

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Factors Influencing Likelihoods and Priors in Multisensory Perception

**Permalink**

<https://escholarship.org/uc/item/0vh7j8wv>

**Author**

Odegaard, Brian Andrew

**Publication Date**

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Factors Influencing Likelihoods and Priors in Multisensory Perception

A dissertation submitted in partial satisfaction of the  
requirements for the degree of Doctor of Philosophy  
in Psychology

by

Brian Andrew Odegaard

2015

© Copyright by

Brian Andrew Odegaard

2015

## ABSTRACT OF THE DISSERTATION

Factors Influencing Likelihoods and Priors in Multisensory Perception

By

Brian Andrew Odegaard

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2015

Professor Ladan Shams, Chair

GOAL: The goal of this dissertation is to investigate several factors influencing multisensory processing using a well-established Bayesian computational framework. Specifically, this project explores how systematic biases in human spatial perception, recalibration of the tendency to integrate sensory signals, and selective attention may (or may not) influence both likelihoods and priors in Bayesian models that can account for human performance quite well in a number of sensory tasks. These explorations are motivated by the following aims:

AIM 1: Exploring computational mechanisms underlying spatial biases in human visual, auditory, and audiovisual perception of space. Many investigations over the past twenty years have shown that human localizations along azimuth in both the visual and auditory modalities are often biased *away* from the true locations of the sensory stimulus. However, these investigations have often used small sample sizes and suboptimal paradigms to explore the nature of these spatial biases. In the first study of this dissertation, a dataset of unprecedented size (384 subjects) is used to explore the presence, magnitude, and direction of biases in humans' visual, auditory, and audiovisual perception. Following this behavioral exploration, several versions of a Bayesian causal inference model are tested to see whether these spatial biases are best accounted for by biases in sensory representations of space (i.e. likelihoods), pre-existing biases for localization (i.e. priors), or a combination of these mechanisms. The best-fitting model from this investigation is then used as a computational tool in subsequent investigations.

AIM 2: Investigating how to recalibrate the tendency to integrate multisensory signals. It has previously been demonstrated in several multisensory investigations that human perception of auditory space can shift based on previous multisensory experiences, and that this effect can be computationally accounted for in a shift in the auditory likelihood distributions. However, little (if any) evidence has shown whether or not the *a priori* tendency to integrate multisensory signals can be recalibrated by previous sensory experiences. In regards to this second aim, six experiments are conducted to determine (1) whether the tendency to integrate can change to due simple, brief (10-minute) exposure tasks, and if so, (2) what the spatial and temporal rules are that the brain uses to update its tendency to bind.

AIM 3: Examining the effects of selective and divided attention on sensory integration.

While the phenomenon of selective attention has been explored extensively by previous research, little consensus exists regarding the mechanism that enables enhanced perception for an attended stimulus. In this investigation, several experiments are conducted requiring subjects to allocate attention either selectively to one modality (vision or audition), or to divide attention across modalities, while performing either (1) a spatial task, or (2) a numerosity task. Data from both the spatial and temporal tasks are modeled to determine whether selective attention impacts the tendency to integrate multisensory stimuli, the sensory representations, or spatial priors. Leaving behind the antiquated idea of attention simply serving as a Bayesian prior, this series of studies demonstrates that attention primarily impacts sensory representations, and does not significantly impact the tendency to integrate multisensory stimuli.

The dissertation of Brian Odegaard is approved.

James Bisley

Mark Frye

Hongjing Lu

Dario Ringach

Ladan Shams, Committee Chair

University of California, Los Angeles

2015

I would like to dedicate the time, persistence, and diligent effort that went into this dissertation to my parents. Many thanks to my father, who left his work on the family farm at a young age to pursue a career in science, providing opportunities and opening doors for his children that would never have otherwise been possible; many thanks to my mother, who supported me through all of the ups and downs during the past ten years of my pursuits in higher education. Without both of your positive influences and guidance, none of this would have been possible.



## TABLE OF CONTENTS

I.	Abstract of the Dissertation .....	ii
II.	List of Figures .....	x
III.	List of Tables.....	xi
IV.	Acknowledgements .....	xii
V.	Vita .....	xiii
VI.	General Introduction.....	1
	A. Computational Models of Multisensory Integration: Towards a Complete Theoretical Account .....	3
	B. Dissertation Overview .....	5
VII.	Biases in Visual, Auditory, and Audiovisual Perception of Space .....	9
	A. Abstract .....	9
	B. Author Summary .....	10
	C. Introduction .....	11
	D. Methods .....	17
	E. Bayesian Causal Inference Model.....	21
	F. Results .....	25
	G. Discussion .....	34
	H. References for the project.....	41
VIII.	Exposure to discrepant audiovisual signals changes the tendency to integrate.....	49
	A. Abstract .....	49
	B. Introduction .....	50
	C. Methods & Model .....	54

D.	Results .....	63
i.	Experiment 1: Control.....	63
ii.	Experiment 2: Spatially Congruent & Temporally Uncorrelated Exposure Task Stimuli .....	64
iii.	Experiment 3: Spatially Congruent & Temporally Congruent Exposure- Task Stimuli .....	65
iv.	Experiment 4: Spatially Congruent & Temporally Uncorrelated Exposure- Task Stimuli .....	66
v.	Experiment 5: Spatially Incongruent & Temporally Congruent Exposure- Task Stimuli .....	67
vi.	Experiment 6: Spatially Incongruent & Temporally-Correlated Exposure- Task Stimuli .....	70
vii.	Summary of $P_{\text{Common}}$ Results .....	71
E.	Discussion .....	72
F.	References for the project.....	76
IX.	The Effects of Selective and Divided Attention on Sensory Integration .....	83
A.	Abstract .....	83
B.	Introduction .....	84
C.	Experiment 1: Single Task vs. Dual Task.....	87
i.	Methods.....	87
ii.	Model .....	90
iii.	Results .....	91
D.	Experiment 2: Single Response Only.....	93

i. Methods.....	94
ii. Model .....	94
iii. Results.....	94
E. Experiment 3: Beep-Flash Numerosity Judgment Task.....	97
i. Methods.....	97
ii. Model .....	99
iii. Results.....	100
iv. Discussion .....	101
v. References for the project .....	105
X. General Discussion and Conclusions .....	111
A. Summary .....	111
B. Future Directions.....	112

## LIST OF FIGURES

Figure 1. Possible underlying mechanisms for the biases observed in subjects' localization responses .....	16
Figure 2. The spatial localization paradigm .....	20
Figure 3. The Causal Inference Model .....	22
Figure 4. Biases present in localization on unisensory trials .....	27
Figure 5. Visual and auditory biases on spatially congruent bisensory trials .....	28
Figure 6. Bisensory trials, classified by inferring one common cause, or two independent Causes.....	30
Figure 7. Standard deviations for all five stimulus positions .....	32
Figure 8. The spatial localization paradigm and exposure task .....	57
Figure 9. The eight parameters fit to individual subjects' data .....	58
Figure 10. The Causal Inference Model .....	59
Figure 11. One representative subject's data and model fits from the pre-test and post-test localization phases.....	69
Figure 12. Design schematic depicting the conditions and blocks from experiment 1 .....	87
Figure 13. Design schematic for experiment 2.....	93
Figure 14. Design schematic for experiment 3.....	97

## LIST OF TABLES

Table 1. Glossary table of all potential parameters .....	24
Table 2. Average log likelihood and BIC values for tested models.....	33
Table 3. Hypotheses regarding the PCommon parameter.....	63
Table 4. Optimized parameter values from Experiment 1 .....	64
Table 5. Optimized parameter values from Experiment 2 .....	65
Table 6. Optimized parameter values from Experiment 3 .....	66
Table 7. Optimized parameter values from Experiment 4 .....	67
Table 8. Optimized parameter values from Experiment 5 .....	68
Table 9. Optimized parameter values from Experiment 6 .....	70
Table 10. Summary of experimental changes in the PCommon parameter .....	71
Table 11. Optimized parameter values from Attention Experiment 1 .....	92
Table 12. Optimized parameter values from Attention Experiment 2 .....	94
Table 13. Optimized parameter values from Attention Experiment 3 .....	100

## ACKNOWLEDGEMENTS

This research was supported by the following:

UCLA Distinguished University Fellowship

As the three main chapters of this manuscript are versions of the following articles, I would like to thank the co-authors for their contributions:

Odegaard, B., Wozny, D., Shams, L., (Submitted) Biases in Visual,  
Auditory, and Audiovisual Perception of Space

Odegaard, B., Wozny, D., Shams, L., (In Preparation) Exposure to Discrepant  
Audiovisual stimuli increases the tendency to bind

Odegaard, B., Wozny, D., Shams, L., (In Preparation) The effects of selective  
attention on sensory integration

I would also like to thank the following individuals for their vital support and feedback on this project series and throughout my six years at UCLA:

Graduate advisor, Prof. Ladan Shams; co-author Dr. David R. Wozny; dissertation committee, Prof. James Bisley, Prof. Mark Frye, Prof. Hongjing Lu, and Prof. Dario Ringach; Prof. Hakwan Lau, Prof. Miguel Eckstein, Prof. Ed Vul, Prof. Chris Stecker, Prof. Wei Ji Ma, Prof. Laurent Itti, Dr. Cristiano Cuppini, and labmate Majed Samad

Finally, I would like to extend a tremendous amount of thanks to many dedicated research assistants that have worked on my projects in the Visual and Multisensory Perception Lab. With their help, we have acquired data from nearly 700 subjects during my time as a graduate student:

Fred Foos, Danielle de Bruin, Tyler Toueg, Rahul Iyengar, Jason Carpenter,  
Siliang Wang, Sy Clark, Seong Hah Cho, Fabian Alvarez, Yoav Zimmerman, Michael Bani, Shannon Gordon, Eden Sayed, Andrew Marin, Jing Liu, Travis Steinberg, Rob Eriksen, and Yasamin Boulourian

## VITA

- 2009        B.A. Psychology  
              B.A. Music (Piano Performance)  
              Calvin College, Grand Rapids, Michigan
- 2009        UCLA Distinguished University Fellowship
- 2011        M.A., Psychology  
              University of California, Los Angeles, Los Angeles, California
- 2013        Accepted to the UCLA Teaching Practicum Program
- 2014        Teaching Fellow, UCLA GE Cluster Program

## PUBLICATIONS AND PRESENTATIONS

- Odegaard, B., Wozny, D., Shams, L., (Submitted Manuscript, 2015) Biases in Visual, Auditory, and Audiovisual Perception of Space
- Salomon, D., Martin-Harris, L., Mullen, B., **Odegaard, B.**, Zvinyatskovsky, Z., Chandler, S.H., (2015). Brain Literate: Making Neuroscience Accessible to a Wider Audience of Undergraduates. *Journal of Undergraduate Neuroscience Education* (JUNE).
- Odegaard, B.**, Wozny, D., Shams, L., (Poster) Exposure to discrepant auditory-visual stimuli increases the tendency to bind auditory-visual signals. *International Multisensory Research Forum*, Amsterdam, The Netherlands, June 2014.
- Odegaard, B.**, Wozny, D., Shams, L., (Poster) Visual, Auditory, and Auditory-Visual Spatial Biases. *International Multisensory Research Forum*, Amsterdam, The Netherlands, June 2014.
- Odegaard, B.**, Wozny, D., Shams, L., (Poster) Visual and auditory biases in spatial perception and their interactions. *Society for Neuroscience Meeting*, San Diego, California, November 2013.
- Odegaard, B.**, Wozny, D., Shams, L., (Poster) The effects of selective and divided attention on sensory integration. *Vision Sciences Society Conference*, Naples, Florida, May 2012.

## MENTOR FOR UNDERGRADUATE PROJECTS

Clark, S., **Odegaard, B.**, Peters, M.A.K., Shams, L., (Poster) The Computational Consequences of Noninvasive Electrical Stimulation on Humans' Audio-Visual Perception. *Stanford Undergraduate Psychology Conference*, Palo Alto, California, May 2014.

Carpenter, J., **Odegaard, B.**, Shams, L., (Poster) Frames of Reference in Sensory Integration: How eye position influences audiovisual computations. *UCLA Psychology Undergraduate Research Conference*, Los Angeles, California, May 2014.

## TEACHING AND STUDENT RATINGS

### Instructor

2014-2015	General Education Cluster Seminar Sex, Drugs, and Rock & Roll: How Romance, Psychoactive Substances, and Music Change the Brain	4.8/5
2013	Introductory Psychobiology	9.0/9.0

### Teaching Fellow

2014-2015	General Education Cluster Mind Over Matter: The History, Philosophy, and Neuroscience of the Brain	8.9/9.0
-----------	--	---------

### Teaching Assistant

2013	Fundamental of Behavioral Neuroscience	8.9/9.0
	Introduction to Research Methods	8.8/9.0
2012	Behavioral Neuroscience Laboratory Course	8.8/9.0
	Sensation & Perception	8.9/9.0
	Psychological Statistics	8.3/9.0
2011	Behavioral Neuroscience Laboratory Course	8.3/9.0
	Introduction to Research Methods	8.9/9.0
2010	Fundamentals of Behavioral Neuroscience	8.8/9.0



## VI.

### GENERAL INTRODUCTION

In order to survive in a complex and changing world, an organism's perceptual estimates about events occurring in a particular region of *space* and at a particular point in *time* must be sufficiently accurate to avoid harmful predators, find plentiful sources of food, and identify potential mates to ensure long-term success for the species. In humans, various sensory modalities have evolved to be particularly adept at representing these vital types of information. Vision, for example, has the most refined capacity to encode spatial information from the surrounding environment; audition provides precise temporal information about when events occurred; modalities such as olfaction and gustation provide less reliable information about the dimensions of space and time, but can be particularly useful when determining whether or not a food source is something you would want to ingest. Every sensory system appears to have adapted to be particularly skilled at providing at least one type of information critical for navigating and acting in a dangerous world.

Considering each modality's specific capacities, it might seem reasonable to assume that the neural systems in the brain comprising each sensory system operate in isolation from one another. Indeed, for decades in twentieth century neuroscience, studies focused on investigating single sensory modalities in simple experimental paradigms. In the last twenty-five years, however, there has been a renaissance in investigations probing how our senses influence and interact with one another in terms of processing and storing information (Chen & Vroomen, 2013; Ernst & Bühlhoff, 2004; Stein & Meredith, 1993; Stein & Stanford, 2008). When considering how spatial and temporal information is represented in the brain, particularly under noisy or ambiguous conditions, it has become clear that the auditory and visual senses

substantially interact with one another in an attempt to provide the most reliable estimate of the surrounding world.

For instance, when briefly presented audiovisual stimuli are presented synchronously with varying amounts of *spatial* discrepancy, it is now well established that the estimate of the auditory stimulus can be influenced by the location of the visual stimulus (Bertelson & Aschersleben, 1998; Bertelson & Radeau, 1981; Chen & Vroomen, 2013; Hairston et al., 2003; Slutsky & Recanzone, 2001). This illusion, commonly called the “Ventriloquist Illusion” (Pick, Warren, & Hay, 1969; Rock & Victor, 1964), has been shown to be modulated by the spatial and temporal relationships between the auditory stimuli (Wallace et al., 2004). Similarly, when briefly presented audiovisual stimuli are presented with varying amounts of *temporal* discrepancy, estimates about when, how many, or the rate at which visual stimuli occurred in the environment can be influenced by the presence of auditory stimuli (Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Scheier, Nijhawan, & Shimojo, 1999; Shams, Kamitani, & Shimojo, 2000). These findings indicate that the modality that with the higher degree of resolution for representing information in a given dimension exerts an influence over noisier estimates from other modalities, and opens several critical questions which need to be answered by future research: what are the computational principles underlying how our sensory modalities interact to represent spatial and temporal information from the surrounding world? Are these principles static, or can they be modified based on previous sensory experiences? And can these principles change based on how we selectively attend to features of stimuli in the environment?

## **A. Computational Models of Multisensory Integration: Towards a Complete Theoretical Account**

Numerous behavioral demonstrations of interactions between the auditory and visual senses have led to the development and creation of several theoretical and computational accounts of the principles governing how modalities exert influences over one another in space and time. One oft-cited example is Welch and Warren's description of the *modality appropriateness* or *modality precision* hypotheses (1980). According to their account, discrepancies between senses are resolved by relying on either the more appropriate or more precise sensory modality in a given dimension. This idea offered an intriguing hypothesis about the underlying principles governing sensory integration, but one question that remained is what the *goal* of integrating sensory information should be in the first place.

