A simple and efficient method to enhance audiovisual binding tendencies

Brian Odegaard[1], David R. Wozny[1] and Ladan Shams[1,2,3]

[1] Department of Psychology, University of California, Los Angeles, Los Angeles, CA, United States
[2] Department of Bioengineering, University of California, Los Angeles, Los Angeles, CA, United States
[3] Neuroscience Interdepartmental Program, University of California-Los Angeles, Los Angeles, CA, United States

ABSTRACT

Individuals vary in their tendency to bind signals from multiple senses. For the same set of sights and sounds, one individual may frequently integrate multisensory signals and experience a unified percept, whereas another individual may rarely bind them and often experience two distinct sensations. Thus, while this binding/integration tendency is specific to each individual, it is not clear how plastic this tendency is in adulthood, and how sensory experiences may cause it to change. Here, we conducted an exploratory investigation which provides evidence that (1) the brain's tendency to bind in spatial perception is plastic, (2) that it can change following brief exposure to simple audiovisual stimuli, and (3) that exposure to temporally synchronous, spatially discrepant stimuli provides the most effective method to modify it. These results can inform current theories about how the brain updates its internal model of the surrounding sensory world, as well as future investigations seeking to increase integration tendencies.

INTRODUCTION

In many studies investigating human perception of concurrent visual and auditory stimuli, individuals exhibit different degrees of multisensory integration; for the same set of stimuli, some individuals integrate often, while others do not integrate at all (*Mcgurk & Macdonald, 1976*; *Hairston et al., 2003*; *Tremblay et al., 2007a*; *Stevenson, Zemtsov & Wallace, 2012*; *Nath & Beauchamp, 2012*; *Hass et al., 2012*; *Magnotti & Beauchamp, 2014*; *Mallick, Magnotti & Beauchamp, 2015*; *Gurler et al., 2015*). Recently, we have shown that this variability across individuals in sensory tasks is at least partially due to differences in the *tendency to bind* crossmodal signals (*Odegaard & Shams, 2016*). However, little is currently known about how experience may contribute to an individual's binding tendency in a given task. Additionally, while many studies have shown that interactions between the senses change over the course of development (*Neil et al., 2006*; *Tremblay et al., 2007b*; *Gori et al., 2008*; *Barutchu, Crewther & Crewther, 2009*; *Barutchu et al., 2010*; *Hillock, Powers & Wallace, 2011*; *Hillock-Dunn, Grantham & Wallace, 2016*), the degree to which an individual's binding tendency remains plastic in adulthood is unknown.

Corresponding author
Brian Odegaard,
odegaard.brian@gmail.com,
odegaard@ucla.edu

Therefore, we conducted the following investigation to explore two questions: (1) can brief sensory experiences alter the brain's tendency to bind audiovisual signals in adulthood; (2) If so, what are the spatial and temporal relationships between visual and auditory stimuli that cause it to change? In pursuit of answers to these questions, we investigated the plasticity of the binding tendency in the context of audiovisual spatial perception. Spatial perception is a fundamental brain function involved in many tasks that the brain has to perform frequently, and sophisticated computational tools have been developed that can be utilized for rigorously and quantitatively measuring both unisensory processing and the binding tendency in spatial tasks (*Körding et al., 2007*; *Wozny, Beierholm & Shams, 2010*; *Odegaard, Wozny & Shams, 2015*). Specifically, human performance in audiovisual spatial tasks is modeled quite effectively by Bayesian causal inference, where an observer's perception of simultaneous audiovisual signals depends not only on the reliability of unisensory encoding (modeled by sensory likelihoods), but also upon the inference of whether a common cause is the source of both the auditory and visual signals (which is influenced by a prior bias for binding the senses) (*Körding et al., 2007*; *Beierholm, Quartz & Shams, 2009*; *Wozny, Beierholm & Shams, 2010*; *Rohe & Noppeney, 2015a*; *Odegaard, Wozny & Shams, 2015*; *McGovern et al., 2016*; *Odegaard, Wozny & Shams, 2016*; *Odegaard & Shams, 2016*). This prior governing integration and segregation, referred to as the "binding tendency" (*Odegaard, Wozny & Shams, 2016*; *Odegaard & Shams, 2016*) is quantifiable, stable over time, and provides a true measure of the brain's capacity to bind audiovisual information which is not confounded by unisensory encoding (*Odegaard & Shams, 2016*); however, it remains an open question of how quickly and under what conditions it may change.

Previous research has indicated that some forms of audiovisual spatial processing are highly plastic and update on a scale of milliseconds (*Wozny & Shams, 2011a*), with these changes best accounted for by shifts in sensory likelihoods (*Wozny & Shams, 2011b*). As sensory spatial processing appears to be modifiable in adulthood by short-term sensory experience (*Chen & Vroomen, 2013*), we investigated whether the binding tendency could also be modified by brief exposures to audiovisual stimuli.

To explore this question, we conducted six different experiments to gain insights into the conditions that could cause an update of this prior. In all experiments, the basic design was as follows: first, the binding tendency for each observer was quantified by having subjects perform a spatial localization task with a variety of stimulus conditions, which included both unisensory (auditory or visual) and bisensory (simultaneous audiovisual) stimuli (Fig. 1A). Next, the observers were presented with a brief "exposure phase" in which they were passively exposed to stimuli with various spatial and temporal relationships. Figure 1B shows an example arrangement of stimuli during the exposure phase, and Table 1 shows how the temporal and spatial relationships between the auditory and visual stimuli in the exposure phase varied across the six experiments (but see Materials and Methods for additional details).

The main part of the exposure phase lasted approximately ten minutes. Finally, the participants completed the same localization task as the one prior to exposure, to quantify any changes in the binding tendency following the exposure phase. To avoid having the effects of the main exposure phase wash out over this period, we employed brief refresher
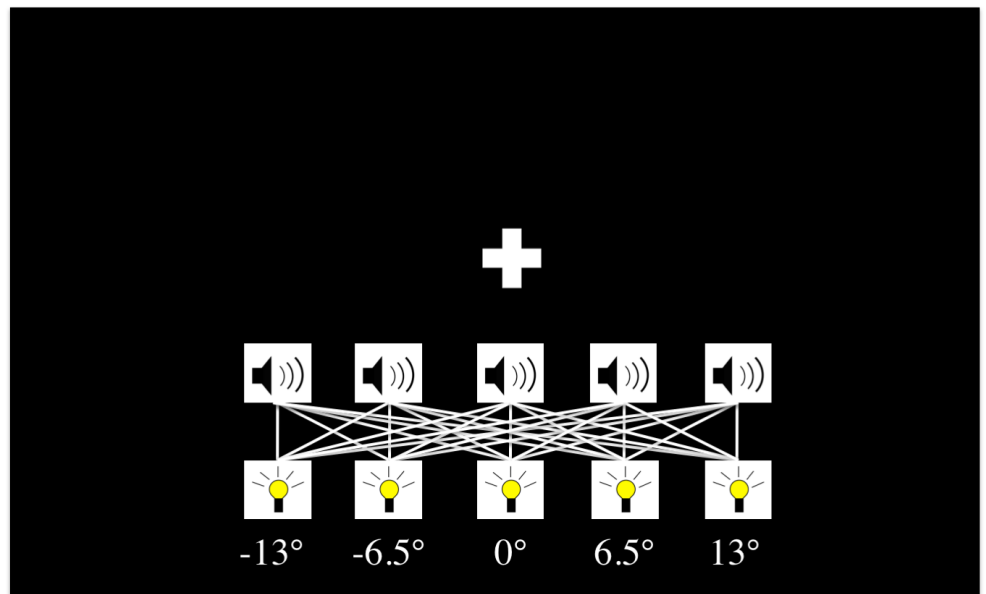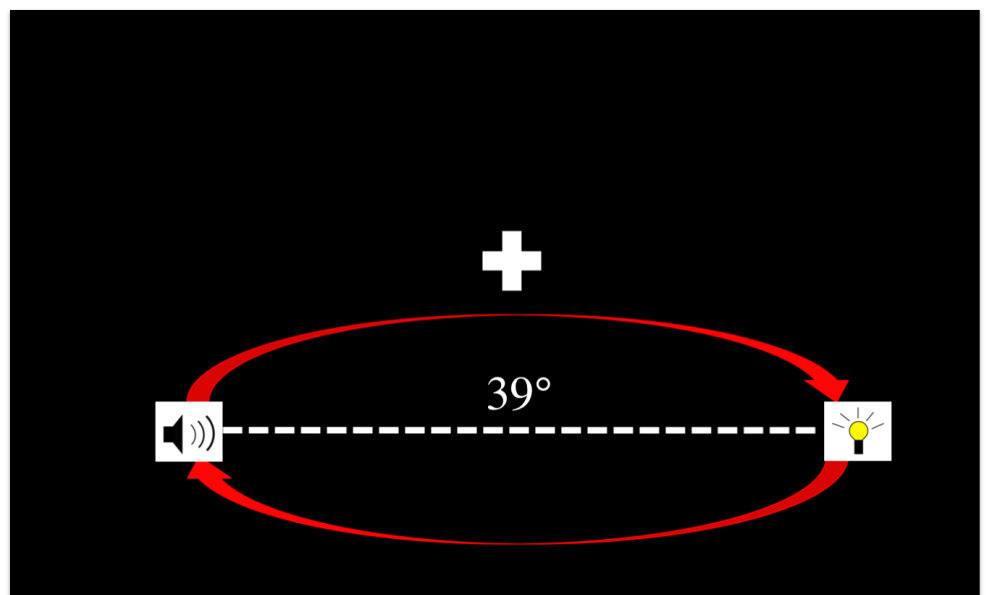
Figure 1 **The spatial localization paradigm and exposure tasks.** (A) The localization task. Stimuli could be presented from one of five locations, ranging from −13 to +13 degrees. On any given trial, subjects could be presented with only an auditory stimulus, only a visual stimulus, or both an auditory and visual stimulus. The white lines in the figure designate all of the different spatial combinations of the bisensory trials, and the white "+" symbol denotes the fixation cross. (B) The exposure task paradigm. This was interleaved with localization blocks in the post-test phase. Depicted above is the stimulus arrangement from experiments #2, #5, and #6 where auditory and visual stimuli were always presented from 39 degrees apart, and alternated positions with each correct detection of the brighter flash. In experiments #3 and #4, the auditory and visual stimuli always were presented from the same location (−19.5 or +19.5 degrees), and alternated together from one side to the other. In experiment #1, no exposure task was implemented, and subjects simply listened to music during blocks when the normal exposure task occurred in other experiments.

