Recent cross-modal statistical learning influences visual perceptual selection

Elise A. Piazza
Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA
Vision Science Graduate Group, University of California, Berkeley, Berkeley, CA, USA
Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA

Rachel N. Denison
Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA
Department of Psychology and Center for Neural Science, New York University, New York, NY, USA

Michael A. Silver
Vision Science Graduate Group, University of California, Berkeley, Berkeley, CA, USA
Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA
School of Optometry, University of California, Berkeley, Berkeley, CA, USA

Incoming sensory signals are often ambiguous and consistent with multiple perceptual interpretations. Information from one sensory modality can help to resolve ambiguity in another modality, but the mechanisms by which multisensory associations come to influence the contents of conscious perception are unclear. We asked whether and how novel statistical information about the coupling between sounds and images influences the early stages of awareness of visual stimuli. We exposed subjects to consistent, arbitrary pairings of sounds and images and then measured the impact of this recent passive statistical learning on subjects' initial conscious perception of a stimulus by employing binocular rivalry, a phenomenon in which incompatible images presented separately to the two eyes result in a perceptual alternation between the two images. On each trial of the rivalry test, subjects were presented with a pair of rivalrous images (one of which had been consistently paired with a specific sound during exposure while the other had not) and an accompanying sound. We found that, at the onset of binocular rivalry, an image was significantly more likely to be perceived, and was perceived for a longer duration, when it was presented with its paired sound than when presented with other sounds. Our results indicate that recently acquired multisensory information helps resolve sensory ambiguity, and they demonstrate that statistical learning is a fast, flexible mechanism that facilitates this process.

Introduction

To accurately perceive the natural world, we must learn that cues from different sensory modalities point to the same object: a friend's face and her voice, a rose and its fragrance, a blackberry and its flavor. Through repeated exposure to consistent couplings between sensory features, we learn to associate information from various senses. Once these associations are established, they can help resolve information that is ambiguous or impoverished for one of the senses.

The process by which the brain chooses a conscious percept from alternative conflicting interpretations of an ambiguous image is known as visual perceptual selection (Lumer, Friston, & Rees, 1998; Meng & Tong, 2004; Mitchell, Stoner, & Reynolds, 2004). Many studies have demonstrated the influence of long-established multisensory associations on this process, often employing binocular rivalry, in which two

Citation: Piazza, E. A., Denison, R. N., & Silver, M. A. (2018). Recent cross-modal statistical learning influences visual perceptual selection. *Journal of Vision*, *18*(3):1, 1–12, https://doi.org/10.1167/18.3.1.

https://doi.org/10.1167/18.3.1 Received August 8, 2017; published March 1, 2018 ISSN 1534-7362

incompatible images are presented separately to the two eyes at overlapping retinal locations. Even if the stimuli are unchanging in binocular rivalry displays, observers perceive a spontaneous alternation between the images (reviewed in Alais & Blake, 2005; Blake & Wilson, 2011). This dissociation between stimulus and percept is very useful for studying contextual influences on conscious awareness (Bressler, Denison, & Silver, 2013).

Extensive exposure to particular multisensory associations throughout life can cause input from nonvisual modalities to enhance visual perceptual selection (typically by increasing the relative dominance) of rivalrous images that are compatible or congruent with the established multisensory context. These contexts can include auditory (Conrad et al., 2013; Guzman-Martinez, Ortega, Grabowecky, Mossbridge, & Suzuki, 2012; Lee, Blake, Kim, & Kim, 2015; Lunghi, Morrone, & Alais, 2014; van Ee, van Boxtel, Parker, & Alais, 2009; Vidal & Barrès, 2014), tactile (Blake, Sobel, & James, 2004; Conrad, Vitello, & Noppeney, 2012; Lunghi, Binda, & Morrone, 2010), and even olfactory input (Zhou, Jiang, He, & Chen, 2010).

For example, a sound that is congruent with one of two rivalrous images in flicker frequency (Kang & Blake, 2005), motion direction (Conrad, Bartels, Kleiner, & Noppeney, 2010), or semantic content (Y.-C. Chen, Yeh, & Spence, 2011) increases predominance of the rivalrous congruent image over the incongruent image. For observers who can read music (a skill acquired by the subjects well before the experiment), listening to a specific melody increases the dominance of congruent musical notation in rivalry (Lee et al., 2015). Cross-modal cues can also promote the perceptual dominance of congruent images that are suppressed, either through binocular rivalry (Lunghi & Alais, 2015; Lunghi & Morrone, 2013) or continuous flash suppression (Salomon, Lim, Herbelin, Hesselmann, & Blanke, 2013).

In these examples, the congruencies that influence perceptual selection are explicit, obvious to the subject, and well learned through a lifetime of experience with cross-modal associations. However, the mechanisms for forming new multisensory associations that may then influence the contents of conscious perception are unclear. One possibility is that long-term exposure to multisensory couplings is required. Alternatively, the brain may have more flexible mechanisms for constraining perceptual interpretations that make use of recently encountered multisensory couplings. One study (Einhäuser, Methfessel, & Bendixen, 2017) found that rivalry was influenced by previous explicit multisensory learning that was induced through an active learning task. Here, we investigated whether and how passive exposure to statistical associations between sounds and images contributes to the resolution of ambiguity in the visual environment.

Associations between stored representations of sensory cues can be established quickly through probabilistic inference (Aslin & Newport, 2012). Both adult (Fiser & Aslin, 2001) and infant (Saffran, Aslin, & Newport, 1996) observers can rapidly learn sequential or spatial patterns of sensory stimuli (e.g., abstract shapes, natural images, or sounds) through passive exposure. This phenomenon, known as statistical learning, is thought to be crucial for detecting and forming internal representations of regularities in the environment. In the exposure phase of a typical statistical learning experiment, subjects passively view or hear long streams of stimuli that contain sequences of two or more items that appear in the same temporal order (Aslin & Newport, 2012; Saffran, Johnson, Aslin, & Newport, 1999). Afterward, when asked to judge which of two sequences is more familiar, subjects are more likely to choose sequences that were presented during the exposure phase than random sequences of stimuli that had not previously been presented in that order.

Statistical learning can also occur for multisensory sequences. Following exposure to streams of bimodal quartets (each containing two consecutive audiovisual, or AV, pairs), subjects performed above chance on familiarity judgments for cross-modal (AV) as well as unimodal (AA and VV) associations (Seitz, Kim, van Wassenhove, & Shams, 2007). Moreover, rapid learning of arbitrary visuo-haptic correspondences (between the luminance and stiffness of objects) has been shown to impact perceptual thresholds (Ernst, 2007). Statistical learning is therefore a fast, flexible associative mechanism that could conceivably constrain perceptual interpretations in one sensory modality based on information in another modality. However, whether multisensory statistical learning can influence perceptual selection, as opposed to recognition memory, reaction time, or perceptual thresholds, has not been investigated.