As noted by the famous Vision scientist David Marr, any sensory modality could potentially be thought of as an information processing system, and thus there are three levels of analysis that must be investigated to fully understand the system: the *computational* level specifies what problem needs to be solved, what the system accomplishes and why it does certain things; the *representational/algorithmic* level specifies the processes and manipulations of representations that must be implemented, and the *physical* level specifies how the computations and algorithms are physically realized in the existing system (Marr, 1982). In terms of the combination of sensory systems, according to Ernst and Bühlhoff, one possible computational goal is to minimize the overall variance of the combined bisensory estimate (2004). This process, known commonly as "Maximum Likelihood Estimation" (MLE), assumes that each sensory modality is estimated with a certain amount of precision, and that the combined estimate

weights each individual cue by the inverse of its variance (Cuppini, Magosso, Bolognini, Vallar, & Ursino, 2014). Previous behavioral studies have indeed shown that when sensory cues are presented with small amounts of discrepancy, these MLE computational models account for human observers' behaviors quite well (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002). However, these models failed to account for conditions with increased discrepancy between modalities in a given dimension, leaving a need for more advanced computational models to be developed to provide a more complete account of the statistical principles underlying multisensory processing.

In pursuit of this goal, Körding et al. (2007) developed a model that not only specified the equations for the optimal integration and segregation of sensory information, but also formalized the process of inference about whether sensory signals come from one or two sources. Utilizing Bayes' Rule, the posterior probability of a single cause could be inferred based on a combination of the sensory likelihoods from the visual and auditory modalities, multiplied by a prior probability for inferring a single cause of sensory information. This inferred posterior could then be used to weight the optimal estimates for inferring one or two causes, and this strategy of model averaging was shown to more effectively account observers' audiovisual localization behaviors than any other previous attempts, including Maximum Likelihood Estimation. Extensions of this framework have been developed to account for multisensory numerosity judgments (Wozny, Beierholm, & Shams, 2008), observers' capacities to localize stimuli of various intensities (Beierholm, Quartz, & Shams, 2009), different perceptual strategies used by observers when localizing multisensory stimuli (Wozny, Beierholm, & Shams, 2010), and well-known aftereffects produced by successively-presented audiovisual stimuli (Wozny & Shams, 2011). One recent study has even demonstrated that various computational elements in this

Bayesian Causal Inference model are encoded in distinct regions of the brain, providing evidence for the involvement of several brain areas implementing the hierarchical computations underlying sensory integration (Rohe & Noppeney, 2015).

Considering the increasing consensus about Bayesian Causal Inference underlying the computations governing multisensory integration in the brain, it becomes an open question how far this framework can extend its explanatory power. On the one hand, it is always good to keep a reasonable degree of skepticism regarding the full efficacy of any specific computational account of a given process. As noted by the statistician George E.P. Box, “essentially, all models are wrong, but some are useful” (Box & Draper, 1987, p.424), and currently, much is still unknown regarding how Bayesian Causal Inference may be implemented at the level of neural circuits (Pouget, Beck, Ma, & Latham, 2013). However, considering the ability of this Bayesian Causal Inference framework to account for human multisensory behaviors across a wide range of tasks and conditions, it becomes an intriguing tool to exploit in pursuit of providing new insights into elusive neuroscientific phenomena. Thus, further exploration using this modeling framework is warranted to provide additional computational insights about how the brain constructs a coherent multisensory representation of the surrounding world.

## **B. Dissertation Overview**

This dissertation is divided into three core chapters, which explore factors influencing likelihoods and priors in the aforementioned Bayesian Causal Inference framework. The first chapter is a behavioral and computational exploration of the nature of human spatial perception in the visual and auditory modalities. It begins with a review of relevant behavioral literature

from the past twenty years, which has demonstrated how humans' perception of space in the visual and auditory modalities is often biased and inaccurate. Interestingly, many studies show that the two modalities are biased in opposite directions: vision is biased towards the center of space, while audition is biased towards the periphery (Adam, Davelaar, Gouw, & Willems, 2008; Cui, Razavi, O'Neill, & Paige, 2010; Fortenbaugh & Robertson, 2011; Lewald & Ehrenstein, 1998; Müsseler, Heijden, Mahmud, Deubel, & Ertsey, 1999; Parise, Spence, & Ernst, 2012). The direction and extent of these spatial biases, however, can depend on the specific paradigm used (i.e. whether the response mechanism was visible at all times, whether the eyes were free to move, whether accuracy or speed was emphasized, etc.). To provide a more definitive account of these potential biases, the chapter proceeds to describe a spatial localization task that was implemented to assess the accuracy of visual, auditory, and audiovisual localizations along azimuth in a dataset of unprecedented size (384 subjects). Subjects' accuracy and precision in the visual and auditory modalities were quantified, and the reported results again provide evidence that spatial perception is, on average, biased away from the true location of the stimulus in both the visual and auditory modalities.

As mentioned previously, it has been shown that human observers' unisensory and multisensory localizations along azimuth are well accounted for by a Bayesian Causal Inference model. However, earlier versions of this model do not attempt to account for the aforementioned spatial biases, and in a Bayesian framework, these inaccuracies could potentially be due to either biases in the sensory representations (i.e. likelihood distributions) or *a priori* biases over space (i.e. priors). Thus, to conclude this chapter, six different models are applied to our large dataset to determine whether spatial biases are best accounted for by changes in likelihoods, changes in priors, or a combination of these two potential mechanisms.

Following this methodological innovation, the second chapter explores the underlying spatial and temporal principles that govern sensory recalibration. Previous research has shown that when visual and auditory signals are repeatedly presented at the same point in time but at different locations in space, subsequent auditory judgments can be pulled in the direction of where the visual stimulus occurred (Bertelson, Frissen, Vroomen, & de Gelder, 2006; Frissen, Vroomen, de Gelder, & Bertelson, 2003, 2005; Kopco, Lin, Shinn-Cunningham, & Groh, 2009; Lewald, 2002; Recanzone, 1998; Slutsky & Recanzone, 2001). Research implementing Bayesian models has demonstrated that this effect can be accounted for by a shift in the auditory likelihoods (Sato, Toyozumi, & Aihara, 2007; Wozny & Shams, 2011). Here in our study, we modify the standard paradigm to adapt observers to visual-auditory spatial discrepancies in two different directions during an exposure task, and note a surprising change: this effect is not accounted for by shifting auditory representations of space, but instead in a *change in the prior for binding sensory information*. By conducting further experiments testing varying degrees of spatial and temporal discrepancies during the exposure task, we discover that the brain updates this prior for binding when sensory information is (1) spatially discrepant, and (2) related (either congruent or correlated) in time. We discuss the implications of this finding in the context of a predictive coding framework (Friston, 2010; Klemen & Chambers, 2012; Talsma, 2015).

The final chapter again utilizes this updated Bayesian Causal Inference Model to explore how conditions of selective and divided attention influence its various computational components (i.e. parameters). In this context, “selective attention” refers to conditions requiring the allocation of endogenous (volitional) attention to *one* sensory modality (either vision or audition) in a multisensory environment, and “divided attention” refers to allocating attention in a bisensory manner, across the visual and auditory modalities. Experiments probed subjects’

abilities to make spatial localization judgments and numerosity judgments while allocating attention in these two manners, and investigated whether attention influences the prior probability of inferring a single source of sensory information, the sensory representations (i.e. likelihoods), or any other *a priori* biases over space or time. Results suggest that the “rich get richer,” as sensory representations that have a certain amount of reliability are *enhanced* by selective attention, but extremely noisy sensory representations fail to see any benefit with selective processing. Interestingly, the prior capturing the tendency to integrate is not altered by selective attention, providing evidence that integration processes operate independently from attentional processes in the brain. Theoretical implications of these findings, and how they build upon previous models of attention, are discussed in the chapter’s conclusion.



## VII.

### **Biases in Visual, Auditory, and Audiovisual Perception of Space**

#### **Abstract**

Localization of objects and events in the environment is critical for survival, as many perceptual and motor tasks rely on estimation of spatial location. Therefore, it seems reasonable to assume that spatial localizations should generally be accurate. Curiously, some previous studies have reported biases in visual and auditory localizations, but these studies have used small sample sizes and the results have been mixed. Therefore, it is not clear (1) if the reported biases in localization responses are real (or due to outliers, sampling bias, or other factors), and (2) whether these putative biases reflect a bias in sensory representations of space or *a priori* expectations (which may be due to the experimental setup, instructions, or distribution of stimuli). Here, to address these questions, a dataset of unprecedented size (obtained from 384 observers) was analyzed to examine presence, direction, and magnitude of sensory biases, and quantitative computational modeling was used to probe the underlying mechanism(s) driving these effects. Data revealed that, on average, observers were biased towards the center when localizing visual stimuli, and biased towards the periphery when localizing auditory stimuli. Moreover, quantitative analysis using a Bayesian Causal Inference framework suggests that while pre-existing spatial biases for central locations exert some influence, biases in the sensory representations of both visual and auditory space are necessary to fully explain the behavioral data. How are these opposing visual and auditory biases reconciled in conditions in which both auditory and visual stimuli are produced by a single event? Potentially, the bias in one modality could dominate, or the biases could interact/cancel out. The data revealed that when integration occurred in these conditions, the visual bias dominated, but the *magnitude* of this bias was

*reduced* compared to unisensory conditions. Therefore, multisensory integration not only improves the *precision* of perceptual estimates, but also the *accuracy*.

### **Author Summary**

Almost all daily tasks performed by humans require localizing objects and events. Since spatial localization is critical for survival, it is expected that the brain performs this task accurately. We tested the accuracy of localizing simple sounds and sights in hundreds of human observers. We found that, on average, localizations of sounds and sights were *not* accurate, and observers made systematic errors: flashes of light were perceived to be *nearer* the center of the visual field than they actually were, and bursts of noise were perceived farther *away* from center. Surprisingly, computational analyses revealed that these biases in perceived location of sounds and sights are at least partly due to a bias in how sights and sounds are encoded by the sensory systems; the visual representation of the world is compressed towards the center and the auditory representation is expanded away from the center. When flashes and noise bursts were presented at the same location simultaneously and the observers perceived a common source for both stimuli, the bias in the perceived location was smaller, showing that synthesizing information across modalities helps increase accuracy in the perceived location of objects and events, compared to when only one sensation is available.

## **Introduction**

Nearly every function critical for human survival depends directly on accurate localization of objects and events in the environment. For example, our capacity to locate food, find potential mates, avoid predators, navigate new terrain, avoid obstacles, act upon objects, and orient towards sudden sounds requires a skilled aptitude to detect stimuli in the surrounding world. Spatial localization is thus a fundamental task that the brain has to solve at any given moment, and evolution has had millions of years to refine this function. Therefore, one would expect that the representation of space in all sensory modalities is generally accurate. However, the existing literature on human spatial localization contains several findings that are at odds with this assumption: not only do representations of space in different sensory modalities appear to be biased, but different sensory modalities exhibit different biases in how they encode spatial information.

For instance, many studies investigating localizations of simple visual stimuli along azimuth have shown that visual localizations often show a bias towards the center of visual space (Adam, Davelaar, Gouw, & Willems, 2008; Fortenbaugh & Robertson, 2011; Mateeff & Gourevich, 1983; Müsseler, Heijden, Mahmud, Deubel, & Ertsey, 1999; Osaka, 1977; Parise, Spence, & Ernst, 2012), but one recent investigation showed that either a central or peripheral bias may emerge, depending on the eccentricity of the target (Dobрева, O'Neill, & Paige, 2012). In the auditory domain, findings of various studies have been even more mixed: some indicate that, on average, auditory localizations are often biased increasingly farther away from center as eccentricity increases (Cui, O'Neill, & Paige, 2010; Lewald & Ehrenstein, 1998; Parise et al., 2012; Razavi, O'Neill, & Paige, 2007), while others have reported central biases for average auditory localizations (Dobрева et al., 2012; Zahn, Abel, & Dell'Osso, 1978). Thus, research has

often shown central biases in the visual modality, but previous results on auditory spatial biases are mixed.

The previous investigations on spatial localization have used relatively small sample sizes, and consequently, variability in the reported findings could be due to sampling bias and/or the undue influence of outliers. Similar issues also apply to the previous reports of *precision* of spatial localization. Several earlier studies have reported that visual precision decreases with increasing eccentricity of transient visual targets (Dobrevá et al., 2012; Hairston et al., 2003). In the auditory domain, findings have been mixed. One study reported that the standard deviation of auditory localizations remains relatively consistent across peripheral regions in space (Hairston et al., 2003), but two recent investigations found that auditory precision declines for peripheral targets (Charbonneau, Véronneau, Boudrias-Fournier, Lepore, & Collignon, 2013; Dobrevá et al., 2012).

Here, we present findings from an extraordinarily large dataset (384 subjects) to obtain more definitive answers regarding not only the direction and magnitude of biases in both the visual and auditory modalities, but also the precision in spatial localizations over a range of spatial positions. By analyzing responses from our immense dataset, we aim to clarify discrepant results in the literature regarding the accuracy and precision of both visual and auditory localizations.

Additionally, while previous studies on unisensory biases and variances have enhanced an overall understanding of how the senses process information from the surrounding world, in real-world settings, objects usually produce signals that simultaneously stimulate multiple sensory modalities. In conditions where auditory and visual stimuli co-occur in close spatial proximity, it has been well-documented that auditory localizations are often pulled in the

direction of a visual stimulus (Bertelson & Aschersleben, 1998; Bertelson & Radeau, 1981; Hairston et al., 2003; Lewald & Guski, 2003; Slutsky & Recanzone, 2001; Wallace et al., 2004). Known as the “Ventriloquist illusion” (Pick, Warren, & Hay, 1969; Rock & Victor, 1964), this effect has been demonstrated in many experiments, lending support to the idea that the most reliable sense in a given dimension exerts a strong influence over other noisier senses as estimates about the world are produced (Welch & Warren, 1980).

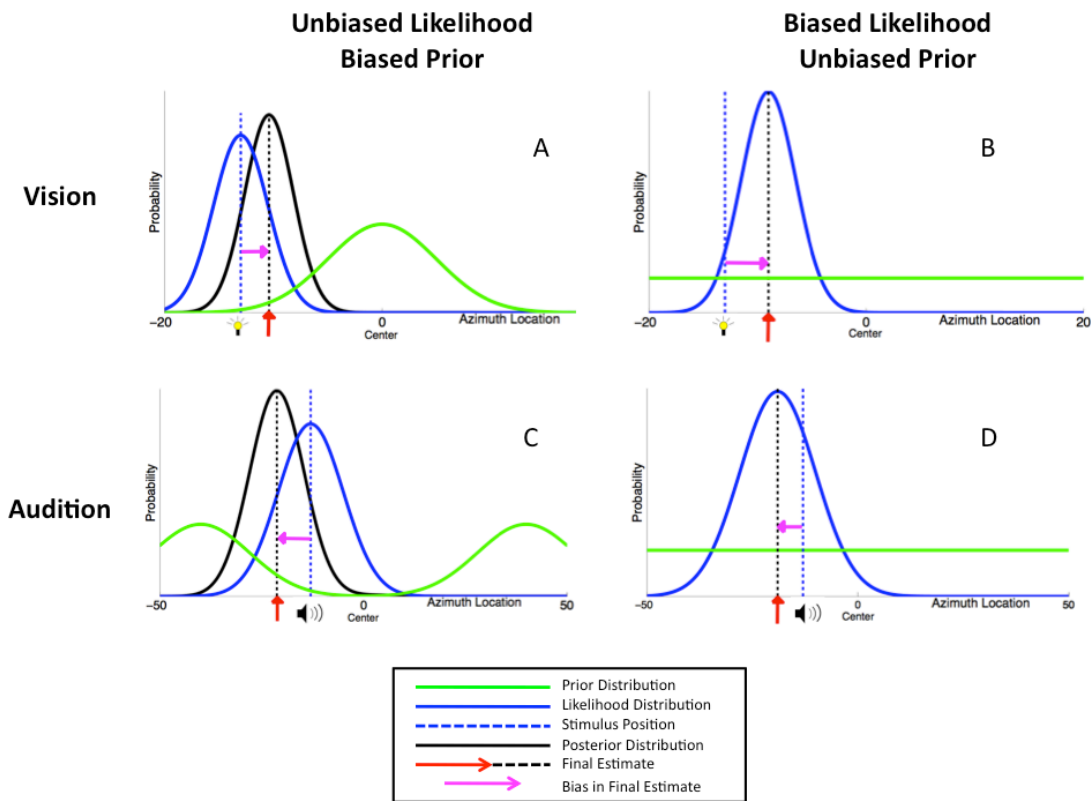
While many studies have focused on how far this “spatial window” of audiovisual integration extends (Chen & Vroomen, 2013), one important question remains: when visual and auditory stimuli *co-occur* at the *same* spatial location (which is generally the case when an event produces both visual and auditory stimulation), which modality’s bias emerges during localization? Is it the auditory bias, the visual bias, or something else? If indeed visual localization is biased towards the center, and auditory localization is biased towards the periphery, three possibilities exist: vision dominates on most trials, and a central bias emerges; the auditory bias dominates, and a peripheral bias emerges; the biases interact, and a hybrid bias emerges in the bimodal localizations. It has been shown previously that even when the stimuli are congruent in both space and time, they may be perceived to have a common cause or independent causes. Therefore, it is also important to explore whether the nature and presence of bias in multisensory conditions is influenced by the inference of common cause.

