The role of cross-modal associations in statistical learning

Arit Glicksohn · Asher Cohen

Published online: 29 May 2013
© Psychonomic Society, Inc. 2013

Abstract Our environment is richly structured, with objects producing correlated information within and across sensory modalities. A prominent challenge faced by our perceptual system is to learn such regularities. Here, we examined statistical learning and addressed learners' ability to track transitional probabilities between elements in the auditory and visual modalities. Specifically, we investigated whether cross-modal information affects statistical learning within a single modality. Participants were familiarized with a statistically structured modality (e.g., either audition or vision) accompanied by different types of cues in a second modality (e.g., vision or audition). The results revealed that statistical learning within either modality is affected by cross-modal information, with learning being enhanced or reduced according to the type of cue provided in the second modality.

Keywords Statistical learning · Cross-modal associations · Grouping

Introduction

Our environment is richly structured, with objects producing correlated information across different sensory modalities. For example, a roaring bear will be perceived as “big” by both our auditory and visual modalities. A prominent challenge faced by our perceptual system is to learn such regularities. Indeed, numerous studies have revealed both infants' and adults' exceptional ability to learn various types of regularities (e.g., Bahrick & Lickliter, 2000; Cohen, Ivry, & Keele, 1990; Ernst, 2007; Reber, 1967). During our initial development, cross-modal information benefits certain forms of learning (Gibson, 1969), and this cross-modal advantage is thought to be preserved throughout our lives (Kim, Seitz, & Shams, 2008).

Here we focus on a paradigm termed *statistical learning*, in which learners track transitional probabilities (TPs) between elements in the auditory and visual modalities (Fiser & Aslin, 2002a, 2002b; Saffran, Aslin, & Newport, 1996). For example, after listening to a concatenated speech stream, with certain syllables forming reoccurring “words,” infants are able to discriminate between those “words” and the sequences of syllables that never appeared successively (Saffran et al., 1996). Similar findings were obtained in the visual modality (Fiser & Aslin, 2002a, 2002b). However, most research to date has focused on statistical learning within single modalities (hence, “unimodal” statistical learning). Here, we will ask whether cross-modal information affects statistical learning, and whether learning is enhanced or reduced according to the type of association between modalities.

Cross-modal information has been found to play a prominent role in different perceptual and cognitive processes. For example, participants are faster at localizing auditory–visual targets than at localizing either auditory or visual targets alone (Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994). Information presented from two modalities is also remembered better: Following a period of encoding in which participants were presented with either picture–sound pairs or pictures alone, they subsequently showed better retrieval of the pictures that had formerly been paired with sounds, as compared to the pictures that had not (Lehmann & Murray, 2005).

A cross-modal benefit is observed for various types of learning, as well, for both infants and adults. Gibson (1969) long ago suggested that during early development redundant information across two or more modalities (properties that are not specific to one modality, but are redundant across multiple modalities, such as size or rhythm) is “foregrounded” from the environment by having attentional precedence over modality-specific information. In support of this claim, it was found that five-month-old infants learned an abstract rule (e.g., ABA) or rhythm inferred from both auditory and visual stimuli, but they did not learn the same underlying structure when it was presented through stimuli from one modality alone (Bahrick & Lickliter, 2000; Frank, Slemmer, Marcus, & Johnson,

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

2009). Thus, for young infants, learning benefits from correlated cross-modal stimuli.

For adults, training with cross-modal information benefits learning of modality-specific information, as is evident in perceptual-learning tasks (Ernst, 2007; Kim et al., 2008; Shams, Wozny, Kim, & Seitz, 2011). Kim et al. found that participants performed better on a visual motion detection task following an audiovisual training protocol, in which a visual motion display was accompanied by a tone sounding as if it was moving in the same direction as the display, as compared with a protocol containing a tone sounding as if moving in the opposite direction, or a visual-only training protocol. The researchers suggest that compatible audiovisual information during training promotes better encoding, and/or better retention of the visual information, leading to better performance even in the subsequent absence of the auditory stimulus.