To address this question, we examined the influence of statistical learning of arbitrary auditory–visual associations on subsequent visual perceptual selection. We asked whether initial perceptual interpretations during binocular rivalry could be rapidly updated based on recent multisensory experience or whether learners might instead require days or even years of exposure to joint probabilities between sounds and images before these congruencies begin to influence perceptual selection. Specifically, we tested whether formation of associations between particular sounds and images during an 8-min exposure phase would cause the presentation of a given sound to alter initial perceptual dominance (i.e., visual awareness) of its associated image during subsequent binocular rivalry.

Discovering this type of impact of cross-modal statistical learning on binocular rivalry would indicate that recent and passive acquisition of probabilistic information about the conjunctions of sounds and images influences the early resolution of conflicting visual information.

Methods

Participants

Twenty participants (ages 18–39, 14 female) completed this study. All subjects provided informed consent according to the Declaration of Helsinki, and all experimental protocols were approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley. We originally collected data from 30 participants but excluded 10 subject data sets from analysis. Of the excluded subjects, one was unable to align the stereoscope to position the two monocular stimuli at corresponding retinal locations, two were missing data due to incorrect response key mapping, and seven were excluded for having incorrect responses on more than 25% of the catch trials (see Procedure section).

Stimuli

All visual stimulus displays were generated on a Macintosh PowerPC (Apple, Inc., Cupertino, CA) using MATLAB (MathWorks, Natick, MA) and Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and were displayed on a gamma-corrected MultiSync FE992 CRT monitor (NEC, Tokyo, Japan) with a refresh rate of 60 Hz at a viewing distance of 100 cm. All images were presented at the fovea and had 100% contrast and the same mean luminance as the neutral gray screen background (59 cd/m^2). Each participant's head position was fixed with a chin rest throughout the entire experimental session.

The stimulus set included six images and six sounds (Figure 1a). The stimuli were designed to be simple and easily discriminable from one another within a given modality. The images were selected so that they could be grouped into three pairs (orthogonal sine wave gratings with ±45° orientations; two hyperbolic gratings with 0° and 45° orientations; and a polar radial grating and a polar concentric, or bull's-eye, grating), such that the members of each pair would rival well with each other (Figure 2). We refer to these image pairs as "rivalry pairs." During the rivalry test (see Procedure section), the two images making up a given rivalry pair were always presented together, one in each eye. The sounds included two sine wave puretones (D5, B6) and four chords composed of sine wave puretones (two distinct dissonant clusters, an A-flat major chord, and an F-minor chord). The sounds were presented through headphones at a comfortable volume.

Procedure

Each subject completed a 1-hr session composed of three phases: exposure, recognition test, and rivalry test.

Exposure phase

Each participant passively viewed and heard an 8-min stream of sounds and images. This exposure period falls within the range of durations typically used in the adult statistical learning literature (e.g., 7 min in Fiser & Aslin, 2001; 8 min in Seitz et al., 2007; 21 min in Saffran et al., 1999). All images were presented in the center of the screen during this phase. Participants were instructed to attend to the stimuli and fixate the images but were not required to perform any task, and the experimenters did not disclose the existence of any patterns in the AV streams. Each sound was presented for 500 ms and then continued playing while an image was presented for another 500 ms, after which the sound and image presentations ended simultaneously (Figure 1b). This was followed by a 500-ms blank interval before the onset of the next sound. Each participant was exposed to a total of 180 of these AV presentations in a continuous stream (Figure 1c).

For every participant, three images (one from each rivalry pair) and three sounds were randomly chosen to be consistently paired during the exposure phase (Figure 1a). Each of these three selected sounds was always presented with its paired image, corresponding to a total of three "AV pairs." Ninety of the 180 AV presentations (frames with dotted borders, Figure 1c) during the exposure phase corresponded to one of these consistent AV pairs for a total of 30 identical presentations of each pair. In the other 90 AV presentations (frames with solid borders, Figure 1c), random combinations of the remaining, unpaired images and sounds (three of each) were presented, so there was no consistent mapping between any of these images and any of the sounds. Selection of the images and sounds for pairing was counterbalanced across subjects to eliminate possible bias due to any inherent congruence between the stimuli; thus, across the group, the subset of possible AV combinations that were chosen as pairings was fully randomized and arbitrary.

Because the subset of images that were reliably paired with sounds during the exposure phase (three out of a total of six images; Figure 1a) was randomly