Finally, another important question that has not been previously explored is the following: if indeed the visual and/or auditory localization of stimuli in space is biased and inaccurate, what underlies this bias? Is this bias in localization a result of bias in the sensory representation of the space, a result of general a priori expectation of location of events in the world, or the result of a combination of these two possibilities? In the former case, the sensory

encoding of the stimuli is compressed towards the center of the visual space, and this early sensory bias leads to a mislocalization of stimuli towards the center without any effect of prior expectations for the location of the stimuli. In the latter case, the sensory representation of the visual space may be perfectly accurate and unbiased, but due to the task demands or expectations about the location of the stimuli (in this task or in general), the localization may be biased towards the center. The third possibility is that the localization bias stems from a combination of both of these mechanisms. Teasing apart these three different scenarios requires quantitative computational modeling of the data, as explained below.

The behavioral findings in this study showed that both visual and auditory localizations were indeed biased, with visual localizations on average biased towards the center and auditory localizations on average biased away from the center. To address whether these biases in spatial perception stem from a bias in sensory representations or from a prior expectation of location, we employed the causal inference model of multisensory perception (Körding et al., 2007; Wozny, Beierholm, & Shams, 2010; Wozny & Shams, 2011a) to quantitatively characterize both sensory representations and prior expectations of each individual observer. This model has been very successful in accounting for human observers' data in a variety of multisensory tasks (Beierholm, Quartz, & Shams, 2009; Körding et al., 2007; Samad, Chung, & Shams, 2015; Wozny, Beierholm, & Shams, 2008; Wozny et al., 2010; Wozny & Shams, 2011a) and a recent brain imaging study has provided further support for the brain utilizing this computation in the spatial localization task that is used in this study (Kayser & Shams, 2015; Rohe & Noppeney, 2015). Importantly, this model can reliably provide a quantitative estimate of several components of perceptual processing for each individual observer.

In a Bayesian framework, the final perceptual estimate is based upon a combination of the sensory representation (i.e. likelihood distribution) and pre-existing expectation (i.e. prior). Fig. 1 shows different kinds of underlying mechanisms that could produce biases similar to those shown by most subjects in our localization task. In the visual domain, either (1) an *a priori* bias for center in expectation of location of visual stimuli (Fig. 1a), or (2) a bias in the sensory representation of visual stimuli towards the center (the likelihood means shifted towards center with the amount of shift being proportional to the degree of stimulus eccentricity) could account for biased perception (Fig. 1b). In the auditory domain, either (1) an *a priori* bias towards periphery (Fig. 1c), or (2) a bias away from the center in sensory representations (Fig. 1d; shifts of likelihood means away from the center with the degree of shift proportional to stimulus eccentricity) could account for participants' behavior. The exhibited behavioral biases may also be due to a combination of biased likelihood and biased priors. To investigate which of these options is indeed the mechanism underlying the observed biases, we implemented all three types of mechanisms (biased likelihoods, bias priors, a combination of the two) into a Bayesian Causal Inference model and performed quantitative model comparisons to determine which computational mechanism best accounts for the behavioral data.



**Fig. 1.** Possible underlying mechanisms for the biases observed in subjects' localization responses. Zero on the x-axis represents the center. In panels (B) and (D), the likelihood and posterior functions are overlapping and are shown in blue. In the visual domain, (A) a prior distribution located at the center of visual space could draw localization estimates towards the center, or (B) the likelihood distributions themselves could be biased, drawing sensory estimates towards a central location. In the auditory domain, (C) a prior bias for periphery could push perception of peripheral target estimates further away from the center of space, or (D) the likelihood distributions themselves maybe biased away from the center. Various combinations of these computational mechanisms were tested in the different proposed models (see Model Comparisons section).

All previous models of multisensory spatial localization assume that sensory representations are unbiased, and that the main benefit in integrating stimuli is improvement in the overall *precision* of the combined estimate. For example, maximum likelihood estimation models have proposed that auditory and visual signals can be represented by distributions centered at the true location of each stimulus, with the combined audiovisual estimate exhibiting a smaller variance than the unisensory estimates (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002). The reduced variance of the combined estimate has been



considered to be the main reason why it may be advantageous for the brain to integrate redundant sensory information in a complex world (Ernst & Bühlhoff, 2004; Ghahramani, 1995). Here, we explore the effect of integration on the *bias* of the estimates and examine whether localization estimates become less biased, more biased or remain the same as a result of integration as compared to unisensory conditions.

In order to address all of these questions, we analyzed psychophysical data and quantitatively characterized the perceptual components involved in the spatial perceptual process using computational modeling. Specifically, the study had five aims: (1) to quantify unisensory spatial biases and variances in the visual and auditory modalities using a large dataset, (2) to investigate how biases are reconciled when subjects localize spatially congruent bisensory stimuli, (3) to determine whether the biases that emerge in spatially congruent bisensory trials depend on an observer's inference about a common cause, and (4) to determine whether the biases in spatial localization are due to a bias in the sensory representations, prior expectations, or both, and (5) to examine how the biases in bisensory conditions compare with biases in unisensory conditions.

## **Methods**

### *Participants*

A total of 412 subjects (ages 18-55) participated in our experiment; since our measurement of localization biases are means that can be influenced by extreme outliers, 28 subjects were excluded because their responses for each of the five unisensory conditions in either the visual or auditory modalities were more than three times the inter-quartile standard deviation for the tested location. This exclusion criterion ensured that we would avoid analyzing data that may have been due to sloppiness or negligence with the response device. 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. Each subject also signed a 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 52cm from a projection screen. The screen was a black, acoustically transparent cloth subtending much of the visual field ( $134^\circ$  width  $\times$   $60^\circ$  height). Behind the screen were 5 free-field speakers (5 x 8 cm, extended range paper cone), positioned along azimuth  $6.5^\circ$  apart,  $7^\circ$  below fixation. The middle speaker was positioned below the fixation point, and two speakers were positioned to the right and two to the left of fixation. The visual stimuli were presented overhead from a ceiling mounted projector set to a resolution of 1280 x 1024 pixels with a refresh rate of 75 Hz.

Prior to the presentation of any stimuli in the experiment, participants were required to have their gaze centered on the central fixation point. To ensure that participants' gaze for each trial was starting from the same location, gaze position and fixation time were recorded at 60Hz with a ViewPoint eye tracker (Arrington Research, Scottsdale, AZ) fixed to the chinrest and PC-60 software (version 2.8.5,000). Stimuli were not displayed until the recorded gaze angle was within  $3.0^\circ$  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 presentation of stimuli and recorded behavioral responses using

MATLAB (version 7.6.0, R2008a). A wireless mouse was used to record the behavioral responses.

The visual stimuli used in the experiments were white-noise disks (.41 cd/m<sup>2</sup>) with a Gaussian envelope of 1.5° FWHM, presented 7° below the fixation point on a black background (.07cd/m<sup>2</sup>), 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 of the participants in the experiment.

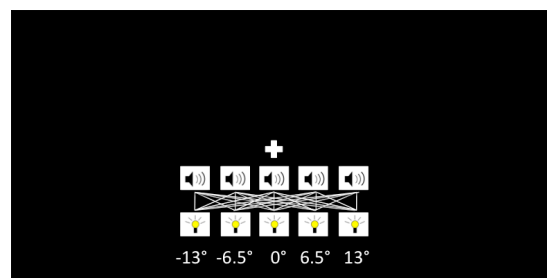
### *Procedure*

While several previous studies have reported a bias towards center in visual localization (Adam et al., 2008; Fortenbaugh & Robertson, 2011; Hubbard & Ruppel, 2000; Kerzel, 2002; Sheth & Shimojo, 2001), a few studies have reported peripheral biases, or in one study, veridical perception (O. Bock, 1986; Otmar Bock, 1993; Bruno & Morrone, 2007; Henriques, Klier, Smith, Lowy, & Crawford, 1998). Speculations about the source of the variability in findings regarding the presence and direction of bias has focused on the specifics of the experimental paradigms, including the role of visibility of response location (Uddin, 2006), the visibility of external boundaries (Fortenbaugh, Sanghvi, Silver, & Robertson, 2012), the memory demands of the task as related to the time before response (Sheth & Shimojo, 2001) and other factors. Given that peripheral localization error in paradigms in which the response mechanism is not visible could be due to motor error, and that this study is not interested in motor error, we implemented a paradigm where the response cursor was visible at all times.

Participants began each session with 10 practice trials requiring localization of unisensory auditory stimuli. This practice session ensured that participants were using the mouse

properly, understood the instructions, and were 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 vertically just above the horizontal line where the stimuli were presented 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. 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, so the fixation requirement was dropped following the presentation of the stimuli.

Following the brief practice session, participants began the localization session, which consisted of 525 trials of interleaved auditory, visual, and audiovisual stimuli presented in pseudorandom order. The stimulus conditions included 5 unisensory auditory locations, 5 unisensory visual locations, and 25 combinations of auditory and visual locations (bisensory conditions), for a total of 35 stimulus conditions, shown in Fig. 2 below. Fifteen trials of each of the 35 conditions were presented in pseudorandom order, lasting about 45 minutes, including breaks.

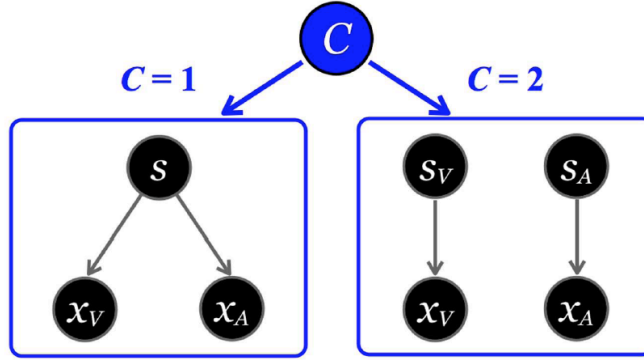


**Fig. 2.** The spatial localization paradigm. Stimuli could be presented from one of five locations, ranging from -13 to +13 degrees. White lines indicate all 25 spatial combinations of bimodal stimuli.

### *Bayesian Causal Inference Model*

The Bayesian causal inference model has been shown in multiple studies to account for the human multisensory spatial perception very well (Beierholm et al., 2009; Körding et al., 2007; Wozny et al., 2010; Wozny & Shams, 2011a) and it has been shown to outperform other proposed models of multisensory spatial perception (Körding et al., 2007). Moreover, a human neuroimaging study recently provided further support for the Bayesian Causal Inference being carried out in human brain in the spatial localization task, using the same task used in current study (Rohe & Noppeney, 2015). Therefore, we used this model to investigate the characteristics and perceptual components of human spatial processing in our localization task. In previous formulations of this model it had been assumed that (1) the likelihood distributions of sensory stimuli are unbiased (centered at the true locations of stimuli in the world), (2) visual and auditory localizations are both impacted by a meta-modal central prior over space, and (3) the variance of sensory representations (likelihood functions) is uniform across azimuth. Considering our behavioral results, which reveal non-zero biases of differing direction and magnitude in the visual and auditory modalities, it appears that these simplifying assumptions are not quite warranted, and the model would need to be enhanced to allow flexibility in the representation of likelihoods and/or priors.

The nervous system does not have access to the true events  $s$  in the world. Instead, it only has access to the noisy sensory representations,  $x_v$  and  $x_a$  (denoting visual and auditory sensations, respectively). Taking into account both the noisy sensory representations (likelihoods) and prior knowledge of the world (priors), the brain makes an inference about whether the sensory signals come from the same source ( $C=1$ ) and should be integrated or the signals come from different sources ( $C=2$ ) and should be segregated (Fig. 3).



**Fig. 3.** The Causal Inference Model (figure from Wozny & Shams, 2011a). One cause,  $s$ , could produce both the auditory and visual signals, or two distinct causes,  $s_V$  and  $s_A$ , could produce the auditory and visual signals. Subjects only have access to the noisy signals  $x_V$  and  $x_A$ , and thus must infer the causal structure,  $C$ , that produced the sensations on a given trial.

Thus, the posterior probability of an event  $s$  is conditioned on the causal structure of the stimuli ( $C=1$  or  $C=2$ ):

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

$$p(s_A | x_A; C = 2) = \frac{p(x_A | s_A)p(s_A)}{p(x_A)} \quad (2)$$

$$p(s_V | x_V; C = 2) = \frac{p(x_V | s_V)p(s_V)}{p(x_V)} \quad (3)$$

The estimate of the location of the stimuli will be based on these posterior probabilities. However, since the causal scenario is not known by the nervous system, this also must be inferred based on the available sensory evidence and prior information. This can also be computed using Bayes' Rule:

$$p(C | x_A, x_V) = \frac{p(x_A, x_V | C)p(C)}{p(x_A, x_V)} \quad (4)$$

Therefore,

$$p(C = 1 | x_A, x_V) = \frac{p(x_A, x_V | C = 1)p_c}{p(x_A, x_V | C = 1)p_c + p(x_A, x_V | C = 2)(1 - p_c)} \quad (5)$$

where  $P_c$  is the prior probability of a common cause, and the likelihood terms can be computed by integrating over the latent variable  $s$ :

$$p(x_A, x_V | C = 1) = \int p(x_A | s)p(x_V | s)p(s)ds \quad (6)$$

$$p(x_A, x_V | C = 2) = \int p(x_A | s_A)p(s_A)ds_A \cdot \int p(x_V | s_V)p(s_V)ds_V \quad (7)$$

The posterior probability of independent causes is computed as follows:

$$p(C = 2 | x_A, x_V) = 1 - p(C = 1 | x_A, x_V) \quad (8)$$

Having calculated the probabilities of each causal structure, and having calculated the optimal estimates for spatial localization under each causal structure, we now need to obtain estimates given *uncertainty* about the causal structure. Previously it has been shown (Wozny et al., 2010) that the vast majority of observers use a probability matching strategy as follows:

$$\hat{s}_A = \begin{cases} \hat{s}_{A,C=1} \text{ if } p(C = 1 | x_A, x_V) > \xi \\ \text{where } \xi \in [0 : 1] \text{ uniform distribution} \\ \hat{s}_{A,C=2} \text{ if } p(C = 1 | x_A, x_V) \leq \xi \\ \text{and sampled on each trial} \end{cases} \quad (9)$$

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

Therefore, here we used this strategy to model the observers' data. We used a generative model to simulate 10,000  $x_A$ 's and  $x_V$ 's for each experiment condition (35 total) using the preceding equations. Trial-to-trial variability is introduced by sampling the likelihoods from a normal distribution centered at the true sensory location, plus a bias term that scales linearly with eccentricity of the stimulus. Thus, if subjects' sensory representations are not centered at the true

locations, this bias term could potentially reflect this systematic shift in the likelihoods. The variability in the normal distribution from which the likelihoods are sampled scales with eccentricity of the stimulus. The prior described in the above equations changes depending on the model version, and can either be the biased central prior, the biased peripheral prior, or both, (see Fig. 1).

We quantitatively compared six models to determine which model could best account for participants' data. Different mechanisms were incorporated in the different models, resulting in different combinations of free parameters. All of these parameters are shown in Table 1, with a brief description of their corresponding mechanism.

Parameter	Description
$P_c$	The prior probability of integrating visual and auditory signals; the tendency to bind
$\sigma_V$	Standard deviation of the visual likelihood; the noise in the visual sensory representation
$\sigma_A$	Standard deviation of the auditory likelihood; noise in the auditory sensory representation
$\Delta x_V$	The shift in the center of the visual likelihood; the bias in the visual representation
$\Delta x_A$	The shift in the center of the auditory likelihood; the bias in the auditory representation
$\Delta \sigma_V$	The increase in visual likelihood std dev with each increase in eccentric position
$\sigma_P$	Standard deviation of the metamodal spatial prior
$x_P$	The mean of the metamodal spatial prior
$\sigma_{VP}$	Standard deviation of the visual spatial prior
$x_{VP}$	The mean of the visual spatial prior
$\sigma_{AP}$	Standard deviation of the auditory spatial prior
$x_{AP}$	The mean(s) of the auditory spatial prior

**Table 1.** Glossary table of all potential parameters.

Three parameters were included in every model: the prior probability of a common cause ( $P_c$ ), the standard deviation of the visual likelihood ( $\sigma_V$ ), and the standard deviation of the auditory likelihood ( $\sigma_A$ ). Two models incorporated the biased likelihood mechanisms, including a symmetric shift in the visual likelihood mean that increases with eccentricity ( $\Delta x_V$ ), a symmetric shift in the auditory likelihood mean that increases with eccentricity ( $\Delta x_A$ ); one of these models included a change in the visual likelihood variance which increases with



eccentricity ( $\Delta\sigma_V$ ), while the other did not. Two models incorporated the biased prior mechanisms, including a visual central prior to capture a central bias ( $\sigma_{VP}, x_{VP}$ ) and a peripheral spatial prior to capture a peripheral auditory bias (with the mean and variance,  $\sigma_{AP}, x_{AP}$ , reflecting the symmetric aspects of each part of the bimodal distribution); one of these included an additional metamodal spatial prior, while the other did not. We also included the original model (Körding et al., 2007), which does not include either the biased likelihood nor the biased prior mechanisms, as the baseline for comparison. Finally, we included a hybrid model that incorporated all biased likelihood mechanisms ( $\Delta x_V, \Delta x_A, \Delta\sigma_V$ ) as well as a metamodal spatial prior. These models are summarized in the results section. Since the different models have different numbers of free parameters, in addition to reporting their model fits (in log likelihood) we also report the BIC value of each model, which is a measure of goodness of fit that penalizes models with larger number of free parameters. The six models described above were applied to the data from each of the 384 observers, and log likelihood and BIC values were computed for each model's fit to each observer.

