BRILL

Multisensory Research 30 (2017) 207–234

MULTISENSORY RESEARCH

brill.com/msr

1851

Cue Integration for Continuous and Categorical Dimensions by Synesthetes

Kaitlyn R. Bankieris [1,*], **Vikranth Rao Bejjanki** [2] **and Richard N. Aslin** [1]

[1] Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA
[2] Department of Psychology, Hamilton College, Clinton, NY, USA

Received 28 September 2016; accepted 16 February 2017

Abstract

For synesthetes, sensory or cognitive stimuli induce the perception of an additional sensory or cognitive stimulus. Grapheme–color synesthetes, for instance, consciously and consistently experience particular colors (e.g., fluorescent pink) when perceiving letters (e.g., *u*). As a phenomenon involving multiple stimuli within or across modalities, researchers have posited that synesthetes may integrate sensory cues differently than non-synesthetes. However, findings to date present mixed results concerning this hypothesis, with researchers reporting enhanced, depressed, or normal sensory integration for synesthetes. In this study we *quantitatively* evaluated the multisensory integration process of synesthetes and non-synesthetes using Bayesian principles, rather than employing multisensory illusions, to make inferences about the sensory integration process. In two studies we investigated synesthetes' sensory integration by comparing human behavior to that of an ideal observer. We found that synesthetes integrated cues for both continuous and categorical dimensions in a statistically optimal manner, matching the sensory integration behavior of controls. These findings suggest that synesthetes and controls utilize similar cue integration mechanisms, despite differences in how they perceive unimodal stimuli.

Keywords

Synesthesia, cue integration, audiovisual integration

1. Introduction

For synesthetes, one sensory or cognitive stimulus causes the perception of another sensory or cognitive stimulus that is not physically present. For instance, a grapheme–color synesthete may automatically and consistently see the color lilac when viewing the number 4. This phenomenon occurs in approximately

* To whom correspondence should be addressed. E-mail: kbankieris@gmail.com

DOI:10.1163/22134808-00002559

4% of the population and manifests itself in up to 61 different varieties (Day 2005, 2009). Although synesthesia has been documented for over a century (e.g., Calkins, 1893; Claparede, 1903; Jewanski *et al*., 2009), its underlying cause remains largely unknown. Neurological theories hypothesize that synesthesia arises due to additional or disinhibited neural connections (e.g., Bargary and Mitchell, 2008; Grossenbacher and Lovelace, 2001; Hubbard *et al*., 2011; Ramachandran and Hubbard, 2001), but the specificity and nature of these connections is debated. That is, some researchers believe that the neural connections giving rise to synesthesia are qualitatively different from cross-modal mechanisms present in the general population. An alternative view is that synesthesia is an exaggerated form of normal cross-modal processing, with synesthetic associations being one manifestation of more widespread differences in brain connectivity and function. In the current study, we quantitatively examine synesthetes' audio-visual integration abilities to determine if synesthetes have a general exaggeration of multisensory processing abilities.

Across various neuroimaging and behavioral studies, there is some evidence that synesthetes have widespread multisensory processing differences unrelated to their particular form of synesthesia. In the neuroimaging literature, multiple studies have demonstrated that structural and functional connectivity differs between synesthetes and non-synesthetes. Interestingly, these differences emerge not only in regions of the brain directly related to synesthetic experiences (e.g., V4 for induced color associations) but also extend to parietal regions of the brain generally associated with multisensory processing or binding (e.g., Hänggi *et al*., 2011; Jäncke *et al*., 2009; O'Hanlon *et al*., 2013; Rouw and Scholte, 2007, 2010; Tomson *et al*., 2013; Weiss and Fink, 2009; see Hupé and Dojat, 2015 or Rouw *et al*., 2011 for a review). These neuroimaging results suggest that synesthetes' hallmark associations may be the behavioral manifestation of widespread neural differences, including generally altered multisensory processing.

Behavioral studies that have investigated synesthetes' multisensory integration have produced conflicting findings. The majority of these behavioral studies have examined synesthetes' susceptibility to multisensory illusions in order to draw conclusions regarding the mechanism underlying multisensory integration. The most commonly studied is the Double Flash Illusion (Shams *et al*., 2000). This illusion occurs when a single visual flash paired with two auditory beeps gives rise to the perception of two visual flashes. Parameters of multisensory integration are inferred by examining susceptibility to the illusion across various temporal delays between the two auditory beeps. Several studies have tested synesthetes on this illusion and have found inconsistent results. Grapheme–color synesthetes have been reported to have greater susceptibility (Brang *et al*., 2012), reduced susceptibility (Neufeld *et al*., 2012), or no difference in susceptibility to this illusion (Whittingham *et al*., 2014)

compared to non-synesthetes. Unfortunately, these studies differ along multiple dimensions (e.g., mean age of synesthetes, alignment of first beep and visual flash, additional types of synesthesia experienced) making it difficult to determine the reason for the conflicting findings.

The McGurk illusion has also been used to investigate multisensory integration in synesthetes. This illusion arises when the visual cue (e.g., video of a mouth producing /ga/) and auditory cue (e.g., audio of /ba/) to a phoneme utterance conflict with one another and give rise to an intermediate percept (e.g., /da/). Sinke *et al*. (2012) found that synesthetes were less susceptible to this illusion, providing evidence against the hypothesis that synesthetes have increased sensory integration in general. The authors additionally tested participants' ability to identify auditory words in noise with or without the added visual cue of matching articulatory movements. They found that synesthetes benefited less than non-synesthetes from this additional visual cue, and interpreted this as evidence that synesthetes have decreased rather than increased multisensory integration. However, since this study did not evaluate performance on this task with a visual only condition, their conclusion is subject to an alternative explanation. The observed multisensory benefit of controls could reflect a similar integration process for each group if synesthetes perform worse on a visual-only version of this task. That is, the larger difference between the audiovisual and audio-only conditions for controls could be due to superior performance with visual information alone compared to synesthetes. If synesthetes did perform worse than controls in a visual-only condition, the difference between audiovisual and audio-only conditions would be smaller than for controls if both groups used the same integration mechanism. Therefore, any conclusions drawn regarding cue integration without data from a visual-only condition are premature. Overall, the results from the existing literature on synesthetes' susceptibility to multisensory illusions do not provide convincing evidence for a generally heightened multisensory ability.

Examining multisensory integration with the previously described illusions allows one to evaluate the *outcome* of integration. That is, the dependent measure assesses whether or not integration occurred (e.g., Körding *et al*., 2007). Taking a slightly different approach and examining the *benefit* of integration, Brang and colleagues (2012) tested grapheme–color synesthetes' and non-synesthetes' reaction time for detecting audio, visual, and audiovisual stimuli. In this paradigm, true multisensory integration predicts that reaction times (RTs) to audiovisual stimuli will be faster than RTs to either unimodal stimulus alone and will exceed the statistical prediction of summing the two targets (i.e., the Race Model; e.g., Hershenson, 1962; Miller, 1982; Laurienti *et al*., 2006). Results demonstrated that both synesthetes and controls had faster reaction times than predicted by the Race Model, reflecting sensory integration. This difference between the Race Model predictions and

participants' observed audiovisual RTs was only marginally greater for synesthetes than non-synesthetes, suggesting that synesthetes may have benefitted more from the multisensory stimulus than non-synesthetes. The results of this study, which quantitatively investigated the outcome of multisensory integration in a 'natural' environment (as opposed to within an illusion), lend weak support for the hypothesis that synesthetes' general multisensory capabilities may be different from those of non-synesthetes. Taken as a whole, the literature examining multisensory integration in synesthesia provides inconsistent evidence regarding the outcome of integration.

Here, rather than making inferences about the multisensory mechanism from the outcome of such integration, we sought to examine the *process* of integration itself. In two studies, we evaluate *how* synesthetes integrate multiple cues. Specifically, we investigate synesthetes' audiovisual integration from the perspective of Bayesian cue integration to determine whether or not synesthetes combine cues in a statistically efficient manner, as has been observed previously with non-synesthetes. In Experiment 1, we assess audiovisual integration with a spatial localization task, which relies on the continuous dimension of azimuth. Experiment 2 examines newly learned categories that are defined by two continuous dimensions (auditory frequency and visual numerosity). Evidence from both studies demonstrates that like non-synesthetes, synesthetes integrate audiovisual cues in a manner indistinguishable from the behavior of an ideal observer, suggesting that both synesthetes and non-synesthetes integrate cues in a statistically-optimal manner.

2. Experiment 1: Audiovisual Localization

In Experiment 1, we use a spatial localization task (which relies on the continuous dimension of azimuth) to investigate synesthetes' audiovisual integration from the perspective of Bayesian cue integration. Studies evaluating cue combination across such continuous dimensions have demonstrated that humans (presumably about 96% non-synesthetes) integrate multiple sources of information efficiently, following the statistically optimal strategy of weighting sensory cues based on their variability (e.g., Ernst and Banks, 2002; Hillis *et al*., 2002; Jacobs and Fiune, 1999; Knill and Saunders, 2003; Körding and Wolpert, 2004; Körding *et al*., 2007; Michel and Jacobs, 2008; Van Beers *et al*., 1999). When locating a chirping bird, for example, this statistically efficient approach predicts that humans should weight visual cues to the location of the bird more heavily than auditory cues because the human visual system more reliably encodes spatial location in comparison to the auditory system. Moreover, if this task is performed at night when visual information is degraded, we should expect a greater reliance on auditory cues.