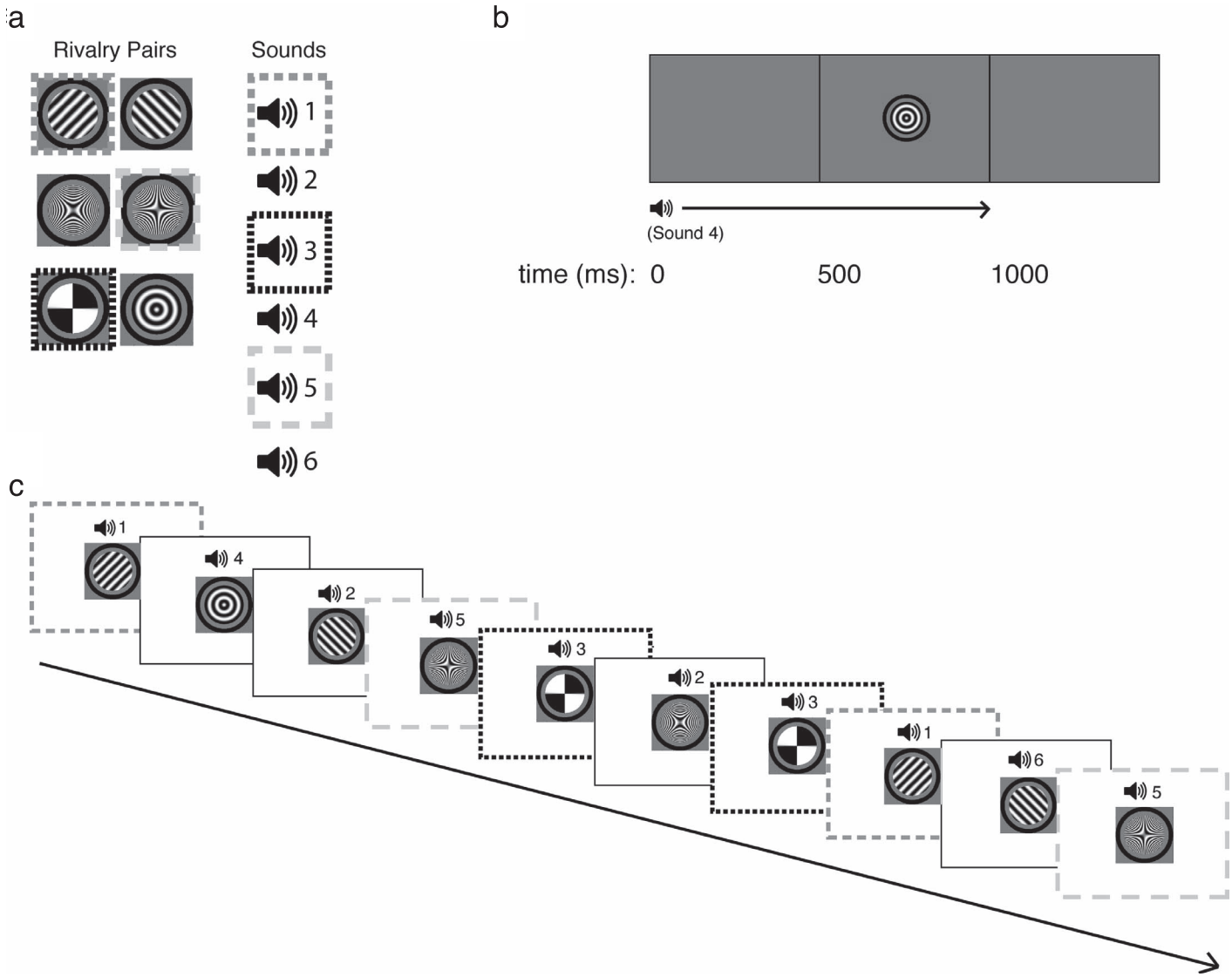

Figure 1. Exposure phase. (a) The full set of images and sounds. Borders are used to indicate paired stimuli for an example subject but were not presented to the subjects. For each subject, one image from each of the three rivalry pairs was randomly chosen to be a "paired image," which means that it was consistently paired with a particular sound during the exposure phase (an "AV pair"). For each paired image, an associated sound was randomly chosen from the set of six sounds for each subject. The remaining three sounds were left unpaired, meaning they could be presented with any of the unpaired images on different individual AV presentations. (b) The time course of a single example AV presentation during the exposure phase. (c) Example sequence of AV presentations during the exposure phase.

assigned across participants and still well balanced after subject exclusion, any group-level systematic effects on perceptual selection during the rivalry test cannot be explained by differences between the two images in a given rivalry pair in baseline dominance. For example, for the rivalry pair shown in Figure 2, the bull's-eye was consistently paired with a sound during exposure for some subjects whereas the radial grating was consistently paired with a sound during exposure for other subjects. Specifically, each image was assigned as "paired" to between nine and 11 subjects, with 10 representing perfectly equal assignment, and we found no statistically significant difference in this likelihood of pairing across images (one-way ANOVA), $F(5, 30) = 0.05$, $p = 0.99$.

Recognition test

After conclusion of the exposure phase, we tested participants' recognition memory for the AV pairs that were presented during exposure. On each trial, one of the six sounds was presented, followed immediately by a display of all six images presented simultaneously and randomly arranged in a row on the screen. Participants

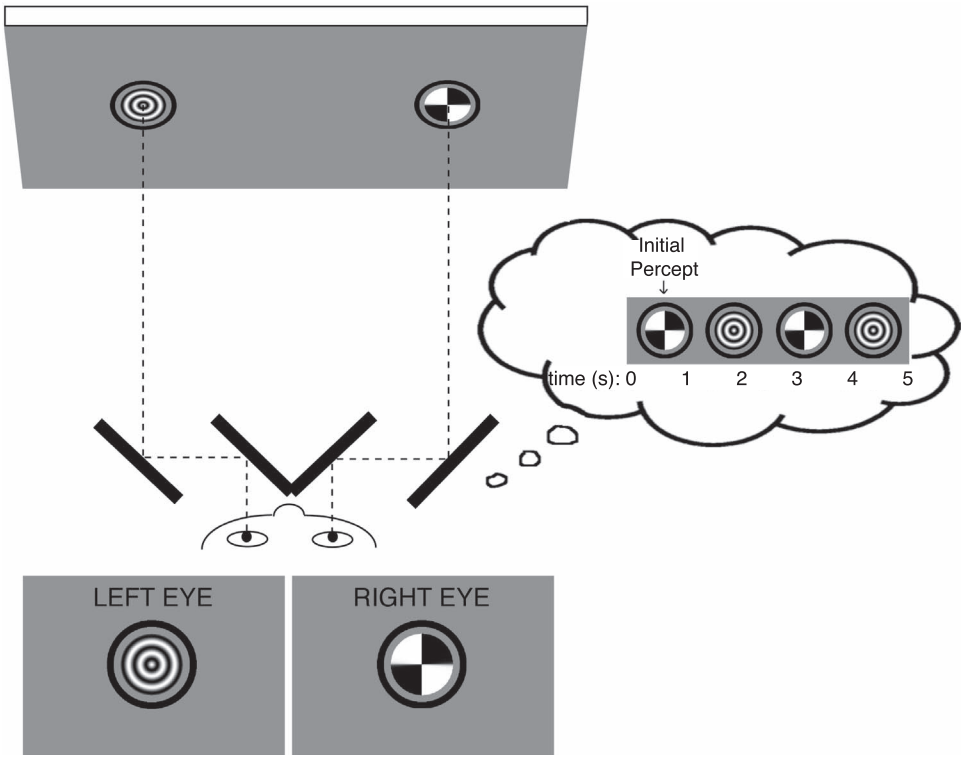

Figure 2. Example images presented separately to the two eyes (at corresponding retinal locations) through a mirror stereoscope and schematic sequence of possible percepts in a given trial of the rivalry test. Although the presented images did not change during a given trial, subjects typically perceived a continuous alternation between the two images throughout the 5-s stimulus presentation.

were instructed to select the image that they thought had been typically presented with that sound during the exposure phase. Each participant completed 36 trials, with six repetitions of each of the six sounds, for a total of approximately 5 min. The recognition test was conducted immediately after the end of the exposure phase to measure participants' learning at its peak, before any possible contamination of the learned associations by administration of the rivalry test (in which all possible combinations of sounds and images were repeatedly presented).

Rivalry test

Subjects viewed pairs of images (the rivalry pairs described above with the two images presented separately to the two eyes) through a mirror stereoscope with their heads stabilized by a chin rest. Each image was a circular patch 1.8° in diameter, surrounded by a black annulus with a diameter of 2.6° and a thickness of 0.2°. Binocular presentation of this annulus allowed it to serve as a vergence cue to stabilize eye position. The two images in a rivalry pair were tinted red and blue during the rivalry test, thereby allowing participants to report their perception during binocular rivalry using only two ("red" or "blue") instead of six (each of the images) response categories. The color and the member of the rivalry pair presented to each eye were fully counterbalanced and randomly intermixed across trials. We asked participants to report color, a feature not present in the exposure phase and therefore unrelated to AV learning, to reduce the likelihood of response bias. The use of colored image tints also served to increase the exclusivity of rivalry (decrease piecemeal percepts), and employing color as a response variable is standard in binocular rivalry research (Alais & Blake, 2005).