## **Results**

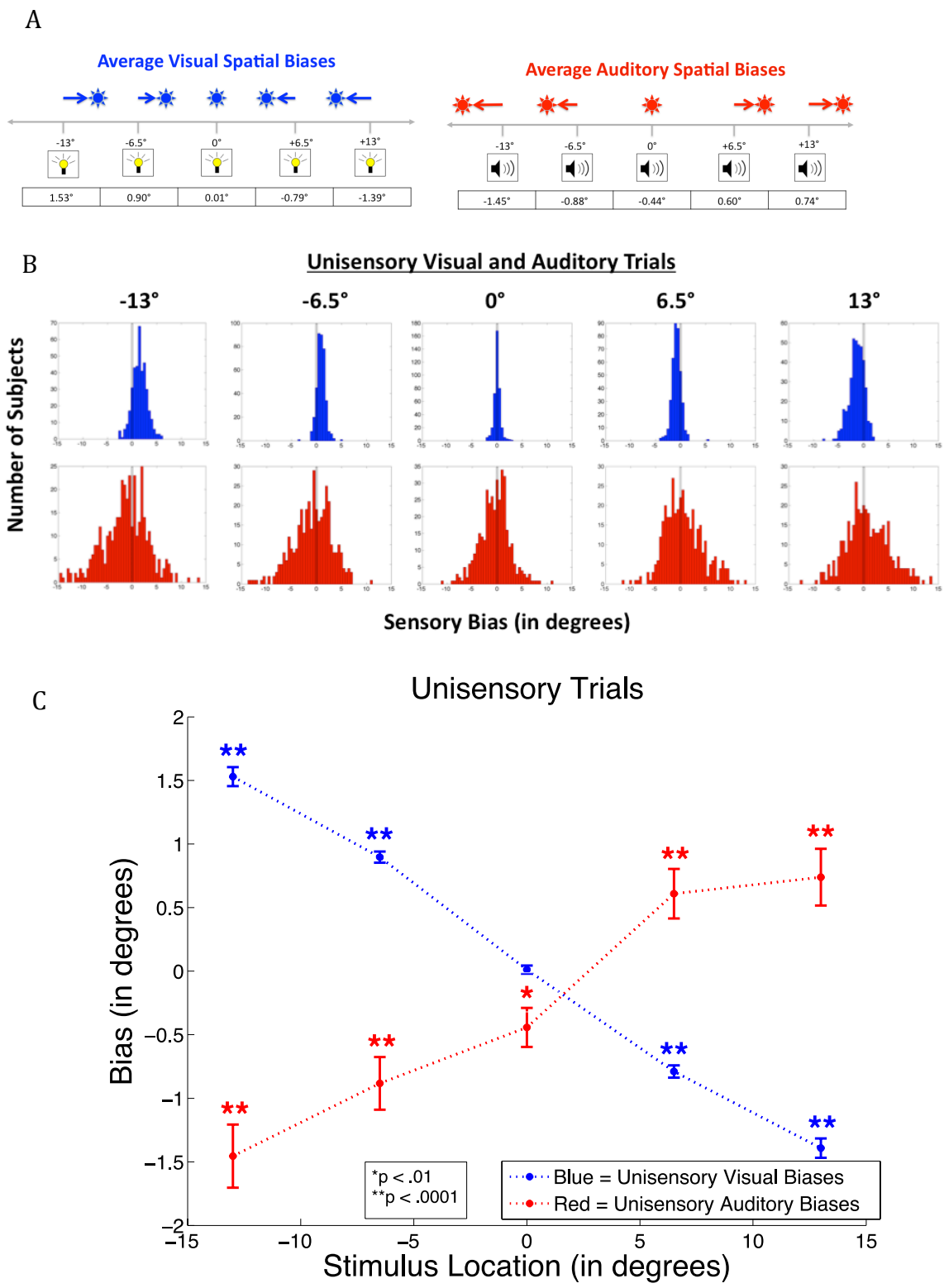
The trials in which only a visual stimulus or only an auditory stimulus was presented were analyzed for characterizing localization in “unisensory” conditions. The trials in which both visual and auditory stimuli were presented and occurred in the same location (congruent audiovisual trials) were used to characterize the localization behavior in the “bisensory” condition.

### *Accuracy of localization in unisensory conditions*

For each observer, the average error in localization (across 15 trials) in the unisensory trials was calculated for each spatial position and for each modality. The mean and distribution

of these biases across observers are shown in Fig. 4. Participants' average localizations for unisensory visual stimuli exhibited a bias for localizing peripheral stimuli closer to the center of visual space. Analysis of unisensory-visual trials showed that 63% of subjects (242) exhibited central biases for all four peripheral locations of visual stimuli, and 84% of subjects (319) exhibited central biases for three out of the four peripheral locations. This trend for a central bias tended to increase as the eccentricity of the visual stimulus increased. In the auditory domain, subjects' average localizations for each location revealed the opposite trend: average localizations for eccentric stimuli exhibited a peripheral bias for localizing the stimuli (Fig. 4a,c). However, further analysis revealed heterogeneity among subjects for this trend: 25% of the subjects exhibited consistent peripheral biases for all four eccentric locations, but 15% of participants exhibited consistent central biases. Overall, though, 45% of the participants exhibited a peripheral bias for three out of the four locations, indicating that a substantial proportion of subjects localized eccentric auditory stimuli further away from the true locations of the sounds.

Bonferonni-corrected one-sample t-tests for all peripheral localizations in both modalities were all highly significant ( $p < .0001$ ), indicating the responses in both the visual and auditory modalities for peripheral locations were biased away from the true location of the stimulus. For the central ( $0^\circ$ ) localizations, there was no bias in the visual modality ( $t(383) = 1.5, p > .05$ ), but a small bias to the left for the auditory modality was observed ( $t(383) = -2.9, p = .0042$ ). Previous research has indicated that participants often display an auditory localization bias contralateral to the preferred hand (Ocklenburg, Hirnstein, Hausmann, & Lewald, 2010), so this slight bias could be explained by the near certain-assumption that the majority of our participants were right-handed.

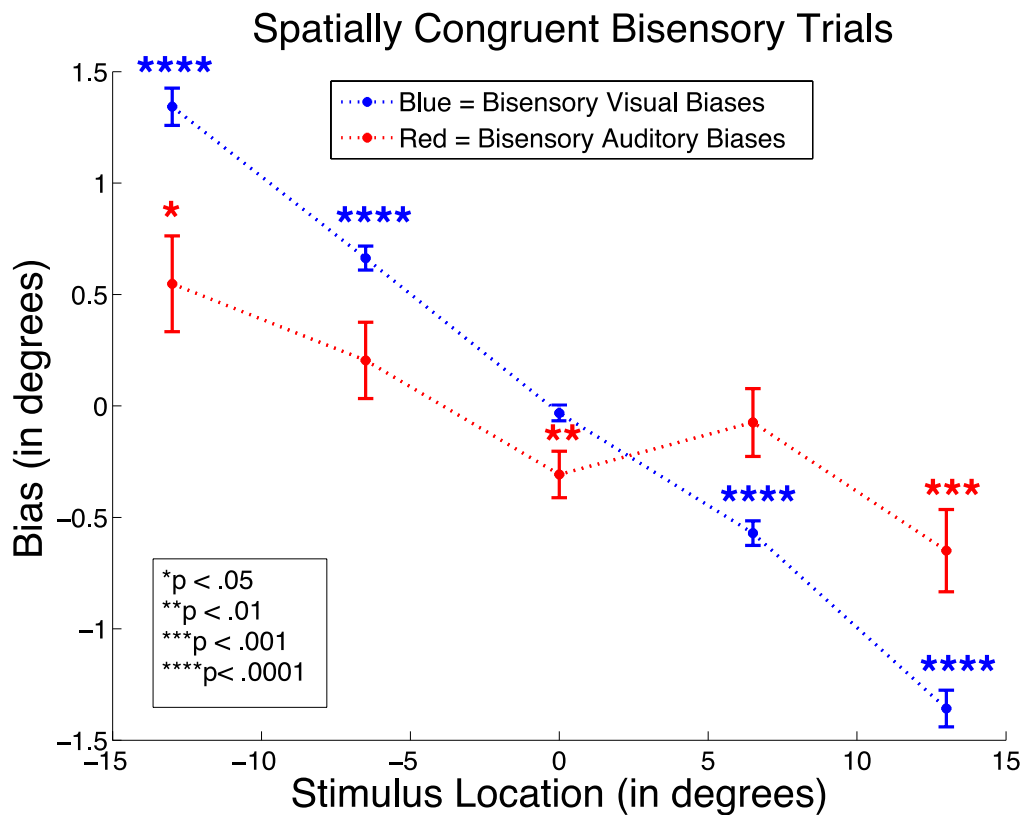


**Fig. 4.** Biases present in localizations on unisensory trials. A) Average biases across subjects. B) The distribution of subjects' individual biases (based on each subjects' mean across 15 trials) for each location. Visual biases are shown in blue, and auditory biases are shown in red. C) Positive numbers indicate rightward biases, and negative

numbers indicate leftward biases. SEM bars (computed over all subjects) are shown around the mean bias in each modality for each location.

### *Accuracy of localization in bisensory conditions*

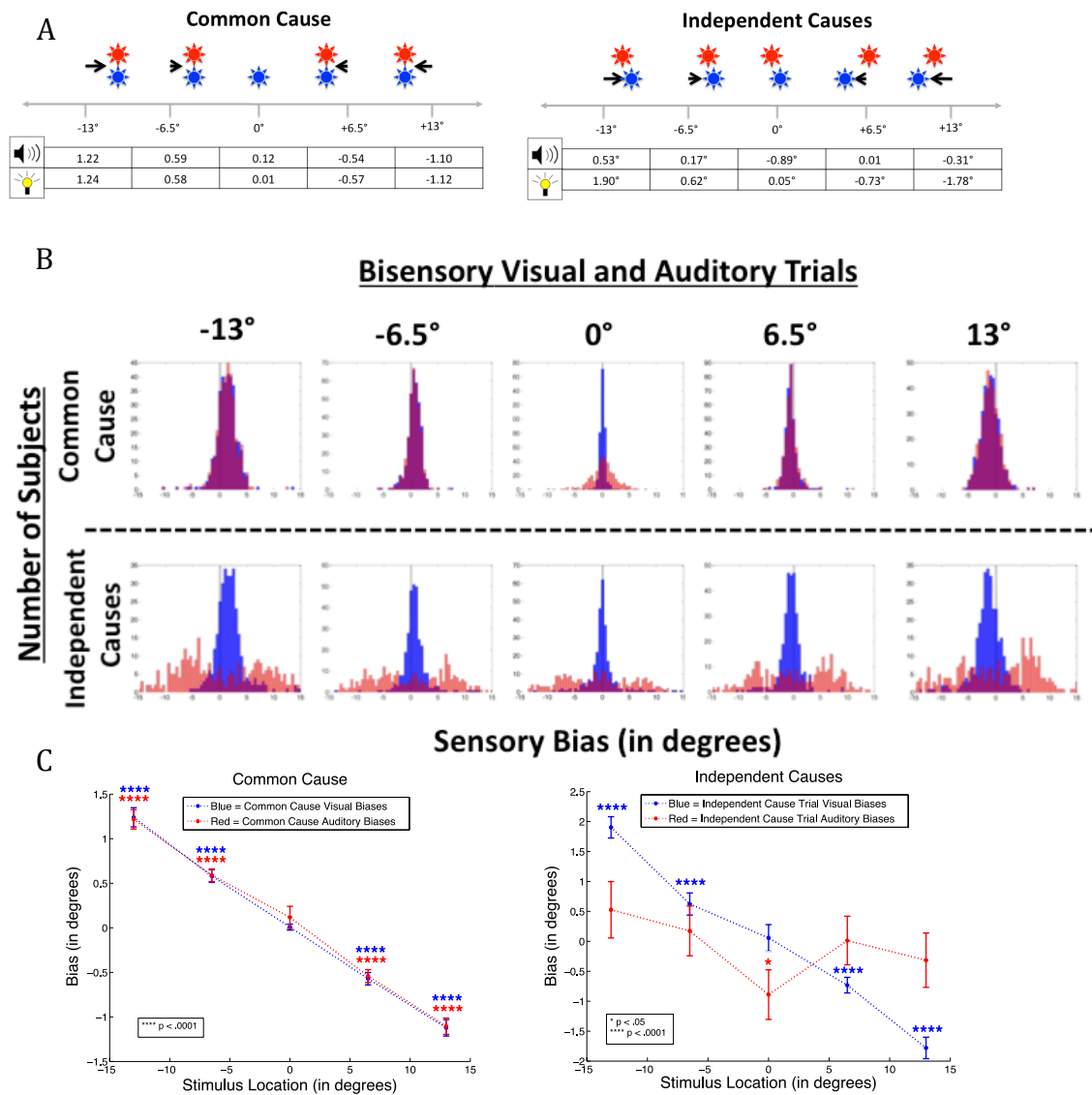
Given that the average biases in the visual and auditory modalities are in opposite directions, an intriguing question arises: in which direction is this conflict resolved when auditory and visual stimuli co-occur at the same location? Analysis of congruent bisensory trials (i.e., trials on which the visual and auditory stimulus co-occurred at the same spatial location) revealed that on average, both the visual and auditory modality exhibited *central* biases (Fig. 5). However, as shown in Fig. 6, further analysis of these trials revealed that the bias that emerged in the auditory modality was dependent on whether the observer inferred a common cause or distinct causes for the stimuli.



**Fig. 5.** Visual and auditory biases on spatially congruent bisensory trials. First, each subject's average bias (over 15 trials) was computed for each of the five locations. Then, the mean across 384 subjects was calculated. Error bars represent standard error of the mean across subjects' averages.

Previous research investigating auditory localizations has shown that the bias that emerges in the auditory domain is contingent upon whether observers perceive a simultaneous visual stimulus as unified with or independent from the auditory stimulus (see fig. 5 in Wallace et al., 2004). Observers' explicit reports indicate that even with spatially and temporally congruent flashes and sound bursts, sometimes a common cause is perceived, but at other times discrepant causes are perceived, and this inference can affect the amount of bias that emerges during the localization task. This phenomenon has been explained quantitatively by a Bayesian causal inference model (Körding et al., 2007). Furthermore, recalibration of auditory space by vision has been shown to strongly depend on whether or not a common cause is perceived for the two stimuli (Wozny & Shams, 2011b).

Therefore, we investigated whether the inference of a common cause influences the bias in bisensory conditions. As in Wozny & Shams (2011b), we classified the congruent bisensory trials into "common-cause" and "independent-cause" trials based on the observer's responses. If the visual and auditory localizations were within two degrees of each other, we considered the trial as a common-cause trial (as the stimuli are perceived to have the same location, thus reflecting the inference of a common cause). If the visual and auditory localizations were more than five degrees apart, we considered the trial to be generated by independent causes (as this degree of discrepancy between the two percepts is inconsistent with the inference of a common cause, and is unlikely to be due to motor error). Trials in which the discrepancy between visual and auditory responses were in between 2 and 5 degrees were excluded from analysis (due to uncertainty about the inference of causal structure). The results are shown in Fig. 6.



**Fig. 6.** Bisensory trials, classified by inferring one common cause, or two independent causes. Panel A shows average biases and panel B displays the distribution of subjects' individual biases for each location. In Panel C, positive numbers indicate rightward biases, and negative numbers indicate leftward biases. SEM bars are shown around the mean bias in each modality for each location.

As can be seen in Fig. 6c, on “common cause” trials both modalities show a considerable central bias, which increases as eccentricity of the target increases. Interestingly, comparing these common cause trials with the corresponding unisensory trials in each modality reveals an important effect: when trials are integrated, the amount of bias *decreases* in each modality, and the localizations become closer to the veridical location of the stimulus ( $p < .005$  for t-tests for

all eccentric locations). For example, in the visual modality, unisensory localizations at the -13 degree location exhibited a bias towards the center of 1.53 degrees, but when visual localizations are integrated with an auditory stimulus at this location, the bias decreases to 1.24 degrees. This reveals an important new finding: multisensory integration not only increases precision in the final spatial estimates of stimuli (which was previously well-established), but also increases the *accuracy* of the estimates by reducing biases and bringing the final estimates closer to the true stimulus location.