Formally, we can represent the information provided by an individual sensory signal A about a stimulus S in the world as a likelihood function, $p(\mathrm{A}|S)$. The value of S that maximizes this likelihood function can be thought of as the estimate of S suggested by A, $\hat{S}_\mathrm{A}$. Given two sensory stimuli A and B that are conditionally independent (e.g., the sensory uncertainty associated with each modality is independent), the information provided by the combination of both the cues can be written as $p(\mathrm{A}, \mathrm{B}|S) = p(\mathrm{A}|S)p(\mathrm{B}|S)$. With the assumption that the individual cue likelihood functions are Gaussian, the peak of the combined likelihood function can be written as a weighted average of the peaks of the individual likelihood functions. Formally, the combined estimate of the stimulus is a weighted linear combination of the estimates suggested by the two sensory signals:

$$\hat{S} = w_\mathrm{A}\hat{S}_\mathrm{A} + w_\mathrm{B}\hat{S}_\mathrm{B} \tag{1}$$

where

$$w_\mathrm{A} = \frac{\frac{1}{\sigma_\mathrm{A}^2}}{\frac{1}{\sigma_\mathrm{A}^2} + \frac{1}{\sigma_\mathrm{B}^2}} \text{ and } w_\mathrm{B} = \frac{\frac{1}{\sigma_\mathrm{B}^2}}{\frac{1}{\sigma_\mathrm{A}^2} + \frac{1}{\sigma_\mathrm{B}^2}} \tag{2}$$

and σ_A^2 and σ_B^2 are the variances of $p(\mathrm{A}|S)$ and $p(\mathrm{B}|S)$, respectively. The variance of the combined likelihood $p(\mathrm{A}, \mathrm{B}|S)$ is given by:

$$\sigma_\mathrm{AB}^2 = \frac{\sigma_\mathrm{A}^2\sigma_\mathrm{B}^2}{\sigma_\mathrm{A}^2 + \sigma_\mathrm{B}^2} \tag{3}$$

These equations [(1)–(3)] describe the behavior of an ideal observer when combining two cues lying along *continuous* dimensions for a given sensory stimulus, such as spatial location or size, because this approach minimizes the variance of the resulting estimate (Ernst and Banks, 2002). In Experiment 1, we ask whether synesthetes' integration behavior conforms to these Bayesian ideal-observer principles using a spatial localization task.

2.1. Methods

2.1.1. Participants

Eleven linguistic–color synesthetes experiencing colors for letters, numbers, days of the week, and/or months of the year were recruited from our existing database of Rochester area synesthetes. Additionally, ten non-synesthetes were recruited from the Rochester area. All participants had normal or corrected-to-normal vision, no known hearing problems, were fluent in English, and were compensated $10/h for their participation. One synesthete was excluded from analyses because she failed to maintain focus on the fixation cross (self-reported that she could not perform the task with the stimuli in

her periphery, so she did not try). In addition, one synesthete and one non-synesthete were excluded from analyses due to poor performance on the detection task (see Procedure). Nine synesthetes (mean age = 24.3, SD = 8.6, two males) and nine non-synesthetes (mean age = 22.2, SD = 4.7, four males) were included in our analyses. Ethical approval was obtained from the University of Rochester Research Subjects Review Board.

All recruited synesthetes' self-reported experiences were previously confirmed with an objective test of genuineness — consistency over time — presented via the diagnostic website synesthete.org (see Eagleman *et al.*, 2007 for methods). This test identifies synesthetes based on replicated findings that synesthetes are significantly more consistent when repeatedly choosing synesthetic colors for the stimuli eliciting them (e.g., letters) compared to non-synesthetes. Our synesthetes experienced colors in response to graphemes ($n = 7$), days of the week ($n = 2$), and/or months of the year ($n = 1$) as confirmed by mean standardized scores of 0.55 (SD = 0.19), 0.65 (SD = 0.49), and 0.38 (SD = 0), respectively, where a score below 1.0 confirms synesthesia (see Eagleman *et al.*, 2007 for details). Seven synesthetes experienced colors for graphemes only; one had synesthetic colors for days of the week and months of the year; and one experienced colors for only days of the week. Non-synesthetes completed a synesthesia questionnaire (see synesthete.org) on paper, indicated no synesthetic experiences, and were further verbally questioned to ensure a complete lack of such experiences.

2.1.2. Stimuli

The visual stimulus was a 20° × 4° rectangle with a Gaussian luminance profile along the x-axis as seen in Fig. 1(b). We created two additional 'noise' levels of this visual stimulus by decreasing the brightness to 50% and 20% of the maximum luminance, thereby reducing the peak-trough difference (contrast) in the Gaussian luminance profile. The auditory stimulus was a 400 ms long recording of popcorn kernels being shaken at 5 Hz in a pill bottle. This auditory stimulus is ideal for localization as it represents a wide range of frequencies and has several onsets and offsets (Muir *et al.*, 1989). To mimic the temporal dynamics of the auditory stimulus and encourage integration, we added flicker to the presentation of the visual stimulus. The visual stimulus appeared for two frames (approximately 32 ms), then disappeared for two frames, and repeated this pattern for the duration of the auditory stimulus.

2.1.3. Procedure

Participants were tested individually in a dark, quiet room over a span of two sessions on consecutive days, with each session lasting approximately 75 min. After adapting to the dark for approximately six minutes, participants were instructed that they were serving as a nighttime boat lookout in a world where giant insects and sharks were the main dangers. Participants' main task was

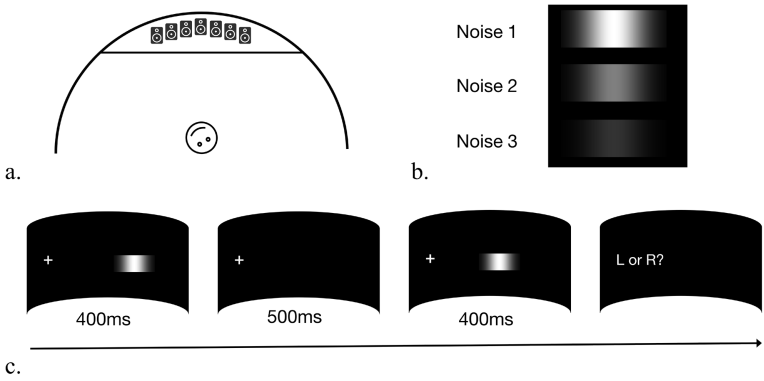

Figure 1. Experiment setup. (a) Schematic depiction of apparatus. Viewing the setup from above, the black curved line represents the screen onto which visual stimuli were projected. The gray semi-circle indicates the custom-built table upon which seven speakers sat (see text for details). The smiley face indicates a participant sitting at the center of the setup and facing ±45°. (b) Visual stimuli, showing the three noise levels used in the experiment. (c) Example of a visual only trial. Note that the visual stimuli flickered, which is not depicted in this figure.

to indicate the *direction* that the giant insects (audio, visual, or audiovisual stimuli) were moving. Additionally, participants had a filler task of detecting 'sharks' (denoted with the caret symbol: ^). Participants completed ten practice trials with feedback and with the experimenter present in the room to ensure that the task was understood before beginning the experimental trials.

Participants sat seven feet in front of a 180° curved projection screen as depicted in Fig. 1(a). Seven speakers located at 0°, ±4°, ±8°, and ±12° were placed on a custom-built curved, foam-lined table directly in front of the projection screen. This entire table was cloaked in black fabric to prevent participants from seeing the location of the speakers. Participants faced and focused on a fixation cross located at ±45° (the front of the boat), counterbalanced between participants. Accordingly, all auditory and visual stimuli for localization (i.e., the giant insects) occurred in the periphery.