Before starting the rivalry test, each subject adjusted the stereoscope mirrors until the two eyes' images (with gratings replaced by identical figures in both eyes for this adjustment phase) were fused and the subject perceived only one annulus with binocular viewing. All subjects completed 10 practice trials before starting the test to ensure that they were using the correct response keys and that the stereoscope was properly aligned.

In each trial of the rivalry test, one of the six (randomly selected) sounds was presented, followed by the static visual images (Figure 2). The timing of the onset of stimulus presentation was the same as in the exposure phase, but here the images were presented continuously for 5 s instead of 500 ms (persisting for 4.5 s beyond the termination of the sound). There was a 1-s blank interval (consisting of only the binocular annulus) between trials. Throughout each trial, subjects

could press one of two keys to indicate their percept: either the red- or blue-tinted image. Subjects were instructed to continuously press a key for as long as the corresponding percept was dominant and to not press any key for ambiguous percepts.

Previous studies on the effects of cross-modal processing on binocular rivalry (Conrad et al., 2010; Einhäuser et al., 2017; Kang & Blake, 2005) reported mean dominance duration across the entire trial, but we focused on the initial response (the first reported percept in a given trial). In our study, sounds were only presented at the beginning of the trial (reflecting the timing of the exposure stimuli), so we expected any cross-modal effects on rivalry to be strongest early in the rivalry presentation. Moreover, effects of prediction on the initial rivalry response have been previously demonstrated (Attarha & Moore, 2015; Denison, Piazza, & Silver, 2011; Denison, Sheynin, & Silver, 2016). Our relatively short (5 s) trial durations were not designed for analysis of perceptual reports after the initial response; across all participants, only about 50% of all trials contained at least one full response (a response that did not persist until the end of the trial) following the initial response. However, this stimulus duration allows rivalry to fully resolve (i.e., become sufficiently unambiguous for the subject to report perceptual dominance of one of the stimuli) on nearly every trial and is therefore different from the very brief (1 s) rivalry presentations used in a paradigm sometimes known as "onset rivalry" (Carter & Cavanagh, 2007).

We collected data from a total of 216 rivalry trials, divided across three blocks and lasting approximately 30 min, from each participant. Each block contained 72 rivalry trials (12 trials per sound) and 12 randomly interleaved catch trials for which the images were identical in both eyes for a total of approximately 10 min of testing per block. Catch trials were considered to be incorrect if they contained any responses that did not correspond to the presented image. These trials were included to ensure that participants were attending to the stimuli, correctly understood the task instructions, and could distinguish the images based on color tint. We included only participants who responded accurately (i.e., made no incorrect key presses) on at least 75% of the catch trials in the first block (mean accuracy = 92.3%).

Results

After exposing participants to streams of sounds and images (with half of the sounds consistently paired with half of the images), we measured the impact of this passive exposure to the associated AV pairs on perceptual selection during binocular rivalry. We also measured the degree of learning of the pairings in a separate recognition test. All AV pairings were randomly chosen for each subject; therefore, any effects of cross-modal associations on rivalry at the group level were due to learning that occurred during the 8-min exposure phase.

Rivalry test

Each rivalry pair had two images: the paired image and the unpaired image (see Methods). During the exposure phase, each paired image was consistently presented with a particular sound, and each unpaired image was presented with any one of the three unpaired sounds on each AV presentation (Figure 1). In the rivalry test, we presented all possible combinations of six sounds and three rivalry pairs (18 combinations total).

We computed the effect of cross-modal learning on rivalry, a measure of how much the initial dominance of an image is enhanced when it is presented with its paired sound, relative to when it is presented with all other sounds. Our main analysis (Figure 3a, b) focused on the likelihood of initially perceiving the paired image (i.e., the proportion of trials for which that image is the first reported percept in the trial). We computed this proportion for each subject and rivalry pair for two types of trials (Figure 3a): (a) when the paired sound was presented and (b) when one of the other five sounds was presented. We averaged each of these proportion values across the three rivalry pairs and then calculated the within-subject difference between the two mean values to quantify the effect of a concurrently presented paired sound on perception during rivalry for each subject. Figure 3b shows the mean effect (i.e., mean difference score) across subjects.

Comparing perceptual selection for the same image in different auditory contexts controlled for possible baseline differences in dominance of the two images in a rivalry pair due to physical stimulus factors. A difference score above zero indicates that, when a given image was accompanied by its paired sound from the exposure phase, it was more likely to be initially perceived during binocular rivalry compared to when it was accompanied by any other sound. The difference score, therefore, quantifies the effect of prior cross-modal learning on rivalry (Figure 3b).

We hypothesized that the strength of the effect of cross-modal learning on rivalry would decline over the course of the rivalry blocks due to a general dissipation of learning across time and/or violation of the particular AV pairings established during the 8-min exposure phase by presentation of many combinations of stimuli during the rivalry test that interfered with