Table 1 **Spatial and temporal relationships of exposure-task stimuli in the six experiments.** The "time" column denotes whether the stimuli in the exposure task were temporally congruent (i.e., synchronous), uncorrelated (with timing of the auditory stimulus drawn from a uniform distribution of −250 ms to +250 ms of when the visual stimulus occurred) or correlated (always separated by a fixed, 400 ms offset, with the visual stimulus preceding the auditory stimulus). The "space" column refers to whether exposure phase stimuli were incongruent (presented from 39 degrees apart, alternating positions with each correct detection of the brighter flash), or congruent (presented from the same location, alternating sides with each correct detection of the brighter flash).

Experiment	Time	Space
1	None	None
2	Uncorrelated	Incongruent
3	Congruent	Congruent
4	Uncorrelated	Congruent
5	Congruent	Incongruent
6	Correlated	Incongruent

blocks of the exposure task in a "top-up" design every 40 trials in the post-test localization session. By comparing differences in the binding tendency in the localization task between the pre-test and post-test phases, we could examine whether any of the exposures caused the brain to update its tendency to bind spatially discrepant audiovisual signals in the post-test localization phase.

In our simple spatial localization task, two dimensions are relevant in the brain's inference of whether or not signals should be integrated or segregated: *space*, and *time*. In each experiment's exposure phase, subjects passively observed audiovisual stimuli with various spatial and temporal relationships. Given that brief flashes and sounds that are presented simultaneously and at the same location are generally perceived to have a common cause (*Wallace et al., 2004*; *Rohe & Noppeney, 2015b*), repeated exposure to spatio-temporally congruent flashes could potentially serve as a cue to update the binding tendency. Similarly, given that flashes and sounds with a large spatial discrepancy or temporal discrepancy are often perceived as originating from independent sources, this could also serve as a cue to update the binding tendency (*Wallace et al., 2004*; *Rohe & Noppeney, 2015b*). While it is not immediately clear *how* these cues would serve to update the binding tendency, as the conditions necessary to cause an update in multisensory priors are currently unknown, general ideas from sensory research motivated our two hypotheses about when sensory experience may cause our internal models of the world to change.

First, research on statistical learning in human observers has shown that exposure to sensory stimuli with statistical regularities (e.g., joint probabilities, conditional probabilities) causes learning of the statistical regularities (*Saffran, Aslin & Newport, 1996*; *Fiser & Aslin, 2016*; *Fiser & Aslin, 2002*). If the binding tendency, which is equivalent to the *a priori* probability of a common cause, is indeed plastic and can be updated by brief experience, it may change to ultimately mirror the statistics of the environment. Therefore, it seems possible that repeated exposure to sounds and sights that are congruent in the relevant dimensions for sensory inference in this task (space and time) would indicate these signals come from a common cause, and should therefore result in an *increase* in the

binding tendency (*Parise, Spence & Ernst, 2012*). Conversely, repeated exposure to sounds and sights that are incongruent in a given dimension would indicate these signals come from separate sources, and should result in a *decrease* in the binding tendency (Fig. 2A).

However, another prominent idea in sensory perception research, predictive coding, presents an alternative idea about when the brain may update priors that influence perception of the world. One of the main tenets of predictive coding is that mismatches between what the brain predicts and actual sensory experience leads to an updating of the brain's internal model of the world (*Rao & Ballard, 1999*; *Friston, 2003*; *Friston, 2012*; *Talsma, 2015*). According to its internal model of the world, the brain would not predict that sensory signals that are incongruent in a given dimension would have a common cause and would repeatedly co-occur. However, when repeatedly presented with temporally coincident but otherwise discrepant stimuli, enough evidence is provided to the brain to signal an association between the stimuli (i.e., a common cause), and this mismatch between the expectation/prediction and sensory experience produces an error signal. According to this idea, the nervous system would modify its internal model of the world to allow for these discrepant signals to be perceived as unified (Fig. 2B).

Thus, interestingly, these ideas lead to different predictions about when the binding tendency may get updated as a result of our exposure tasks. For example, shown in Fig. 2C is our first hypothesis. In this framework, the internal model of the world is a generative model of how sensory signals are produced by causes in the outside world. For any two sensory signals (x_A and x_V), they are either caused by the same source (s), or by two independent sources (s_A and s_V). The prior probability of the common cause hypothesis is the "binding tendency," which is the *a priori* bias of the nervous system to expect a common cause for any two sensory (here, audiovisual) signals. In Fig. 2C, the light blue box depicts four (out of many) possible types of sensory experience, and these correspond to the four types of exposures used in the current study across the six experiments. In this framework, the statistics of sensory stimuli with a common cause would lead to an increase of prior for the common cause (or binding tendency), and vice versa. On each exposure-phase stimulus presentation where the audiovisual stimulus leads to inference of a common cause (because they are spatiotemporally congruent), it would slightly increase the binding tendency (shown by the red arrow). On each exposure presentation where the audiovisual stimulus leads to an inference of two causes (because the audiovisual stimuli are too discrepant in either space or time), it would slightly decrease the binding tendency (shown by the black arrow). Over the course of many trials, these changes would accumulate. The change in the binding tendency would occur on every single trial, but the change would accumulate across trials and become more evident as more experience is acquired.

However, our second hypothesis (as shown in Fig. 2D) makes different predictions about when changes should occur to the binding tendency. For example, neurophysiological studies provide evidence that *time* may be a more important factor than *space* in driving plasticity of multisensory integration capacities, and that strong temporal associations between audiovisual signals can cause an increase in the capacity to integrate spatially discrepant stimuli (*Wallace & Stein, 2007*). Therefore, in our experiments, only one type
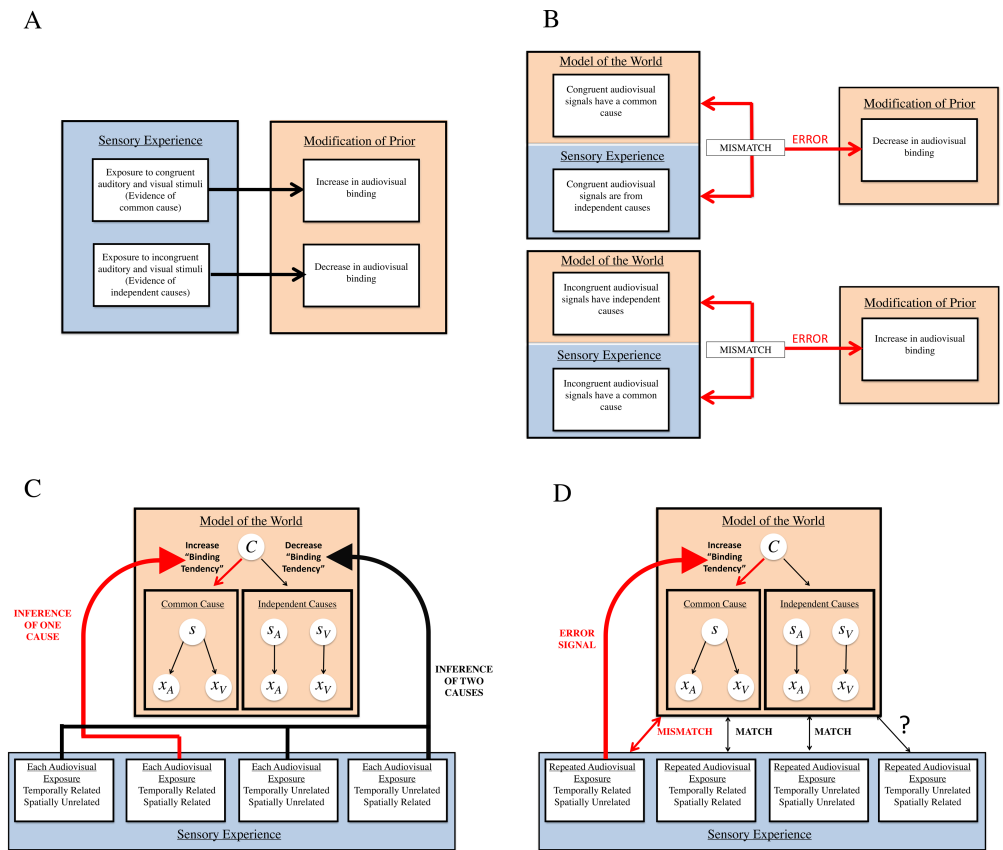
Figure 2 **Our two hypotheses of how the binding tendency may change: general predictions, and specific applications to Bayesian causal inference.** (A) Hypothesis #1, general predictions: priors are updated to reflect statistics of stimuli. Within our first hypothesis, all types of sensory experience are informative of relationships between auditory and visual signals in the environment, and therefore should cause a change in audiovisual binding. For instance, exposure to audiovisual stimuli that are *congruent* (either in space or in time) should lead to an increase in audiovisual binding, as the brain learns the association between the stimuli and updates its model of the world. Similarly, exposure to *incongruent* audiovisual stimuli should lead to a decrease in audiovisual binding. (B) Hypothesis #2, general predictions: mismatch between internal model and sensory experience requires modification. Within our second hypothesis, *mismatches* between the current model of the world and sensory experience are most informative regarding novel relationships between auditory and visual signals that the brain must take into account. Thus, whenever sensory experience indicates that congruent audiovisual stimuli actually originate from independent causes, this should change the brain's internal model to decrease audiovisual binding. Similarly, if experience indicates that incongruent audiovisual stimuli actually originate from a common cause, the brain should increase binding of audiovisual information. (C) Hypothesis #1 and Bayesian causal inference. In the Bayesian causal inference framework, according to the first hypothesis, each audiovisual exposure serves to update the prior for binding (i.e., binding tendency). Evidence of a common cause by each exposure to congruent signals accumulates in this prior, increasing the tendency to bind, and evidence of independent causes by each exposure to incongruent signals decreases the tendency to bind. (D) Hypothesis #2 and Bayesian causal inference. According to our second hypothesis, the brain may only update its internal model of the world when there is a mismatch between sensory evidence and the current model. In our experiments, only conditions which present repeated evidence (i.e., strong temporal relationships between audiovisual stimuli) that signals thought to originally have independent causes actually originate from a common cause produce an error signal, and thus require an update of the internal model of the world. As evidence indicates that spatial relationships are not as strong of a cue as temporal relationships in updating integration tendencies (*Wallace & Stein, 2007*), it is unlikely that repeated presentation of temporally unrelated and spatially related stimuli would be potent enough to modify the brain's tendency to bind.

of sensory experience (which was used in Experiment 5 and Experiment 6), wherein the auditory and visual signals are presented with strong temporal consistency and spatial inconsistency, would violate the predictions of the internal model. Repeated temporal coincidence of audiovisual signals can provide strong evidence of unity, which can be at odds with large spatial discrepancies. Because of the mismatch between the model prediction and the sensory data over repeated trials, an error signal is produced, which is then used to modify the internal model by adjusting the binding tendency. This process is shown by the red arrow. By increasing the binding tendency, the nervous system becomes more tolerant towards spatial discrepancies between the two sensory signals in perceiving unity. Alternatively, signals that are either fully related or unrelated do not produce any mismatch with repeated presentation, and thus an update of the internal model of the world is not warranted.