On each trial participants were presented with two sequential stimuli (Fig. 1(c)). The task was to indicate whether the stimulus was moving to the left or to the right using the left and right shoulder buttons of a gaming controller, respectively. Auditory only, visual only, and audiovisual (aligned or misaligned) trials were randomly intermixed and presented in blocks of 100 with mandatory one-minute breaks in between blocks. Trials presented a 'standard' stimulus at 0° and a 'probe' stimulus at 0°, ±4°, ±8°, or ±12°, yielding seven different positions for the unimodal trial conditions. Unimodal trials were presented to ascertain the reliability of individual participants' auditory and visual performance. Audiovisual trials were presented to determine how participants integrated audio and visual cues to azimuthal location. Crucially,

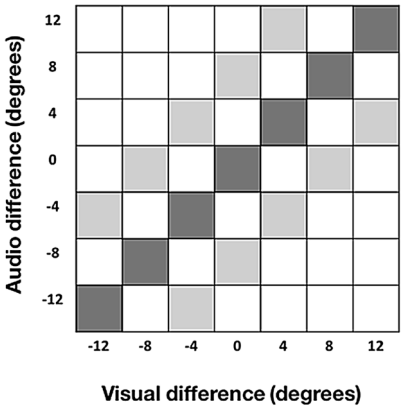

Figure 2. Audiovisual trials. Axes indicate the location of the probe with respect to the standard audiovisual stimulus (which was always aligned and presented at 0°). Dark grey = aligned, light grey = misaligned. Twenty-five repetitions of each stimulus were presented.

a subset of the audiovisual trials slightly misaligned the audio and visual cues. Introducing such discrepancies (i.e., cue conflicts) is crucial for quantitatively measuring cue weights during the integration process. Figure 2 displays the 17 audiovisual trial positions, which were either aligned (audio and visual stimuli presented at the same location, dark grey grid locations) or misaligned (audio and visual stimuli for the probe separated by ±8°, light grey grid locations). Presentation order of the standard and probe was counterbalanced across trials within participants. For audiovisual and visual only trials, the noise level (1–3) that varied the reliability of the visual cue was also randomized. Each individual trial type was presented 25 times, yielding a total of 1975 trials across two sessions. In addition to the localization task, participants completed an embedded detection task at the fixation location. On approximately 5% of trials, the fixation cross briefly changed to a caret (^) and participants pressed a button with their right thumb to indicate this occurrence. We used performance on this task as a measure of attention and motivation (excluding participants whose detection rates were below 85%).

2.2. Results

In Experiment 1, synesthetes performed a spatial localization task when presented audio only, visual only, and audiovisual stimuli. Crucially, a subset of the audiovisual stimuli presented slightly conflicting cues regarding the location of the stimulus, which allowed us to estimate the auditory and visual weights used during the cue combination process (i.e., the extent to which they relied on auditory and visual information, respectively). We also manipulated the signal to noise ratio in the visual signal in order to test whether decreasing visual signal reliability leads synesthetes to decrease their visual

weights, as predicted by the statistically optimal use of the two sources of sensory information.

Before comparing synesthetes' and non-synesthetes' cue integration behavior to a statistically optimal model, we fit psychometric functions to characterize their behavior in our task (see Appendix for fitting procedure details). First, we estimated unimodal sensory variances (audio and visual) for each participant by fitting psychometric curves to their localization performance in each of the four unimodal conditions (three noise levels of visual only and one noise level of auditory only). Calculating sensory variance for multiple noise levels of visual only stimuli while keeping the auditory only noise level constant allows us to test the prediction that participants should weight visual sensory information as a function of visual sensory variance relative to auditory sensory variance, when combining the two sources of information. Fitting participants' unimodal labeling data with cumulative Gaussian distributions (Fig. 3) yielded the point of subject equality (PSE) and variance (slope) associated with the participants' representation of each unimodal cue condition. Next, we fit synesthetes' localization data during each of the three audiovisual conditions (noise 1–3) with psychometric curves and simultaneously ascertained the weights that participants actually assigned to each modality (Fig. 4).

After calculating participants' visual weights during our cue combination task, we examined the extent to which their behavior conformed to an ideal observer using all sensory information available (predictions generated by equations (1)–(3); see Fig. 4). If synesthetes used both auditory and visual information efficiently, their visual weights should align with the predictions of the ideal observer. To determine whether synesthetes utilized sensory information efficiently in our audiovisual localization task, we conducted a mixed-effects linear regression predicting visual weight from weight type (observed, predicted), noise level (1–3), and full random effects (i.e., intercepts and slopes by participant). In line with the predictions of the ideal observer, synesthetes' visual weights decreased as visual noise increased; $\beta = -0.09$, $SE = 0.01$, $p < 0.001$. Moreover, the rate at which synesthetes' weights changed as a function of noise was indistinguishable from that predicted by the ideal observer; $\beta = 0.02$, $SE = 0.03$, ns. These results suggest that synesthetes integrate audio and visual cues to azimuthal location efficiently — that is, consistent with statistically optimal behavior. To our knowledge, these findings are the first quantitative investigation of cue weighting in synesthetes and demonstrate synesthetes' statistically efficient use of auditory and visual information during azimuthal localization.

Lastly, we evaluated controls' cue integration behavior and compared it to that of synesthetes. To determine whether controls (as has been shown previously in the literature) also utilized sensory information efficiently in our audiovisual localization task, we conducted a mixed-effects linear regression

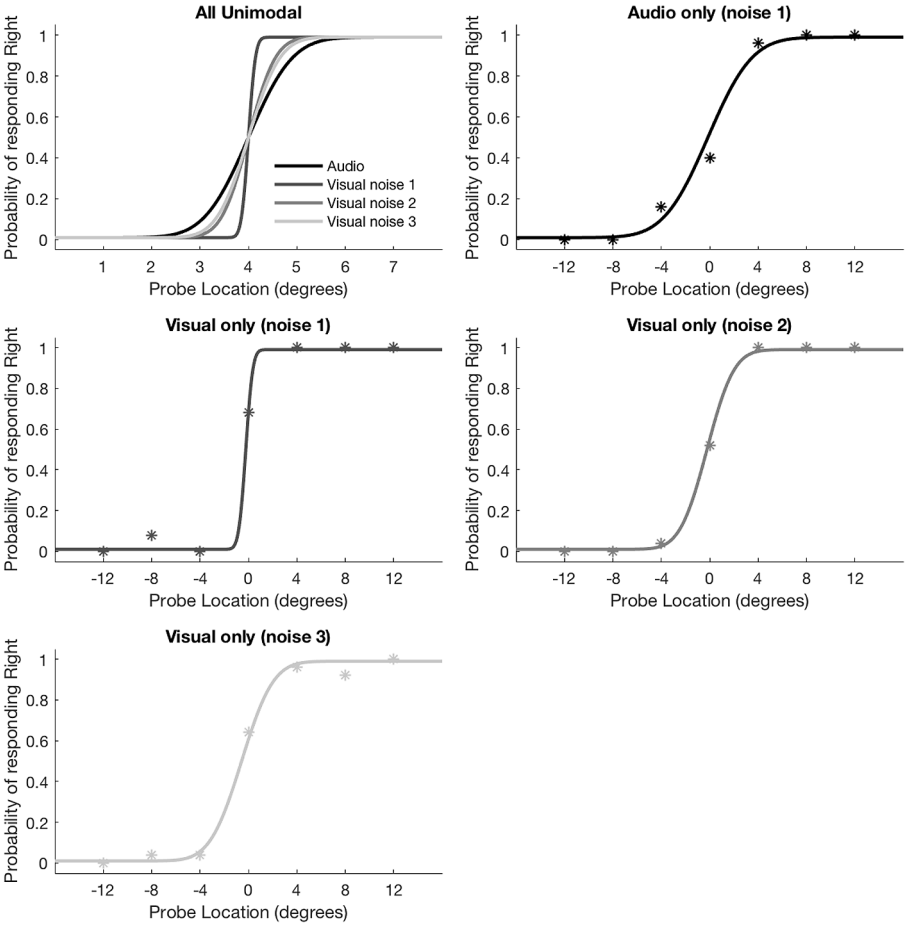

Figure 3. Cumulative Gaussian fits of unimodal trials for a representative synesthete. The top left panel plots all four unimodal cumulative Gaussian fits with the PSE equalized for descriptive purposes, to allow for easier slope comparison. The remaining panels plot cumulative Gaussian fits along with data for each unimodal condition separately. The standard is always presented at 0°.

predicting visual weight from weight type (observed, predicted), noise level (1–3), and full random effects (i.e., intercepts and slopes by participant). Matching the findings for synesthetes, this analysis revealed that controls' visual weights decreased as visual noise increased; $\beta = -0.15$, $SE = 0.03$, $p < 0.001$. Additionally, the rate at which controls' weights changed as a function of noise was indistinguishable from that predicted by the statistically optimal ideal observer; $\beta = 0.07$, $SE = 0.04$, ns. These findings further support the extensive literature demonstrating that a random sample of humans

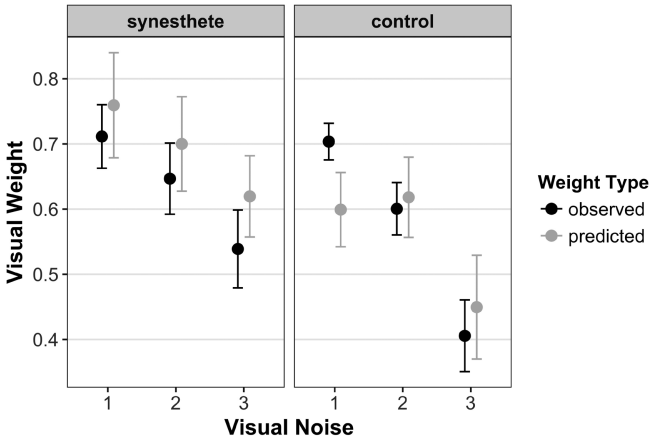

Figure 4. Observed and predicted visual weights for audiovisual trials. Note that neither synesthetes' nor controls' observed visual weights differ from the predicted visual weights. Error bars are standard error.