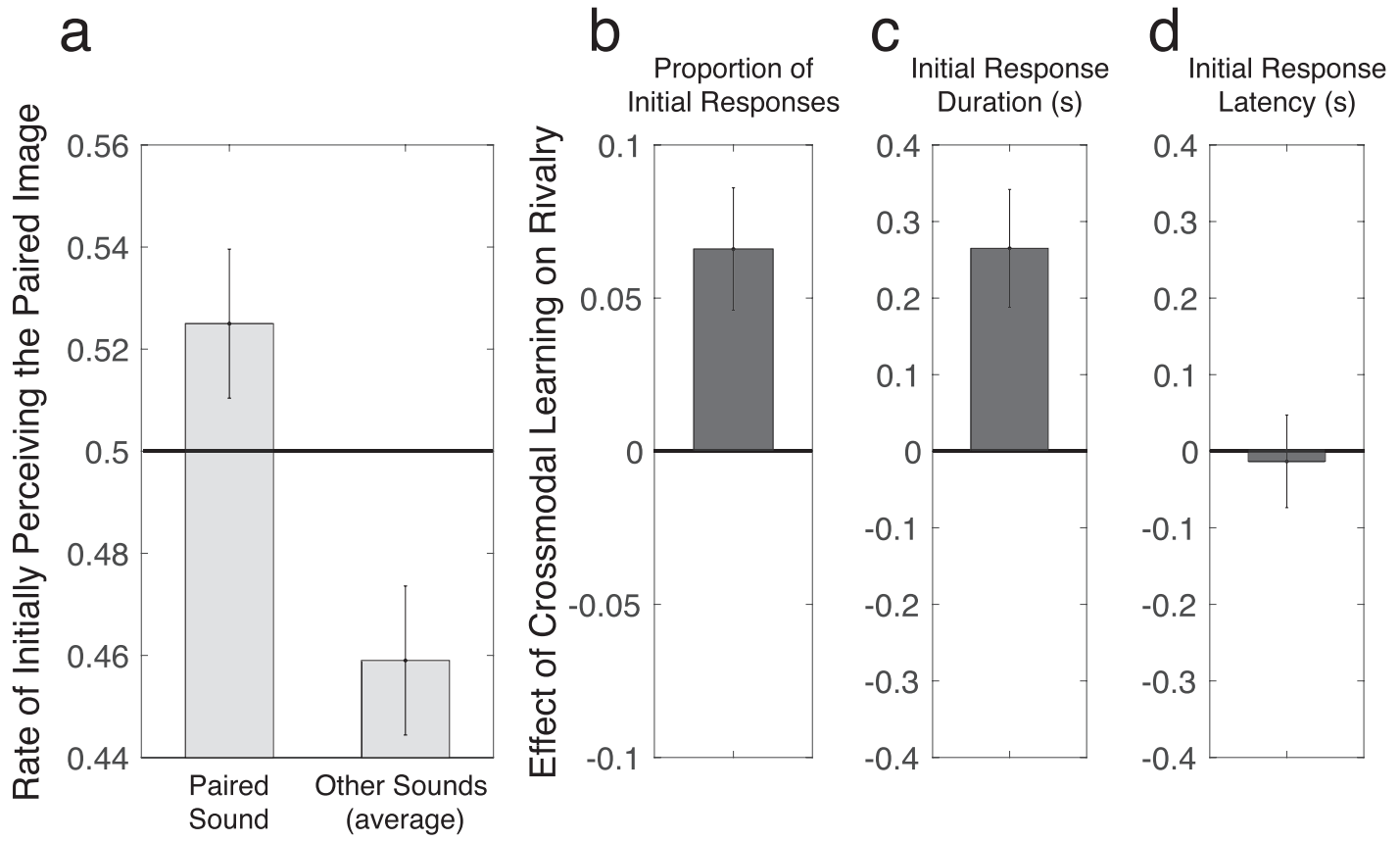

Figure 3. Effects of statistical learning of AV pairs on subsequent binocular rivalry. (a) For each subject, the proportion of trials in which a given image was initially perceived following the onset of binocular rivalry was measured in two conditions: trials in which that image was presented with its paired sound and trials in which it was presented with any other sound (averaged across all five other sounds). For each of these two conditions, we averaged the values across the three rivalry pairs in the first block and plotted the mean across subjects. Data from the first block of the rivalry test are presented. Error bars are within-subject *SEM* (Morey, 2008). (b) The within-subject difference between the two conditions in panel a quantified the influence of cross-modal statistical learning on initial perceptual selection in binocular rivalry. The bar shows the mean across subjects in the first block; error bar is *SEM*. (c, d) Effects of statistical learning of AV pairs on the duration and latency of the initial response, respectively. The procedure for computing these effects is the same as the one used in panel b. $N = 20$.

prior statistical learning. Indeed, the effect of statistical learning was only significant in the first of the three blocks (two-tailed t tests): first block, $t(19) = 3.18$, $p < 0.01$; second block, $t(19) = -0.27$, $p = 0.79$; third block, $t(19) = 1.07$, $p = 0.30$. Therefore, we present results from rivalry test data only from the first block. Moreover, we did not find a significant effect of rivalry pair category (i.e., sine wave gratings, hyperbolic gratings, and bull's-eye/radial) on the size of the AV learning effect on rivalry (ANOVA), $F(2, 18) = 0.75$, $p = 0.47$, $\eta^2_p = 0.04$, so all analyses of rivalry test data are collapsed across the three rivalry pairs.

The significant effect of cross-modal learning on the initial likelihood of perceiving an image (Figure 3b; Cohen's $d = 0.71$, 95% CI [0.03, 0.11]) indicates that exposure to novel, arbitrary AV pairs influenced subsequent perceptual selection during binocular rivalry. This effect may be driven more by suppression than facilitation: the facilitative effect of hearing a paired sound was not significantly different from a chance rate of 0.5 (Figure 3a, left bar): two-tailed t test, $t(19) = 1.03$, $p = 0.32$, whereas the suppressive effect of hearing any other sound was marginally significant (Figure 3a, right bar): $t(19) = -1.91$, $p = 0.07$.

We also conducted separate analyses of rivalry trials containing "other-paired" sounds (i.e., sounds that were always paired with a visual stimulus during the exposure phase that was neither of the presented rivalrous images on a given trial; for example, Sound 3 for the tilted gratings in Figure 1a) and rivalry trials containing "unpaired" sounds (i.e., sounds that were not consistently paired with a specific visual image but were presented together with one of the images of each rivalry pair on one third of the presentations of a given sound during exposure; for example, Sound 2 and the left-tilted grating in Figure 1a).

These analyses showed that an image was significantly more likely to be initially perceived in paired

sound trials compared to other-paired sound trials (two-tailed t test), $t(19) = 2.22$, $p < 0.05$. This indicates that paired sounds boost initial perceptual selection relative to other-paired sounds (which were never paired with, and were therefore irrelevant to, both images in the rivalry trial). In addition, an image was also more likely to be initially perceived in paired compared to unpaired sound trials, $t(19) = 3.26$, $p < 0.01$.

Previous work has demonstrated perceptual enhancement of items that are predictable (due to statistical learning) even when those items are not predicted on a given trial (Barakat, Seitz, & Shams, 2013). We therefore analyzed those rivalry trials in which one of the rivaling images had been paired with a sound during exposure and one had not, but the sound presented during rivalry was not the paired sound from the exposure phase. We found that the proportion of initial responses for the grating that had been paired with a different sound was indistinguishable from chance levels, $t(19) = -1.31$, $p = 0.21$.

In addition to assessing effects of cross-modal learning on the likelihood of initial perceptual selection, we also assessed its effects on initial response duration (Figure 3c) and initial response latency (Figure 3d). When analyzing initial response duration, we excluded responses that were truncated by the end of the trial (12.5% of all responses across participants). We found a significant effect of presentation of the paired sound, compared to all other sounds, on the mean duration of the initial response (Figure 3c; two-tailed t test): $t(19) = 3.44$, $p < 0.01$, Cohen's $d = 0.77$, 95% CI [0.10, 0.43]. It is of interest that cross-modal learning affects both the identity of the initial response and the duration of that response given that perceptual selection and maintenance in binocular rivalry have been shown to be dissociable (Bressler et al., 2013; Levelt, 1965; Silver & Logothetis, 2004). We found no significant effect of presentation of the paired sound, compared to other sounds, on the latency of initial responses (Figure 3d; two-tailed t test): $t(19) = -0.22$, $p = 0.83$, Cohen's $d = -0.05$, 95% CI [–0.14, 0.11].