MATERIALS AND METHODS

Participants

A total of 189 subjects (ages 18–55) participated in this study. Twenty-one subjects were excluded due to either negligence with the response device or eyetracker malfunctions during the task. Thus, analyses for each experiment included between 22–31 subjects. A subset of the "pre-test" data in these experiments was used in a previous paper to compare the performance of various Bayesian models to account for multisensory spatial localization abilities (*Odegaard, Wozny & Shams, 2015*). All participants verbally reported that they did not have a history of any neurological conditions (seizures, epilepsy, stroke), had normal or corrected-to-normal vision and hearing, and had not experienced head trauma. The University of California North General IRB granted ethical approval to carry out the study in our lab (IRB#13-000476), and each subject provided written consent on the consent form approved by the UCLA IRB.

Apparatus and stimuli

Eligible participants sat at a desk in a dimly lit room with their chins positioned on a chinrest 52 cm from a screen. The screen was a black, acoustically transparent cloth subtending much of the visual field (134° width° × 60° height). Behind the screen were seven free-field speakers (5 × 8 cm, extended range paper cone), positioned along azimuth 6.5° apart, 7° below where the fixation cross was displayed. The middle speaker was positioned below the fixation point, and three speakers were positioned to the right and three to the left of fixation.

The visual stimuli were presented overhead from a ceiling mounted projector set to a resolution of 1,280 × 1,024 pixels with a refresh rate of 75 Hz. The visual stimuli used in the experiments were white-noise disks (.41 cd/m2) with a Gaussian envelope of 1.5° FWHM, presented 7° below the fixation point on a black background (.07 cd/m2), for 35 ms. The center of visual stimuli overlapped the center of one of the five speakers behind the screen positioned at −13°, −6.5°, 0°, 6.5°, and 13°. Auditory stimuli were ramped white noise bursts of 35 ms measuring 59 dB(A) sound pressure level at a distance of 52 cm. The speaker locations were unknown to all participants in the experiment.

Prior to the presentation of any stimuli in the experiment, participants were required to have their gaze centered on a central fixation point, which was a small, white cross. To ensure that participants' gaze for each trial was starting from the same location, gaze position and fixation time were recorded at 60 Hz with a ViewPoint eye tracker (Arrington Research, Scottsdale, AZ, USA) fixed to the chinrest and PC-60 software (version 2.8.5,000). Stimuli were not displayed until the recorded gaze angle was within 3.0° of the fixation point and the fixation time was greater than 250 ms. Viewing of the stimuli was binocular, although only movements of the right eye were tracked. The eye tracker was adjusted for each participant before the test session to ensure that the entire eye was being monitored, and a calibration task was performed before trials for the experiment began. A separate computer controlled stimuli presentation and recorded behavioral responses using MATLAB (version 7.6.0, R2008a). A wireless mouse was used to record the behavioral responses.

Bayesian causal inference model

Following collection of each individual subject's data, responses were divided into the pre-test and post-test datasets, and a Bayesian Causal Inference model with 8 free parameters (*Odegaard, Wozny & Shams, 2015*; *Odegaard, Wozny & Shams, 2016*) was fit to the data (see Supplemental Information for further details). The free parameters included the "binding tendency" (b_t), which in this Bayesian framework is the prior probability of inferring a common cause, a prior distribution over space characterized by a mean (x_p) and variance (σ_p), and five parameters characterizing the sensory likelihood distributions. These likelihood parameters included the standard deviation of the visual likelihood (σ_v), how the visual likelihood variance scales with eccentricity ($\Delta\sigma_v$), how bias in the visual likelihood mean scales with eccentricity (Δx_v), the standard deviation of the auditory likelihood distributions (σ_a), and how bias in the auditory likelihood mean scales with eccentricity (Δx_a).

We included these likelihood parameters to determine whether any changes in participants' localization behaviors were due to a change in *binding tendency* (i.e., the prior probability of inferring a common cause, or "p_{common}"), or a change in the sensory representations of space. Previously, it has been well established that repeated exposure to spatially discrepant audiovisual stimuli can result in subsequent auditory localizations being mislocalized towards the direction of the previous visual stimulus (*Canon, 1970*; *Radeau & Bertelson, 1977*; *Recanzone, 1998*; *Lewald, 2002*; *Frissen et al., 2003*; *Frissen et al., 2005*; *Bertelson et al., 2006*; *Frissen, Vroomen & De Gelder, 2012*), and that this effect can be accounted for by a shift in the auditory likelihoods (*Sato, Toyoizumi & Aihara, 2007*; *Wozny & Shams, 2011b*). In order to allow for this possible explanation of the data, we needed to include these additional parameters to determine whether any potential changes induced by the exposure task were due to changes in *sensory representations*, or changes in the *binding tendency*. In addition to the aforementioned parameters, the perceptual decision-making strategy (probability matching strategy, model averaging strategy, or model selection strategy (*Wozny, Beierholm & Shams, 2010*)) was also fitted to the data for each individual participant, and parameters from the best-fitting strategy were analyzed.
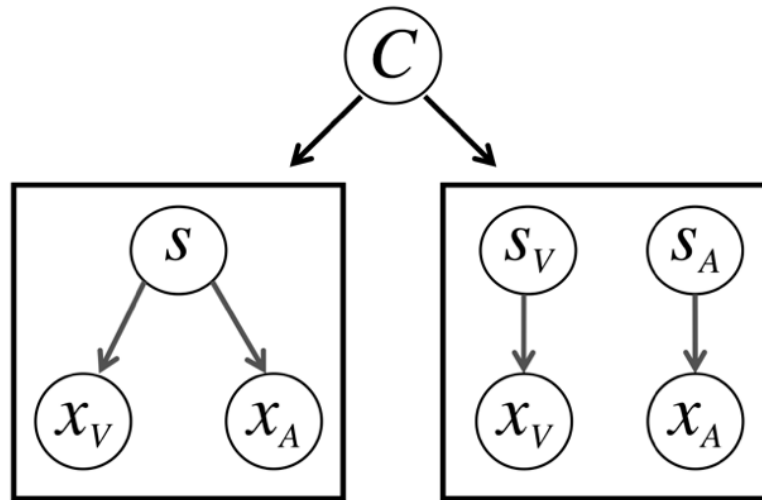
Figure 3 The causal inference model. On some trials, the observer may infer that both the auditory (x_A) and visual (x_V) signals come form a single source (s) and should be integrated. On other trials, the observer may infer that the signals come from two sources (s_V and s_A) and should be segregated. The latent variable C determines which model generates the data.

Shown in Fig. 3, the general idea behind the model is that in this task, observers only have access to sensory signals x_A and x_V, and must determine whether or not the signals come from one source and should be integrated, or come from two sources and should be segregated.

The noisy sensory signals that the subjects have access to are modeled by likelihood distributions for the visual and auditory modalities. The means for these likelihood distributions (x_A and x_V) are sampled from a normal distribution around the true location of the stimulus (s_A and s_V) to simulate the corruption of sensory channels by the noise parameters σ_V and σ_A, which capture the inherent precision of encoding in each sensory modality. A delta term is also included to capture biases in the two sensory channels' representations. This delta terms scales with eccentricity, based on the spatial position K of each stimulus; each value of K corresponds to a specific spatial position (i.e., $-13°$ corresponds to -2, $-6.5°$ corresponds to -1, $0°$ corresponds to 0, $6.5°$ corresponds to 1, and $13°$ corresponds to 2). If the delta term is negative, this will account for central biases in the sensory representations; if the delta term is positive, this will account for peripheral biases in the sensory representations. We also include a delta term for the visual likelihood noise that scales with stimulus eccentricity, as shown in the equations below:

$$x_V \sim N(s_V + (\Delta x_V K), \sigma_V + |\Delta \sigma_V K|) \tag{1}$$
$$K = \{-2, -1, 0, 1, 2\}$$
$$x_A \sim N(s_A + (\Delta x_A K), \sigma_A) \tag{2}$$
$$K = \{-2, -1, 0, 1, 2\}.$$

In this model, observers' final estimates are not only influenced by noise in the sensory likelihoods and biases in how those distributions are represented, but also by pre-existing

a priori biases over space. Thus, we also include a spatial prior, as this element has been shown to significantly improve the model fits.

$$p(s) = N(\mu_p, \sigma_P) \tag{3}$$

The likelihoods and spatial prior are combined to compute the posterior probability of a sensory event. The more similar the sensory signals are, the more likely scenario is that they have originated from the same source and thus should be integrated; with more discrepant sensory signals, it is more likely they originated from separate sensory signals and thus should be segregated. Thus, the posterior probability of a sensory event *s* is conditioned on the causal structure of the stimuli, with the optimal combinations according to each causal scenario listed below:

$$p(s|x_A, x_V; C=1) = \frac{p(x_A|s)p(x_V|s)p(s)}{p(x_A, x_V)} \tag{4}$$

$$p(s_A|x_A; C=2) = \frac{p(x_A|s_A)p(s)}{p(x_A)} \tag{5}$$

$$p(s_V|x_V; C=2) = \frac{p(x_V|s_V)p(s)}{p(x_V)} \tag{6}$$

The maximum a posteriori estimate of these computed posteriors reflects the optimal combination of the sensory signals given one or two causes. (Note that Eqs. (5) and (6) are also used to model the unisensory visual and unisensory auditory data.) Since subjects do not know whether or not the signals were generated by one or two sources, this must be inferred, incorporating both the sensory likelihoods and the prior information as well. This process of inference can be computed according to Bayes' Rule:

$$p(C|x_A, x_V) = \frac{p(x_A, x_V|C)p(C)}{p(x_A, x_V)} \tag{7}$$

In order to compute the posterior probability of a single cause, this can be done in the following manner, with p_c denoting the prior probability of the common cause. This prior probability of a common cause, p_c, is the measure of interest in this manuscript (i.e., it is the binding tendency). Values near 1 indicate that nearly all bisensory stimuli, regardless of spatial discrepancy, will be integrated, while values near 0 indicate that nearly all bisensory stimuli will be segregated:

$$p(C=1|x_A, x_v) = \frac{p(x_A, x_V|C=1)p_c}{p(x_A, x_V|C=1)p_c + p(x_A, x_V|C=2)(1-p_c)} \tag{8}$$

In this equation, the likelihood terms can be computed by integrating over the latent variable *s*:

$$p(x_A, x_V|C=1) = \int p(x_A|s)p(x_V|s)p(s)ds \tag{9}$$

$$p(x_A, x_V|C=2) = \int p(x_A|s_A)p(s_A)ds_A \cdot \int p(x_V|s_V)p(s_V)ds_V \tag{10}$$

