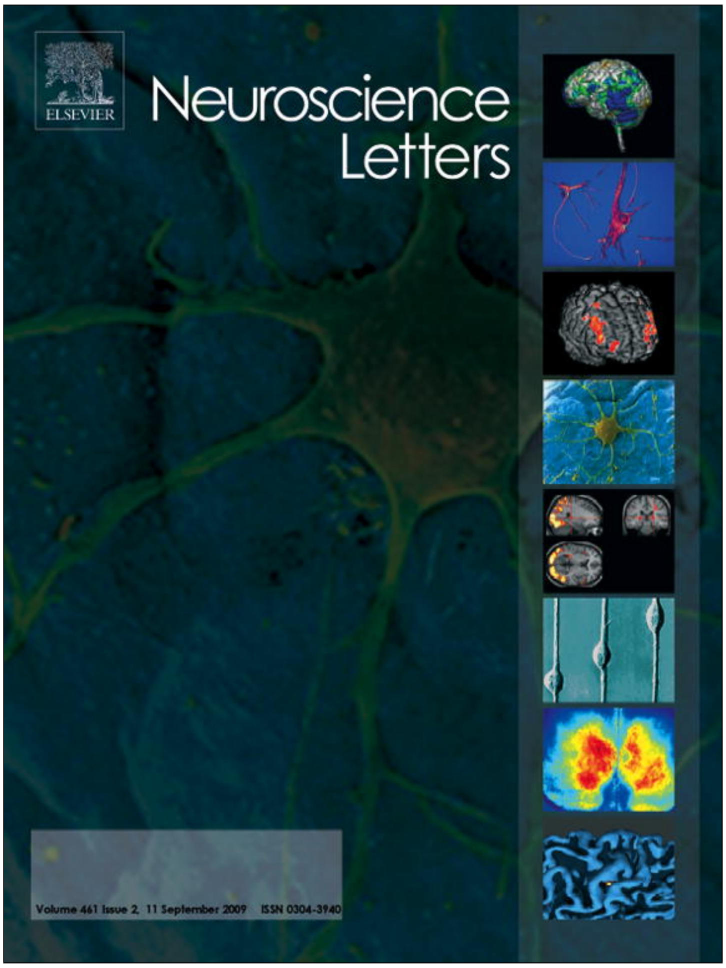
ELSEVIER
Neuroscience Letters
Volume 461 Issue 2, 11 September 2009 ISSN 0304-3940

Neuroscience Letters 461 (2009) 145–149

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Testing assumptions of statistical learning: Is it long-term and implicit?

Robyn Kim [a], Aaron Seitz [b], Heather Feenstra [a], Ladan Shams [a,*]

[a] *Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095, USA*
[b] *Department of Psychology, University of California, Riverside, USA*

ARTICLE INFO

Article history:
Received 6 January 2009
Received in revised form 5 May 2009
Accepted 4 June 2009

Keywords:
Statistical learning
Implicit
Long-term

ABSTRACT

Statistical learning has been studied as a mechanism by which people automatically and implicitly learn patterns in the environment. Here, we sought to examine general assumptions about statistical learning, including whether the learning is long-term, and whether it can occur implicitly. We exposed participants to a stream of stimuli, then tested them immediately after, or 24 h after, exposure, with separate tests meant to measure implicit and explicit knowledge. To measure implicit learning, we analyzed reaction times during a rapid serial visual presentation detection task; for explicit learning, we used a matching questionnaire. Subjects' reaction time performance indicated that they did implicitly learn the exposed sequences, and furthermore, this learning was unrelated to explicit learning. These learning effects were observed both immediately after exposure and after a 24-h delay. These experiments offer concrete evidence that statistical learning is long-term and that the learning involves implicit learning mechanisms.

Statistical learning has been studied as a mechanism by which people automatically discover patterns in the environment. In typical statistical learning studies, people learn arbitrary associations between stimuli based on the statistics of inter-stimulus contingencies, without necessarily intention or effort [11,19]. In this paradigm, subjects are passively exposed to stimulus configurations that have been organized into regular patterns. After only a brief period of passive exposure, observers can correctly identify which patterns were repeatedly presented during the exposure phase, despite reporting no conscious awareness of this knowledge. This type of learning has been demonstrated with a wide range of stimuli (e.g., visual spatial arrangements [11], visual temporal sequences [10], auditory tone sequences [20], audiovisual sequences [23], and tactile sequences [5], in both adults and infants [12,19]. In addition, people can learn multiple patterns (from different input sources) in parallel [6,23].