In the present study, we explored whether cross-modal information affects adults' learning of regularities across time, specifically focusing on statistical learning (Fiser & Aslin, 2002a, 2002b; Saffran et al., 1996). In their seminal study, Saffran et al. demonstrated that infants are able to learn TPs between syllables. Eight-month-old infants were exposed to a continuous stream of syllables for a duration of 2 min. The stream consisted of 12 syllables divided into triplets, with the three syllables of each triplet always appearing one after the other. Thus, the TPs between syllables within a triplet equaled 1. The order of the triplets was randomized. In a subsequent test phase, infants were able to discriminate between triplets and foil triplets—that is, three syllables that never appeared successively (i.e., a TP of 0 between syllables). Saffran et al. suggested that such a capacity for auditory statistical learning helps infants segment a speech stream into smaller “word-like” units, thus providing a bootstrapping mechanism for word learning.

Similar findings were obtained in the visual modality. Fiser and Aslin (2002a, 2002b) found that following a familiarization phase in which novel shapes were combined into visual triplets, both infants (2002b) and adults (2002a) learned the visual triplets. These similar findings across different modalities are compatible with a claim that statistical learning is an amodal learning mechanism (Fiser, 2009; Kirkham, Slemmer, & Johnson, 2002) giving rise to higher-order primitives in both audition and vision (e.g., words and visual objects; Turk-Browne, Scholl, Chun, & Johnson, 2009). Statistical learning is thought to scaffold more complex types of regularity learning, such as those prevalent in natural languages (Aslin & Newport, 2012; Frost, 2012). In line with this claim, statistical learning in the auditory as well as the visual modality correlates with language abilities (Arciuli & Simpson, 2012; Frost, Siegelman, Narkiss, & Afek, *in press*; Kidd, 2012; Mirman, Magnuson, Graf Estes, & Dixon, 2008).

Several studies have explored how unimodal statistical learning occurs when the input consists of information from

multiple modalities. In these studies, one modality stream consists of triplets, whilst the other modality stream consists of elements that are either correlated or uncorrelated with the triplets. For example, Thiessen (2010) found that visual information correlated with auditory triplets enhanced auditory statistical learning. Participants were presented with a stream of reoccurring auditory triplets, with each auditory triplet being accompanied by a single visual shape. For the experimental group, each shape was uniquely associated with a specific auditory triplet, whilst for the control group the shapes appeared randomly. Participants showed better auditory statistical learning when the shapes were associated with the auditory triplets than when they were random. Similar results were obtained by Cunillera, Laine, Cámara, and Rodríguez-Fornells (2010), using a within-subjects design in which half of the auditory triplets were accompanied by uniquely associated shapes whilst half were accompanied by random shapes. Here, participants learned only those auditory triplets that were associated with shapes, and they did not learn those presented with random shapes. Together, these findings suggest that auditory statistical learning is enhanced by the presence of correlated visual information, and is disrupted, or even nonexistent, in the presence of random visual information.

Auditory statistical learning may also be enhanced when visual information marks the auditory-triplet boundaries. Cunillera, Cámara, Laine, and Rodríguez-Fornells (2010) presented participants with an auditory stream accompanied by random shapes, with the shapes appearing at the onset of either the first, second, or third syllable of each auditory triplet. Learning was enhanced when the shape coincided with the first or third syllable of the auditory triplet (but not with the second syllable), as compared with when the auditory stream was presented alone. The researchers proposed that when the shapes coincide with changes in TPs between syllables, attention is drawn to the boundaries of the auditory triplets (with the first syllable denoting the beginning and the third syllable the end of a triplet).

The importance of the timing of cross-modal information was further evident in another study exploring visual statistical learning (Neath, Guérard, Jalbert, Bireta, & Surprenant, 2010). Participants presented with reoccurring visual triplets accompanied by a stream of random syllables that were not synchronized with the presentation of the shapes showed reduced visual learning relative to when the visual stream was presented alone. The researchers proposed that serial information regarding the auditory stream interfered with that of the visual stream, resulting in impaired learning of the visual triplets.

What may be the importance of cross-modal information for statistical learning? It is well documented that processing within each modality may be limited by a host of perceptual and attentional processes (see, e.g., Pashler, 1998). A

learning system that is tuned to multiple modalities can be influenced by a larger set of stimuli. Moreover, regularities across modalities may be more accessible to high-level executive functions. It is possible, for example, that cross-modal regularities are attended and/or encoded better during familiarization than are unimodal regularities. Subsequently, at test, unimodal regularities that have previously been part of cross-modal regularities are retrieved more easily and accurately. In support of this claim, participants revealed a higher learning rate of triples that had previously been accompanied by consistent rather than random cues in another modality (Cunillera et al., 2010; Thiessen, 2010). A second possibility highlights the role of temporal information entangled in cross-modal information, with elements that appear simultaneously capturing attention. This claim is in line with previous studies showing that elements appearing simultaneously and consistently with triplets enhanced learning of the triplets (Cunillera et al., 2010), whilst elements appearing at random intervals reduced learning (Neath et al., 2010).