Following integration over the latent variable *s*, the posterior probability of a common cause can be computed, and thus, the posterior probability of two causes can also be

computed:

$$p(C=2|x_A,x_V)=1-p(C=1|x_A,x_V). \tag{11}$$

Finally, it must be determined how this inferred posterior probability can be used to weight the various sensory estimates. Three possible perceptual strategies exist, as shown in the following section:

Model Selection:

$$\hat{s}_A = \begin{cases} \hat{s}_{A,c=1} \text{ if } p(C=1|x_A,x_V) > .5 \\ \hat{s}_{A,c=2} \text{ if } p(C=1|x_A,x_V) \le .5 \end{cases}$$

$$\hat{s}_V = \begin{cases} \hat{s}_{v,c=1} \text{ if } p(C=1|x_A,x_V) > .5 \\ \hat{s}_{V,c=2} \text{ if } p(C=1|x_A,x_V) \le .5 \end{cases} \tag{12}$$

Model Averaging:

$$\hat{s}_A = p(C=1|x_A,x_V)\hat{s}_{A,C=1} + p(C=2|x_A,x_V)\hat{s}_{A,C=2}$$
$$\hat{s}_V = p(C=1|x_A,x_V)\hat{s}_{V,C=1} + p(C=2|x_A,x_V)\hat{s}_{V,C=2} \tag{13}$$

Probability Matching:

$$\hat{s}_A = \begin{cases} \hat{s}_{A,c=1} \text{ if } p(C=1|x_A,x_V) > \xi \\ \text{where } \xi \in [0:1] \text{ uniform distribution} \\ \hat{s}_{A,c=2} \text{ if } p(C=1|x_A,x_V) \le \xi \\ \text{and sampled on each trial} \end{cases}$$

$$\hat{s}_V = \begin{cases} \hat{s}_{V,c=1} \text{ if } p(C=1|x_A,x_V) > \xi \\ \text{where } \xi \in [0:1] \text{ uniform distribution} \\ \hat{s}_{V,c=2} \text{ if } p(C=1|x_A,x_V) \le \xi \\ \text{and sampled on each trial.} \end{cases} \tag{14}$$

For each subject, we fit eight free parameters to individual subjects' data, and also determined whether subjects data were best classified by a probability matching strategy, model averaging strategy, or model selection strategy (*Wozny, Beierholm & Shams, 2010*, see study #1 for details). This was done simulating 10,000 trials using MATLAB's *fminsearchbnd* function (*Mathworks, 2009*), maximizing the likelihood of the parameters of the model. The parameter fits from the best-fitting strategy for each experiment are shown in the tables in the Supplemental Information.

Procedure

Participants began each session by being exposed to 10 practice trials with only unimodal auditory stimuli presented. This practice session ensured that participants were accurately using the mouse, understood the instructions, and were accurately fulfilling the fixation requirements for each trial. Each trial started with the fixation cross, followed after 750 ms (if the subject was fixating properly) by the presentation of stimuli. 450 ms after the stimuli, fixation was removed and a cursor appeared on the screen just above the horizontal line where the stimuli were presented, and at a random horizontal location in order to minimize response bias. The cursor was controlled by the trackball mouse placed

in front of the subject, and could only be moved in the horizontal direction (i.e., to the left and right). Participants were instructed to "move the cursor as quickly and accurately as possible to the exact location of the stimulus and click the mouse." This enabled the capture of continuous responses with a resolution of 0.1 degree/pixel. No feedback about the correctness of responses was given. Participants were allowed to move their eyes as they made their response.

Following the brief practice session, participants began the "pre-test" session, which consisted of 525 trials of interleaved auditory, visual, and audiovisual stimuli presented in pseudorandom order, which took about 45 min to complete, including break screens that would arise approximately every 9–10 min. The stimulus conditions included five unisensory auditory locations, five unisensory visual locations, and 25 combinations of auditory and visual locations (bisensory conditions), for a total of 35 stimulus conditions (15 trials per condition). In the bisensory conditions, the stimuli could either be spatially congruent (composing five of the 25 bisensory conditions) or spatially incongruent (composing 20 of the 25 bisensory conditions).

The localization task required subjects to localize all stimuli that were presented on every trial. Specifically, on unisensory auditory trials, subjects had to localize the sound; on unisensory visual trials, subjects had to localize the visual stimulus; on bisensory trials, observers had to localize both the visual *and* the auditory stimulus. Subjects were instructed before the start of the experiment that on bisensory trials, "sometimes the sound and light come from the same location, but sometimes the sound and light come from different localizations." On these bisensory trials, when the response cursor first appeared, it included a letter inside the cursor that denoted which stimulus the subject should localize first. For half of the subjects, following presentation of bisensory stimuli, the cursor first appeared with a red "S" inside it to indicate the sound was to be localized first; for the other half of subjects, the cursor first appeared with a blue "L" to indicate the light was to be localized first. Following the subjects' first localization response, the cursor immediately randomized to a new position and displayed the other letter. Thus, on every bisensory trial, we obtained subjects' estimates for the spatial locations of both the visual and auditory stimuli. While the response order was always the same (e.g., first localize the sound, then light, or vice versa) for a given subject, the order was counterbalanced across subjects, and our previous studies have revealed that response order does not impact localization responses in this task.

Following the "pre-test" session, participants were allowed to take a 10-minute break, and then began the "post-test" session, which employed a "top-up" design with interleaved blocks of an exposure task, and post-test localization trials. The first part of the exposure task lasted for approximately ten minutes, and then would re-appear for brief two-minute sessions every 40 localization trials in the post-test localization phase.

In the exposure task, in experiments #2–#6, a visual flash and auditory burst of sound were presented (on average) every 600 ms. Between every 5th to 15th presentation of the stimuli, the visual flash would get noticeably brighter (from 0.41 to 3.2 cd/m^2), and participants would have to detect this change in brightness by clicking the mouse. If they

correctly detected the change, the stimuli would change spatial positions, with the visual and auditory stimuli switching to opposite sides of the midline.

The *spatial* relationships between the repeated visual and auditory stimuli in the exposure task were as follows: in three experiments (experiments 2, 5, and 6), the visual and auditory stimuli were *incongruent* and presented from spatially discrepant locations (either $+19°$ for vision and $-19°$ for audition, or $+19°$ for audition and -19 for vision, depending on the trial). In two experiments (experiments 3 and 4), they were presented from the same location (either $+19°$ and $-19°$), and alternated the side they were presented on at the same time to maintain spatial congruency. If participants missed the brightness change, they had to wait through another cycle for a brighter flash to be presented. Participants were forced to maintain fixation at $0°$, and if fixation was broken (as detected by the eyetracker), the stimuli paused until fixation at the proper location was regained.

The *temporal* relationships between the repeated visual and auditory stimuli in the exposure task were the following: in two of the experiments (experiment #2 and #4), the temporal stimuli were "uncorrelated" during the exposure task, as the auditory stimulus was drawn from a uniform distribution from -250 ms to $+250$ ms of when the visual flash was presented. In experiments #3 and #5, the visual and auditory stimuli were "congruent" and were presented synchronously. In experiment #6, the visual stimulus always preceded the auditory stimulus by 400 ms, and thus was perfectly "correlated" with the auditory signal, but was always presented with a fixed temporal discrepancy. (See Table 1 for a summary of the spatial and temporal relationships between stimuli in different experiments).

In the post-test phase, the initial exposure task block lasted for 35 correct detections (approximately ten minutes), and was followed by a localization block consisting of 40 trials. Then, for the remainder of the post-test phase, exposure-task blocks (requiring eight detections) and localization blocks (40 trials) alternated until 525 localization trials were completed. Typically, the post-test phase lasted between 60 and 75 min.

RESULTS

We fit a Bayesian causal inference model (*Odegaard, Wozny & Shams, 2015*; *Odegaard, Wozny & Shams, 2016*) with eight free parameters (see Materials and Methods for details, and the Supplemental Information for the parameter tables) to each individual subjects' data to acquire an estimate of the binding tendency in the pre-test and post-test localization phases. The average fit for each model across subjects was quite reliable for both the pre-test and the post-test datasets (Pre-Test R^2 Mean $= .82$, SD $= .19$; Post-Test R^2 Mean $= .81$, SD $= .24$) (*Nagelkerke, 1991*).

Visualization of changes in the binding tendency on a subject-by-subject basis can be informative to reveal both overall trends and heterogeneity in the potential effects produced by different experiments. Figure 4 displays how the binding tendency for each individual subject changed in each experiment. As can be seen in the figure, with the exception of a few outliers, binding tendency values appear relatively stable across sessions in experiments 1–4. (For a violin plot of this same data showing means and medians of the binding
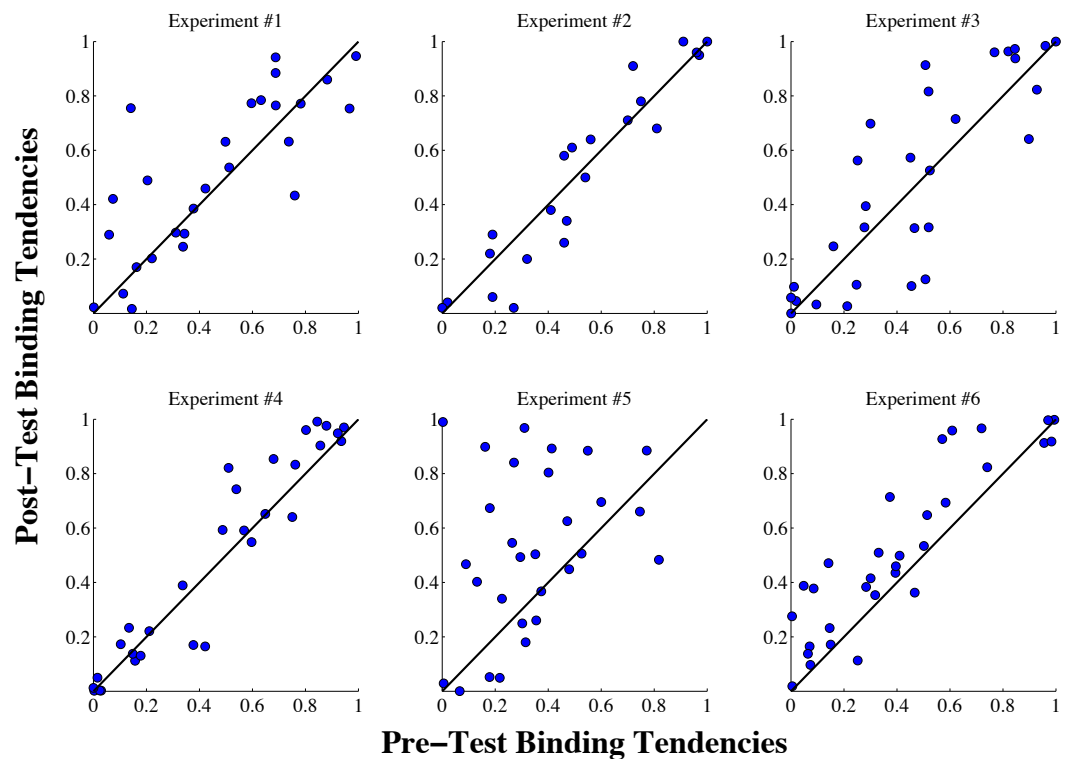
Figure 4 **Individual binding tendency scores from the pre-test and post-test sessions for all six experiments.** The blue dots show the binding tendency values for each individual subject, and the black lines show the identity line for each plot, indicating perfect stability of the binding tendency from the pre-test to the post-test session. As can be seen in the figure, in experiments #1–4, with the exception of a few outliers, most subjects exhibited consistent binding tendency scores from the pre-test to the post-test sessions. In experiment #5, most subjects displayed an increase in the binding tendency was displayed in the post-test phase, with some heterogeneity among subjects in this trend. Similarly, an increase in the binding tendency was exhibited by most subjects in experiment #6, but the overall trend was smaller than in Experiment #5.

tendency changes for each experiment, see Fig. S1 in the Supplemental Information). The most intriguing trend based on visualizing individual data appears in experiment 5, where a majority (17) of participants displayed an *increase* in the binding tendency from the pre-test to the post-test session. Interestingly, however, 12 of the subjects displayed binding tendency values that are either at or slightly below the pre-test value, indicating there may be heterogeneity in the efficacy of this effect.