The pervasiveness of statistical learning phenomena is consistent with the proposal that it is a domain-general learning mechanism that may underlie our ability to automatically adapt to our environment without intention or awareness (note that some suggest that attention to the stimuli may be required for successful statistical learning [2,26,28]). It has been suggested that this type of learning may contribute to language learning [19,21], acquisition of scene and object representations [11], and the hierarchical coding of objects (i.e. chunking) [9,14]. It has also been pointed out that statistical learning shares properties with the artificial grammar learning (AGL) introduced by Reber [18], the most obvious similarity being their implicit nature [6,16]. Indeed, Reber suggested that implicit learning was likely a general mechanism and thus should emerge in many different contexts and with different types of stimuli [17].

If statistical learning is indeed tapping into adaptive mechanisms to help us adjust to our environment, it should persist across long delays. However, to date, statistical learning has only been tested closely after exposure. Thus far, long-term implicit learning has been well documented in the area of perceptual learning [3,7,22,24,29]. More relevantly, AGL has been demonstrated to persist across 2 years [1], as measured by performance on a yes/no grammaticality judgment task, and over 14 days on a visuo-motor serial reaction time test [27]. However, long-term visual statistical learning has not been reported.

Statistical learning has largely been assumed to be an implicit mechanism in that the learning can occur without supervision (instructions or feedback) and without conscious awareness of such learning. Notably, statistical learning studies typically use a two-alternative forced choice test in which subjects report which stimuli are more "familiar" (e.g. [11,20,23]) and argue based upon debriefing reports that familiarity reports reveal implicit knowledge. However, inasmuch as it requires *recognition* of previously exposed patterns, the familiarity test may be more akin to an explicit learning test (e.g., recognition in the memory literature); and in fact, some statistical learning researchers consider familiarity a measure of explicit rather than implicit learning [2]. Furthermore, as Turk-Browne et al. [28] comment, the familiarity measure is "an odd sort of dependent measure to use, because it essentially asks

* Corresponding author. Tel.: +1 310 206 3630; fax: +1 310 267 2141.
E-mail address: ladan@psych.ucla.edu (L. Shams).

0304-3940/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2009.06.030

observers to make an explicit judgment about implicitly learned relationships."

The evidence for lack of conscious knowledge of learning in the face of accurate familiarity judgments typically comes from informal verbal reports [5,9,11]. However, aside from such subjective reports, explicit knowledge has not been quantitatively measured and compared with implicit knowledge. Thus, results could be tainted by a degree of (ineffable) explicit knowledge. For the long-term artificial grammar learning studies (discussed earlier), subjects demonstrated explicit as well as implicit learning [1,27]; interestingly, both studies also reported a greater degree of decay for explicit compared to implicit knowledge, supporting a dissociation between implicit and explicit learning.

Statistical learning has proven a valuable research paradigm, but as described above, some commonly held but critical assumptions need to be more concretely addressed. The first goal of this paper is to investigate long-term statistical learning. In Experiment 1, we test people immediately after exposure; in Experiment 2, we test a separate group of people approximately 24 h after exposure to test for long-term statistical learning. Secondly, we verify previous reports of implicit learning independent of explicit knowledge. To measure implicit learning, we use an alternative test of learning to familiarity, a rapid serial visual presentation (RSVP) reaction time test [28]. To measure explicit learning, we use an item matching test. Also, we examine the possibility that consolidation or forgetting has differing effects on implicit vs. explicit knowledge [8,27] by comparing long-term with short-term evaluations.

Experiments 1 and 2 were implemented identically, except that all tests for Experiment 1 were conducted immediately after exposure, while tests for Experiment 2 were conducted the following day.

Twenty-four undergraduates, aged 18–35 (12 in each Experiment) took part in the study, and received course credit for their participation. All participants were naïve to the purpose of the study and participated in only one experimental condition. The experiments were conducted in accordance with the IRB approved by the Committee on Human Research of the University of California and the Declaration of Helsinki.