Recognition memory test

To test participants' recognition memory of the AV pairings presented during the exposure phase, we measured the proportion of trials (for each of the three AV pairs) in which each subject correctly identified the image that was paired with the presented sound during the exposure phase. Mean recognition performance was 3.23 out of six ($SEM = 0.36$) or 54%. This recognition rate was significantly above chance (one out of six or 17%; two-tailed t test), $t(19) = 9.03$, $p < 0.0001$, Cohen's $d = 2.02$, 95% CI [0.41, 0.66], indicating that, on average, participants learned the AV pairings during the exposure phase. There was no significant effect of rivalry pair category on recognition performance (ANOVA), $F(2, 18) = 2.80$, $p = 0.08$, $\eta^2_p = 0.13$. In our sample of 20 subjects (Figure 4), only three had nearly perfect performance (an average of at least five out of six correct responses across the image pairs), indicating that they may have explicitly learned the pairings from the exposure phase. Even after removing these three subjects, the effect of statistical learning was maintained for both the proportion of initial rivalry responses, $t(16) = 2.58$, $p < 0.05$, Cohen's $d = 0.63$, 95% CI [0.01, 0.10], and initial response duration, $t(16) = 2.61$, $p < 0.05$, Cohen's $d = 0.63$, 95% CI [0.04, 0.40].

The correlation between recognition memory and the effect of statistical learning on proportion of initial rivalry responses across individual subjects was not significant (Figure 4a), $r(18) = 0.37$, $p = 0.11$, 95% CI [–0.08, 0.70]. There was a significant correlation between recognition memory and the effect of learning on the duration of the initial response (Figure 4b), $r(18) = 0.57$, $p < 0.01$, 95% CI [0.17, 0.81]. It is therefore possible that some degree of explicit learning contributed to the effects of cross-modal learning on binocular rivalry, at least for initial response duration. However, when we conducted a within-subject comparison of each participant's most explicitly learned versus least explicitly learned pair (with degree of explicit learning defined as performance on the recognition task), we found no significant difference between these two pairs in the effect of statistical learning on either the proportion of initial rivalry responses, $t(19) = 1.16$, $p = 0.26$, 95% CI [–0.08, 0.29], or initial response duration, $t(19) = -0.09$, $p = 0.93$, 95% CI [–0.44, 0.40].

Discussion

Here, we provide the first demonstration that recently, passively formed statistical associations between sounds and images impact what we initially see when the visual environment is ambiguous. Specifically, we found that a given image was more likely to be initially perceptually selected and maintained in awareness during binocular rivalry when it was preceded (and accompanied) by its paired sound from the exposure phase than by other sounds, indicating that cross-modal statistical learning influenced which of two competing images first reached conscious awareness.

Formation of multisensory associations is a high-level form of associative learning, requiring the integration of very different types of sensory information via potentially arbitrary mappings. Well-established, explicit multisensory associations (following

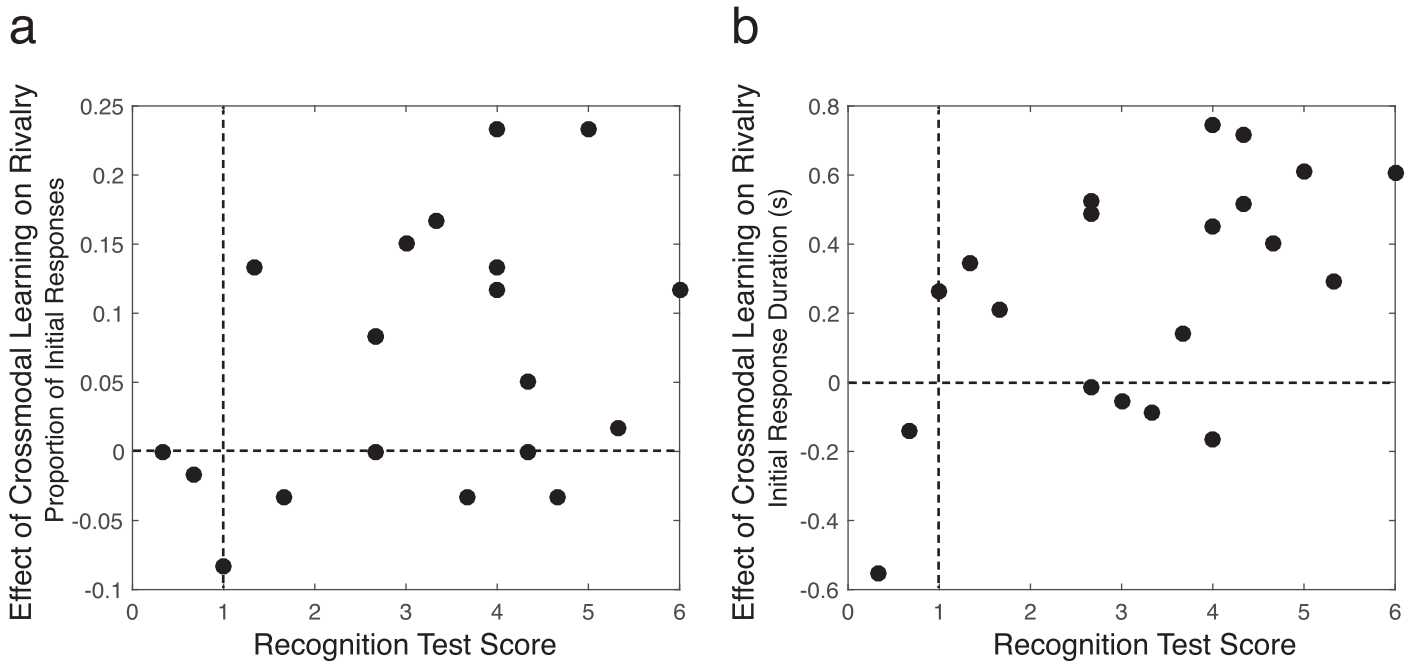

Figure 4. Relationship between recognition memory and effects of cross-modal learning on the (a) proportion and (b) duration of initial rivalry responses. Each point represents data from a single subject. Dotted lines indicate no effects of cross-modal learning (i.e., horizontal line indicates no effect of learning on rivalry and vertical line indicates chance performance on the recognition test). $N = 20$.

years of experience with the mapping) can bias the perceptual interpretation of ambiguous stimuli (Y.-C. Chen et al., 2011; Conrad et al., 2010; Kang & Blake, 2005; Sekuler, Sekuler, & Lau, 1997). Our findings demonstrate that even brief, passive exposure to novel arbitrary multisensory pairings influences visual perceptual selection. Our experimental procedures provided tight control over the associations that were formed as well as the timescale of learning, allowing us to directly link perceptual changes to rapid associative learning. A surprisingly brief (8-min) exposure was sufficient to observe these effects, indicating that cross-modal associative learning can continually update perceptual experience.