To definitively determine whether any significant differences were produced by the exposure tasks in our six experiments, we used a bootstrap sampling approach to evaluate changes in the binding tendency between the pre-test and post-test phases. Our first analysis involved computing 10,000 sample means (with replacement) of the binding tendency differences scores (post test—pre test) for each experiment to construct a 95% confidence interval, and determine whether this confidence interval contained 0, which would indicate that the binding tendency did not significantly change from the pre-test to the post-test phases. To ensure our findings would not be influenced by extreme outliers

Table 2 Summary of experimental changes in the binding tendency parameter, based on the temporal and spatial characteristics of the exposure-task stimuli in different experiments.

Experiment	Time	Space	Binding tendency: actual change
1	None	None	None
2	Uncorrelated	Incongruent	None
3	Congruent	Congruent	None
4	Uncorrelated	Congruent	None
5	Congruent	Incongruent	Increase
6	Correlated	Incongruent	Marginal increase

(as the scatterplots in Fig. 4 reveal a few anomalous subjects), we excluded any data points that fell beyond the value of four times the mean of Cook's distance for a given experiment.

The values from this initial bootstrapped analysis for our first four experiments were the following: Experiment 1: Mean = .034, 95% CI [−.026, .093]; Experiment 2: Mean = 0, 95% CI [−.043, .042]; Experiment 3: Mean = .037, 95% CI [−.034, .106]; Experiment 4: Mean = .029, 95% CI [−.002, .061]. All bootstrapped confidence intervals for the first four experiments contained 0, indicating the binding tendency did not change across sessions. Even in experiment 4, where the confidence interval barely contained 0, the overall magnitude of the increase was quite small, indicating a negligible effect. Interestingly, however, this same comparison for Experiments 5 and 6 yielded the following results: Experiment 5: Mean = .155, 95% CI [.057, .253]; Experiment 6: Mean = 0.113, 95% CI [.065, .162]. This analysis provided initial evidence that the binding tendency changes were considerably higher (and different from 0) in Experiments 5 and 6.

Our second analysis, though, focused on a more rigorous test: comparing the results of Experiments 5 and 6 to the original control experiment. To this end, we sampled (with replacement) the difference scores from Experiments 1, 5 and 6 simultaneously 10,000 times, and on each iteration calculated the mean of the differences between the Experiment 5/6 scores and the control scores. We then computed means and confidence intervals on these distributions, which yielded the following results: Experiment 5 vs. Experiment 1: Mean = .12, 95% CI [0.012, .26]; Experiment 6 vs. Experiment 1: Mean = .079, 95% CI [.005, .152]. For Experiment 5, the confidence interval did not contain 0, indicating a significant increase in Experiment 5 when compared to the control. This was also true of experiment 6, but by a smaller margin; the lower end of the confidence interval was almost exactly 0.

Interestingly, none of the other experiments differed significantly from the control experiment. This same analysis for the other three experiments yielded the following results: Experiment 2 vs. Experiment 1: Mean = −.033, 95% CI [−.099, .032], Experiment 3 vs. Experiment 1: Mean = 0.003, 95% CI [−0.085, 0.092], Experiment 4 vs. Experiment 1: Mean = −.005, 95% CI [−.068, .059]. Therefore, while Experiment 5 and Experiment 6 provide evidence for increases in the binding tendency from the pre-test to the post-test phase, none of the other experiments revealed a change in this parameter. All of these results are summarized in Table 2.

The increase in the binding tendency in Experiment 5 can be seen in the change in many subjects' behavioral data and model fits from the pre-test and post-test phases. Shown in Fig. 5 is all of the data from subjects that were best fit by the probability-matching strategy in our model, as well as the modeling results when fit to this aggregate dataset. In the conditions with spatially discrepant audiovisual stimuli, the increase in the overlap between the visual and auditory response demonstrates the increase in tendency to integrate spatially discrepant visual and auditory signals.

DISCUSSION

One of the important questions in a Bayesian account of perception is how priors can change over time as new sensory experience is acquired (*Seriès & Seitz, 2013*). While previous studies have shown that priors can be modified by sensory experience (e.g., *Chalk, Seitz & Seriès, 2010*; *Adams, Kerrigan & Graf, 2010*; *Sotiropoulos, Seitz & Seriès, 2011*), and that associations between arbitrarily paired signals across modalities can be learned (*Ernst, 2007*; *Shams et al., 2011*; *Altieri et al., 2015*), it is not currently clear to what degree brief sensory experiences cause the audiovisual binding tendency to change, and what the conditions are that result in an update of this mechanism the brain. Our results provide preliminary evidence that the prior tendency to bind/integrate audiovisual signals *can* change substantially due to brief sensory experiences, and that this tendency seems to increase after repeated exposure to spatially *discrepant* audiovisual stimuli that are presented with fixed temporal relationships, but not after exposure to spatiotemporally concordant stimuli. Specifically, repeated exposure to auditory and visual stimuli that were mismatched in space (in a variable fashion, reversing positions frequently, Fig. 1B) but had a synchronous temporal relationship resulted in an increase in the tendency to bind on localization trials in the post-test phase.

The result from experiments 5 (and strong trend in experiment 6) seems consistent with the framework of our second hypothesis (see Figs. 2B and 2D). We note that both of the original hypotheses discussed in this paper can be interpreted as reflecting error correction mechanisms, but operating on *different kinds of errors*. To clarify the distinction between these two types of errors and clarify the hypotheses motivated by each construct, we focus on Experiment 5 (which contained our main result) and contrast the predictions of the two hypotheses for this specific experiment.

Explanation of the original hypotheses about how priors are updated

Regarding our first hypothesis (Figs. 2A and 2C), let us assume that the observer starts the experiment with an *a priori* binding tendency of .25, which would predict that, on average, flashes and beeps have a common cause 25% of the time. During the Exposure phase, the observer is exposed to many trials with high spatial discrepancy. If 95% of these trials result in an inference of independent causes, then the observer experiences a 5% frequency of common cause during this phase which is at odds with the *a priori* expectation of 25%. According to this hypothesis, this should cause a "correction" of binding tendency by reducing it. In this framework, the nervous system treats its inference about spatially
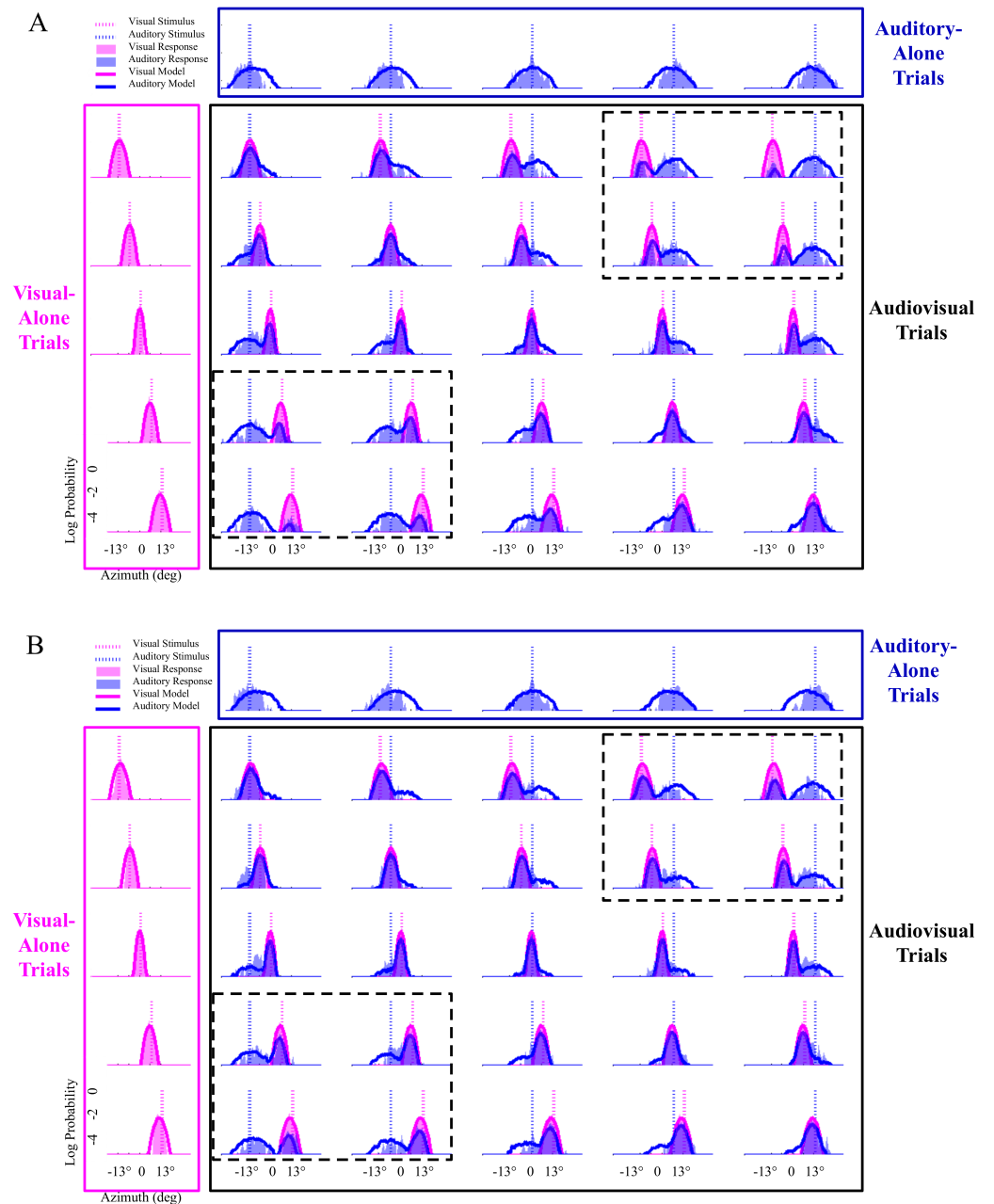
Figure 5 **Modeling results with data from all subjects in experiment #5 that were best characterized by the probability-matching strategy.** (A) Pre-test localization data from twenty subjects are shown, with the model fit to all of these subjects' data at once. Visual-only conditions are shown in the pink boxes, auditory-only conditions in the dark blue boxes, and audiovisual conditions in the black boxes. Audio-visual trials with stimuli played at the same location are along the diagonal in the black box. The bisensory conditions with the highest amount of spatial discrepancy between the visual and auditory stimuli are shown in the black boxes marked by the dotted lines in each figure. (B) Post-test localization data. The conditions in the black boxes with dotted lines clearly reveal the increase in binding tendency from the pre-test to the post-test phase, as the amount of overlap between the pink (visual) and blue (auditory) response distributions increases from the pre-test to the post-test phase, showing that a larger percentage of auditory stimuli were localized where the visual stimuli occurred on these spatially-discrepant trials.

discrepant sensory signals as ground truth, and adjusts its prior accordingly to match the statistics of the world.