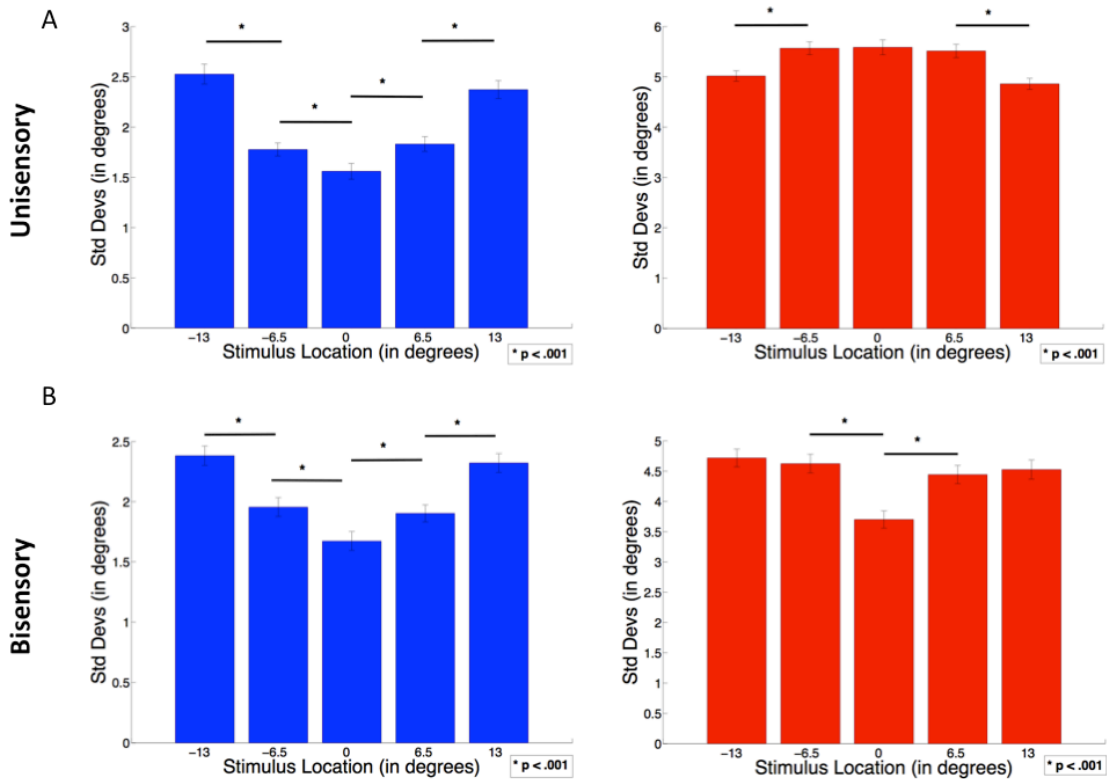
Localizations on trials where independent causes were inferred are much more variable: while the visual modality shows a central bias that appears to increase as eccentricity increases, the bias that emerges in the auditory modality is irregular, and the trend less clear. This variability is apparent in Fig. 6c, where the irregularity in auditory localizations on trials where independent causes were inferred is evident. Thus, to summarize: in spatially congruent audiovisual trials in which a common cause is perceived, the visual bias dominates, however, the degree of bias is smaller relative to unisensory trials. In contrast, on segregated trials, while the visual modality still exhibits a central bias, auditory biases are extremely variable.

### *Precision of localization*

Analysis of the variability in unisensory trials (Fig. 7) showed the following: the standard deviation of visual localizations was not equivalent across conditions ( $F(4,1532) = 39.467$ ,  $p < .001$ ), and increased as the eccentricity of the stimulus increased ( $p < .001$  for all paired-samples t-tests between adjacent conditions). The standard deviation of auditory localizations also revealed differences contingent on stimulus eccentricity ( $F(4, 1532) = 16.32$ ,  $p < .001$ ), but in a different way. That is to say, variability in the two auditory standard deviations surrounding the zero-degree location (+6.5 and -6.5) degrees) were not significantly different from the variability

at the central location ( $p > .05$  for both paired-samples t-tests), but variability in the peripheral locations (-13 and +13 degrees) were significantly different from the variability in localizations at the spatially adjacent neighbors ( $p < .001$  for both paired-samples t-tests).

We also evaluated the precision in visual and auditory localizations on trials where the visual and auditory stimuli co-occurred at the same locations in space. In the visual modality, similar to the unisensory trials, the standard deviations for visual localizations increased as eccentricity increased ( $F(4,1532) = 29.89, p < .001$ ). In the auditory modality, although a significant effect emerged ( $F(4,1532) = 22.73, p < .001$ ), none of the peripheral locations were significantly different from one another, and the only significant difference was that variance in peripheral locations was larger than that of the center ( $p < .001$  for all paired-samples t-tests).



**Fig. 7.** Standard deviations for all five stimulus positions for (A) unisensory visual (blue) and unisensory auditory (red) localizations, and (B) bisensory visual (blue) and bisensory auditory (red) localizations for each modality.



## Model Comparisons

Table 2 shows the results of model fits for each of the 6 models. The first two columns describe each model, and columns 3 and 4 show the average of goodness of fits for each model averaged across all observers. The lower the log likelihood, the better the fit to the data. As can be seen, the 8-parameter model accounts for the observers' data the best, as it has the lowest average log likelihood value, and also has the lowest average BIC value (i.e. the best account of the data even after penalizing it for having larger number of free parameters than some of the other models).

Model Description	Model Parameters	Log Likelihood	BIC
<u>4-Param Original Model</u> (Körding et al., 2007) Meta-modal Central Bias prior	$P_c, \sigma_V, \sigma_A, \sigma_P$	-2416.0	4857.1
<u>5-Param Likelihood Model</u> Shift in Likelihood Means	$P_c, \sigma_V, \sigma_A, \Delta X_V,$ $\Delta X_A$	-2357.9	4747.2
<u>6-Param Likelihood Model</u> Shift in Likelihood Means, Variable Visual Likelihood Variance	$P_c, \sigma_V, \sigma_A, \Delta X_V,$ $\Delta X_A, \Delta \sigma_V$	-2333.9	4705.3
<b><u>8-Param Hybrid Model</u></b> <b>Shift in Likelihood Means,</b> <b>Variable Visual Likelihood</b> <b>Variance, Meta-Modal Central</b> <b>Bias Prior</b>	<b><math>P_c, \sigma_V, \sigma_A, \Delta X_V,</math></b> <b><math>\Delta X_A, \Delta \sigma_V, \sigma_P, X_P</math></b>	<b>-2311.8</b>	<b>4673.7</b>
<u>7-Param Prior Model</u> Visual Central Prior & Auditory Peripheral Prior	$P_c, \sigma_V, \sigma_A, \sigma_{VP},$ $X_{VP}, \sigma_{AP}, X_{AP}$	-2352.3	4748.4
<u>9-Param Prior Model</u> Visual Central Prior, Auditory Peripheral Prior, & General Spatial Prior	$P_c, \sigma_V, \sigma_A, \sigma_{VP},$ $X_{VP}, \sigma_{AP}, X_{AP}, \sigma_P,$ $X_P$	-2351.7	4765.7

**Table 2.** Average log likelihood and BIC values for all models that were tested.

A one-way repeated-measures ANOVA applied to the BIC values (Sato & Kording, 2014) showed that the model fits were significantly different from one another ( $F(5) = 248.1, p < .001$ ). Furthermore, a post-hoc t-test between the two best models (the 6-parameter and 8-

parameter model) showed that the 8-parameter model was significantly superior ( $t(383) = 10.2$ ,  $p = .001$ ). Thus, a model that incorporates biases in the sensory representations and a general (metamodal) prior bias for center best captures observers' performance in our multisensory localization task. Comparing the 6-parameter likelihood model with the 7-parameter and 9-parameter prior models shows that the likelihood models outperform the prior models despite having smaller number of parameters. Therefore, it appears that a shift in the likelihood centers, (i.e. a sensory representation bias) is critical in explaining the data.

In the winning hybrid model (8 parameters), the optimized value of the mean for the general meta-modal prior was almost zero, reaffirming the central bias that had been assumed in the original 4-parameter model (Körding et al., 2007; Wozny et al., 2010). Therefore, the only difference between the winning model and the original 4-parameter model is in the bias in visual and auditory sensory representations, and the increase in visual likelihood variance as a function of eccentricity.

## **Discussion**

### *Auditory and visual perception of space is indeed biased*

Previous studies have suggested that humans' visual and auditory spatial localizations of space are biased and inaccurate (Adam et al., 2008; Cui et al., 2010; Fortenbaugh & Robertson, 2011; Razavi et al., 2007). Why should such a fundamental function that is critical to our survival be inaccurate? If such biases are truly present in human perception of space, they would pose a riddle and warrant further investigation into their characteristics and function. However, the previous studies have been based on relatively small sample sizes that render the results prone to outliers and other statistical irregularities. To examine whether these biases are real (or reflect noise or other phenomena), we used a very large sample size (384 subjects) and

an experimental design minimizing the potential for induction of any perceptual or response biases. The experiment consisted of a large variety of unisensory and bisensory stimulus conditions presented in pseudo-random order with a uniform distribution across space, and a fairly large number of trials per condition. Furthermore, the positions of the auditory and visual sources were unknown to the observers, no feedback was provided, and the initial position of the response cursor was randomized. Thus, the experimental design could not induce any perceptual or response bias for any given position in space. By using rigorous experimental design and acquiring a large dataset, we aimed to obtain a more definitive answer to the question of the existence and nature of any biases in the visual and auditory perception of space along azimuth.

The results demonstrated that when observers were asked to localize briefly presented visual and auditory targets along azimuth, on average, localizations of unisensory visual targets displayed a central bias that increased with eccentricity, and localizations of unisensory auditory targets displayed a peripheral bias that increased with eccentricity. Our finding of a central bias in visual localization is consistent with most studies in this area, especially those in which the cursor was visible during response (which thus minimizes motor error) (Adam et al., 2008; Cui et al., 2010; Parise et al., 2012; Razavi et al., 2007). While the vast majority of observers in our sample showed a clear and consistent bias towards center in vision, the biases exhibited by observers in the auditory modality were more variable and less consistent, but still on average displayed a peripheral bias. This finding is also consistent with those of several previous studies of auditory spatial localizations (Cui et al., 2010; Lewald & Ehrenstein, 1998; Parise et al., 2012).

Localizations for spatially congruent bisensory stimuli exhibited biases that were contingent upon the subject's inference of the causal structure of stimuli: when a common cause

was inferred, the visual bias dominated. When independent causes were inferred, vision continued to exhibit a bias towards the center of space, but the auditory modality on average showed a peripheral bias (as in the unisensory auditory condition). These findings are quite consistent with research requiring explicit judgments of unity or discrepancy (Wallace et al., 2004), indicating that sensory systems implicitly perform inference processes, and that the biases that emerge are contingent upon whether or not sensory signals were integrated or segregated. Interestingly, on trials for which a common cause was perceived, the opposing biases were reconciled in the following manner: the visual bias dominated, although its magnitude was reduced. This finding strongly suggests that the observed biases in the unisensory conditions stem from a bias in sensory representations. Because visual representations are significantly more precise (and reliable) than the auditory representations, in bisensory conditions the visual signals dominate, and therefore, the visual bias dominates the auditory bias. However, in principle, any bias in perception could originate from either a bias in sensory representations (modeled by likelihood functions) or a bias in the expectations or model of the world (prior).

*The biases in localization largely reflect biases in sensory representations*

While the findings discussed above suggest that biases in bisensory conditions are more consistent with a bias in the sensory representations, to address this question rigorously, one needs to quantitatively model these scenarios and perform model comparisons to determine which mechanism(s) account for the behavioral data the best. To do this, we used the Bayesian Causal Inference model that has been shown to account for human observers' multisensory perception quite effectively (Beierholm et al., 2009; Körding et al., 2007; Rohe & Noppeney, 2015; Samad et al., 2015; Wozny et al., 2008, 2010; Wozny & Shams, 2011a). We compared models incorporating a bias in likelihoods to those incorporating a bias in prior expectations, and

also tested one model that incorporated both mechanisms. We fitted the model parameters to each observer's data, and compared the models based on data from 384 observers. Consistent with the qualitative observation discussed above, the quantitative model comparison results revealed that indeed a bias in auditory and visual spatial sensory representations is necessary to account for the participants' data. But over and above these sensory biases, a general (metamodal) prior bias for center appears to also exist in perception of space.

### *Why do biases exist?*

Why should the visual system be biased in perceiving locations towards the center? Several possibilities exist; one plausible hypothesis is that this bias is an epiphenomenon of the superior acuity in the fovea. Due to the much larger number of neurons representing the fovea compared to peripheral regions (Connolly & Van Essen, 1984), the representation of space may be skewed in that direction if spatial representation involves a population code (Fischer & Peña, 2011). In our task, since we enforce fixation at a central point, it seems plausible that this cortical magnification effect may skew visual localizations towards the center. Another possible explanation is that an uneven distribution of attention across the visual field causes the biases. Evidence for this idea has been demonstrated in previous experiments (Fortenbaugh & Robertson, 2011), and these authors note that distributed attention across an axis “leads to focal baseline changes at the point of fixation that dissipate with increased eccentricity” (p. 826). Additionally, a recent review (Anton-Erxleben & Carrasco, 2013) describes neural models accounting for spatial biases caused by attention (see Box 3), and particularly relevant for our study is a model (Compte & Wang, 2006) which incorporates attention's spatial modulation of receptive fields in visual cortex (see Fig. 1 and Fig. 3). Thus, while we can only speculate on the

neural mechanisms underlying bias in the visual modality, future neurophysiological studies should investigate the origin of this phenomenon.

The findings regarding peripheral auditory biases also raise the question of why these biases should exist in the first place. As noted by Razavi et al. (2007), prolonged fixation in an eccentric direction shifts the entire map of auditory space in the direction of the fixation point, even for auditory stimuli that are at the same location as fixation. While auditory localizations still exhibit a slight peripheral bias when central fixation is maintained, it appears that when the eyes are allowed to move, auditory space is shifted in the direction of the eye movement. Thus, in our task, as eye position was not constrained, the auditory bias may be at least partially due to this mechanism. The tremendous amount of plasticity in auditory representations has been demonstrated not only in the context of eye movements, but also as a function of momentary exposure to visual and auditory-visual discrepancy (Wozny & Shams, 2011b). Therefore, auditory representation of the space appears to be highly dynamic.

Moreover, much is still unknown about how auditory spatial perception is encoded in the human brain. One current idea is that rather than having explicit auditory spatial *maps*, some form of a distributed population code is built by neurons that may have broad tuning curves (Lee & Groh, 2014). As noted in a recent review (Salminen, Tiitinen, & May, 2012), this population rate code may be able to be characterized as a “hemifield code,” where the firing rates of the relevant auditory neuron population are highest for stimuli in the periphery (-90 and +90 degrees) that maximize the ITD and ILD information available. This idea has substantial support from experimental findings indicating that sound source locations could be encoded by the relative firing rates of the right-tuned or left-tuned auditory cortical neurons (Stecker, Harrington, & Middlebrooks, 2005; Werner-Reiss & Groh, 2008) and has been further supported

by recent neuroimaging work (Salminen, May, Alku, & Tiitinen, 2009; Salminen, Tiitinen, Yrttiaho, & May, 2010). Thus, it seems possible that when peripheral stimuli are presented, enhanced activity in the auditory population code skews perception of auditory space towards eccentric locations. Future research will need to determine exactly how activity in both cortical and subcortical regions contributes to humans' perception of auditory space.

#### *Multisensory integration reduces the bias in spatial perception*

Multisensory integration may be one of the strategies the nervous system attempts to reduce the problem of unisensory biases. Our behavioral data showed that biases in spatial perception were much smaller in bisensory congruent trials (in which both sound and light were presented at the same location), relative to either the unisensory visual or unisensory auditory trials. Therefore, it appears that multisensory integration not only benefits perception by improving precision (reducing variability), but also by reducing bias; a win-win approach.

#### *Open Questions*

While the present work provides valuable insight into the computational mechanism underlying spatial biases in the visual and auditory systems, future research must address several pertinent questions. For example, future studies will need to address the extent of these biases in more complex environments, including more naturalistic settings. For instance, it has been reported that the localization of objects is often biased towards other elements in the visual field (Kerzel, 2002). Known as the “landmark attraction effect” (Hubbard & Ruppel, 1999, 2000), this effect is particularly relevant for the current paradigm; while this investigation illuminates the biases present in each modality in a sparse visual scene (i.e. with only a fixation cross present

at the start of each trial), the question remains as to how biases change as objects that are potential sources of the relevant auditory and visual signals are introduced to the environment.