(presumably 96% non-synesthetes) combine cues in proportion to their reliability.

Finally, we conducted a mixed-effects linear regression including both synesthetes' and controls' observed weights to investigate group differences. This analysis revealed a significant group by noise interaction, with synesthetes' visual weights decreasing more slowly than controls' weights as a function of noise; $\beta = 0.06$, $SE = 0.03$, $p < 0.05$. It is important to note that this interaction does not bear on each group's performance with regard to statistically optimal performance, *given* their own actual visual and auditory weights. However, this interaction does imply that synesthetes are less affected by visual noise than controls. Our speculation about this finding is that synesthetes may more effectively build an internal model of the noise in the stimuli, thereby overcoming to a greater extent than controls the influence of this noise on the estimate of the signal (i.e., the actual location of the combined auditory-visual location).

2.3. Discussion

With an audiovisual localization task, we quantitatively investigated synesthetes' cue integration behavior. We used individual participants' unimodal performance during our task to generate predictions from a model that efficiently uses all sensory information available. Comparing synesthetes' actual cue weights to those predicted by the ideal observer, we found that synesthetes weighted cues in a manner consistent with statistically optimal integration. Specifically, synesthetes' visual weights decreased as a function of increasing visual noise at a rate that was indistinguishable from the model's predictions.

In line with the large body of existing research examining cue integration in the general population, we found that our non-synesthetes predictably integrated audiovisual cues in the same manner. Therefore, our results suggest that synesthetes rely on computational strategies for cue integration that are similar to those of non-synesthetes.

Additionally, this experiment highlights the importance of considering unimodal performance when investigating bimodal integration. That is, comparing synesthetes' and controls' bimodal performance in this experiment alone may lead one to conclude that synesthetes and controls integrate cues in different manners given the group differences in visual weight across noise levels. However, comparing each group's bimodal performance to a model that incorporates their unimodal data reveals that both groups integrate audiovisual cues in our task in accordance with the predictions of a statistically optimal observer. Accordingly, our results demonstrate that it is necessary to consider synesthetes' and controls' sensitivity to individual cues when investigating potential group differences during cue combination.

3. Experiment 2: Audiovisual Categorization

Experiment 1 investigated synesthetes' integration of audio and visual sensory cues to a *continuous* variable — spatial location. However, much of our world is hierarchically structured from sensory input into categorical representations, and ultimately to abstract semantic dimensions (Ahissar and Hochstein, 2004). Thus, sensory cues are not the only source of information relevant for cue combination. Deciding whether a beverage is coffee or tea, for example, may require integrating color, smell, and taste values of the beverage along with knowledge of the categories 'coffee' and 'tea'. Changing the relationship between these categories (e.g., discriminating apple juice and coffee) or the variance of each category (e.g., discriminating English breakfast tea from Starbucks' dark roast) should influence the weights assigned to each sensory cue.

Previous studies have theorized about, and investigated, how humans integrate information in this more complex scenario, arguing that the precise distributional properties of task-relevant categories should be utilized during cue combination (e.g., Bejjanki *et al*., 2011; Feldman *et al*., 2009). That is, the mean and variance (assuming Gaussian distributions) of the task-relevant categories, in addition to sensory information, should influence how cues are combined. This complex cue integration problem across categorical dimensions could, in principle, be solved by extension of the continuous linear cue integration model used in Experiment 1.

Formally, when categorizing a multisensory stimulus, an ideal learner constructs a discriminant vector linearly connecting the means of each category,

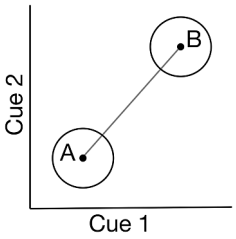

Figure 5. Cue combination involving categorization. A depiction of the categorization problem where each category is defined by two cues. The x and y axes represent the strength of each sensory cue. The circles labeled A and B represent the mean and covariance of each cue for categories A and B for a given participant. The grey diagonal line represents the linear discriminant vector D that an optimal categorizer projects the received bi-cue signal onto (see text).

and projects the stimulus onto this vector (Bejjanki *et al*., 2011; see Fig. 5). This projection of the stimulus onto the discriminant vector is the decision variable D, which determines the categorization of the stimulus based on some criterion:

$$D = w_{\mathrm{A}}\hat{S}_{\mathrm{A}} + w_{\mathrm{B}}\hat{S}_{\mathrm{B}} \tag{4}$$

where the estimates generated from the two cues are represented by $\hat{S}_{\mathrm{A}}$ and $\hat{S}_{\mathrm{B}}$. The weights for each cue are then given by:

$$w_{\mathrm{A}} = \frac{\frac{\Delta\mu_{\mathrm{A}}}{\sigma^2_{\mathrm{A,sense}}+\sigma^2_{\mathrm{A,cat}}}}{\frac{\Delta\mu_{\mathrm{A}}}{\sigma^2_{A,\mathrm{sense}}+\sigma^2_{\mathrm{A,cat}}} + \frac{\Delta\mu_{\mathrm{B}}}{\sigma^2_{\mathrm{B,sense}}+\sigma^2_{\mathrm{B,cat}}}}\ \text{and}$$

$$w_{\mathrm{B}} = \frac{\frac{\Delta\mu_{\mathrm{B}}}{\sigma^2_{\mathrm{B,sense}}+\sigma^2_{\mathrm{B,cat}}}}{\frac{\Delta\mu_{\mathrm{B}}}{\sigma^2_{\mathrm{B,sense}}+\sigma^2_{\mathrm{B,cat}}} + \frac{\Delta\mu_{\mathrm{A}}}{\sigma^2_{\mathrm{A,sense}}+\sigma^2_{\mathrm{A,cat}}}} \tag{5}$$

where $\sigma^2_{\mathrm{A,sense}}$ and $\sigma^2_{\mathrm{B,sense}}$ are sensory uncertainty variances for the two signals and $\sigma^2_{\mathrm{A,cat}}$ and $\sigma^2_{\mathrm{B,cat}}$ represent the variability in the distribution of the sensory signals occurring in the categories. Lastly, $\Delta\mu_{\mathrm{A}}$ and $\Delta\mu_{\mathrm{B}}$ represent the difference between category means along each cue dimension. The formalization of cue combination in categorization thus posits that an ideal observer should incorporate not only sensory information, but also the precise distributional properties of the task relevant categories when combining the cues.

Previous research has investigated the extent to which human performance is qualitatively consistent with the predictions of this ideal model, using real world categories such as phonemes (e.g., Bejjanki *et al*., 2011; Clayards *et al*., 2008; Feldman *et al*., 2009) and we recently developed a paradigm for *quantitatively* investigating humans' ability to optimally integrate cues and category information by teaching participants novel audiovisual categories (Bankieris

et al., subm.). Our findings demonstrated that non-synesthetes' behavior is quantitatively indistinguishable from a statistically optimal model that integrates both sensory and categorical information. In the present experiment, we use the same paradigm to test whether synesthetes' cue integration also adheres to these statistically optimal principles.

3.1. Methods

3.1.1. Participants

Eight linguistic–color synesthetes served as participants (six of whom participated in Experiment 1). They had no known hearing problems and normal or corrected-to-normal vision, were recruited from our existing database of Rochester area synesthetes, and were compensated $10/h for their participation. To compare synesthetes' performance to that of non-synesthetes, we also report data from 15 non-synesthetes who previously participated in this experiment (Bankieris *et al*., subm.). One additional non-synesthete participated and was excluded from group analyses because his performance for unimodal auditory trials at noise level 4 was indistinguishable from chance across all auditory steps and thus could not be fit with a psychometric function. Ethical approval was obtained from the University of Rochester Research Subjects Review Board.

We confirmed our synesthetes' self-reported experiences using the same on-line test as that used in Experiment 1 (synesthete.org; see Eagleman *et al*., 2007 for methods). Our linguistic–color synesthetes experienced colors in response to letters and/or numbers ($n = 5$), days of the week ($n = 4$), and/or months of the year ($n = 3$) as confirmed by mean standardized scores of 0.56 (SD = 0.23), 0.57 (SD = 0.20), and 0.45 (SD = 0.11), respectively, where a score below one confirms synesthesia (see Eagleman *et al*., 2007 for details). Four synesthetes experienced color for graphemes only, one synesthete experienced colors for days of the week only, one experienced colors for days of the week only, and three experienced colors for graphemes, days of the week, and months of the year.