On the other hand, according to our second hypothesis, the brain arguably has a model of the world in which sensations with highly discrepant spatial locations generally have independent causes, and the degree of "tolerance" for the discrepancy is determined by the *a priori* expectation of a common cause (or binding tendency). The larger the binding tendency, the higher the tolerance to discrepancy in signals for inferring a common cause. Now, given a binding tendency of .25, highly discrepant auditory and visual signals (e.g., of 39 degrees apart) are inferred to have independent causes on most trials. However, the observance of repeated temporal coincidence across trials provides independent and strong evidence of a common cause. Given the good temporal precision in both modalities and poor spatial precision for auditory signals, the repeated temporal coincidence provides an evidence of common cause that overrides the spatial discrepancy (which is less reliable) and results in an interpretation of common cause for these stimuli after repeated exposure. This inference contradicts the model of the world ("highly discrepant signals have independent causes") and leads to "correction" of the model by increasing the binding tendency.

Therefore, while both of our hypotheses rely on statistics of the observed stimuli, because they make different assumptions about how the causal inference proceeds across trials, they make opposite predictions for learning. The results from Experiment 5 could be interpreted to be consistent with the predictions of the a "predictive coding" mechanism in the brain (*Rao & Ballard, 1999*; *Friston, 2003*; *Friston, 2005*; *Friston, 2012*; *Friston & Kiebel, 2009*). That is, because the apparent mechanism that updates the prior relies on production of an error between model prediction and sensory observation, it could be considered to be compatible with previous ideas about how predictive coding may be implemented.

We can only speculate why the mechanism specified by our second hypothesis appears to operate, rather than the mechanism specified by the first hypothesis; we posit that the correction (or updating) of the binding tendency, which is a prior, cannot depend on constructs that depend on that prior. To correct a variable, information may need to come from a source which is independent from that variable; that is, to correct this prior, the information can come from sensory observations, but not from an inference that is derived from that prior. In the scenario specified by the first hypothesis, the inference of independent causes on each trial can only occur based on a low binding tendency; it cannot be entirely based on the stimuli themselves, because while the stimuli are spatially discrepant, they are temporally consistent. Therefore, an inference which is itself based on the binding tendency cannot be a reliable source for updating the binding tendency. Whereas under the second hypothesis, the evidence of a common cause is merely based on the sensory observations; namely, the repeated temporal coincidence (and the low spatial resolution of auditory signals). Therefore, this is a source of information that is independent of the prior (binding tendency), and thus, can serve as the basis to correct the prior.

This remodeling of the brain's internal model of the world appears to occur because the repeated co-occurrence of otherwise unrelated/discrepant stimuli is treated by the nervous system as compelling evidence of unity or association between the signals (Figs. 2B and 2D).

Unless the nervous system relaxes its criterion for perception of unity, it cannot account for the repeated co-occurrence of the stimuli. This modification of the prior in order to better account for compelling sensory regularity has been observed in other phenomena (*Chalk, Seitz & Seriès, 2010*; *Sotiropoulos, Seitz & Seriès, 2011*), and one study has referred to it as the "oops" factor (*Adams, Kerrigan & Graf, 2010*). While it is possible that the mechanism characterized by hypothesis #1 may play a relevant role on each trial in incrementally updating multisensory priors, it appears that here, the modification of the priors is due to large error signals, and thus the mechanism specified by hypothesis #2 likely underlies this substantial change in the prior for binding sensory signals across space.

Relating the present results to previous empirical findings

Considering our main empirical findings, it is important to comment on how our results relate to previous research on audiovisual recalibration. Previous studies have shown that repeated exposure to auditory and visual stimuli with fixed spatial discrepancy results in a modification of auditory spatial perception (*Canon, 1970*; *Radeau & Bertelson, 1977*; *Recanzone, 1998*; *Lewald, 2002*; *Frissen et al., 2003*; *Frissen et al., 2005*; *Kopco et al., 2009*; *Eramudugolla et al., 2011*). This phenomenon is known as the "ventriloquist aftereffect" and has been computationally accounted for by a shift in the auditory sensory representations (likelihood functions) rather than a change in spatial prior expectations (*Wozny & Shams, 2011b*). Here, for the first time, we report that repeated exposure to auditory-visual stimuli with *variable* spatial discrepancy (i.e., discrepancy which alternated between +39° and −39°) results in a modification of the binding tendency. Therefore, it appears that repeated exposure to spatially discrepant auditory-visual stimuli always results in an update in perceptual processing, with the type of the update depending on relationship between the adapting stimuli. If the spatial relationship in the adapting stimuli is fixed, the nervous system appears to treat the more precise signal as the "teaching" signal and recalibrate the less precise representations by shifting them to reduce the discrepancy. If the relationship is variable, the sensory discrepancy cannot be rectified by a simple shift in sensory representations, and the nervous system does not try to "correct" either sensory representation. Instead, it attempts to "correct" its model of the world by relaxing its criterion for the perception of unity.

It is also important to note the relationship between our present results and previous findings in the literature regarding how sensory experiences can influence integration on subsequent encounters. Recent studies have primarily investigated this question in the context of multisensory speech perception (*Nahorna, Berthommier & Schwartz, 2012*; *Nahorna, Berthommier & Schwartz, 2015*; *Gau & Noppeney, 2016*). These results show that prior exposure to asynchronous or incongruent audiovisual speech combinations decrease participants' tendency to perceive the McGurk illusion. At first glance, these findings may appear to contradict the results reported here, but a closer look at these studies reveals important differences in the stimuli and paradigms. For instance, in the strongly incongruent condition in *Nahorna, Berthommier & Schwartz (2015)*, there is no temporal consistency between the auditory and visual modalities that could potentially provide evidence of a common cause; in the phonetically incongruent conditions in studies by

Nahorna, Berthommier & Schwartz (2012) and *Nahorna, Berthommier & Schwartz (2015)* and the incongruent condition in *Gau & Noppeney (2016)*, which are more similar to incongruent condition in our study, the "incongruent" speech stimuli used syllables that are *never* integrated to produce the McGurk illusion. It is possible that if the stimuli are along independent or orthogonal dimensions, then exposure to this type of completely unrelated stimuli, which *cannot possibly* be integrated, may activate the mechanism from our first hypothesis and cause a decrease in the tendency to integrate. In contrast, in our study, the sounds and sights are along a dimension that brain can integrate over, depending on the binding tendency. Therefore, modification of binding tendency in this scenario *can* reduce the prediction error, and the mechanism from our second hypothesis can operate.

Another study relevant for our present results is by *Van Wanrooij, Bremen & Van Opstal (2010)*, which investigated the effects of frequency of auditory-visual spatial alignment on multisensory spatial integration. In this study, integration was measured primarily by reaction times, and the main finding was that spatially aligned audiovisual stimuli elicited the fastest reaction times to targets (i.e., the visual component of the AV stimuli) in blocks when the stimuli were *always* aligned (100–0 condition) compared to blocks with 50–50 or 10–90 ratios for AV alignment. While these results suggest that the binding tendency in the 100–0 condition was higher than that in the other two conditions, we are not certain if this difference can be entirely attributed to the statistics of and exposure to the stimuli in the blocks, since the instructions given to subjects in the three conditions differed: in the 100-0 condition, subjects were instructed that the audiovisual stimuli were presented in spatiotemporal alignment and that the flash was the target (pp. 1764–1765), while in the other conditions, subjects were told to ignore the auditory stimuli, and the methods do not indicate that subjects had any *a priori* knowledge about the stimulus distributions.

We think it is critical that future studies include baseline measures of integration before evaluating how context plays a role, and include plots to show how individual data points may be impacted by context and exposure. As has been recently noted, using bar graphs alone can conceal important trends in data (*Rousselet, Foxe & Bolam, 2016*); thus, we strongly recommend that future studies plot individual data points to make the effects of context and exposure on integration more interpretable. We also acknowledge it is possible that there are factors that are at work that we are not considering that could explain these differences in findings. To ascertain what factors drive changes in binding tendency, and what the underlying mechanisms are, future studies will need to specify a taxonomization of influences on perceptual priors by noting how tasks, stimuli, instructions, and other factors can impact this mechanism.

We consider this study to be one piece of evidence to demonstrate that the binding tendency can change quickly, but additional research is needed to catalogue the wide array of possible influences on binding tendency. We think that the binding tendency (as with most other priors) is likely formed and influenced by both "low-level" and "high-level" factors. The low-level factors would include prior experience of the relationship between certain sounds and sights. This factor is influenced by the stimulus statistics and regularities and does not involve top-down decisional factors *per se*. On the other hand, higher-level factors such as explicit cognitive knowledge about the stimuli (e.g., knowledge obtained

through instructions given by the experimenter) can also influence the expectation of a common cause and hence, the binding tendency. Many perceptual learning studies have shown that perceptual learning can occur without feedback; therefore, we do not find it surprising that an update of the binding tendency can occur implicitly and without feedback.