A recent paper (Einhäuser et al., 2017) reported effects of AV learning on the perceptual dominance of a grating during binocular rivalry as measured using optokinetic nystagmus. However, there are several critical differences between the learning procedures employed by Einhäuser et al. (2017) and in our study. First, we used a passive statistical learning paradigm (with no task during the exposure phase) whereas Einhäuser et al. induced associations using explicit training during exposure (subjects performed a go/no-go task in which they responded whenever the trained stimulus pair appeared and were required to achieve near-perfect accuracy). Our subjects, on the other hand, were not explicitly told about the pairings (or even that there were pairings), and as a group, they did not demonstrate strong explicit knowledge of the pairings as assessed by the recognition test. Second, Einhäuser et al.'s participants received substantially more exposure to the associated items: their training phase was longer (20 min vs. 8 min) and included fewer paired items (two colors with two pitches vs. three images and three sounds) with no neutral, nonassociated stimuli. Notably, their observers experienced between 192 and 240 repetitions of each multisensory pairing in total across multiple learning phases versus our 30 total repetitions per pairing.

Although we found that performance on the recognition test was significantly correlated with the effects of learning on initial response duration, a within-subject comparison yielded no significant difference between the initial response durations of each participant's most explicitly learned versus least explicitly learned pair. One explanation for this apparent discrepancy is that the between-subjects correlations captured individual differences in participants' overall learning ability, which may have influenced both implicit and explicit learning whereas the within-subject comparisons reflect relative differences across learned stimulus pairs within the same learner. Our results are also consistent with possible contributions of both implicit and explicit learning to participants' above-chance performance on the recognition memory test. Thus, further research (possibly involving additional explicit learning measures, including tests of free recall)

is needed to clarify the relative influences of explicit and implicit learning on perceptual selection.

The facilitative effect of statistical learning on binocular rivalry that we report here, in which an associated sound at the onset of rivalrous stimulus presentation influences initial visual competition, is consistent with previous evidence of influences of predictive information on perceptual selection (reviewed in Panichello, Cheung, & Bar, 2013). For example, briefly seeing an image increases its likelihood of being subsequently perceived during rivalry (Brascamp, Knapen, Kanai, van Ee, & van den Berg, 2007) as can merely imagining an image prior to rivalry (Pearson, Clifford, & Tong, 2008) or maintaining a "perceptual memory" trace of a dominant image across temporal gaps in stimulus presentation (X. Chen & He, 2004; Leopold, Wilke, Maier, & Logothetis, 2002). In addition, a grating is more likely to be initially selected during rivalry when the rivalrous pair is immediately preceded by a stream of rotating gratings whose motion trajectory predicts that grating (Attarha & Moore, 2015; Denison et al., 2011).

Statistical learning of natural image sequences has recently been found to reduce perceptual selection of statistically predicted images (Denison et al., 2016), an effect of visual statistical learning that is the opposite of what we found here for AV pairing. Differences between these studies include within-modality versus cross-modal associations, sequential versus concurrent presentation of the associated stimuli, and the use of complex natural images versus simpler geometric stimuli. Further studies are required to test the relative contributions of each of these factors to the effects of statistical learning on rivalry.

Although a few studies have demonstrated the impact of recent cross-modal learning on visual motion perception (Kafaligonul & Oluk, 2015; Teramoto, Hidaka, & Sugita, 2010), these studies did not manipulate the probability of AV pairings in a statistical learning paradigm or investigate perceptual selection from multiple stimuli competing for conscious awareness. Learning to associate a neutral face with negative gossip enhances dominance of that face during rivalry (Anderson, Siegel, Bliss-Moreau, & Barrett, 2011), but this effect is likely mediated by social and emotional saliency and not by multisensory learning per se. Our study is therefore the first to demonstrate that the process of selecting visual information for awareness is continually updated by new, passively learned statistical information regarding the relationships between images and sounds in our environment.

Although the effects we observed are reliable, they are small in magnitude and may require sensitive psychophysical measures or particular stimuli to observe. For instance, previous studies have shown that associative learning of arbitrary AV mappings—between a particular direction of 3-D rotation of a Necker cube and either auditory pitch (Haijiang, Saunders, Stone, & Backus, 2006) or mechanical sound identity (Jain, Fuller, & Backus, 2010)—did not bias visual perceptual interpretation of the Necker cube. Future studies should attempt to reconcile the effectiveness of multimodal statistical learning in biasing perceptual selection during binocular rivalry (as shown here and in Einhäuser et al., 2017) but not during Necker cube viewing. Possible explanations include differences between mapping sounds to images that are distinct enough to be represented as separate objects (rivalry studies) versus images that provide competing perspectives on the same object (Necker cube studies) and differences between auditory influences on interocular visual competition versus perceptual selection of bistable stimuli that are binocularly congruent.

As organisms gather information about the statistics of the natural world through sensory experience, they form and hone associations between sounds and images. The consistency of relationships between particular sounds and images in the environment modulates these associations, influencing the likelihood of predicting the presence of one stimulus based on the occurrence of another. For example, when we see a brown, furry animal far in the distance, as soon as we hear it bark, we are relieved to know it is more likely to be a dog than a bear. Our results show that rapid probabilistic learning can transform arbitrarily linked object features in different sensory modalities into automatic associations that can in turn shape perception and help resolve visual ambiguity.

In conclusion, we demonstrate that arbitrary associations between sounds and images that are acquired through brief passive statistical learning bias subsequent visual perceptual selection during binocular rivalry. These results suggest that statistical information about recently experienced patterns of sounds and images helps resolve ambiguities in sensory information by influencing competitive interactions between visual representations.

Keywords: statistical learning, multisensory integration, binocular rivalry, perceptual selection, conscious awareness

Acknowledgments

The authors thank Jacob Sheynin and Maxwell Schram for assistance with data collection. This work was supported by the Department of Defense through a National Defense Science and Engineering Graduate Fellowship and C.V. Starr Postdoctoral Fellowship awarded to Elise A. Piazza, NIH NEI F32 EY025533 and National Science Foundation Graduate Research

Fellowship awarded to Rachel N. Denison, and NEI Core grant EY003176.

Commercial relationships: none.
Corresponding author: Elise A. Piazza.
Email: elise.piazza@gmail.com.
Address: Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA.

References

Alais, D., & Blake, R. (Eds.). (2005). *Binocular rivalry*. Cambridge: MIT Press.

Anderson, E., Siegel, E. H., Bliss-Moreau, E., & Barrett, L. F. (2011, June 17). The visual impact of gossip. *Science*, *332*, 1446–1448.

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.

Attarha, M., & Moore, C. M. (2015). Onset rivalry: Factors that succeed and fail to bias selection. *Attention, Perception, & Psychophysics*, *77*, 520–535.

Barakat, B. K., Seitz, A. R., & Shams, L. (2013). The effect of statistical learning on internal stimulus representations: Predictable items are enhanced even when not predicted. *Cognition*, *129*, 205–211.