3.1.2. Stimuli

We created novel categories defined by two cues: number of dots and auditory pitch (see Fig. 6; Bankieris *et al*., subm.). The number of dots spanned from 11 to 47 in 15 perceptually linear steps. These steps fall along a mathematically logarithmic scale which creates linear steps in perceptual space because number is perceived according to Weber's law. As seen in Fig. 7, black dots were positioned pseudorandomly within a predefined square area to create a specific level of numerosity with no dot overlap. Pitch stimuli were pure tones with frequencies ranging from 264 to 502 Hz in 15 perceptually linear steps. We created three additional noise levels of auditory

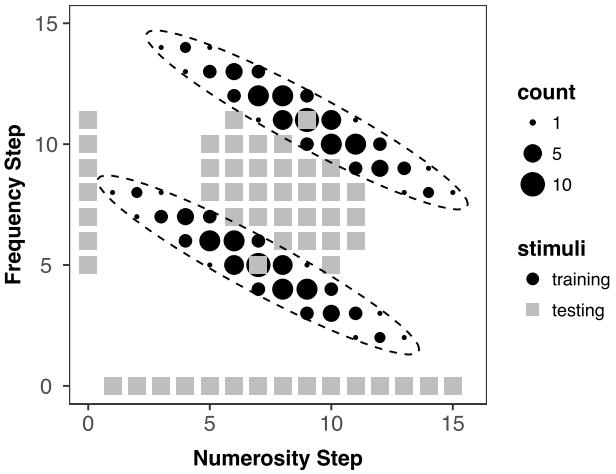

Figure 6. Training and test stimuli. Black circles represent the occurrence of exemplars of the two-cue stimuli during training. The elliptical clusters of black circles represent the Gaussian distributions of the two task-relevant categories. The size of each circle represents the number of exemplars of each stimulus that were presented during one learning block. Grey squares represent testing stimuli (bimodal in center, unimodal along the x- and y-axes). Twenty-five repetitions of each testing stimulus were presented. Category labels (taygoo and dohkah) and locations (as above or rotated 90°) were counterbalanced across participants.

stimuli by adding pink noise (a signal in which power is inversely proportional to the frequency of the signal: $1/f$) to the pure tones. Noise level 1 stimuli were 100% pure tones with 0% pink noise added; noise levels 2–4 were composed of pure tones with 83.3%, 93.8%, and 96.8% pink noise added, respectively, and normalized for overall acoustic energy (available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2VGOOY). Novel categories were defined as two-dimensional Gaussian distributions in the auditory-visual space of the two cues (with the frequency of occurrence of each stimulus rounded to integers). Importantly, these categories cannot be separated using only one of the cues. That is, no horizontal or vertical line drawn in Fig. 6 will perfectly separate these two categories, which necessitates the use of both cues for successful categorization. Half of the participants learned the categories depicted in Fig. 6 (small number and low pitch, large number and high pitch) and the other half learned these categories rotated 90° (small number and high pitch, large number and low pitch).

3.1.3. Procedure

Participants were tested individually in a quiet room over a span of four sessions on consecutive days, with each session lasting approximately one hour. On the first day, participants were told that scientists had just discovered two new species and their task was twofold: (1) to become an expert at classifying

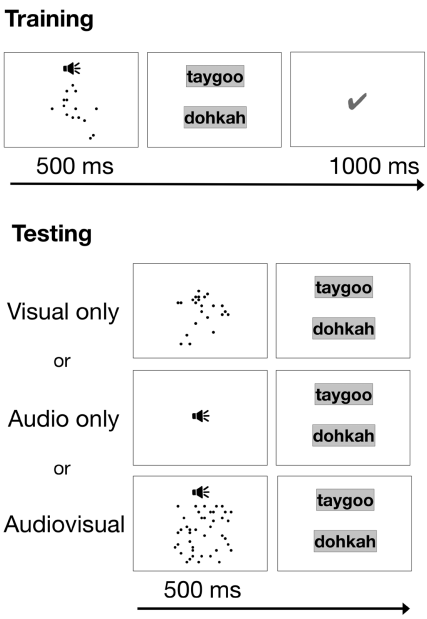

Figure 7. Trial structure. *Training*: example of audiovisual training trials with feedback. *Testing*: example of visual only, audio only, and audiovisual testing trials without feedback.

exemplars and (2) to help the scientists categorize unclassified exemplars. We informed participants that the two species, labeled with the nonsense words *taygoo* and *dohkah*, could be discriminated using both the pitch of their calls (i.e., pitch frequency) and the number of droppings they produce (i.e., number of dots).

3.1.4. Training

Each of the four sessions began with a training phase composed of a variable number of blocks, depending on each participant's learning rate. Each training block presented the full distribution of audiovisual category stimuli (103 of *taygoo* and 103 of *dohkah*) to ensure that all participants experienced the same category statistics. Participants completed as many blocks as necessary to reach 90% classification accuracy, with a limit set at four training blocks. As seen in Fig. 7, each trial within a training phase block presented an audiovisual stimulus for 500 ms drawn without replacement from the two-dimensional Gaussian category distributions (see Fig. 6). Two buttons labeled 'taygoo' and 'dohkah' then appeared on the touch screen and participants touched a button to submit their classification response. Feedback indicating whether their choice was correct or incorrect was displayed on the screen for 1000 ms before the next trial began. Category and button labels were counterbalanced across participants.

3.1.5. Testing

After 90% classification accuracy was reached, participants progressed to the test phase, where audio only, visual only, and audiovisual trials were presented (Fig. 7). Eight blocks of approximately 130 testing trials were completed during each session — six blocks presenting audiovisual stimuli and two blocks presenting audio only and visual only trials intermixed. The order of these blocks (unimodal or bimodal first) was counterbalanced within participant across day. Each test trial displayed a visual stimulus (or a speaker icon in the case of audio only trials) for 500 ms while an auditory stimulus of equal length was played for audiovisual and audio only trials. As in the practice trials, participants then selected their answer by touching one of two buttons but did not receive feedback. A blank screen was presented for 500 ms before the next trial began. We concentrated our unimodal auditory test items on the seven steps in the middle of the auditory frequency range (steps 5–11), from the mean of one category to the mean of the other category. Since the purpose of these unimodal trials was to ascertain a full psychometric function for each cue individually and the difference between category means on the numerosity cue is only two steps, we included all 15 numerosity steps in the visual only trials. The audiovisual trials consisted of 31 unique combinations of audio and visual cues (central gray squares in Fig. 6), designed to introduce slight discrepancies between individual cues. For most audiovisual stimuli, therefore, the likelihood that the visual component was a 'taygoo' was not equal to the likelihood that the auditory component was a 'taygoo.' Introducing such discrepancies (i.e., cue conflicts) is crucial for quantitatively measuring cue weights during the integration process. Auditory stimuli in audiovisual trials and audio only trials were presented in four different noise levels randomly interleaved throughout the test phase. We presented 25 repetitions of each of these test stimuli, yielding a total of 4175 test trials across four sessions.

3.2. Results

As in Experiment 1, we fit psychometric functions to characterize participant behavior before comparing synesthetes' bimodal behavior to the categorical model. First, we estimated unimodal sensory variances (audio and visual) for each synesthete by fitting psychometric curves to their categorization performance in each of the five unimodal conditions (four noise levels of auditory only and one noise level of visual only). Fitting participants' unimodal labeling data with cumulative Gaussian distributions yielded the point of subject equality (PSE) and variance (slope) associated with participants' representation of the sensory information available in each unimodal cue condition (Fig. 8). Next, we fit synesthetes' labeling data during each of the four audiovisual conditions (noise 1–4) with psychometric curves and simultaneously