Future directions for studying the binding tendency

The current findings demonstrate that this plastic capacity of the brain can occur in an extremely short period of time (less than ten minutes), revealing an adaptive ability of the system to learn about new associations between sensory signals in the environment, even in adulthood. Considering that deficits in sensory integration are thought to accompany several disorders, including dyslexia (*Hairston et al., 2005*; *Harrar et al., 2014*; *Hahn, Foxe & Molholm, 2014*), autism (*Stevenson et al., 2015*; *Stevenson et al., 2014a*; *Stevenson et al., 2014b*; *Wallace & Stevenson, 2014*; *Baum, Stevenson & Wallace, 2015*), and schizophrenia (*Ross et al., 2007*; *Szycik et al., 2009*; *Williams et al., 2010*; *Zvyagintsev et al., 2013*; *Stekelenburg et al., 2013*), by illuminating basic principles underlying how the tendency to bind sensory signals across space can be modified, we think these findings represent a key contribution as the field attempts to develop effective remedial or educational protocols for enhancing multisensory integration in clinical populations (*Wallace & Stevenson, 2014*).

For instance, it has recently been shown that an auditory training protocol can improve higher-level cognitive functioning in young individuals with early-onset schizophrenia (*Fisher et al., 2015*). Indeed, recent data from our laboratory indicates that individuals with low binding tendency scores on a spatial localization task report higher numbers of psychometrically-defined prodromal features in their everyday lives (*Odegaard & Shams, in press*). Thus, if a protocol can be developed to improve lower-level sensory binding in individuals with schizophrenia, it seems possible that downstream, higher-level functions could also improve. Currently, we can only speculate whether increasing the binding tendency can ameliorate symptoms in clinical disorders, but we think recent results demonstrating increased executive functioning with sensory training protocols in individuals with schizophrenia are promising (*Fisher et al., 2015*; *Dale et al., 2016*), and that our findings can help inform future investigations looking to exploit multisensory techniques to modify how audiovisual information is integrated in the brain.

Limitations of this investigation

Finally, while the six experiments presented here represent an important, first demonstration of audiovisual binding tendency modification, it is important to note limitations of the current study and what is currently unknown regarding modification of this mechanism in the brain. First, future studies will be needed to uncover further specifics about the phenomenon, including how the increase in the spatial binding tendency may scale with the spatial discrepancy between stimuli presented in the exposure task, how long the increase in binding may last following exposure, and whether the principles reported here generalize to other domains and stimuli. For example, would a 10-minute exposure

to McGurk speech stimuli, which are consistent in time but discrepant in syllabic content, increase the tendency to bind audiovisual speech signals? Considering the current findings, this is an intriguing hypothesis. The question of how exposure to stimuli that have a fixed temporal relationship but *vary* in terms of the magnitude of spatial discrepancy also remains unanswered. We posit that this type of condition would still cause a slight increase in the tendency to bind; indeed, one may interpret the slight increase in binding tendency in our control experiment in support of this hypothesis, as exposure to our localization task (with 20 out of 35 conditions containing spatially discrepant stimuli) appears to slightly increase the tendency to bind, but this question will need to be addressed more thoroughly by future work.

Second, we recently demonstrated that the brain's tendency to bind auditory and visual sensory signals does not generalize across tasks (*Odegaard & Shams, 2016*). Therefore, we do not expect that the changes in binding tendency observed in Experiment 5 and 6 would generalize to temporal tasks. However, the question of how general the changes in binding tendency induced in this fashion are, and to what other stimuli, tasks, perceptual domains and contexts they would generalize to is a very important question, both from the perspective of underlying mechanisms and the perspective of translation to clinical and educational applications. Another open question is whether the same type of plasticity would operate in the temporal domain; namely, whether repeated exposure to sensory signals that are temporally discrepant but spatially congruent increase the tendency to bind sensory signals in a temporal task, More broadly, the question of which spatial and temporal relationships between audiovisual stimuli in the exposure task would cause an update in the temporal binding tendency in the brain. These important questions remain unanswered, and will need to be investigated by future studies.

Third, the fact that none of our experiments resulted in a *decrease* in the binding tendency is also interesting, and the conditions that cause a *reduction* in the tendency to integrate audiovisual information across space remain unknown. It remains possible that a protocol in which there is repeated congruency in temporal and spatial dimensions, but compelling evidence that the signals originate from different sources (thus creating a mismatch between model predictions and sensory experience) would lead to a decrease in binding tendency. Also, recent studies have provided preliminary evidence that active forms of sensory training (for example, training subjects to discriminate audiovisual signals with greater precision) can decrease the tendency to bind (*McGovern et al., 2016*). This contrasts our results here, which are based on a more passive form of exposure. Therefore, future studies should further examine the conditions in which the binding tendency may decrease by probing whether semantic incongruency, attentional factors, or other indicators that the signals originate from separate sources are necessary to induce this type of change.

Lastly, the neural mechanism underlying the binding tendency currently remains unknown, and additional work will be needed to uncover how this tendency is encoded in the brain. Recent work using functional magnetic resonance imaging reveals some initial insight into how the binding tendency might be encoded in the brain (*Rohe & Noppeney, 2015a*), but work using additional methods, such as neural network modeling, will be necessary to uncover the specific neural mechanisms (e.g., the specific type of connectivity,

neural architecture, circuitry and physiology) that determine binding tendency in the brain.

To conclude, we think these findings from our large dataset from six experiments represent a pivotal first step in learning about how the binding tendency can be modified, and we hope future investigations will build upon these results in answering many questions that remain as the study of multisensory processing continues to investigate both translational and clinical applications of this work.

ACKNOWLEDGEMENTS

We would like to thank Uri Maoz, Majed Samad, James Bisley, Megan Peters, Dario Ringach, Hongjing Lu, and Mark Frye for their constructive comments and criticisms on this project. We would also like to thank Seong Hah Cho, Eden Sayed, Michael Bani, and Shannon Gordon for their help with data collection.

Funding

This work was supported by NSF grant BCS-1057969. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
NSF: BCS-1057969.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Brian Odegaard conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- David R. Wozny conceived and designed the experiments, contributed reagents/materials/analysis tools, prepared figures and/or tables, reviewed drafts of the paper.
- Ladan Shams conceived and designed the experiments, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The University of California North General IRB granted ethical approval to carry out the study in our lab (IRB#13-000476).

Data Availability

The raw data has been supplied as a Supplementary File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.3143#supplemental-information.

REFERENCES

Adams WJ, Kerrigan IS, Graf EW. 2010. Efficient visual recalibration from either visual or haptic feedback: the importance of being wrong. *The Journal of Neuroscience* **30**:14745–14749 DOI 10.1523/JNEUROSCI.2749-10.2010.

Altieri N, Stevenson RA, Wallace MT, Wenger MJ. 2015. Learning to associate auditory and visual stimuli: behavioral and neural mechanisms. *Brain Topography* **28**:479–493 DOI 10.1007/s10548-013-0333-7.

Barutchu A, Crewther DP, Crewther SG. 2009. The race that precedes coactivation: development of multisensory facilitation in children. *Developmental Science* **12**:464–473 DOI 10.1111/j.1467-7687.2008.00782.x.

Barutchu A, Danaher J, Crewther SG, Innes-Brown H, Shivdasani MN, Paolini AG. 2010. Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology* **105**:38–50 DOI 10.1016/j.jecp.2009.08.005.

Baum SH, Stevenson RA, Wallace MT. 2015. Testing sensory and multisensory function in children with autism spectrum disorder. *Journal of Visualized Experiments* **98**:e52677 DOI 10.3791/52677.

Beierholm UR, Quartz SR, Shams L. 2009. Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision* **9(5)**:23 DOI 10.1167/9.5.23.

Bertelson P, Frissen I, Vroomen J, De Gelder B. 2006. The aftereffects of ventriloquism: patterns of spatial generalization. *Perception & Psychophysics* **68**:428–436 DOI 10.3758/BF03193687.

Canon LK. 1970. Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology* **84**:141–147 DOI 10.1037/h0028925.

Chalk M, Seitz AR, Seriès P. 2010. Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision* **10**:2–2 DOI 10.1167/10.8.2.

Chen L, Vroomen J. 2013. Intersensory binding across space and time: a tutorial review. *Attention, Perception, & Psychophysics* **75**:790–811 DOI 10.3758/s13414-013-0475-4.

Dale CL, Brown EG, Fisher M, Herman AB, Dowling AF, Hinkley LB, Subramaniam K, Nagarajan SS, Vinogradov S. 2016. Auditory cortical plasticity drives training-induced cognitive changes in schizophrenia. *Schizophrenia Bulletin* **42(1)**:220–228 DOI 10.1093/schbul/sbv087.

Eramudugolla R, Kamke MR, Soto-Faraco S, Mattingley JB. 2011. Perceptual load influences auditory space perception in the ventriloquist aftereffect. *Cognition* **118**:62–74 DOI 10.1016/j.cognition.2010.09.009.

Ernst MO. 2007. Learning to integrate arbitrary signals from vision and touch. *Journal of Vision* **7(5)**:7 DOI 10.1167/7.5.7.

Fiser J, Aslin RN. 2002. Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology* **28**:458 DOI 10.1037/0278-7393.28.3.458.

Fiser J, Aslin RN. 2016. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science* **12(6)**:499–504 DOI 10.1111/1467-9280.00392.

Fisher M, Loewy R, Carter C, Lee A, Ragland JD, Niendam T, Schlosser D, Pham L, Miskovich T, Vinogradov S. 2015. Neuroplasticity-based auditory training via laptop computer improves cognition in young individuals with recent onset schizophrenia. *Schizophrenia Bulletin* **41(1)**:250–258 DOI 10.1093/schbul/sbt232.

Frissen I, Vroomen J, De Gelder B. 2012. The aftereffects of ventriloquism: the time course of the visual recalibration of auditory localization. *Seeing and Perceiving* **25**:1–14 DOI 10.1163/187847611X620883.

Frissen I, Vroomen J, De Gelder B, Bertelson P. 2003. The aftereffects of ventriloquism: are they sound-frequency specific? *Acta Psychologica* **113**:315–327 DOI 10.1016/S0001-6918(03)00043-X.

Frissen I, Vroomen J, De Gelder B, Bertelson P. 2005. The aftereffects of ventriloquism: generalization across sound-frequencies. *Acta Psychologica* **118**:93–100 DOI 10.1016/j.actpsy.2004.10.004.

Friston K. 2003. Learning and inference in the brain. *Neural Networks* **16**:1325–1352 DOI 10.1016/j.neunet.2003.06.005.

Friston K. 2005. A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **360**:815–836 DOI 10.1098/rstb.2005.1622.

Friston K. 2012. Prediction, perception and agency. *International Journal of Psychophysiology* **83**:248–252 DOI 10.1016/j.ijpsycho.2011.11.014.

Friston K, Kiebel S. 2009. Cortical circuits for perceptual inference. *Neural Networks* **22**:1093–1104 DOI 10.1016/j.neunet.2009.07.023.

Gau R, Noppeney U. 2016. How prior expectations shape multisensory perception. *NeuroImage* **124(Part A)**:876–886 DOI 10.1016/j.neuroimage.2015.09.045.