Visual stimuli consisted of 12 arbitrary black and white figures (Fig. 1A) adapted from Fiser and Aslin [11]. Images were sized to 128 × 128 pixels and subtended 3 squared degrees of visual angle, and were each presented for 200 ms. Visual stimuli were presented on a 19″ Cathode Ray Tube monitor with resolution of 1024 × 768 pixels and refresh rate of 75 Hz.

Stimuli were presented one at a time during both parts of the experiment ('Exposure' and 'Testing,' see below). Each figure was uniquely assigned to one of four 'triplets' (a sequence of three visual stimuli, see Fig. 1 for details) [10]. The stimulus make-up for these triplets was randomly assigned for each subject.

The experiments took place in a dimly lit room. Participants sat 57 cm away from the monitor with their heads stabilized using a chin-rest. Stimuli were presented using custom software written with use of the Psychophysics Toolbox [4,15] for Matlab™ (Natick, MA) on a Macintosh G4 computer.

Exposure: During the first phase of the experiment (~5 min long), participants were presented with a rapid serial presentation of a continuous stream of four visual triplet sequences presented 100 times each, in a pseudorandom order with the constraint that a given triplet could not appear twice in immediate succession. Triplets could not be segmented based on any temporal or spatial cues as the ISI of the stimuli was fixed within and across triplets (30 ms), and each stimulus appeared at the same central location on the screen. Stimuli were presented for 200 ms. The same ISI was used both for Exposure and Testing. Subjects were asked to carefully watch and listen to the stimuli at the time of Exposure, and participants were not informed about the subsequent Testing phase.

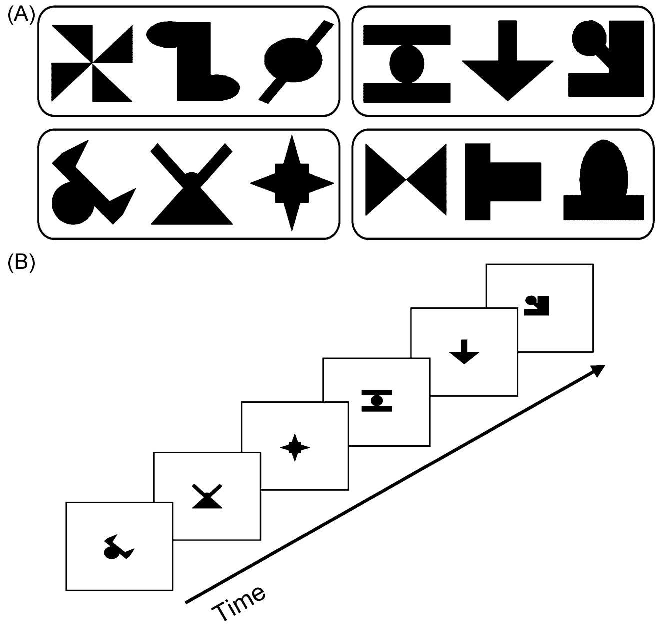

Fig. 1. Stimuli. (A) The 12 shapes used in the experiments are shown; sample triplets are encircled. The actual subcomponents of the triplets were randomly assigned for each participant. (B) Sample exposure stream using triplets from A.

Testing: One group (Experiment 1) was tested immediately after Exposure, while another group (Experiment 2) was tested approximately 24 h after Exposure. Participants performed two tests. The first consisted of a rapid serial visual presentation (RSVP) paradigm, in which the task was to press a key as soon as they detect a visual target within a stream of stimuli. The stream of stimuli consisted of one presentation each of the triplets they experienced during Exposure, in random order. Each stimulus was presented for the same duration and with the same ISI as in the Exposure period. The trials proceeded as follows: a target stimulus was presented at the top of the screen. Once the subject was ready to start the test trial, he/she pressed a key to begin the test stream and the target disappeared. Then, each triplet was presented once in a pseudorandom order, with the constraint that the triplet containing the target stimulus could not occur first or last in the stream. When the subject saw the target stimulus, he/she pressed the space bar as quickly as possible, and the reaction time was recorded. Each of the 12 visual stimuli was presented as a target eight times, for a total of 96 trials. This test lasted about 10 min. If subjects did learn the sequences during Exposure, reaction times to items later in each sequence (i.e., the second and third items in each triplet) should be shorter than they were to the first items in each triplet, since the later items got primed by the first items in each sequence, whereas the first items appeared unpredictably within the stream. If, however, subjects did not learn the triplets, reaction times should not differ among the item positions.