However, contradictory findings, suggesting that learning proceeds independently within modalities and is unaffected by information in other modalities, also exist. In a study by Seitz, Kim, van Wassenhove, and Shams (2007), participants were presented with simultaneous auditory and visual streams, with the auditory stream being composed of reoccurring tone pairs and the visual stream of reoccurring shape pairs. Importantly, each tone was always presented with a consistently associated shape. In a subsequent test, participants showed learning of the unimodal pairs, choosing auditory and visual pairs over random tones or shapes, with learning rates being comparable between the modalities. Furthermore, learning of unimodal pairs following the cross-modal familiarization was comparable to unimodal learning following unimodal familiarization (an auditory or visual stream only). Together, these findings suggest that participants independently learned the unimodal pairs, with learning being unaffected by the presence of cross-modal regularities. Comparable findings were obtained with a different learning paradigm known as *artificial-grammar learning* (Conway & Christiansen, 2006).

In our study, we addressed two issues: (1) whether unimodal statistical learning is affected by simple stimuli in another modality, and (2) what types of associations between modalities affect unimodal learning. Toward this aim, we designed a systematic study of the types of cross-modal associations that may affect unimodal learning, with comparable conditions and stimuli across audition and vision. Participants were familiarized with a stream composed of reoccurring triplets in one modality (e.g., audition) accompanied by one of various types of cues in a second modality (e.g., vision). Across participants, we used five conditions: (1) a *consistent cue*, in which a single identical element from one modality was consistently

associated with all of the elements of a triplet from the other modality throughout the familiarization phase; (2) an *inconsistent cue*, in which each appearance of a triplet was always accompanied by a single element from the other modality, but the identity of the accompanying element could change over different appearances of the triplet; (3) a *random cue with associations*, in which each element of each triplet was consistently associated with an element of the other modality (thus creating three audiovisual associations for each triplet); (4) a *random cue*, in which accompanying elements were completely random; or (5) *no cue* at all. Participants were subsequently tested on a recognition test assessing learning of the triplets. Our general hypothesis was that we would observe cross-modal effects on learning of the stream. More specifically, depending on the nature of the cross-modal cues, we hypothesized that we would find a pattern of either facilitation or hindrance of learning (in comparison to the no-cue condition) according to the cross-modal information. The consistent-cue condition should enhance learning of triplets because it segmented them, whereas the random cue with associations should reduce learning because it could hinder associations between elements within triplets. The other two cue conditions would be intermediate: The inconsistent cue should both facilitate segmentation (within elements of the triplet) and hinder segmentation (due to differences between different appearances of the same triplet); the random cue should simply add cross-modal noise.

Method

Participants

A total of 320 participants from the Hebrew University were given either course credit or payment for their participation. Thirty-two of the participants took part in each of the five conditions within each modality.

Learning in the auditory modality

Auditory stimuli

A set of 40 syllables were created using a Pratt synthesizer and were matched on pitch (~76 Hz), volume (~60 dB), and duration (250–350 ms). Because human listeners are sensitive to perceptual similarity in the auditory domain (e.g., preferring a triplet such as *balada* over *balide*; see Bregman, 1990; Newport, & Aslin, 2004), the auditory triplets and their matching foil triplets were created so as to avoid either consonant or vowel repetitions within each structure. Two “mini-languages” (corresponding to two conditions) were created, each of which was composed of 24 syllables organized into eight auditory triplets (see Table 1). The two auditory conditions were counterbalanced across participants.

Table 1 Auditory triplets and foil triplets

Condition A		Condition B	
Triplets	Foil Triplets	Triplets	Foil Triplets
balodi	bagile	gaduni	gabelo
nagiku	namobi	labeti	laginu
kamole	kagudi	tagilo	takeni
negubi	neloku	bokenu	boduti
temudo	temino	dimeno	digumo
tamike	tagebo	maguto	makile
ligeno	damuke	dekimo	debuno
datubo	litudo	kabule	kameto

Visual stimuli

The visual stimuli were eight circles colored in blue, purple, turquoise, orange, yellow, green, red, and pink. The circles were presented at the center of the screen and had a diameter of 2.8° viewed from a distance of 60 cm.