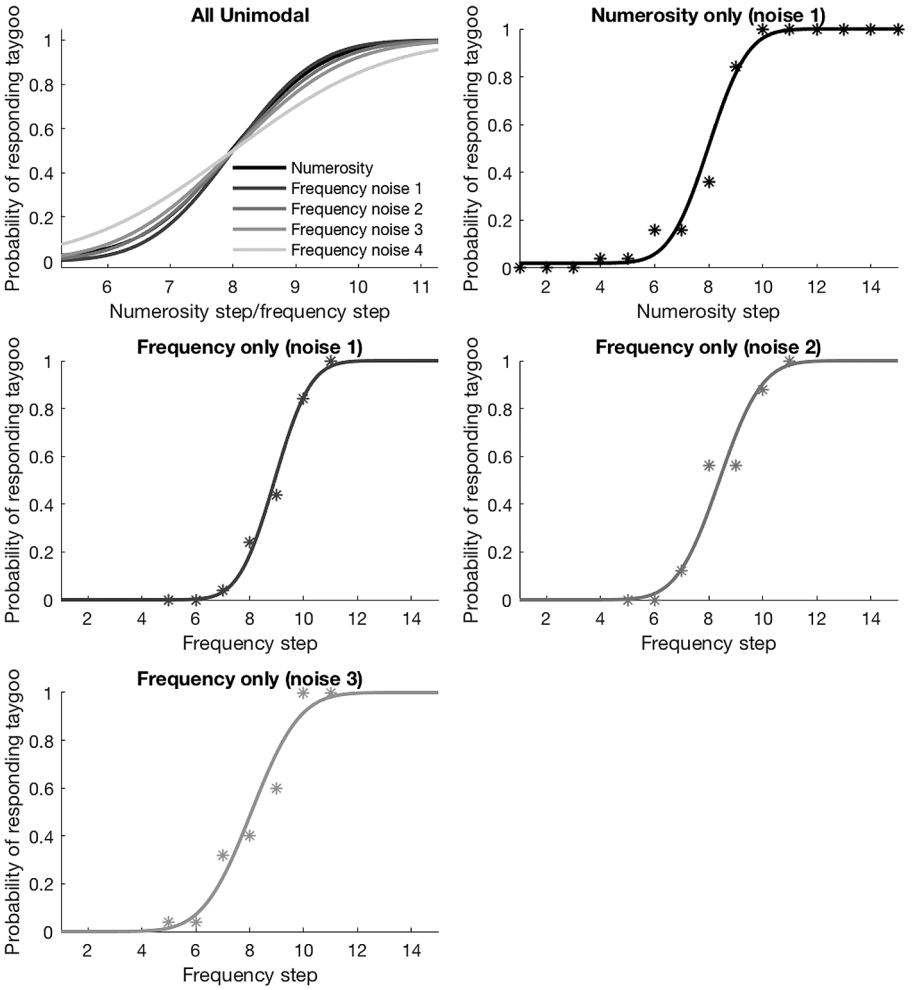

Figure 8. Cumulative Gaussian fits of unimodal trials for a representative synesthete. The top left panel plots all five unimodal cumulative Gaussian fits with the PSE equalized for descriptive purposes, to allow for easier slope comparison. The remaining panels plot cumulative Gaussian fits along with data for each unimodal condition separately.

ascertained the weights that participants actually assigned to each modality (Fig. 9).

After calculating synesthetes' auditory weights during the audiovisual categorization task, we examined the extent to which their behavior conformed to an ideal observer using both sensory and category information (see Fig. 9; predictions generated by equations (4) and (5) and the inclusion of a standard correction for within category cue correlation; Oruç *et al.*, 2003). As a comparison, we also generated predictions from the continuous model (with

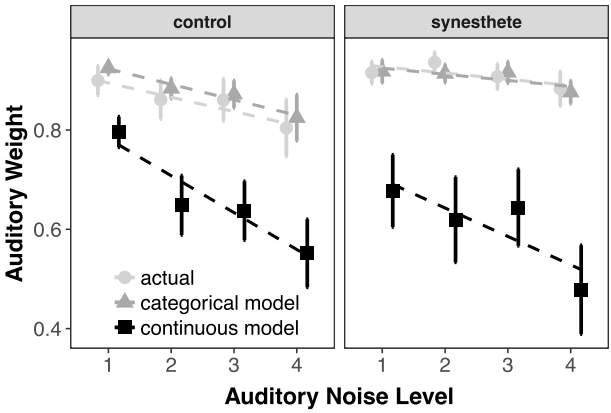

Figure 9. Observed auditory weights for audiovisual trials alongside predictions from the categorical model and the continuous model. Note that synesthetes' and controls' actual auditory weights differ from the continuous model's predictions but are indiscriminable from the categorical model's predictions. Error bars are standard error. Lines are linear fits generated for visualization purposes only.

the correlation correction included) which considers only sensory information (equations (1)–(3)). If synesthetes are using only sensory information, auditory weights should decrease as sensory uncertainty is added to the auditory cue (as we found in Experiment 1). However, if synesthetes are using category information in addition to sensory information during their judgments, then we should see two patterns in their data:

(1) Auditory weights should be higher than those predicted by the continuous model.

(2) The amount by which auditory weights decrease as a function of auditory noise should be less than predicted by the continuous model.

3.2.1. Prediction 1: Higher Auditory Weights Compared to Continuous Model

If synesthetes were appropriately considering category information while performing this cue integration task, their auditory weights should align with the predictions of the categorical model, which are higher than those of the continuous model. While the continuous model uses only sensory information to determine auditory weights, our participants (and the ideal categorical model) have access to the distributional information of the categories. The fact that the category means in this task have a greater distance between them along the auditory frequency dimension makes frequency more informative than numerosity at the category level. Likewise, there is less frequency variance within a category compared to numerosity variance within a category, again making auditory frequency information more reliable at the category level. To

examine whether participants used this category information, we fit a mixed-effect linear regression predicting auditory weights from weight type (actual, category model predictions, continuous model predictions) and noise (1–4) with a full random effects structure (i.e., random intercepts and weight type by noise slopes per participant). With participants' observed weights as the reference level (i.e., coded as 0), the beta coefficients for the two other levels of weight type (category model predictions, continuous model predictions) indicate whether or not the participants' weights differ from each of these model predictions. Our analysis found that synesthetes' actual auditory weights were significantly higher than the continuous model's predictions ($\beta = -0.19$, $SE = 0.08$, $p < 0.05$) but did not differ from the category model's predictions; $\beta = -0.01$, $SE = 0.04$, ns. These results demonstrate that synesthetes' auditory weights did not align with the predictions of a model using only sensory information, and were quantitatively indiscriminable from the predictions of a model that incorporates both category and sensory information during cue combination. This finding supports the hypothesis that synesthetes are sensitive to the distributions of categories during cue combination.

3.2.2. Prediction 2: Smaller Effect of Noise on Auditory Weights

The second prediction made by the category model of cue combination is that auditory weights will decrease as a function of auditory noise, but by a smaller amount than predicted by the continuous model. That is, the effect of noise on auditory weights should be smaller if participants are using category information in addition to sensory information. This prediction arises because in addition to sensory information, the category model is utilizing information regarding the category distributions, which does not change as a function of noise. If synesthetes used category information in an ideal manner during our task, their auditory weights should align with the predictions of the category model and not the continuous model. Using the mixed-effects linear regression described above, we investigated the amount by which synesthetes' auditory weights decreased as a function of noise. With synesthetes' observed weights as the reference level (i.e., coded as 0), the beta coefficients for the interaction of noise (1–4) and the two other levels of weight type (category model predictions, continuous model predictions) indicate whether or not the noise effect for synesthetes' weights differs from each of these models' predictions. This analysis revealed that synesthetes' auditory weights decreased as a function of noise by a significantly smaller amount than the continuous model's predictions ($\beta = -0.04$, $SE = 0.02$, $p < 0.05$) but were indistinguishable from the category model's predictions; $\beta = 0.00$, $SE = 0.01$, ns. These results demonstrate that the amount by which synesthetes down-weighted auditory information across noise levels (i.e., as reliability decreases) occurs exactly in the manner predicted by the category model rather than the continuous model.

Synesthetes did not use sensory variance as the sole factor influencing their cue weights, but additionally integrated the information provided by category structure into their cue weights. Taken together, our findings represent the first set of evidence demonstrating that synesthetes qualitatively and quantitatively integrate both sensory and category information during cue combination across categorical dimensions in a manner consistent with a statistically optimal model.

3.2.3. Group Differences

Comparing the synesthetes to a set of 15 non-synesthetes who previously participated in this same task (Bankieris *et al.*, subm.), we found that both synesthetes and non-synesthetes integrated sensory and category information in a manner that is indistinguishable from a statistically optimal observer. To quantitatively compare group behavior we conducted a mixed effects linear regression, modelling observed weights from group (synesthete, control) and noise level (1–4). Results from this analysis found no significant group differences, suggesting that synesthetes and controls quantitatively combine cues along a categorical dimension using the same computational principles. Supporting our findings from Experiment 1, this cue combination task involving categories demonstrates that synesthetes — like non-synesthetes — integrate cues according to Bayesian principles.

3.3. Discussion

In Experiment 2, we examined the computational principles underlying synesthetes' cue combination when categorical dimensions are involved. To do so, we introduced novel multisensory categories to synesthetes and, after these categories were learned, we quantitatively analyzed their cue integration behavior during a categorization task in which cue conflicts were present. Critically, an ideal observer model performing such a cue combination task over categorical dimensions predicts that environmental variability of the categories themselves (specifically separation of categories along each cue dimension and category variance along each cue dimension) in addition to sensory variability should influence cue weighting. While it is very difficult to have knowledge of a given participant's category distributions for natural categories (i.e., speech phonemes), creating novel audiovisual categories allowed us to have strict control over synesthetes' exposure to these categories. Thus, we were able to quantitatively compare synesthetes' behavior to that of a statistically optimal observer who utilizes both category and sensory information. Our results demonstrate that synesthetes' behavior, like that of non-synesthetes, quantitatively matched a statistically optimal observer who ideally integrates cues based on both sensory and category information.