Gori M, Del Viva M, Sandini G, Burr DC. 2008. Young children do not integrate visual and haptic form information. *Current Biology* **18**:694–698 DOI 10.1016/j.cub.2008.04.036.

Gurler D, Doyle N, Walker E, Magnotti J, Beauchamp M. 2015. A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception, & Psychophysics* **77**:1333–1341 DOI 10.3758/s13414-014-0821-1.

Haas B de, Kanai R, Jalkanen L, Rees G. 2012. Grey matter volume in early human visual cortex predicts proneness to the sound-induced flash illusion. *Proceedings of the Royal Society of London B: Biological Sciences* **279**:4955–4961 DOI 10.1098/rspb.2012.2132.

Hahn N, Foxe JJ, Molholm S. 2014. Impairments of multisensory integration and cross-sensory learning as pathways to dyslexia. *Neuroscience & Biobehavioral Reviews* **47**:384–392 DOI 10.1016/j.neubiorev.2014.09.007.

Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT. 2005. Altered temporal profile of visual–auditory multisensory interactions in dyslexia. *Experimental Brain Research* **166**:474–480 DOI 10.1007/s00221-005-2387-6.

Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA. 2003. Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience* **15**:20–29 DOI 10.1162/089892903321107792.

Harrar V, Tammam J, Pérez-Bellido A, Pitt A, Stein J, Spence C. 2014. Multisensory integration and attention in developmental dyslexia. *Current Biology* **24**:531–535 DOI 10.1016/j.cub.2014.01.029.

Hillock AR, Powers AR, Wallace MT. 2011. Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia* **49**:461–467 DOI 10.1016/j.neuropsychologia.2010.11.041.

Hillock-Dunn A, Grantham DW, Wallace MT. 2016. The temporal binding window for audiovisual speech: children are like little adults. *Neuropsychologia* **88**:74–82 DOI 10.1016/j.neuropsychologia.2016.02.017.

Kopco N, Lin I-F, Shinn-Cunningham BG, Groh JM. 2009. Reference frame of the ventriloquism aftereffect. *The Journal of Neuroscience* **29**:13809–13814 DOI 10.1523/JNEUROSCI.2783-09.2009.

Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. 2007. Causal inference in multisensory perception. *PLOS ONE* **2**:e943 DOI 10.1371/journal.pone.0000943.

Lewald J. 2002. Rapid adaptation to auditory-visual spatial disparity. *Learning & Memory* **9**:268–278 DOI 10.1101/lm.51402.

Magnotti JF, Beauchamp MS. 2014. The noisy encoding of disparity model of the McGurk effect. *Psychonomic Bulletin & Review* **22**:701–709 DOI 10.3758/s13423-014-0722-2.

Mallick DB, Magnotti JF, Beauchamp MS. 2015. Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review* **22**:1299 DOI 10.3758/s13423-015-0817-4.

Mathworks. 2009. MATLAB 7.9. Natick: The MathWorks, Inc. *Available at http://www.walkingrandomly.com/?p=4767*.

McGovern DP, Roudaia E, Newell FN, Roach NW. 2016. Perceptual learning shapes multisensory causal inference via two distinct mechanisms. *Scientific Reports* **6**:24673 DOI 10.1038/srep24673.

Mcgurk H, Macdonald J. 1976. Hearing lips and seeing voices. *Nature* **264**:746–748 DOI 10.1038/264746a0.

Nagelkerke NJD. 1991. A note on a general definition of the coefficient of determination. *Biometrika* **78**:691–692 DOI 10.1093/biomet/78.3.691.

Nahorna O, Berthommier F, Schwartz J-L. 2012. Binding and unbinding the auditory and visual streams in the McGurk effect. *The Journal of the Acoustical Society of America* **132**:1061–1077 DOI 10.1121/1.4728187.

Nahorna O, Berthommier F, Schwartz J-L. 2015. Audio-visual speech scene analysis: characterization of the dynamics of unbinding and rebinding the McGurk effect. *The Journal of the Acoustical Society of America* **137**:362–377 DOI 10.1121/1.4904536.

Nath AR, Beauchamp MS. 2012. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage* **59**:781–787 DOI 10.1016/j.neuroimage.2011.07.024.

Neil PA, Chee-Ruiter C, Scheier C, Lewkowicz DJ, Shimojo S. 2006. Development of multisensory spatial integration and perception in humans. *Developmental Science* **9**:454–464 DOI 10.1111/j.1467-7687.2006.00512.x.

Odegaard B, Shams L. 2016. The brain's tendency to bind audiovisual signals is stable but not general. *Psychological Science* **27(4)**:583–591 DOI 10.1177/0956797616628860.

Odegaard B, Shams L. 2017. The relationship between audiovisual binding tendencies and prodromal features of schizophrenia in the general population. *Clinical Psychological Science* In Press.

Odegaard B, Wozny DR, Shams L. 2015. Biases in visual, auditory, and audiovisual perception of space. *PLOS Computational Biology* **11**:e1004649 DOI 10.1371/journal.pcbi.1004649.

Odegaard B, Wozny DR, Shams L. 2016. The effects of selective and divided attention on sensory precision and integration. *Neuroscience Letters* **614**:24–28 DOI 10.1016/j.neulet.2015.12.039.

Parise CV, Spence C, Ernst MO. 2012. When correlation implies causation in multisensory integration. *Current Biology* **22**:46–49 DOI 10.1016/j.cub.2011.11.039.

Radeau M, Bertelson P. 1977. Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics* **22**:137–146 DOI 10.3758/BF03198746.

Rao RPN, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* **2**:79–87 DOI 10.1038/4580.

Recanzone GH. 1998. Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proceedings of the National Academy of Sciences of the United States of America* **95**:869–875 DOI 10.1073/pnas.95.3.869.

Rohe T, Noppeney U. 2015a. Cortical hierarchies perform bayesian causal inference in multisensory perception. *PLOS Biology* **13**:e1002073 DOI 10.1371/journal.pbio.1002073.

Rohe T, Noppeney U. 2015b. Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision* **15**:22–22 DOI 10.1167/15.5.22.

Ross LA, Saint-Amour D, Leavitt VM, Molholm S, Javitt DC, Foxe JJ. 2007. Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophrenia Research* **97**:173–183 DOI 10.1016/j.schres.2007.08.008.

Rousselet GA, Foxe JJ, Bolam JP. 2016. A few simple steps to improve the description of group results in neuroscience. *European Journal of Neuroscience* **44**:2647–2651 DOI 10.1111/ejn.13400.

PeerJ

Saffran JR, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. *Science* **274**:1926–1928 DOI 10.1126/science.274.5294.1926.

Sato Y, Toyoizumi T, Aihara K. 2007. Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Computation* **19**:3335–3355 DOI 10.1162/neco.2007.19.12.3335.

Seriès P, Seitz AR. 2013. Learning what to expect (in visual perception). *Frontiers in Human Neuroscience* **7**:668 DOI 10.3389/fnhum.2013.00668.

Shams L, Wozny DR, Kim R, Seitz A. 2011. Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology* **2**:264 DOI 10.3389/fpsyg.2011.00264.

Sotiropoulos G, Seitz AR, Seriès P. 2011. Changing expectations about speed alters perceived motion direction. *Current Biology* **21**:R883–R884 DOI 10.1016/j.cub.2011.09.013.

Stekelenburg JJ, Maes JP, Van Gool AR, Sitskoorn M, Vroomen J. 2013. Deficient multisensory integration in schizophrenia: an event-related potential study. *Schizophrenia Research* **147**:253–261 DOI 10.1016/j.schres.2013.04.038.

Stevenson RA, Segers M, Ferber S, Barense MD, Wallace MT. 2015. The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Frontiers in Psychology* **5**:379 DOI 10.3389/fpsyg.2014.00379.

Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, Wallace MT. 2014a. Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience* **34**:691–697 DOI 10.1523/JNEUROSCI.3615-13.2014.

Stevenson RA, Siemann JK, Woynaroski TG, Schneider BC, Eberly HE, Camarata SM, Wallace MT. 2014b. Evidence for diminished multisensory integration in autism spectrum disorders. *Journal of Autism and Developmental Disorders* **44**:3161 DOI 10.1007/s10803-014-2179-6.

Stevenson RA, Zemtsov RK, Wallace MT. 2012. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology* **38**:1517–1529 DOI 10.1037/a0027339.

Szycik GR, Münte TF, Dillo W, Mohammadi B, Samii A, Emrich HM, Dietrich DE. 2009. Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study. *Schizophrenia Research* **110**:111–118 DOI 10.1016/j.schres.2009.03.003.

Talsma D. 2015. Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience* **9**:19 DOI 10.3389/fnint.2015.00019.

Tremblay C, Champoux F, Bacon BA, Theoret H. 2007a. Evidence for a generic process underlying multisensory integration. *Open Behavioral Science Journal* **1**:1–4.

Tremblay C, Champoux F, Voss P, Bacon BA, Lepore F, Théoret H. 2007b. Speech and non-speech audio-visual illusions: a developmental study. *PLOS ONE* **2**:e742 DOI 10.1371/journal.pone.0000742.

Van Wanrooij MM, Bremen P, Van Opstal A. 2010. Acquired prior knowledge modulates audiovisual integration. *European Journal of Neuroscience* **31**:1763–1771 DOI 10.1111/j.1460-9568.2010.07198.x.

Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA. 2004. Unifying multisensory signals across time and space. *Experimental Brain Research* **158**:252–258 DOI 10.1007/s00221-004-1899-9.

Wallace MT, Stein BE. 2007. Early experience determines how the senses will interact. *Journal of Neurophysiology* **97**:921–926 DOI 10.1152/jn.00497.2006.

Wallace MT, Stevenson RA. 2014. The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia* **64**:105–123 DOI 10.1016/j.neuropsychologia.2014.08.005.

Williams LE, Light GA, Braff DL, Ramachandran VS. 2010. Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia* **48**:3128–3136 DOI 10.1016/j.neuropsychologia.2010.06.028.

Wozny DR, Beierholm UR, Shams L. 2010. Probability matching as a computational strategy used in perception. *PLOS Computational Biology* **6**:e1000871 DOI 10.1371/journal.pcbi.1000871.

Wozny DR, Shams L. 2011a. Recalibration of auditory space following milliseconds of cross-modal discrepancy. *The Journal of Neuroscience* **31**:4607–4612 DOI 10.1523/JNEUROSCI.6079-10.2011.

Wozny DR, Shams L. 2011b. Computational characterization of visually induced auditory spatial adaptation. *Frontiers in Integrative Neuroscience* **5**:75 DOI 10.3389/fnint.2011.00075.

Zvyagintsev M, Parisi C, Chechko N, Nikolaev AR, Mathiak K. 2013. Attention and multisensory integration of emotions in schizophrenia. *Frontiers in Human Neuroscience* **7**:674 DOI 10.3389/fnhum.2013.00674.