Procedure

Familiarization phase Each syllable was presented for 250–350 ms, with a stimulus onset asynchrony (SOA) of 1,000 ms. Since the syllables differed in duration, the interstimulus interval (ISI) differed as well, lasting from 650 to 750 ms. Note that, in contrast to previous, typical auditory statistical-learning studies, in which syllables have appeared one after the other with no pause, here we inserted an ISI in order to equate the auditory and visual tasks, as in the latter task shapes typically appear one after the other following a brief pause.

The stream of syllables was constructed as follows: The eight auditory triplets were randomly selected without replacement, with each auditory triplet consisting of three unique syllables appearing in succession (see Fig. 1). The eight auditory triplets were repeated 24 times, and the same randomization procedure applied to each repetition, with the constraint that two identical triplets did not appear in succession (i.e., the same triplet did not appear as the last and first of two successive repetitions).

Each syllable was accompanied in most conditions by the simultaneous appearance of a colored circle at the center of the screen for 800 ms with an SOA of 1,000 ms. Five conditions were used, four of which included both syllables and circles (see Fig. 2 for an example of the various conditions

- In the *consistent-cue* condition, the three syllables of each triplet were accompanied by the same three identical

colored circles throughout the familiarization phase. Thus, each triplet was uniquely associated with a specific color.

- In the *inconsistent-cue* condition, each auditory triplet was accompanied by three identical colored circles, but with each repetition of the auditory triplet, the color of the circles was randomly determined. The comparison between the consistent- and inconsistent-cue conditions would allow us to determine the role of consistency in statistical learning.
- In the *random-cue-with-associations* condition, each syllable was uniquely associated with one of the eight possible colors throughout the familiarization phase. Thus, there was no association between a single auditory triplet and a color, as each triplet could appear with one to three different colors, which might lead to hindrance in grouping the elements of the sequence into a single triplet.
- In the *random-cue* condition, the syllables appeared with random colors. The comparison between the random-cue conditions with and without associations would allow us to determine the role of simultaneous audiovisual associations in statistical learning.
- In the baseline *no-cue* condition, the syllables appeared alone.

Auditory test phase The second part of the experiment consisted of a two-alternative forced choice auditory test. On each test trial, participants heard two three-syllable sequences, with each syllable being presented for 250–350 ms with an SOA of 1,000 ms. One of the test sequences was an auditory triplet, whereas the second test sequence was an auditory foil triplet. We used a method originated by Turk-Browne, Jungé, and Scholl (2005) to create the foil triplets. The eight auditory triplets were divided into two groups, each consisting of four triplets (e.g., triplets with the syllables *ABC*, *DEF*, *GHI*, and *JKL*). Then, foil triplets were constructed using three syllables from different auditory triplets from this group, with the constraint that each syllable appeared in its original relative position (e.g., *AEI*, *DHL*, *GKC*, and *JBF*). Each auditory triplet appeared once with each foil triplet, resulting in 16 trials per group. This was done for the second group of auditory triplets as well, for a total of 32 trials.

Between the first and second sequences was a 1,000-ms break. Participants were instructed to indicate in a nonspeeded manner which of the two sequences seemed more familiar to them by pressing one of two buttons. The

Fig. 1 Example of an auditory stream composed of syllables. Each color (or gray shade) denotes a different triplet

Fig. 2 Examples of an auditory triplet according to condition, in two separate appearances of the triplet

test trials were presented in a random order, and the order of triplets versus foil triplets in individual test trials was counterbalanced for each participant.

Learning in the visual modality

Visual stimuli

A total of 24 nonsense shapes adapted from Turk-Browne et al. (2005) were presented at the center of the screen with a white background. Viewed from a distance of 60 cm, the height and width of each shape was approximately 1.9°. The 24 shapes were organized into eight visual triplets, each defined as three unique shapes that always appeared successively (see Fig. 3). The organization of shapes into triplets was randomly determined for each participant.

Auditory stimuli

The auditory stimuli were eight syllables (ba, di, lo, te, ma, ko, ne, and ge).