4. General Discussion

Using ideas from the cue integration literature, we sought to examine the computational mechanisms underlying cue combination in synesthetes. Specifically, we investigated synesthetes' audiovisual integration from the perspective of Bayesian cue integration to determine whether synesthetes rely on the same computational principles as non-synesthetes for cue integration. In Experiment 1, we assessed audiovisual integration with a spatial localization task. Results indicated that synesthetes, like non-synesthetes, integrated audio and visual sensory cues to location in a statistically optimal manner. Experiment 2 added the additional layer of category to the integration problem, examining audiovisual integration in the context of newly learned artificial categories. Again, results demonstrated that synesthetes' cue integration behavior mimicked that of non-synesthetes', as it was indistinguishable from an ideal observer model incorporating both sensory and category information. To our knowledge, these findings represent the first quantitative evidence demonstrating that synesthetes integrate audiovisual cues in a manner indistinguishable from an ideal observer, thereby suggesting that synesthetes and non-synesthetes use similar computational principles for cue integration. Our findings also help to clarify the rather confusing literature on sensory integration in synesthetes, which has provided highly variable evidence (from inferior integration to superior integration: Brang *et al*., 2012; Neufeld *et al*. 2012; Whittingham *et al*., 2014).

As our studies were designed to reveal the computational principles of integration itself, our results do not answer questions regarding the conditions under which integration occurs or the downstream effects of integration (e.g., changes in reaction time). However, given our findings that synesthetes integrate cues according to the same computational principles as controls, studies examining these additional questions can exclude differential cue integration principles as a confounding factor. Importantly, our findings highlight the necessity of considering group differences related to *unimodal* sensory processing when comparing synesthetes' and non-synesthetes' performance on cue integration tasks. Behavioral differences in group performance on audiovisual illusions, for instance, might arise due to quantitatively different unimodal perceptual abilities, which are then combined in an identical manner compared to controls. That is, group differences in unimodal processing could lead to different patterns of susceptibility to multisensory illusions, but might actually reflect use of the same integration principles. By generating bimodal predictions for each individual participant (based on their unimodal sensitivities), bimodal performance across groups (e.g., synesthetes and non-synesthetes) can be appropriately compared even if group unimodal differences exist. Accordingly,

this work empirically demonstrates the importance of evaluating synesthetes' performance on individual cues when investigating their cue integration.

We have presented evidence from two different experiments demonstrating statistically optimal audiovisual cue integration by synesthetes, but future research is needed to confirm that this ideal integration process also holds for other cues and modalities. As statistically optimal integration has been demonstrated in a variety of tasks and sensory domains with the general population, we expect that the same will be true for synesthetes. It might be particularly interesting, however, to examine whether or not synesthetes integrate their *synesthetic* experiences according to these same computational principles. A strength of our two experiments is that we presented auditory and visual stimuli that did not induce synesthetic experiences (thus eliminating the possibility that our findings are due to an additional synesthetic percept). Future experiments could present synesthesia-inducing stimuli to determine whether the reliability of a synesthetic percept itself influences the cue integration process. Such a design would examine whether synesthetic percepts are governed by the same cue integration principles as nonsynesthetic percepts.

References

Ahissar, M. and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning, *Trends Cogn. Sci.* **8**, 457–464.

Bankieris, K., Bejjanki, V. R. and Aslin, R. (subm.). Sensory cue-combination in the context of newly learned categories.

Bargary, G. and Mitchell, K. J. (2008). Synaesthesia and cortical connectivity, *Trends Neurosci.* **31**, 335–342.

Bejjanki, V. R., Clayards, M., Knill, D. C. and Aslin, R. N. (2011). Cue integration in categorical tasks: insights from audio-visual speech perception, *PLoS One* **6**, e19812. DOI:10.1371/journal.pone.0019812.

Brang, D., Williams, L. E. and Ramachandran, V. S. (2012). Grapheme–color synesthetes show enhanced crossmodal processing between auditory and visual modalities, *Cortex* **48**(5), 630–637.

Calkins, M. W. (1893). A statistical study of pseudochromesthesia and of mental forms, *Am. J. Psychol.* **5**, 439–466.

Claparede, E. (1903). Persistance de l'audition colorée, *C. R. Seances Soc. Biol. Fil.* **55**, 1257–1259.

Clayards, M., Tanenhaus, M. K., Aslin, R. N. and Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues, *Cognition* **108**, 804–809.

Day, S. (2005). Some demographic and socio-cultural aspects of synesthesia, in: *Synesthesia: Perspectives From Cognitive Neuroscience*, L. C. Robertson and N. Sagiv (Eds), pp. 11–33. Oxford University Press, New York, NY, USA.

Day, S. (2009). Types of synesthesia, in: *Synesthesia*. Retrieved from http://www.daysyn.com/Types-of-Syn.html.

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D. and Sarma, A. K. (2007). A standardized test battery for the study of synesthesia, *J. Neurosci. Meth.* **159**, 139–145.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion, *Nature* **415**(6870), 429–433.

Feldman, N. H., Griffiths, T. L. and Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference, *Psychol. Rev.* **116**, 752–782.

Grossenbacher, P. G. and Lovelace, C. T. (2001). Mechanisms of synesthesia: cognitive and physiological constraints, *Trends Cogn. Sci.* **5**, 36–41.

Hänggi, J., Wotruba, D. and Jäncke, L. (2011). Globally altered structural brain network topology in grapheme-color synesthesia, *J. Neurosci.* **31**, 5816–5828.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation, *J. Exp. Psychol.* **63**, 289–293.

Hillis, J. M., Ernst, M. O., Banks, M. S. and Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses, *Science* **298**(5598), 1627–1630.

Hubbard, E. M., Brang, D. and Ramachandran, V. S. (2011). The cross-activation theory at 10, *J. Neuropsychol.* **5**, 152–177.

Hupé, J. M. and Dojat, M. (2015). A critical review of the neuroimaging literature on synesthesia, *Front. Hum. Neurosci.* **9**, 103. DOI:10.3389/fnhum.2015.00103.

Jacobs, R. A. and Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth, *Vis. Res.* **39**, 4062–4075.

Jäncke, L., Beeli, G., Eulig, C. and Hänggi, J. (2009). The neuroanatomy of grapheme–color synesthesia, *Eur. J. Neurosci.* **29**, 1287–1293.

Jewanski, J., Day, S. and Ward, J. (2009). A colorful albino: the first documented case of synaesthesia, by Georg Tobias Ludwig Sachs in 1812, *J. Hist. Neurosci.* **18**, 293–303.

Körding, K. P. and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning, *Nature* **427**(6971), 244–247.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B. and Shams, L. (2007). Causal inference in multisensory perception, *PLoS One* **2**, e943. DOI:10.1371/journal.pone.0000943.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A. and Wallace, M. T. (2006). Enhanced multisensory integration in older adults, *Neurobiol. Aging* **27**, 1155–1163.

Michel, M. M. and Jacobs, R. A. (2008). Learning optimal integration of arbitrary features in a perceptual discrimination task, *J. Vis.* **8**(3), 1–16. DOI:10.1167/8.2.3.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals, *Cogn. Psychol.* **14**, 247–279.

Muir, D. W., Clifton, R. K. and Clarkson, M. G. (1989). The development of a human auditory localization response: a U-shaped function, *Can. J. Psychol.* **43**, 199–216.

Neufeld, J., Sinke, C., Zedler, M., Emrich, H. M. and Szycik, G. R. (2012). Reduced audio–visual integration in synaesthetes indicated by the double-flash illusion, *Brain Res.* **1473**, 78–86.

O'Hanlon, E., Newell, F. N. and Mitchell, K. J. (2013). Combined structural and functional imaging reveals cortical deactivations in grapheme–color synaesthesia, *Front. Psychol.* **4**, 755. DOI:10.3389/fpsyg.2013.00755.

Oruç, I., Maloney, L. T. and Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error, *Vis. Res.* **43**, 2451–2468.

Ramachandran, V. and Hubbard, E. M. (2001). Synaesthesia: a window into perception, thought and language, *J. Consc. Stud.* **8**, 3–34.

Rouw, R. and Scholte, H. S. (2007). Increased structural connectivity in grapheme-color synesthesia, *Nat. Neurosci.* **10**, 792–797.

Rouw, R. and Scholte, H. S. (2010). Neural basis of individual differences in synesthetic experiences, *J. Neurosci.* **30**, 6205–6213.

Rouw, R., Scholte, H. S. and Colizoli, O. (2011). Brain areas involved in synaesthesia: a review, *J. Neuropsychol.* **5**, 214–242.

Shams, L., Kamitani, Y. and Shimojo, S. (2000). Illusions: what you see is what you hear, *Nature* **408**(6814), 788.

Sinke, C., Neufeld, J., Zedler, M., Emrich, H. M., Bleich, S., Münte, T. F. and Szycik, G. R. (2012). Reduced audiovisual integration in synesthesia–evidence from bimodal speech perception, *J. Neuropsychol.* **8**, 94–106.

Tomson, S. N., Narayan, M., Allen, G. I. and Eagleman, D. M. (2013). Neural networks of colored sequence synesthesia, *J. Neurosci.* **33**, 14098–14106.

Van Beers, R. J., Sittig, A. C. and Denier, J. A. N. J. (1999). Integration of proprioceptive and visual position-information: an experimentally supported model, *J. Neurophysiol.* **81**, 1355–1364.