### *Conclusion*

Nearly every function critical for human survival depends directly on accurate localization of objects and events in the environment. While numerous studies have investigated spatial perception in humans, little consensus exists on whether representation of space in the human brain is accurate or biased, and if the latter is true, how so and why.

The findings of this large-scale study revealed that on average, observers' localizations are *inaccurate*, as visual localizations show a bias towards the center of space, while auditory localizations show a bias towards the periphery. Even more surprisingly, the observed biases in localization appear to be at least partly due to biases in sensory representations of space as opposed to *a priori* expectations of the spatial layout of objects in the environment (which may be due to a number of non-sensory factors such as the experimental setup, instructions, learning the distribution of stimuli during experiment, etc.).

In real-world settings, objects frequently produce signals that simultaneously stimulate multiple sensory modalities. Considering the opposing biases in the visual and auditory modalities, when visual and auditory stimuli *co-occur* at the *same* spatial location, which bias emerges during localization? The results show that visual bias dominates, however, the magnitude of the central bias in the visual modality is *reduced* when the visual stimulus is integrated with a co-occurring auditory stimulus, thus revealing an additional advantage of multisensory integration: *the reduction of perceptual biases*.



## References

- Adam, J. J., Davelaar, E. J., Gouw, A. van der, & Willems, P. (2008). Evidence for attentional processing in spatial localization. *Psychological Research*, *72*(4), 433–442.  
<http://doi.org/10.1007/s00426-007-0126-2>
- Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. <http://doi.org/10.1016/j.cub.2004.01.029>
- Anton-Erxleben, K., & Carrasco, M. (2013). Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nature Reviews Neuroscience*, *14*(3), 188–200. <http://doi.org/10.1038/nrn3443>
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*(7), 1391–1397. <http://doi.org/10.1364/JOSAA.20.001391>
- Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*(5), 23.  
<http://doi.org/10.1167/9.5.23>
- Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482–489. <http://doi.org/10.3758/BF03208826>
- Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*(6), 578–584.  
<http://doi.org/10.3758/BF03207374>
- Bock, O. (1986). Contribution of retinal versus extraretinal signals towards visual localization in goal-directed movements. *Experimental Brain Research*, *64*(3), 476–482.  
<http://doi.org/10.1007/BF00340484>

- Bock, O. (1993). Localization of objects in the peripheral visual field. *Behavioural Brain Research*, 56(1), 77–84. [http://doi.org/10.1016/0166-4328\(93\)90023-J](http://doi.org/10.1016/0166-4328(93)90023-J)
- Bruno, A., & Morrone, M. C. (2007). Influence of saccadic adaptation on spatial localization: Comparison of verbal and pointing reports. *Journal of Vision*, 7(5), 16. <http://doi.org/10.1167/7.5.16>
- Charbonneau, G., Véronneau, M., Boudrias-Fournier, C., Lepore, F., & Collignon, O. (2013). The ventriloquist in periphery: Impact of eccentricity-related reliability on audio-visual localization. *Journal of Vision*, 13(12), 20. <http://doi.org/10.1167/13.12.20>
- Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, & Psychophysics*, 75(5), 790–811. <http://doi.org/10.3758/s13414-013-0475-4>
- Compte, A., & Wang, X.-J. (2006). Tuning Curve Shift by Attention Modulation in Cortical Neurons: a Computational Study of its Mechanisms. *Cerebral Cortex*, 16(6), 761–778. <http://doi.org/10.1093/cercor/bhj021>
- Connolly, M., & Van Essen, D. (1984). The representation of the visual field in parvocellular and magnocellular layers of the lateral geniculate nucleus in the macaque monkey. *The Journal of Comparative Neurology*, 226(4), 544–564. <http://doi.org/10.1002/cne.902260408>
- Cui, Q. N., O'Neill, W. E., & Paige, G. D. (2010). Advancing age alters the influence of eye position on sound localization. *Experimental Brain Research*, 206(4), 371–379. <http://doi.org/10.1007/s00221-010-2413-1>
- Dobrev, M. S., O'Neill, W. E., & Paige, G. D. (2012). Influence of age, spatial memory, and ocular fixation on localization of auditory, visual, and bimodal targets by human subjects.

*Experimental Brain Research*, 223(4), 441–455. <http://doi.org/10.1007/s00221-012-3270->

x

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433. <http://doi.org/10.1038/415429a>

Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162–169. <http://doi.org/10.1016/j.tics.2004.02.002>

Fischer, B. J., & Peña, J. L. (2011). Owl's behavior and neural representation predicted by Bayesian inference. *Nature Neuroscience*, 14(8), 1061–1066.

<http://doi.org/10.1038/nn.2872>

Fortenbaugh, F. C., & Robertson, L. C. (2011). When here becomes there: attentional distribution modulates foveal bias in peripheral localization. *Attention, Perception, & Psychophysics*, 73(3), 809–828. <http://doi.org/10.3758/s13414-010-0075-5>

Fortenbaugh, F. C., Sanghvi, S., Silver, M. A., & Robertson, L. C. (2012). Exploring the edges of visual space: the influence of visual boundaries on peripheral localization. *Journal of Vision*, 12(2). <http://doi.org/10.1167/12.2.19>

Ghahramani, Z. (1995). *Computation and psychophysics of sensorimotor integration*.

Massachusetts Institute of Technology. Retrieved from

<http://www.learning.eng.cam.ac.uk/zoubin/papers/thesis.pdf>

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual Localization Ability Influences Cross-Modal Bias. *Journal of Cognitive Neuroscience*, 15(1), 20–29. <http://doi.org/10.1162/089892903321107792>

- Henriques, D. Y. P., Klier, E. M., Smith, M. A., Lowy, D., & Crawford, J. D. (1998). Gaze-Centered Remapping of Remembered Visual Space in an Open-Loop Pointing Task. *The Journal of Neuroscience*, *18*(4), 1583–1594.
- Hubbard, T. L., & Ruppel, S. E. (1999). Representational momentum and the landmark attraction effect. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *53*(3), 242–256. <http://doi.org/10.1037/h0087313>
- Hubbard, T. L., & Ruppel, S. E. (2000). Spatial memory averaging, the landmark attraction effect, and representational gravity. *Psychological Research*, *64*(1), 41–55. <http://doi.org/10.1007/s004260000029>
- Kayser, C., & Shams, L. (2015). Multisensory Causal Inference in the Brain. *PLoS Biol*, *13*(2), e1002075. <http://doi.org/10.1371/journal.pbio.1002075>
- Kerzel, D. (2002). Memory for the position of stationary objects: disentangling foveal bias and memory averaging. *Vision Research*, *42*(2), 159–167. [http://doi.org/10.1016/S0042-6989\(01\)00274-7](http://doi.org/10.1016/S0042-6989(01)00274-7)
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, *2*(9), e943. <http://doi.org/10.1371/journal.pone.0000943>
- Lee, J., & Groh, J. M. (2014). Different Stimuli, Different Spatial Codes: A Visual Map and an Auditory Rate Code for Oculomotor Space in the Primate Superior Colliculus. *PLoS ONE*, *9*(1), e85017. <http://doi.org/10.1371/journal.pone.0085017>
- Lewald, J., & Ehrenstein, W. H. (1998). Auditory-visual spatial integration: A new psychophysical approach using laser pointing to acoustic targets. *The Journal of the Acoustical Society of America*, *104*(3), 1586–1597. <http://doi.org/10.1121/1.424371>

- Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, *16*(3), 468–478.  
[http://doi.org/10.1016/S0926-6410\(03\)00074-0](http://doi.org/10.1016/S0926-6410(03)00074-0)
- Mateeff, S., & Gourevich, A. (1983). Peripheral vision and perceived visual direction. *Biological Cybernetics*, *49*(2), 111–118. <http://doi.org/10.1007/BF00320391>
- Müsseler, J., Heijden, A. H. C. V. D., Mahmud, S. H., Deubel, H., & Ertsey, S. (1999). Relative mislocalization of briefly presented stimuli in the retinal periphery. *Perception & Psychophysics*, *61*(8), 1646–1661. <http://doi.org/10.3758/BF03213124>
- Ocklenburg, S., Hirnstein, M., Hausmann, M., & Lewald, J. (2010). Auditory space perception in left- and right-handers. *Brain and Cognition*, *72*(2), 210–217.  
<http://doi.org/10.1016/j.bandc.2009.08.013>
- Osaka, N. (1977). Effect of Refraction on Perceived Locus of a Target in the Peripheral Visual Field. *The Journal of Psychology*, *95*(1), 59–62.  
<http://doi.org/10.1080/00223980.1977.9915860>
- Parise, C. V., Spence, C., & Ernst, M. O. (2012). When Correlation Implies Causation in Multisensory Integration. *Current Biology*, *22*(1), 46–49.  
<http://doi.org/10.1016/j.cub.2011.11.039>
- Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, *6*(4), 203–205.  
<http://doi.org/10.3758/BF03207017>
- Razavi, B., O'Neill, W. E., & Paige, G. D. (2007). Auditory Spatial Perception Dynamically Realigns with Changing Eye Position. *The Journal of Neuroscience*, *27*(38), 10249–10258. <http://doi.org/10.1523/JNEUROSCI.0938-07.2007>

- Rock, I., & Victor, J. (1964). VISION AND TOUCH: AN EXPERIMENTALLY CREATED CONFLICT BETWEEN THE TWO SENSES. *Science (New York, N.Y.)*, *143*(3606), 594–596.
- Rohe, T., & Noppeney, U. (2015). Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLoS Biol*, *13*(2), e1002073.  
<http://doi.org/10.1371/journal.pbio.1002073>
- Salminen, N. H., May, P. J. C., Alku, P., & Tiitinen, H. (2009). A Population Rate Code of Auditory Space in the Human Cortex. *PLoS ONE*, *4*(10), e7600.  
<http://doi.org/10.1371/journal.pone.0007600>
- Salminen, N. H., Tiitinen, H., & May, P. J. C. (2012). Auditory Spatial Processing in the Human Cortex. *The Neuroscientist*, *18*(6), 602–612. <http://doi.org/10.1177/1073858411434209>
- Salminen, N. H., Tiitinen, H., Yrttiaho, S., & May, P. J. C. (2010). The neural code for interaural time difference in human auditory cortex. *The Journal of the Acoustical Society of America*, *127*(2), EL60–65. <http://doi.org/10.1121/1.3290744>
- Samad, M., Chung, A. J., & Shams, L. (2015). Perception of Body Ownership Is Driven by Bayesian Sensory Inference. *PLoS ONE*, *10*(2), e0117178.  
<http://doi.org/10.1371/journal.pone.0117178>
- Sato, Y., & Kording, K. P. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision*, *14*(13), 13. <http://doi.org/10.1167/14.13.13>
- Sheth, B. R., & Shimojo, S. (2001). Compression of space in visual memory. *Vision Research*, *41*(3), 329–341. [http://doi.org/10.1016/S0042-6989\(00\)00230-3](http://doi.org/10.1016/S0042-6989(00)00230-3)
- Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7–10.

- Stecker, G. C., Harrington, I. A., & Middlebrooks, J. C. (2005). Location Coding by Opponent Neural Populations in the Auditory Cortex. *PLoS Biol*, 3(3), e78.  
<http://doi.org/10.1371/journal.pbio.0030078>
- Uddin, M. K. (2006). Visual spatial localization and the two-process model. *Psychological Research of Kyushu University*, 7, 65–75.
- Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2), 252–258. <http://doi.org/10.1007/s00221-004-1899-9>
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88(3), 638–667. <http://doi.org/10.1037/0033-2909.88.3.638>
- Werner-Reiss, U., & Groh, J. M. (2008). A Rate Code for Sound Azimuth in Monkey Auditory Cortex: Implications for Human Neuroimaging Studies. *The Journal of Neuroscience*, 28(14), 3747–3758. <http://doi.org/10.1523/JNEUROSCI.5044-07.2008>
- Wozny, D. R., Beierholm, U. R., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision*, 8(3). <http://doi.org/10.1167/8.3.24>
- Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability Matching as a Computational Strategy Used in Perception. *PLoS Comput Biol*, 6(8), e1000871.  
<http://doi.org/10.1371/journal.pcbi.1000871>
- Wozny, D. R., & Shams, L. (2011a). Computational Characterization of Visually Induced Auditory Spatial Adaptation. *Frontiers in Integrative Neuroscience*, 5.  
<http://doi.org/10.3389/fnint.2011.00075>

Wozny, D. R., & Shams, L. (2011b). Recalibration of Auditory Space following Milliseconds of Cross-Modal Discrepancy. *The Journal of Neuroscience*, 31(12), 4607–4612.

<http://doi.org/10.1523/JNEUROSCI.6079-10.2011>

Zahn, J. R., Abel, L. A., & Dell'Osso, L. F. (1978). Audio-ocular response characteristics. *Sensory Processes*, 2(1), 32–37.



## VIII.

### **Exposure to discrepant audiovisual signals changes the tendency to integrate**

#### **Abstract**

While many studies have investigated how spatial and temporal factors influence audiovisual integration on single trials, few studies have systematically investigated whether the general perceptual tendency to integrate stimuli is fixed or plastic. In our study, subjects first completed a pre-test phase involving localization along azimuth of simple audiovisual stimuli with varying amounts of spatial discrepancy, and then engaged an exposure task that included audiovisual stimuli with varying spatial and temporal relationships. This exposure task was interleaved throughout a post-test audiovisual localization session. We fit a Bayesian Causal Inference model to both the pre-test and post-data data, and discovered that subjects' estimated prior probability of integrating stimuli increased the most in the post-test phase following exposure to repeated stimuli that were (1) spatially discrepant, and (2) synchronous in time. Exposure to spatially and temporally synchronous stimuli did not alter integration tendencies. The integration prior also increased with exposure to spatially discrepant stimuli that were consistently offset by 400ms, indicating the importance of temporal correlations in updating this prior. These findings indicate the brain can learn to integrate discrepant pieces of information as long as the temporal structure of the signals provides sufficient evidence to warrant inference that the signals come from a common source. This study reveals counterintuitive findings about the factors influencing sensory recalibration, and can inform future investigations looking to manipulate the tendency to bind sensory signals.

## **Introduction**

Our sensory systems are constantly being bombarded by information from various sources in the environment, and the brain must determine whether these signals originate from a single source and should be integrated, or originate from separate sources and should be segregated (Körding et al., 2007; Shams & Beierholm, 2010). In this process of perceptual inference, two of the most influential factors that determine whether or not sensory signals are integrated are (1) the spatial relationship between the signals, and (2) the temporal relationship between the signals.

Investigations of how the spatial and temporal relationships between signals influence multisensory integration on single trials has been well-documented (for a thorough review, see Chen & Vroomen, 2013). For instance, when auditory and visual stimuli are displayed simultaneously from spatially discrepant sources, localization of the auditory signal is often biased towards the location of the visual stimulus (Alais & Burr, 2004; Bertelson & Aschersleben, 1998; Bertelson & Radeau, 1981; Jack & Thurlow, 1973; M Radeau & Bertelson, 1987). The strength of this phenomenon, known as the “ventriloquist effect,” has been shown to decrease with increasing spatial and temporal disparity between the signals (Hairston et al., 2003; Lewald & Guski, 2003; Slutsky & Recanzone, 2001; Wallace et al., 2004).

In addition to this spatial form of ventriloquism, temporal ventriloquism has also been shown by a number of studies (Bertelson & Aschersleben, 2003; Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Vroomen & de Gelder, 2004). In temporal ventriloquism, judgments about the timing of visual stimuli are influenced by auditory stimuli that either precede or follow the visual stimulus, indicating that crossmodal interactions modify multisensory percepts in the temporal domain as well.

However, in addition to these examples of spatial and temporal discrepancies influencing immediate perception of multisensory stimuli, other studies reveal how repeated exposures to stimuli with fixed discrepancies can cause substantial changes in future perceptual experiences. For instance, repeated exposure to spatially discrepant visual and auditory stimuli can result in subsequent auditory localizations being mislocalized towards the direction of the previous visual stimulus (Bertelson, Frissen, Vroomen, & de Gelder, 2006; Canon, 1970; Frissen, Vroomen, & de Gelder, 2012; Frissen, Vroomen, de Gelder, & Bertelson, 2003, 2005; Lewald, 2002; Monique Radeau & Bertelson, 1977; Recanzone, 1998). This form of spatial adaptation is typically referred to as the “ventriloquist aftereffect,” although a related phenomenon exists in the temporal domain, where studies have shown that auditory and visual signals presented with a consistent temporal offset result in recalibration of the point of subjective simultaneity for audiovisual signals (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Machulla, Luca, Froehlich, & Ernst, 2012; Roseboom & Arnold, 2011; Vroomen, Keetels, de Gelder, & Bertelson, 2004).

