israel Science Foundation to A.C.

References

- Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, 15, 460–466. doi:10.1111/j.0956-7976.2004.00702.x
- Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, 1, 295–311. doi:10.1162/neco.1989.1.3.295
- Bellman, R. (1961). *Adaptive control processes: A guided tour*. Princeton, NJ: Princeton University Press.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28–71. doi:10.1006/cogp.1998.0681
- Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 17–30. doi:10.1037/0278-7393.16.1.17
- Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1119–1130. doi:10.1037/0278-7393.30.5.1119
- Driver, J., & Baylis, G. (1998). Attention and visual object segmentation. In R. Parasuraman (Ed.), *The attentive brain* (pp. 299–325). Cambridge, MA: MIT Press.
- Fernandes, T., Kolinsky, R., & Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics*, 72, 1522–1532. doi:10.3758/APP.72.6.1522
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504. doi:10.1111/1467-9280.00392
- Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, 134, 521–537. doi:10.1037/0096-3445.134.4.521
- Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychonomic Bulletin & Review*, 14, 173–178.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning on non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162. doi:10.1016/S0010-0285(03)00128-2
- 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:10.1073/pnas.0708424105
- Palmer, S. (1999). *Vision science: Photons to phenomenology*. 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. doi:10.1126/science.274.5294.1926
- Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97, 25–34. doi:10.1016/j.cognition.2005.01.006
- 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:10.1037/0096-3445.134.4.552