Weiss, P. H. and Fink, G. R. (2009). Grapheme-colour synaesthetes show increased grey matter volumes of parietal and fusiform cortex, *Brain* **132**, 65–70.

Whittingham, K. M., McDonald, J. S. and Clifford, C. W. (2014). Synesthetes show normal sound-induced flash fission and fusion illusions, *Vis. Res.* **105**, 1–9.

Wichmann, F. A. and Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit, *Percept Psychophys.* **63**, 1293–1313.

Appendix

A.1. Experiment 1: Data Analysis

To analyze participants' left/right responses, we fit psychometric curves to participants' localization performance for each of the seven stimulus conditions (one noise level of auditory only trials, three noise levels of visual only trials, and three noise levels of audiovisual trials) in a manner similar to the approach used by Bejjanki *et al*. (2011). For each unique stimulus, the raw response data were organized into arrays specifying the proportion of trials that a participant responded 'right' (out of 25 repetitions). Realizing that individual participants' data did not always span the entire range from 0.0 to 1.0, we used modified cumulative Gaussian psychometric functions including lapse rates to model their behavior more accurately (Wichmann and Hill, 2001). This psychometric function modeled the probability of responding 'right' as a mixture of an underlying Gaussian discrimination process and a random guessing process. We coded participant responses as $y_i = 0$ for a response of 'left' and $y_i = 1$ for a response of 'right'. We used the following psychometric

model:

$$p(y_i = 1 \mid x_i) = \gamma + (1 - \gamma - \lambda)\Gamma(x_i; \mu, \sigma)$$
$$p(y_i = 0 \mid x_i) = 1 - p(y_i = 1 \mid x_i) \tag{A.1}$$

where y_i is the participant's response when presented with stimulus x_i on trial i. μ and σ are the mean and standard deviation of the cumulative Gaussian, respectively. For the current task, μ represents the Point of Subjective equality (PSE) where the two presented locations for the standard and the probe are the same, and σ represents the discrimination threshold. Lapse rate parameters are represented by γ and λ, where γ is the base rate of responding 'right' when there is no evidence that the second stimulus was presented to the right of the first, and λ is the miss rate, i.e., the probability of responding incorrectly regardless of the amount of information for rightward movement from the first to the second stimulus. We constrained the lapse parameters to be between 0.0 and 0.25 and assumed that they were constant across noise levels and conditions (audio only, video only, or audiovisual). We used maximum likelihood fits to estimate the parameters of participants' psychometric functions.

A.2. Estimating Parameters for Unimodal Performance

The audio only stimuli were presented in noise 1 only. Accordingly, the likelihood of a subject making a decision Y_i on audio only trial i, when presented with auditory stimulus x_i can be written as:

$$\begin{aligned} l_i = &\left[\left(\gamma + (1 - \gamma - \lambda)\Gamma(x_i; \mu, \sigma)\right)Y_i\right] \\ &+ \left[\left(1 - \left(\gamma + (1 - \gamma - \lambda)\Gamma(x_i; \mu, \sigma)\right)\right)(1 - Y_i)\right] \end{aligned} \tag{A.2}$$

The likelihood function for the entire set of audio only data for a given subject is then:

$$L_{Aud} = \prod_{i=1}^{N} l_i \tag{A.3}$$

where N is the total number of audio only trials.The visual only trials were presented in three noise levels. Thus, the likelihood of a subject making a decision $Y_{i,j}$ on visual only trial i for noise level j, when presented with visual stimulus $x_{i,j}$ can be written as:

$$\begin{aligned} l_{i,j} = &\left[\left(\gamma + (1 - \gamma - \lambda)\Gamma(x_{i,j}; \mu_j, \sigma_j)\right)Y_{i,j}\right] \\ &+ \left[\left(1 - \left(\gamma + (1 - \gamma - \lambda)\Gamma(x_{i,j}; \mu_j, \sigma_j)\right)\right)(1 - Y_{i,j})\right] \end{aligned} \tag{A.4}$$

The likelihood function for the entire set of visual only trials for a given subject is then given by

$$L_{Vis} = \prod_{j=1}^{3} \prod_{i=1}^{N} l_{i,j} \tag{A.5}$$

where N is the number of visual only trials for each noise level and there are three noise levels.

A.3. Estimating Parameters for Audiovisual Performance

During each trial of the audiovisual localization task, stimuli consisted of both an auditory and a visual cue to location. Crucially, in a subset of the audiovisual stimuli, there were cue conflicts between the two modalities which allows for the estimation of the weights that participants used when combining the two cues. Since we are assuming that cue combination in our task is in a linear regime, we consider the effective stimulus in this task to be a weighted combination of the two stimuli. Parameters for the psychometric model (μ, σ, γ, and λ) and the weights assigned to each modality (w_{a} and w_{v}) for bimodal performance were computed from maximum likelihood fits to the raw bimodal performance data for each participant. Specifically, the audiovisual condition of trials had three noise levels, so the likelihood of a subject making a decision $Y_{i,j}$ on audiovisual trial i for noise level j, where the presented stimulus was $x_{\mathrm{a}_{i,j}}$ in the auditory domain and $x_{\mathrm{v}_{i,j}}$ in the visual domain, can be written as:

$$\begin{aligned} l_{i,j} = &\left[\left(\gamma + (1-\gamma-\lambda)\Gamma\left(\left((1-w_{\mathrm{v}})x_{\mathrm{a}_{i,j}} + w_{\mathrm{v}}x_{\mathrm{v}_{i,j}}\right); \mu_j, \sigma_j\right)\right) Y_{i,j}\right] \\ &+ \left[\left(1 - \left(\gamma + (1-\gamma-\lambda)\Gamma\left(\left((1-w_{\mathrm{v}})x_{\mathrm{a}_{i,j}} + w_{\mathrm{v}}x_{\mathrm{v}_{i,j}}\right); \mu_j, \sigma_j\right)\right)\right.\right. \\ &\left.\left. \times (1 - Y_{i,j})\right]\right. \end{aligned} \tag{A.6}$$

Since w_{a} and w_{v} sum to one, the above expression replaces w_{a} with $1 - w_{\mathrm{v}}$. The likelihood function for the entire set of audiovisual trials for a given subject is then given by:

$$L_{\mathrm{AV}} = \prod_{j=1}^{3} \prod_{i=1}^{N} l_{i,j} \tag{A.7}$$

where N is the number of audiovisual trials in each noise level and there are three noise levels.

A.4. Avoiding Local Maxima When Fitting Psychometric Functions

To avoid converging on local maxima, rather than on the desired global maximum likelihood, we repeated each maximum likelihood fit starting from five randomly chosen initial values for the parameters. We then selected the parameters that corresponded to the fit with the best maximal likelihood value, across the five fitting runs, as the best-fit parameters for the psychometric model.

A.5. Experiment 2: Data Analysis

To analyze categorization behavior, we fit psychometric curves to participants' labeling performance for each of the nine stimulus conditions (one noise level of visual only trials, four noise levels of auditory only trials, and four noise levels of audiovisual trials). For each unique stimulus, the raw response data were organized into arrays specifying the proportion of trials that a participant responded 'taygoo' (out of 25 repetitions). As with Experiment 1, realizing that individual participants' data did not always span the entire range from 0.0 to 1.0, we used modified cumulative Gaussian psychometric functions including lapse rates to model their behavior more accurately (Wichmann and Hill, 2001). This psychometric function modeled the probability of selecting the category 'taygoo' as a mixture of an underlying Gaussian discrimination process and a random guessing process. We coded participant responses as $y_i = 0$ for a response of 'dohkah' and $y_i = 1$ for a response of 'taygoo'. We used the following psychometric model:

$$\begin{aligned} p(y_i = 1 \mid x_i) &= \gamma + (1 - \gamma - \lambda)\Gamma(x_i; \mu, \sigma) \\ p(y_i = 0 \mid x_i) &= 1 - p(y_i = 1 \mid x_i) \end{aligned} \tag{A.8}$$

where y_i is the participant's categorization of stimulus x_i on trial i. μ and σ are the mean and standard deviation of the cumulative Gaussian, respectively. For the current task, μ represents the Point of Subjective equality (PSE) between the two categories, and σ represents the discrimination threshold. Lapse rate parameters are represented by γ and λ, where γ is the base rate of responding 'taygoo' when there is no evidence for category 'taygoo', and λ is the miss rate, i.e., the probability of responding incorrectly regardless of the amount of information for category 'taygoo'. We constrained the lapse parameters to be between 0.0 and 0.25, held them constant across noise levels within a condition (audio only, video only, or audiovisual), but allowed them to vary across conditions. We used maximum likelihood functions to estimate the parameters of participants' psychometric functions.

A.6. Estimating Parameters for Unimodal and Bimodal Performance

The visual only stimuli were presented in noise level 1 only (i.e., maximum reliability) and the auditory stimuli were presented in four noise levels (for auditory only and audiovisual trials). Parameters for unimodal and bimodal performance were estimated using a similar approach to that used in Experiment 1.