Procedure

Familiarization phase This phase consisted of eight visual triplets. In most conditions, each shape was accompanied by the simultaneous appearance of a syllable. Five conditions were presented: In the *consistent-cue* condition, the three shapes of each visual triplet were always accompanied by the same three identical syllables. In the *inconsistent-cue* condition, each visual triplet was accompanied by three identical syllables, but with each repetition of the visual triplet, a different random syllable was determined. In the *random-cue-with-associations* condition, each shape was uniquely associated with one of the eight possible syllables. In the *random-cue* condition, the shapes appeared with random syllables. And in the *no-cue* condition, the shapes appeared alone.

Visual test phase The visual test was similar to the auditory test: Each trial contrasted a visual triplet with a visual foil triplet. On each test trial, participants viewed two three-shape test sequences, with each shape appearing for 800 ms with an SOA of 1,000 ms, as well as a 1,000-ms break between the first and second sequences. Participants chose which one of the two sequences was more familiar.

Results

Across auditory and visual modalities

Statistical analyses according to modality and condition are presented in Table 2. As can be observed from the table, statistical learning was significant for all conditions, as assessed by a one-sample *t* test.

A two-way analysis of variance (ANOVA) with Cue (consistent cue, inconsistent cue, random cue with associations, random cue, or no cue) and Modality (audition or vision) as factors revealed a significant effect of cue, $F(4, 320) = 15.9, p < .0001$. We found no significant effect of modality, $F(1, 310) = 0.1, p = .7$, nor a significant Cue $\times$ Modality interaction, $F(4, 310) = 0.03, p = .13$. Since similar results arose across modalities, we performed a one-way ANOVA with Cue as the factor, as well as four planned contrasts, each comparing two conditions. These revealed that learning was higher in the presence of a consistent cue than of an inconsistent cue, $F(4, 315) = 19.26, p < .0001$, $MSE = 2.1$. No difference in learning emerged between the inconsistent cue and no cue ($p = .09$). In contrast, learning was lower in the presence of a random cue with associations

Fig. 3 Example of a visual stream composed of shapes for one participant. Each color (or gray shade) denotes a different triplet

Table 2 Descriptive statistics for statistical learning in the presence of a consistent cue, an inconsistent cue, a random cue with associations, a random cue, or no cue

	Across Modalities	Audition	Vision
Consistent cue	74.5 % (<i>SD</i> = 17), $t(63) = 11.3, p < .0001$	70.9 % (<i>SD</i> = 16), $t(31) = 7.4, p < .001$	78 % (<i>SD</i> = 18), $t(31) = 8.8, p < .0001$
Inconsistent cue	63 % (<i>SD</i> = 16), $t(63) = 6.7, p < .0001$	66 % (<i>SD</i> = 15), $t(31) = 6.09, p < .001$	60.6 % (<i>SD</i> = 16), $t(31) = 3.6, p < .01$
Random cue with associations	55.7 % (<i>SD</i> = 11), $t(63) = 3.99, p < .0001$	56.15 % (<i>SD</i> = 9.8), $t(31) = 3.5, p < .01$	55 % (<i>SD</i> = 13), $t(31) = 2.3, p < .05$
Random cue	58.7 % (<i>SD</i> = 11.8), $t(63) = 5.9, p < .0001$	59.5 % (<i>SD</i> = 11), $t(31) = 4.8, p < .001$	58 % (<i>SD</i> = 12), $t(31) = 3.5, p < .01$
No cue	62.3 % (<i>SD</i> = 14.5), $t(63) = 6.76, p < .0001$	60.6 % (<i>SD</i> = 11), $t(31) = 5.5, p < .0001$	63.8 % (<i>SD</i> = 17), $t(31) = 4.5, p < .001$

than it was with no cue, $F(4, 315) = 6.6, p < .02$. Also, we found no difference between the conditions with a random cue and a random cue with associations ($p = .2$; see Fig. 4).

One potential caveat of our study concerns the method that we used to create the “mini-languages” in the auditory condition. The vowels and consonants used were not counterbalanced between the first and last syllables of triplets, which could lead to potential biases. For example, the vowel “a” appeared with a high probability in the first syllable, but never in the last one. Such uneven distributions of vowels and consonants could have provided segmentation cues during familiarization. However, these same vowel/consonant distributions existed for the foil triplets as well (as the foil triplets consisted of three syllables in their original relative positions from three different auditory triplets), so the distinction between a triplet and a foil triplet

could not be based on vowel/consonant cues. Moreover, we found essentially the same results in the visual domain, where the triplets were fully randomized.