The presence of these multisensory interactions in the spatial and temporal domains has raised interesting questions about the computational principles underlying multisensory integration. In the last decade, several models using Bayes’ rule have shown that they can computationally characterize immediate audiovisual interactions in both space (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003; Beierholm, Quartz, & Shams, 2009; Körding et al., 2007; Witten & Knudsen, 2005; Wozny, Beierholm, & Shams, 2010) and time (Sato & Aihara, 2009; Yamamoto, Miyazaki, Iwano, & Kitazawa, 2012). Additionally, two studies modeling the ventriloquist aftereffect in a Bayesian framework have shown that spatial shifts in auditory localizations are best captured by shifts in the means of the auditory likelihood distributions (Sato, Toyozumi, & Aihara, 2007; Wozny & Shams, 2011a).

While all of these behavioral and computational accounts have provided valuable insights into how spatial and temporal factors influence multisensory integration, an important question remains unanswered: how does the brain use spatial and temporal information from previous experiences to update its tendencies to integrate multisensory signals? Few studies have probed this question thoroughly, but one important investigation was conducted by Van Wanrooij, Bremen, & John Van Opstal (2010). In their study, subjects were instructed to direct head-pointing responses to the visual component of a spatially misaligned audiovisual target, and completed blocks where the audiovisual stimuli were either always spatially aligned, aligned 10% of the time, or aligned 50% of the time. Upon assessing spatially congruent trials, differences were shown not only between blocks in the experiments (indicating that prior knowledge of trials could modulate integration effects), but from *trial-to-trial*, indicating that the observer's integration tendencies are extremely plastic, and can be influenced on a moment-to-moment basis. This idea regarding the remarkable plasticity of sensory systems is quite congruent with previous work showing that representations of external auditory space are updated on a moment-to-moment basis, based upon previous multisensory experience (Frissen et al., 2012; Wozny & Shams, 2011b).

Other investigations that shed light on this question have been conducted using cats raised in sensory-impooverished environments. For instance, one study revealed that cats reared in an environment containing spatially and temporally coincident audiovisual stimuli exhibited significantly more neurons with integration capabilities in the superior colliculus than a group reared with audiovisual stimuli presented at random locations and times (Xu, Yu, Rowland, Stanford, & Stein, 2012), indicating that spatial and temporal congruence of signals in the environment plays an important role in the development of multisensory integration. These

neurons retained these properties even after living for an entire year in complete darkness, indicating the preservation of integration tendencies even in the absence of subsequent sensory experience. This finding is quite congruent with behavioral evidence in humans, which shows that without counterevidence, sensory adaptation effects remain for an extended period of time (Machulla et al., 2012). Intriguingly, spatial coincidence does not seem to be necessary for these neurons to develop multisensory integration capabilities, as previous work has shown that animals reared in an environment with audiovisual stimuli that are always spatially disparate and temporally coincident develop neurons with unique integration tendencies, responding maximally to stimuli with spatial discrepancy (Wallace & Stein, 2007). This raises an important question: under what conditions do stimuli cause an increase in the tendency to integrate multisensory signals? Is it exposure to spatially congruent and temporally synchronous stimuli, or spatially disparate and temporally synchronous stimuli? Additionally, is temporal *synchrony* necessary to increase the overall tendency to integrate, or will exposure to temporally *correlated* stimuli suffice?

Recent work has shown that spatially congruent, temporally-offset audiovisual signals are integrated optimally when they are presented with correlated timing (Parise, Spence, & Ernst, 2012); audiovisual signals presented with uncorrelated timing are integrated in a suboptimal manner. Additionally, the just-noticeable differences of spatially-offset audiovisual signals are greatest when presented with perfectly correlated timing, compared to conditions with only partially-correlated or uncorrelated timing (Parise, Harrar, Ernst, & Spence, 2013). This finding was interpreted in a Bayesian framework, with a coupling prior that captures the strength of the tendency to bind based on the temporal correlations between signals.

In our investigation, we first computationally characterized observers' integration tendencies by modeling their data from an audiovisual localization paradigm (Wozny & Shams, 2011a), and then sought to manipulate their tendency to integrate with a ten-minute exposure task that included audio and visual signals which varied in their spatial and temporal relationships. Participants were then assessed in a post-test phase using a top-up design to determine how these spatial and temporal factors influenced the prior probability of binding auditory and visual signals in the environment.

## **Methods & Model**

### *Participants and Apparatus*

A total of 189 subjects (ages 18-55) participated in our experiment; 20 subjects were excluded due to either negligence with the response device or eyetracker malfunctions during the task, and thus analyses for each experiment included between 22-31 subjects. 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. Each subject also signed a 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 projection screen. The screen was a black, acoustically transparent cloth subtending much of the visual field (134° width ° x 60° height). Behind the screen were 9 free-field speakers (5 x 8 cm, extended range paper cone), positioned along azimuth 6.5° apart, 7° below fixation. The middle speaker was positioned below the fixation point, and four speakers were positioned to the right and four to the left of fixation. The visual stimuli were presented

overhead from a ceiling mounted projector set to a resolution of 1280 x 1024 pixels with a refresh rate of 75 Hz.

Prior to the presentation of any stimuli in the experiment, participants were required to have their gaze centered on the central fixation point. To ensure that participants' gaze for each trial was starting from the same location, gaze position and fixation time were recorded at 60Hz with a ViewPoint eye tracker (Arrington Research, Scottsdale, AZ) fixed to the chinrest and PC-60 software (version 2.8.5,000). Stimuli were not displayed until the recorded gaze angle was within  $3.0^\circ$  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.

The visual stimuli used in the experiments were white-noise disks (.41 cd/m<sup>2</sup>) with a Gaussian envelope of  $1.5^\circ$  FWHM, presented  $7^\circ$  below the fixation point on a black background (.07cd/m<sup>2</sup>), for 35 ms. The center of visual stimuli overlapped the center of one of the five speakers behind the screen positioned at  $-13^\circ$ ,  $-6.5^\circ$ ,  $0^\circ$ ,  $6.5^\circ$ , and  $13^\circ$ . 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.

### *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

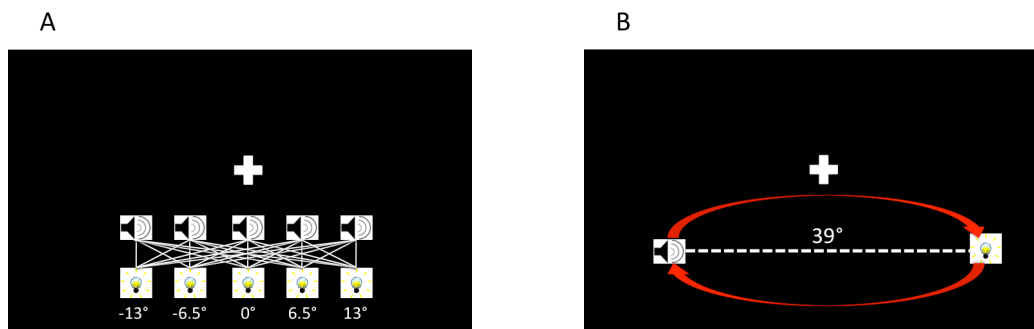
the mouse, understood the instructions, and were accurately fulfilling the fixation requirements for each trial. Each trial started with the fixation cross, followed after 750 ms (if the subject was fixating properly) by the presentation of stimuli. 450 ms after the stimuli, fixation was removed and a cursor appeared on the screen vertically just above the horizontal line where the stimuli were presented 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. 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 minutes to complete. The stimulus conditions included 5 unisensory auditory locations, 5 unisensory visual locations, and 25 combinations of auditory and visual locations (bisensory conditions), for a total of 35 stimulus conditions (15 trials per condition), shown in Figure 8A on the following page. 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 localization trials. In the exposure task, a visual flash and auditory burst of sound (with a fixed spatial and temporal relationship; see individual experiments for details) were presented every 600ms.

Between every 5<sup>th</sup> to 15<sup>th</sup> presentation of the stimuli, the visual flash would get noticeably brighter (from 0.41 to 3.2 cd/m<sup>2</sup>), and participants would have to detect this change in brightness



by clicking the mouse. If they correctly detected the change, the stimuli would change positions, with the visual and auditory stimuli switching to opposite sides of the midline (see Figure 8B). Note that in three experiments (exps. 2, 5, and 6), the visual and auditory stimuli were presented from spatially discrepant locations, but in two experiments (exps. 3 and 4), they were presented from the same location, and alternated sides at the same time to maintain spatial congruency. If subjects missed the brightness change, participants 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 it was regained. This brightness detection task was interleaved throughout the post-test session, so localization blocks and exposure tasks blocks were performed until 525 localization trials (the same as the pre-test session) were completed.



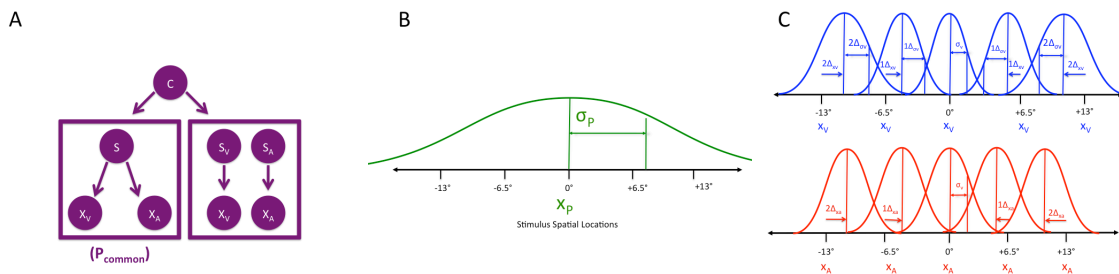
**Fig. 8.** The spatial localization paradigm and exposure task. (A) Stimuli could be presented from one of five locations, ranging from -13 to +13 degrees. (B) The exposure task paradigm. This was interleaved with localization blocks in the post-test phase. Depicted above is the stimulus arrangement 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, and alternated together from one side to the other (see results section for each experiment's details.)

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 8 detections) and

localization blocks (40 trials) alternated until 525 localization trials were completed. Typically, the post-test phase lasted between 60 and 75 minutes.

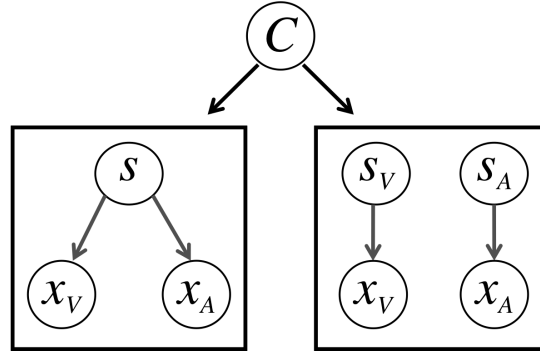
### Model

Following collection of each individual subjects' data, responses were divided into the pre-test and post-test datasets, and a Bayesian Causal Inference model with 8 free parameters (Odegaard et al., submitted) was fit to the data. The 8 free parameters included the prior probability of inferring a common cause ( $p_{\text{common}}$ , or  $p_c$ , see Figure 9A), a prior distribution over space (Figure 9B) characterized by a mean ( $x_p$ ) and variance ( $\sigma_p$ ), and several parameters characterizing the sensory likelihood distributions (Figure 9C). These likelihood parameters included the standard deviation of the visual likelihood variance ( $\sigma_v$ ), how the visual likelihood variance changes with eccentricity ( $\Delta\sigma_v$ ), how bias in the visual likelihood mean changes with eccentricity ( $\Delta x_v$ ), the standard deviation of the auditory likelihood distributions ( $\sigma_a$ ), and how bias in the auditory likelihood mean changes with eccentricity ( $\Delta x_a$ ).



**Fig. 9.** Shown above are the eight parameters fit to individual subjects' data. (A) The inferred causal structure for the presented stimuli. Subjects could either infer the audiovisual signals come from the same source and should be integrated (shown in the left box), or infer that the signals come from disparate sources and should be segregated. Each subject has a unique prior probability of inferring a common cause, which ranges anywhere from 0 to 1. (B) The spatial prior distribution. (C) Graphical representation of the visual and auditory likelihood parameters.

Shown in Figure 10, 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.



**Fig. 10.** The Causal Inference Model. On some trials, the observer may infer that both the auditory ( $x_a$ ) and visual ( $x_v$ ) signals come from a single source and should be integrated (shown on the left). On other trials, the observer may infer that the signals come from two sources and should be segregated. The latent variable  $C$  determines which model generates the data.

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 are sampled from a normal distribution around the true location of the stimulus to simulate the corruption of sensory channels by the noise parameters  $\sigma_v$  and  $\sigma_A$ , 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 term scales with eccentricity (see figure 9C); if the term is positive, this will account for peripheral biases in the sensory representations. If the term is negative, this will account for central biases in the sensory representations.

$$x_v \sim N(s_v + \Delta x_v, \Delta \sigma_v) \tag{1}$$

$$x_A \sim N(s_A + \Delta x_A, \Delta \sigma_A) \tag{2}$$

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 (Odegaard et al., submitted):

$$p(s) = N(\mu_p, \sigma_p) \quad (3)$$

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:

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

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

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

The maximum a posteriori estimate of these computed posteriors reflects the optimal combination of the sensory signals given one or two causes. 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:

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

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:

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

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

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

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

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:

$$p(C = 2 | x_A, x_V) = 1 - p(C = 1 | x_A, x_V) \quad (11)$$

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 below:

Model Selection:

$$\hat{s}_A = \begin{cases} \hat{s}_{A,C=1} & \text{if } p(C = 1 | x_A, x_V) > .5 \\ \hat{s}_{A,C=2} & \text{if } p(C = 1 | x_A, x_V) \leq .5 \end{cases} \quad (12)$$

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

Model Averaging:

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

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

Probability Matching:

$$\hat{s}_A = \begin{cases} \hat{s}_{A,C=1} & \text{if } p(C = 1 | x_A, x_V) > \xi \\ \text{where } \xi \in [0 : 1] \text{ uniform distribution} \\ \hat{s}_{A,C=2} & \text{if } p(C = 1 | x_A, x_V) \leq \xi \\ \text{and sampled on each trial} \end{cases} \quad (14)$$

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

For each subject, we fit the eight free parameters (Figure 9) 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 et al., 2010, see study #1 for details). This was done simulating 10000 trials using MATLAB's *fminsearchbnd* function (Mathworks, 2006), maximizing the likelihood of the parameters of the model. The parameter fits from the best-fitting strategy were then included in analyses shown in the results section.

### *Hypotheses regarding the prior for integrating stimuli*

While we evaluated changes in all parameters between our pre-test and post-test localization sessions, we formulated specific hypotheses regarding how the prior for inferring that stimuli came from a common source (“ $P_{\text{Common}}$ ” or “ $p_c$ ”) might change as a result of the different exposure task conditions. These hypotheses are shown in Table 3 on the following page. Intuitively, it seems reasonable that the brain should learn to integrate stimuli when exposed to repetitive presentations of spatiotemporally congruent signals, indicating that these stimuli could come from the same source (i.e. occur at the same location and at the same time). Similarly, it also seems reasonable to hypothesize that the brain should learn to segregate information when these stimuli are repeatedly presented in a manner that indicates that they originate from different sources (i.e. differ in either location or in time).

Experiment	Time	Space	Expected Change in $P_{\text{Common}}$
Experiment 1	None	None	<b>None</b>
Experiment 2	Uncorrelated	Incongruent	<b>Decrease</b>
Experiment 3	Congruent	Congruent	<b>Increase</b>
Experiment 4	Uncorrelated	Congruent	<b>Decrease</b>
Experiment 5	Congruent	Incongruent	<b>Decrease</b>
Experiment 6	Correlated	Incongruent	<b>Decrease</b>

**Table 3.** Hypotheses regarding the  $P_{\text{Common}}$  parameter. The “time” column denotes whether the stimuli in the exposure task were temporally uncorrelated (drawn from a uniform distribution), temporally correlated (separated in time, but always by a fixed 400ms discrepancy), or temporally congruent (i.e. synchronous).

## **Results**

### *Experiment #1: The Control Experiment (N = 27)*

In this experiment, participants completed a standard pre-test localization session, and then completed a post-test localization session, which did not include the exposure task. Instead, this post-test session included blocks where participants would simply listen to music for brief

periods of time (equivalent to the amount of time the exposure task would have taken). The purpose of this experiment was to determine whether model parameters were consistent from one session to the next, in the absence of including the exposure task.