The second test was a matching questionnaire intended to probe whether participants explicitly learned which stimuli were grouped together. On each trial, one visual stimulus was presented and the observer was asked on the monitor, "Which two images are related to this one?" Underneath the test stimulus, the remaining eleven visual stimuli were displayed in a random order, labeled a–k. Subjects were instructed to type in the letters corresponding to the two items they believed were associated with the test stimulus based on what they observed during the first part of the experiment. We used this test, a type of recognition test rather than a pure recall

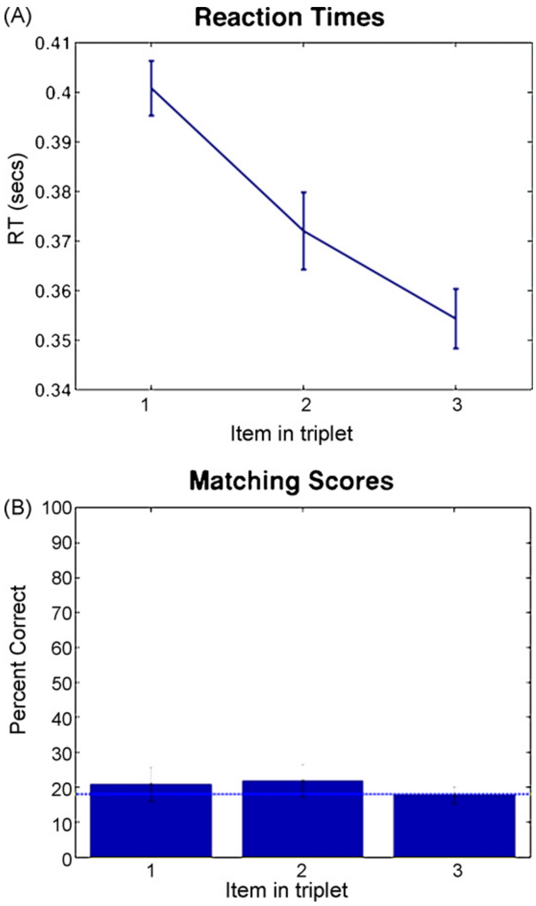

Fig. 2. Experiment 1, short-term learning of visual triplet sequences. (A) Mean reaction times to the items in different positions in the exposed sequences. Error bars indicate between subject standard error of the means. (B) Mean matching scores to the items in different positions in the exposed sequences. Error bars indicate between subject standard error of the means. Dashed line indicates chance level.

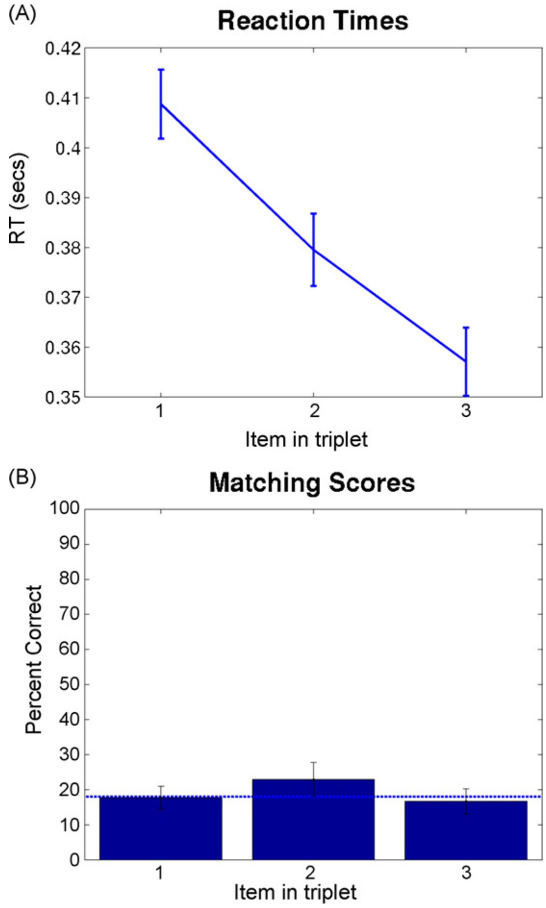

Fig. 3. Experiment 2, long-term learning of visual triplet sequences. (A) Mean reaction times to the items in different positions in the exposed sequences. Error bars indicate between subject standard error of the means. (B) Mean matching scores to the items in different positions in the exposed sequences. Error bars indicate between subject standard error of the means. Dashed line indicates chance level.

test, so that we could more sensitively measure explicit learning. It is possible that subjects do not have enough memory of the stimuli to spontaneously recall each item, yet can report which items are associated when all the choices are provided. Also, this way we can compare performance with a baseline chance performance (2/11).