Blake, R., Sobel, K. V., & James, T. W. (2004). Neural synergy between kinetic vision and touch. *Psychological Science*, *15*, 397–402.

Blake, R., & Wilson, H. (2011). Binocular vision. *Vision Research*, *51*, 754–770.

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

Brascamp, J. W., Knapen, T. H., Kanai, R., van Ee, R., & van den Berg, A. V. (2007). Flash suppression and flash facilitation in binocular rivalry. *Journal of Vision*, *7*(12):12, 1–12, https://doi.org/10.1167/7.12.12. [PubMed] [Article]

Bressler, D. W., Denison, R. N., & Silver, M. A. (2013). High-level modulations of binocular rivalry: Effects of stimulus configuration, spatial and temporal context, and observer state. In S. M. Miller (Ed.), *The constitution of visual consciousness: Lessons from binocular rivalry* (pp. 253–280). Amsterdam, the Netherlands: John Benjamins.

Carter, O., & Cavanagh, P. (2007). Onset rivalry: Brief presentation isolates an early independent phase of perceptual competition. *PLoS One*, *2*, e343.

Chen, X., & He, S. (2004). Local factors determine the stabilization of monocular ambiguous and binocular rivalry stimuli. *Current Biology*, *14*, 1013–1017.

Chen, Y.-C., Yeh, S.-L., & Spence, C. (2011). Cross-modal constraints on human perceptual awareness: Auditory semantic modulation of binocular rivalry. *Frontiers in Psychology*, *2*, 212.

Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision*, *10*(10): 27, 1–15, https://doi.org/10.1167/10.10.27. [PubMed] [Article]

Conrad, V., Kleiner, M., Bartels, A., O'Brien, J. H., Bülthoff, H. H., & Noppeney, U. (2013). Naturalistic stimulus structure determines the integration of audiovisual looming signals in binocular rivalry. *PLoS One*, *8*, e70710.

Conrad, V., Vitello, M. P., & Noppeney, U. (2012). Interactions between apparent motion rivalry in vision and touch. *Psychological Science*, *23*, 940–948.

Denison, R. N., Piazza, E. A., & Silver, M. A. (2011). Predictive context influences perceptual selection during binocular rivalry. *Frontiers in Human Neuroscience*, *5*, 166.

Denison, R. N., Sheynin, J., & Silver, M. A. (2016). Perceptual suppression of predicted natural images. *Journal of Vision*, *16*(13):6, 1–15, https://doi.org/10.1167/16.13.6. [PubMed] [Article]

Einhäuser, W., Methfessel, P., & Bendixen, A. (2017). Newly acquired audio-visual associations bias perception in binocular rivalry. *Vision Research*, *133*, 121–129.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5):7, 1–14, https://doi.org/10.1167/7.5.7. [PubMed] [Article]

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499–504.

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology*, *22*, 383–388.

Haijiang, Q., Saunders, J. A., Stone, R. W., & Backus, B. T. (2006). Demonstration of cue recruitment: Change in visual appearance by means of Pavlovian conditioning. *Proceedings of the National Academy of Sciences, USA*, *103*, 483–488.

Jain, A., Fuller, S., & Backus, B. T. (2010). Absence of cue-recruitment for extrinsic signals: Sounds, spots, and swirling dots fail to influence perceived 3D

rotation direction after training. *PLoS One*, *5*, e13295.

Kafaligonul, H., & Oluk, C. (2015). Audiovisual associations alter the perception of low-level visual motion. *Frontiers in Integrative Neuroscience*, *9*, 26.

Kang, M.-S., & Blake, R. (2005). Perceptual synergy between seeing and hearing revealed during binocular rivalry. *Psichologija*, *32*, 7–15.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1–16.

Lee, M., Blake, R., Kim, S., & Kim, C.-Y. (2015). Melodic sound enhances visual awareness of congruent musical notes, but only if you can read music. *Proceedings of the National Academy of Sciences, USA*, *112*, 8493–8498.

Leopold, D. A., Wilke, M., Maier, A., & Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. *Nature Neuroscience*, *5*, 605–609.

Levelt, W. J. M. (1965). *On binocular rivalry*. Soesterberg, The Netherlands: Institute for Perception RVO-TNO.

Lumer, E. D., Friston, K. J., & Rees, G. (1998, June 19). Neural correlates of perceptual rivalry in the human brain. *Science*, *280*, 1930–1934.

Lunghi, C., & Alais, D. (2015). Congruent tactile stimulation reduces the strength of visual suppression during binocular rivalry. *Scientific Reports*, *5*, 9413.

Lunghi, C., Binda, P., & Morrone, M. C. (2010). Touch disambiguates rivalrous perception at early stages of visual analysis. *Current Biology*, *20*, 143–144.

Lunghi, C., & Morrone, M. C. (2013). Early interaction between vision and touch during binocular rivalry. *Multisensory Research*, *26*, 291–306.

Lunghi, C., Morrone, M. C., & Alais, D. (2014). Auditory and tactile signals combine to influence vision during binocular rivalry. *Journal of Neuroscience*, *34*, 784–792.

Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision*, *4*(7): 2, 539–551, https://doi.org/10.1167/4.7.2. [PubMed] [Article]

Mitchell, J. F., Stoner, G. R., & Reynolds, J. H. (2004, May 27). Object-based attention determines dominance in binocular rivalry. *Nature*, *429*, 410–413.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*, 61–64.

Panichello, M. F., Cheung, O. S., & Bar, M. (2013). Predictive feedback and conscious visual experience. *Frontiers in Psychology*, *3*, 620.

Pearson, J., Clifford, C. W., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology*, *18*, 982–986.

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

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

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

Salomon, R., Lim, M., Herbelin, B., Hesselmann, G., & Blanke, O. (2013). Posing for awareness: Proprioception modulates access to visual consciousness in a continuous flash suppression task. *Journal of Vision*, *13*(7):2, 1–8, https://doi.org/10.1167/13.7.2. [PubMed] [Article]

Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, *36*, 1445–1453.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997, January 23). Sound alters visual motion perception. *Nature*, *385*, 308.

Silver, M. A., & Logothetis, N. K. (2004). Grouping and segmentation in binocular rivalry. *Vision Research*, *44*, 1675–1692.

Teramoto, W., Hidaka, S., & Sugita, Y. (2010). Sounds move a static visual object. *PLoS One*, *5*, e12255.

van Ee, R., van Boxtel, J. J., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *Journal of Neuroscience*, *29*, 11641–11649.

Vidal, M., & Barrès, V. (2014). Hearing (rivaling) lips and seeing voices: How audiovisual interactions modulate perceptual stabilization in binocular rivalry. *Frontiers in Human Neuroscience*, *8*: 677.

Zhou, W., Jiang, Y., He, S., & Chen, D. (2010). Olfaction modulates visual perception in binocular rivalry. *Current Biology*, *20*, 1356–1358.