Discussion

Our study revealed that unimodal statistical learning of transitional probabilities is clearly affected by information in another modality, supporting the existence of an interaction between modalities during learning. This effect takes the forms of both facilitation and reduction in learning. Statistical learning was enhanced in the presence of a consistent as compared to an inconsistent cue, with no difference emerging between the inconsistent- and no-cue conditions. This finding highlights the importance of consistency in

Fig. 4 Learning of triplets in the presence of a consistent cue, an inconsistent cue, no cue, a random cue, or a random cue with associations, across modalities

learning the triplets. In contrast, statistical learning was reduced in the presence of a random cue with associations, as compared to no cue at all. We observed no significant difference between the random-cue conditions with and without associations.

Together, these findings reveal a pattern of facilitation and reduction in unimodal statistical learning as a function of the type of cross-modal information provided. What information is provided by a consistent cue? One possibility is that the circles/syllables offer a grouping cue in the form of temporal information: a shared time frame between modalities for the duration of a triplet, which may have helped the perceptual system extract the triplets. A second possibility is that the circles/syllables direct attention to the boundaries of triplets. Attention has formerly been found to play a prominent role in statistical learning, since auditory statistical learning has been compromised when participants were required to perform a secondary task (Toro, Sinnett, & Soto-Faraco, 2005). In the visual domain, participants learned the statistical regularities of only one of two streams of colored shapes to which their attention was directed by task instructions (Turk-Browne et al., 2005). Finally, as we described in the introduction, a shape coinciding with the first or third syllable of auditory triplets, thus presumably drawing attention to the boundaries of the triplets, enhanced learning (Cunillera et al., 2010). Note, however, that we found no enhancement of learning with an inconsistent cue, with which the boundaries of triplets were clearly marked, but the marker changed across repetitions of the triplet. This added finding suggests that the marker must be consistent across repetitions to serve as either a grouping cue, attention cue, or a segmentation cue.

In contrast, learning is reduced in the presence of random cues with associations as compared with no cue. We propose that a competition over learning resources occurs between learning of simultaneous cross-modal associations (e.g., colored circle–syllable pairs) and learning of unimodal triplets, apparently resulting in impaired learning of the triplets. This finding may have two implications: (a) Cross-modal information has precedence over unimodal information, or alternatively, (b) associations between stimuli appearing at the same time have precedence over associations between successive elements, regardless of their nature (cross- or within-modality elements).

Learning is intrinsically connected to memory processes. In our study, both facilitation and reduction of learning were found at the absence of cues during test, and these occurred across all cue conditions. This suggests that cues affect the stage of encoding: Differences at the level of encoding can occur if a consistent cue enables

better encoding of the triplets as objects, whereas in the presence of random cues with associations, encoding of the triplets is compromised, since resources are drawn to the cross-modal associations. Learning can also be affected at the stage of retrieval, as was made evident in a study conducted by Turk-Browne, Isola, Scholl, and Treat (2008). Participants were familiarized with colored-shape triplets, such that each triplet consisted of three unique shapes and colors. Participants were subsequently tested on (a) the colored-shape triplets or (b) the color or shape triplets alone. Learning was reduced when the shapes or colors were removed at test, as compared with when they were present. This suggests that the colored-shape triplets were encoded as bound objects, and in the absence of a consistent feature such as color, retrieval of the triplets was compromised. Thus, statistical learning may be affected at the encoding and retrieval phases.

Our study demonstrates that a cross-modal association affects learning within a single modality. This was demonstrated using stimuli across modalities that were arbitrarily related (e.g., the color blue associated with three syllables). However, natural cross-modal associations exist (Spence, 2011). For example, both adults and children associate lightness with loudness, with light colors being associated with louder sounds, and darker colors associated with quieter sounds (Bond & Stevens, 1969; Stevens & Marks, 1965). Similarly, high-pitched sounds are associated with brighter and smaller objects (Mondloch & Maurer, 2004). Such natural associations can be viewed as a form of a natural consistent association, as some associations presumably reflect cross-modal dimensions that are correlated in nature and are learned by our perceptual systems (Marks, 2000; Spence, 2011). Would such associations between the auditory and visual modalities also act as grouping cues and enhance unimodal learning? For example, would triplets of dark shapes matched to a low-pitched tone, rather than a high-pitched tone, be learned better? Future research ought to address this issue.