Finally, participants were verbally asked how confident they felt about their responses on both tests, what strategies they used for each, and if they noticed any patterns among the stimuli during exposure.

Fig. 2A plots average reaction times for items in the first, second, and third positions in the triplets. A repeated measures ANOVA revealed a significant difference among reaction times ($F(2,22) = 7.91$, $p < 0.01$). Planned comparison paired t-tests indicated that reaction times for the second and third items in the triplets were significantly faster than for the first items (item 1 vs. item 2: $t(11) = 2.28$, $p < 0.05$; item 1 vs. item 3: $t(11) = 5.23$, $p < 0.01$), but not for second vs. third items ($t(11) = 1.32$, $p = 0.21$).

Since it is possible that some learning could occur within the test session itself, we examined whether performance differed between the first and second half of the testing sessions. Absolute reaction times did not change ($F(2,22) = 1.99$, $p = 0.19$) from first to second half of testing. More importantly, the sequence learning effects, or difference in reaction times (ΔRT) to different item positions (e.g., ΔRT_{12} = RT for item 1 – RT for item 2), did not significantly differ between the first and second halves of the testing session (paired t-tests comparing reaction time differences ΔRT_{12}, ΔRT_{13}, and ΔRT_{23} between first and second halves: $t(11) = .21, .41, .35$; $p = .84, .69, .73$, respectively), suggesting that the reaction time effect did not result from learning within the test session.

As Fig. 2B shows, on the matching questionnaire, subjects performed at chance for all three item positions, with no significant difference between item positions ($F(2,22) = .16$, $p = .86$). Across all items, subjects scored an average of 21% (SE = .03; t-test vs. chance (2/11 or 18%): $t(35) = 1.14$, $p = .28$). Thus, we found no evidence of explicit learning, even with a recognition test.

Although on average there was no evidence of explicit learning, it is possible that the amount of explicit knowledge could affect reaction time results, i.e., the implicit and explicit measures used here might not be independent of each other. However, when comparing reaction time differences between different items with average matching scores, i.e., (% correct for 1st item +% correct for 2nd item)/2), performance was not correlated across all conditions ($r(10) = .01$, $p = .99$), nor for each condition separately ($r(10) = -.31, .02, .30$; $p = .33, .94, .34$, for ΔRT_{12}, ΔRT_{13}, and ΔRT_{23}, respectively).

Fig. 3A plots average reaction times for items in the first, second, and third positions in the triplets. A repeated measures one-way ANOVA revealed a significant difference among the three conditions ($F(2,22) = 8.37$, $p < 0.01$). Planned comparison paired t-tests indicated that reaction times for the second and third items in the triplets were significantly shorter than for the first items (item 1 vs. item 2: $t(11) = 2.25$, $p < 0.05$; item 1 vs. item 3: $t(11) = 4.23$, $p < 0.01$), but not for the second compared to the third items, though they approached significance ($t(11) = 1.75$, $p = 0.11$).

When comparing between the first and second halves of the testing session we found no change in absolute reaction times ($F(2,22) = 0.90$, $p = .36$). More importantly, the statistical learning effects ΔRT_{12} and ΔRT_{23} did not significantly differ between the first and second halves of the testing session (paired t-tests $t(11) = .02$, 1.16; $p = .98$, .27, respectively), suggesting that the reaction time effect did not result from learning within the test session. However, there was a change for items 1 vs. 3 from the first to the second half of the test session (from $\Delta RT_{13} = .0371$ s to $\Delta RT_{13} = .0645$ s, $t(11) = 2.48$, $p < 0.05$), thus there may have been within-test learning of that relationship.

As Fig. 3B shows, on the matching questionnaire, subjects performed at chance for all three item positions, with no significant difference between item positions (RM ANOVA: $F(2,22) = 1.12$, $p = .34$). Across all items, subjects scored an average of 19% (SE = 3%; t-test vs. chance (2/11 or 18%): $t(35) = .41$, $p = .69$).

Performance was not correlated between reaction time differences and matching test scores across all conditions ($r(10) = -.09$, $p = .61$), nor for each condition separately ($r(10) = -.2$, $-.14$, .07; $p = .52$, .66, .83, for ΔRT_{12}, ΔRT_{13}, and ΔRT_{23}, respectively).

Reaction time patterns for short-term and long-term visual statistical learning were very similar. A two-way repeated measures ANOVA on reaction time data with factors Experiment (1 or 2) and item (1, 2, or 3) yielded no significant interaction, and no significant main effect of experiment ($F(1,22) = .1$, $p = .76$), only a significant main effect of item ($F(2,22) = 18.8$, $p < .01$). For the matching task as well, performance was not significantly different between experiments ($F(1,22) = .51$, $p = .49$), nor between items ($F(2,22) = .89$, $p = .44$).

Finally, although the difference between second and third items does not reach significance in each experiment separately, there is a difference if we combine the data from both experiments ($t(23) = 2.217$, $p < 0.05$).

Here we found for the first time that visual statistical learning can be long-term. The amount of implicit learning immediately after exposure was similar to that after a 24-h delay, suggesting that implicit statistical learning can persist without deterioration, at least over 1 day, even with only 5 min of exposure. For learning regular patterns in the environment it is desirable for the learning effect to withstand unrelated intervening stimuli and activities. This robustness over time has been demonstrated in artificial grammar learning [1,27] and now, in visual statistical learning.

We observed implicit statistical learning as measured by reaction time differences between the first and second/third items in the sequences, which is consistent with previous findings [28]. Although the difference between second and third items does not reach significance in each experiment separately, there is a difference if we combine the data from both experiments, consistent with Turk-Browne et al. [28]. This reaction time measure is a good alternative measure of implicit learning to the familiarity test, and is one that does not require subjects to make explicit decisions about recognition.

This implicit learning is obtained in the absence of explicit learning, as measured by our matching questionnaire; furthermore, the observed degree of implicit and explicit learning are not correlated, suggesting that the implicit and explicit learning (if any) are not inter-dependent. Previous statistical learning studies have relied on post-experiment debriefing reports for evidence of explicit learning; however, introspective judgments are an unreliable gauge of learning. It is known that people have poor metacognition in assessing their learning and memory [13,30]. For example, people often confidently judge massed practice as more effective learning regimes than spaced practice, though their test results indicate otherwise [25]. Confirming this lack of insight, a few subjects in our study reported some confidence or moderate knowledge of patterns in the stimuli in verbal post-test interviews, but those assessments were not correlated with their performance on the matching test. This highlights the fact that a structured test of explicit knowledge can provide a more useful metric of the contribution of explicit learning than verbal reports.

These experiments demonstrate that the learned visual associations can persist over 1 day, and offer concrete evidence supporting the assumption that statistical learning involves implicit learning mechanisms. However, our results also highlight the fact that the assumption of implicitness cannot be taken for granted, and should be empirically tested.

Acknowledgments

We thank Jozsef Fiser for kindly providing us with the stimuli used in their previous studies. This research was supported by UCLA Faculty Grants Program. R.K. was supported by a predoctoral NSF fellowship.