In conclusion, patterns of both facilitation and reduction in learning in the auditory and visual modalities were found by varying the type of cross-modal association between the modalities. Further research may explore whether the modalities differ in their sensitivities to cross-modal associations, by investigating modality-specific constraints that may differ between audition and vision.

Author note We thank Anat Hornik and Tali Arad for their assistance in running the experiments. This work was funded by a grant from the Israel Science Foundation to A.C.

References

- Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, 36, 286–304. doi:10.1111/j.1551-6709.2011.01200.x
- Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21, 170–176. doi:10.1177/0963721412436806
- Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, 36, 190–201. doi:10.1037/0012-1649.36.2.190
- Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics*, 6, 337–339. doi:10.3758/BF03212787
- Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge: MIT Press.
- 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
- Cunillera, T., Cámara, E., Laine, M., & Rodríguez-Fornells, A. (2010a). Speech segmentation is facilitated by visual cues. *Quarterly Journal of Experimental Psychology*, 63, 260–274. doi:10.1080/17470210902888809
- Cunillera, T., Laine, M., Cámara, E., & Rodríguez-Fornells, A. (2010b). Bridging the gap between speech segmentation and word-to-word mappings: Evidence from an audiovisual statistical learning task. *Journal of Memory and Language*, 63, 295–305. doi:10.1016/j.jml.2010.05.003
- Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities. *Psychological Science*, 17, 905–912. doi:10.1111/j.1467-9280.2006.01801.x
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7(5), 1–14. doi:10.1167/7.5.7
- Fiser, J. (2009). Perceptual learning and representational learning in humans and animals. *Learning & Behavior*, 37, 141–153. doi:10.3758/LB.37.2.141
- Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458–467. doi:10.1037/0278-7393.28.3.458
- Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822–15826. doi:10.1073/pnas.232472899
- Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps five-month-olds learn abstract rules. *Developmental Science*, 12, 504–509. doi:10.1111/j.1467-7687.2008.00794.x
- Frost, R. (2012). Towards a universal model of reading. *The Behavioral and Brain Sciences*, 35, 263–279. doi:10.1017/S0140525X11001841. disc. 280–329.
- Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (in press). What predicts successful literacy acquisition in a second language? *Psychological Science*. doi:10.1177/0956797612472207
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
- Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual–auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 131–153. doi:10.1037/0096-1523.20.1.131
- Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, 48, 171–184. doi:10.1037/a0025405
- Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One*, 3, e1532. doi:10.1371/journal.pone.0001532
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35–B42. doi:10.1016/S0010-0277(02)00004-5
- Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research*, 24, 326–334. doi:10.1016/j.cogbrainres.2005.02.005
- Marks, L. E. (2000). Synesthesia. In E. Cardena, S. J. Lynn, & S. C. Krippner (Eds.), *Varieties of anomalous experience: Examining the scientific evidence* (pp. 121–149). Washington: American Psychological Association.
- Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, 108, 271–280. doi:10.1016/j.cognition.2008.02.003
- Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch–object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, 4, 133–136. doi:10.3758/CABN.4.2.133
- Neath, I., Guérard, K., Jalbert, A., Bireta, T. J., & Surprenant, A. M. (2010). Irrelevant speech effect and statistical learning. *Quarterly Journal of Experimental Psychology*, 62, 1551–1559. doi:10.1080/17470210902795640
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162. doi:10.1016/S0010-0285(03)00128-2
- Pashler, H. (1998). *The psychology of attention*. Cambridge: MIT Press.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855–863. doi:10.1016/S0022-5371(67)80149-X
- 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
- Shams, L., Wozny, D. R., Kim, R., & Seitz, A. (2011). Influences of multimodal experience on subsequent unimodal processing. *Frontiers in Perception Science*, 2(264), 1–9. doi:10.3389/fpsyg.2011.00264
- Seitz, A., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, 36, 1445–1453. doi:10.1068/p5843
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73, 971–995. doi:10.3758/s13414-010-0073-7
- Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences*, 54, 407–411.
- Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science*, 34, 1093–1106. doi:10.1111/j.1551-6709.2010.01118.x

- Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97, B25–B34. doi:[10.1016/j.cognition.2005.01.006](https://doi.org/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](https://doi.org/10.1037/0096-3445.134.4.552)
- 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. doi:[10.1037/0278-7393.34.2.399](https://doi.org/10.1037/0278-7393.34.2.399)
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, 21, 1934–1945. doi:[10.1162/jocn.2009.21131](https://doi.org/10.1162/jocn.2009.21131)