References

[1] R. Allen, A.S. Reber, Very long term memory for tacit knowledge, Cognition 8 (1980) 175–185.
[2] C.I. Baker, C.R. Olson, M. Behrmann, Role of attention and perceptual grouping in visual statistical learning, Psychol. Sci. 15 (2004) 460–466.
[3] K. Ball, R. Sekuler, Direction-specific improvement in motion discrimination, Vision Res. 27 (1987) 953–965.
[4] D.H. Brainard, The Psychophysics Toolbox, Spat. Vis. 10 (1997) 433–436.
[5] C.M. Conway, M.H. Christiansen, Modality-constrained statistical learning of tactile, visual, and auditory sequences, J. Exp. Psychol. Learn. Mem. Cogn. 31 (2005) 24–39.
[6] C.M. Conway, M.H. Christiansen, Statistical learning within and between modalities: pitting abstract against stimulus-specific representations, Psychol. Sci. 17 (2006) 905–912.
[7] R.E. Crist, W. Li, C.D. Gilbert, Learning to see: experience and attention in primary visual cortex, Nat. Neurosci. 4 (2001) 519–525.
[8] T.M. Ellmore, K. Stouffer, L. Nadel, Divergence of explicit and implicit processing speed during associative memory retrieval, Brain Res. 1229 (2008) 155–166.
[9] J. Fiser, R.N. Aslin, Encoding multielement scenes: statistical learning of visual feature hierarchies, J. Exp. Psychol. Gen. 134 (2005) 521–537.
[10] J. Fiser, R.N. Aslin, Statistical learning of higher-order temporal structure from visual shape sequences, J. Exp. Psychol. Learn. Mem. Cogn. 28 (2002) 458–467.
[11] J. Fiser, R.N. Aslin, Unsupervised statistical learning of higher-order spatial structures from visual scenes, Psychol. Sci. 12 (2001) 499–504.
[12] N.Z. Kirkham, J.A. Slemmer, S.P. Johnson, Visual statistical learning in infancy: evidence for a domain general learning mechanism, Cognition 83 (2002) B35–B42.
[13] N. Kornell, R.A. Bjork, Learning concepts and categories: is spacing the "enemy of induction"? Psychol. Sci. 19 (2008) 585–592.
[14] G. Orban, J. Fiser, R.N. Aslin, M. Lengyel, Bayesian learning of visual chunks by human observers, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 2745–2750.
[15] D.G. Pelli, The VideoToolbox software for visual psychophysics: transforming numbers into movies, Spat. Vis. 10 (1997) 437–442.
[16] P. Perruchet, S. Pacton, Implicit learning and statistical learning: one phenomenon, two approaches, Trends Cogn. Sci. 10 (2006) 233–238.
[17] A.S. Reber, Implicit learning and tacit knowledge, J. Exp. Psychol.: Gen. 118 (1989) 219–235.
[18] A.S. Reber, Implicit learning of artificial grammars, J. Verbal Learn. Verbal Behav. 6 (1967) 855–863.
[19] J.R. Saffran, R.N. Aslin, E.L. Newport, Statistical learning by 8-month-old infants, Science 274 (1996) 1926–1928.
[20] J.R. Saffran, E.K. Johnson, R.N. Aslin, E.L. Newport, Statistical learning of tone sequences by human infants and adults, Cognition 70 (1999) 27–52.
[21] J.R. Saffran, E.L. Newport, R.N. Aslin, R.A. Tunick, S. Barrueco, Incidental language learning: listening (and learning) out of the corner of your ear, Psychol. Sci. 8 (1997) 5.
[22] D. Sagi, D. Tanne, Perceptual learning: learning to see, Curr. Opin. Neurobiol. 4 (1994) 195–199.
[23] A.R. Seitz, R. Kim, V. van Wassenhove, L. Shams, Simultaneous and independent acquisition of multisensory and unisensory associations, Perception 36 (2007) 1445–1453.
[24] A.R. Seitz, J.E. Nanez Sr., S.R. Holloway, T. Watanabe, Perceptual learning of motion leads to faster flicker perception, PLoS One 1 (2006) e28.
[25] D.A. Simon, R.A. Bjork, Metacognition in motor learning, J. Exp. Psychol. Learn. Mem. Cogn. 27 (2001) 907–912.
[26] J.M. Toro, S. Sinnett, S. Soto-Faraco, Speech segmentation by statistical learning depends on attention, Cognition 97 (2005) B25–B34.

[27] R.J. Tunney, Implicit and explicit knowledge decay at different rates: a dissociation between priming and recognition in artificial grammar learning, Exp. Psychol. 50 (2003) 124–130.
[28] N.B. Turk-Browne, J. Junge, B.J. Scholl, The automaticity of visual statistical learning, J. Exp. Psychol. Gen. 134 (2005) 552–564.
[29] T. Watanabe, J.E. Nanez Sr., S. Koyama, I. Mukai, J. Liederman, Y. Sasaki, Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task, Nat. Neurosci. 5 (2002) 1003–1009.
[30] E.B. Zechmeister, J.J. Shaughnessy, When you know that you know and when you think that you know but you don't, Bull. Psychon. Soc. 15 (1980) 41–44.