	Prior Parameters			Likelihood Parameters				
	$p_c$	$\sigma_p$	$x_p$	$\sigma_v$	$\Delta\sigma_v$	$\Delta x_v$	$\sigma_a$	$\Delta x_a$
Pre-Test	0.46 ( $\pm .06$ )	21.57 ( $\pm 2.79$ )	-0.06 ( $\pm 1.71$ )	1.38 ( $\pm .12$ )	0.44 ( $\pm .05$ )	-0.64 ( $\pm .08$ )	15.37 ( $\pm 1.76$ )	4.58 ( $\pm .87$ )
Post-Test	0.51 ( $\pm .06$ )	21.26 ( $\pm 3.38$ )	2.37 ( $\pm 1.41$ )	1.50 ( $\pm .13$ )	0.45 ( $\pm .04$ )	-0.67 ( $\pm .08$ )	17.35 ( $\pm 2.09$ )	5.61 ( $\pm .82$ )

**Table 4.** Optimized Parameter values  $\pm$  standard error parameter estimates from experiment 1.

As shown in Table 4, none of the parameters significantly changed from the pre-test to the post-test in the control experiment, indicating that practice and fatigue effects played a negligible role in participants' localization behaviors over time in this paradigm, and that any changes seen in parameters in other experiments would be due to the effects of the exposure task. The probability of inferring a common cause ( $p_c$ ) increased by only .05 in the post-test phase, and pairwise t-tests conducted to test for differences did not yield any significant findings for this parameter, or any others included in the model ( $p > .05$  for all paired-samples t-tests).

*Exp. #2: Spatially Incongruent & Temporally Uncorrelated Exposure-Task Stimuli (N=22)*

In this experiment, individuals first completed the standard localization session comprised of 525 trials of visual, auditory, and audiovisual trials. Following this, they completed a ten-minute exposure-task session where visual and auditory stimuli were always presented from 39 degrees apart (i.e. when one stimulus was at  $-19.5^\circ$ , the other stimulus was at  $+19.5^\circ$ , and these stimuli flipped positions with each correct detection). Timing of the stimuli during the exposure-task session was randomized, as the presentation times were drawn from



uniform distributions, and thus the two signals were temporally uncorrelated. Two-minute blocks of this task were then interleaved in the post-test localization sessions.

	Prior Parameters			Likelihood Parameters				
	$p_c$	$\sigma_p$	$x_p$	$\sigma_v$	$\Delta\sigma_v$	$\Delta x_v^{**}$	$\sigma_a^*$	$\Delta x_a$
Pre-Test	0.52 ( $\pm .07$ )	24.64 ( $\pm 4.23$ )	0.69 ( $\pm 2.14$ )	1.40 ( $\pm .13$ )	0.59 ( $\pm .06$ )	-0.32 ( $\pm .12$ )	13.38 ( $\pm 1.5$ )	2.71 ( $\pm .86$ )
Post-Test	0.51 ( $\pm .07$ )	31.80 ( $\pm 4.07$ )	1.60 ( $\pm 2.02$ )	1.70 ( $\pm .20$ )	0.49 ( $\pm .03$ )	-0.61 ( $\pm .14$ )	16.66 ( $\pm 1.5$ )	1.59 ( $\pm 1.0$ )

**Table 5.** Optimized parameter values  $\pm$  standard error parameter estimates from experiment 2. \* indicates  $p < .05$ , \*\* indicates  $p < .01$ .

As shown in Table 5, the prior for inferring a common cause ( $p_c$ ) was not significantly different from the pre-test to the post-test session, as these values only changed from 0.52 to 0.51, indicating that subjects were consistent in their tendency to integrate following engagement with exposure-task stimuli. Thus, presentation of these discrepant stimuli in the exposure task did not decrease the tendency to bind; instead, it remained consistent from one session to the next. Two other minor trends are worthy of commentary: the central bias in the visual likelihood parameter increased by three tenths of a degree in the post-test phase ( $t(21) = 4.05$ ,  $p = .001$ ); this is likely due to the fact that fixation was enforced during the interleaved exposure task sessions, thus resulting in a slight “tunnel vision” effect from the eyes being left in one place for extended periods of time. Additionally, the auditory likelihood variance parameter was trending towards significance, but the paired-samples t-test did not survive a Bonferroni correction ( $t(21) = -2.856$ ,  $p > .00625$ ).

### *Experiment #3: Spatially Congruent & Temporally Congruent Exposure-Task Stimuli (N = 30)*

In this experiment, participants completed the standard pre-test localization phase, and then a post-test session with both exposure task blocks and localization blocks. The exposure

task stimuli in this experiment consisted of spatially and temporally congruent visual flashes and auditory bursts of noise. These repeated presentations arose from either +/-19.5°, and would switch to the position on the other side of the midline with each correct detection of the bigger and brighter flash. The purpose of this experiment was to determine how repeated presentations of spatially & temporally synchronous inputs would affect behavior in the post-test localization blocks.

	Prior Parameters			Likelihood Parameters				
	$p_c$	$\sigma_p$	$x_p$	$\sigma_v$	$\Delta\sigma_v$	$\Delta x_v$	$\sigma_a$	$\Delta x_a$
Pre-Test	0.45 (± .06)	21.24 (± 3.28)	0.70 (± 1.26)	1.18 (± .10)	0.64 (± .08)	-0.69 (± .11)	18.80 (± 1.80)	5.07 (± .73)
Post-Test	0.48 (± .07)	29.13 (± 4.69)	1.08 (± 1.42)	1.30 (± .17)	0.55 (± .07)	-0.73 (± .12)	17.34 (± 1.78)	4.17 (± 1.0)

**Table 6.** Optimized parameter values ± standard error parameter estimates from experiment 3.

Displayed in Table 6 above, none of the parameters significantly changed from the pre-test to the post-test localization session. It is interesting to note that repeated exposure to spatiotemporally congruent stimuli did *not* significantly increase the tendency to bind ( $p_c$ ) in the post-test localization session ( $t(29) = -0.72, p = .480$ ), as individuals only slightly increased from 0.45 to 0.48 between the two sessions.

*Experiment #4: Spatially Congruent & Temporally Uncorrelated Exposure-Task Stimuli (N=30)*

In experiment four, participants again first took part in a standard pre-test localization block, and then engaged an exposure task with stimuli that were spatially congruent, but temporally uncorrelated (i.e. drawn from a uniform distribution). As in other experiments, following the initial ten-minute exposure task block, shorter two-minute versions of the exposure task were interleaved throughout the post-test localization session.

	Prior Parameters			Likelihood Parameters				
	$p_c$	$\sigma_p$	$x_p$	$\sigma_v$	$\Delta\sigma_v$	$\Delta x_v^*$	$\sigma_a$	$\Delta x_a$
Pre-Test	0.48 ( $\pm .06$ )	28.80 ( $\pm 4.96$ )	1.88 ( $\pm 1.10$ )	1.29 ( $\pm .09$ )	0.58 ( $\pm .04$ )	0.15 ( $\pm .08$ )	18.65 ( $\pm 1.73$ )	5.88 ( $\pm .74$ )
Post-Test	0.51 ( $\pm .07$ )	30.31 ( $\pm 5.24$ )	1.16 ( $\pm 1.10$ )	1.44 ( $\pm .13$ )	0.52 ( $\pm .03$ )	-0.01 ( $\pm .09$ )	17.70 ( $\pm 1.83$ )	5.41 ( $\pm .72$ )

**Table 7.** Optimized parameter values  $\pm$  standard error parameter estimates from experiment 4. \* indicates  $p < .05$ .

As shown in Table 7, none of the parameters were significantly different from one another between the pre-test and post-test localization sessions, indicating that exposure to spatially congruent, temporally uncorrelated exposure-task stimuli did not change the tendency to integrate, the sensory likelihoods, or the spatial prior. The visual likelihood bias in this group of subjects was trending towards more veridical perception in the post-test localization phase, but this trend was not strong enough to survive a Bonferroni correction ( $p > .00625$ ).

*Experiment #5: Spatially Incongruent & Temporally Congruent Exposure-Task Stimuli (N = 29)*

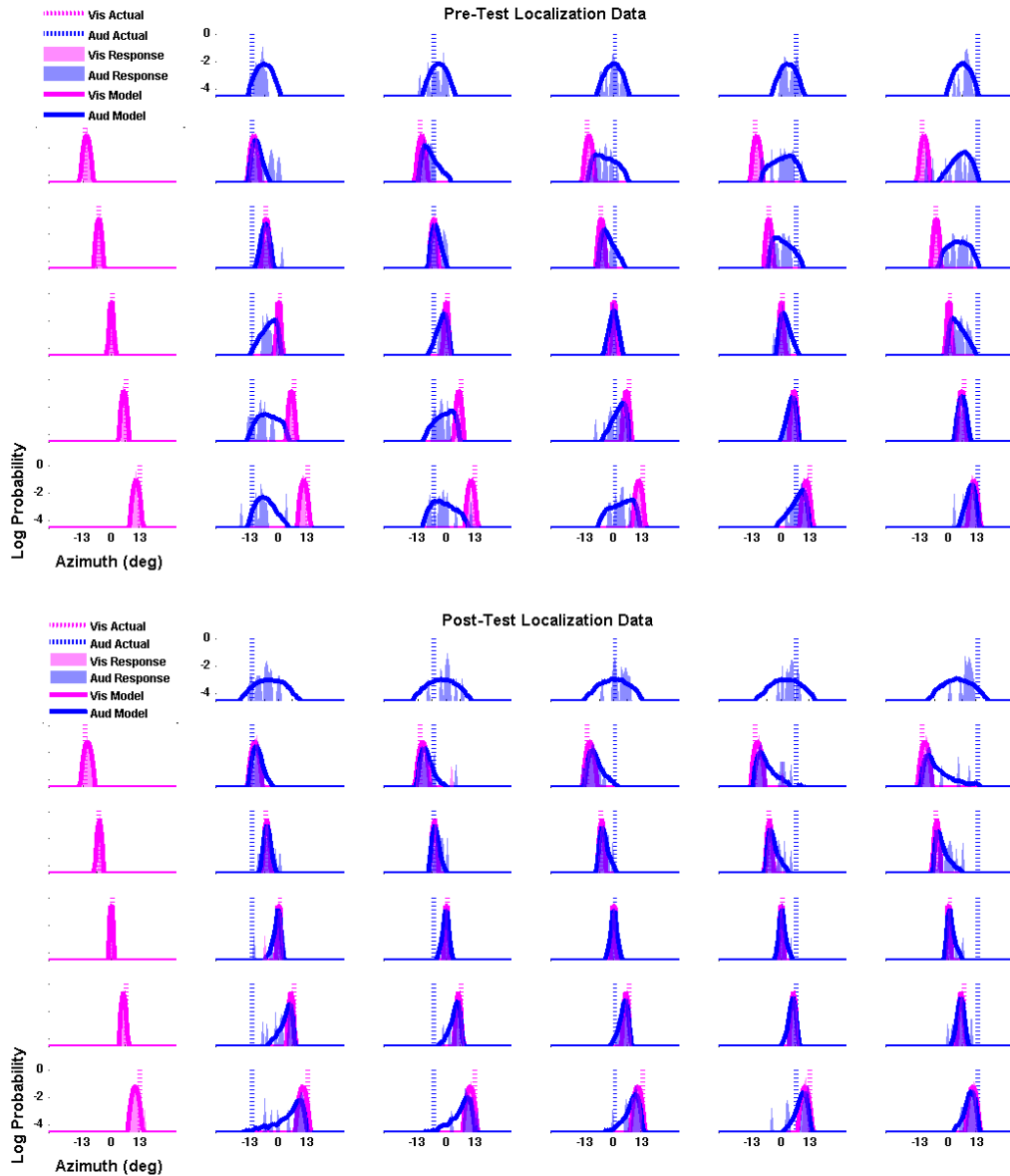
This experiment included the standard pre-test localization phase, and then a post-test session that included exposure-task stimuli that were temporally synchronous, but arose from spatially discrepant locations (see Figure 8B). For instance, during the exposure task, if the visual flash was presented at  $+19.5^\circ$ , the auditory stimulus would be synchronously presented at  $-19.5^\circ$ . Upon a correct detection, these stimuli would switch positions, with each now being present in the position on the opposite side of the midline. Thus, the visual and auditory stimuli in the exposure task were always presented  $39^\circ$  apart. The purpose of this experiment was to investigate how exposure to spatially discrepant stimuli that are presented at the same time impacts subsequent localizations in the experiment.

	Prior Parameters			Likelihood Parameters				
	$p_c^{**}$	$\sigma_p$	$x_p$	$\sigma_v$	$\Delta\sigma_v$	$\Delta x_v^{**}$	$\sigma_a$	$\Delta x_a$
Pre-Test	0.34 ( $\pm .04$ )	19.76 ( $\pm 2.93$ )	-1.30 ( $\pm 1.23$ )	1.20 ( $\pm .08$ )	0.51 ( $\pm .05$ )	-0.44 ( $\pm .08$ )	15.88 ( $\pm 1.73$ )	4.61 ( $\pm 0.90$ )
Post-Test	0.52 ( $\pm .06$ )	24.20 ( $\pm 4.20$ )	-0.70 ( $\pm 1.22$ )	1.23 ( $\pm .09$ )	0.47 ( $\pm .05$ )	-0.60 ( $\pm .08$ )	17.28 ( $\pm 1.76$ )	2.95 ( $\pm 0.86$ )

**Table 8.** Optimized parameter values  $\pm$  standard error parameter estimates from experiment 5. **\*\*** indicates  $p < .01$ .

As shown in Table 8, two important changes to the parameters occurred. First, the prior probability of inferring a common cause significantly increased by 18% in the post-test phase ( $t(28) = -3.203, p = .003$ ). Importantly, this change in binding in subjects was not fully accounted for by increases in unisensory noise, or shifts in the likelihood means; the change was primarily in the prior for integration. Additionally, the delta parameter for visual likelihood was also highly significant ( $t(28) = 3.655, p = .001$ ), indicating that participants' slight central bias in the visual modality in the pre-test phase increased in magnitude in the post-test phase. In other words, participants may have experienced an effect of "tunnel vision" due to the exposure task, as fixation was enforced throughout the exposure task that interleaved localization blocks in the post-test phase, and thus, and their central bias likely increased by two-tenths of a degree due to this factor alone.

The significant increase in the prior probability for a common cause can be seen in the model fits and behavioral data from the pre-test and post-test phase. Shown in Figure 11 on the following page is one representative subject that exhibited a 14% increase in  $P_{\text{Common}}$  from the pre-test to the post-test phase; note the increase in the overlap between the visual and auditory distributions in spatially discrepant bimodal trials, demonstrating the subject's increased tendency to integrate visual and auditory signals.



**Fig. 11.** One representative subject’s data and model fits from the pre-test and post-test localization phases. This subject had a “Pcommon” value of 0.58 in the pre-test phase, which increased by 14% to 0.72 in the post-test phase (this is comparable to the overall average increase of 18%, shown in Table 8). Note the increase in overlap between the visual and auditory distributions in the post-test phase, particularly for the conditions in which the visual and auditory stimuli were presented from spatially disparate locations. This figure visually demonstrates how exposure to spatially discrepant auditory and visual stimuli during the detection task *increased* the tendency to bind sensory signals during the post-test localization phase.

*Experiment #6: Spatially Incongruent & Temporally Correlated Exposure-Task Stimuli (N = 31)*

In this final study, the standard pre-test localization phase was conducted, and then in the post-test phase, the exposure-task stimuli were composed of visual flashes and auditory bursts that were always presented 39° apart. For instance, if the visual stimulus was presented at -19.5°, the auditory stimulus would be presented at +19.5°, and then these stimuli would alternate positions with each correct detection. However, in contrast to experiment #5, the visual stimulus always preceded presentation of the auditory stimulus by 400ms; thus, the stimuli were always temporally incongruent, but were presented with a *consistent temporal correlation*. The purpose of this study was to probe whether temporal *synchrony* was necessary to induce the parameter changes seen in experiment #5, or whether temporal *correlations* could also have the same effect.

	Prior Parameters			Likelihood Parameters				
	$p_c^{**}$	$\sigma_p$	$x_p$	$\sigma_v$	$\Delta\sigma_v$	$\Delta x_v^{**}$	$\sigma_a$	$\Delta x_a$
Pre-Test	0.40 (± .05)	20.04 (± 3.53)	2.31 (± 1.14)	1.20 (± .10)	0.49 (± .04)	-0.46 (± .10)	15.53 (± 1.72)	4.97 (± 0.79)
Post-Test	0.51 (± .05)	26.28 (± 5.17)	1.54 (± 1.12)	1.12 (± .10)	0.52 (± .04)	-0.66 (± .12)	14.26 (± 1.62)	4.04 (± 0.96)

**Table 9.** Optimized parameter values ± standard error parameter estimates from experiment 6. \*\* p < .01

As shown in Table 9, the prior probability of inferring a common cause increased by 11% in the post-test phase, ( $t(30) = -4.503$ ,  $p < .001$ ), indicating that while exposure to synchronous temporal relationships increases the tendency to integrate multisensory signals (exp. #5), temporally-offset, but perfectly *correlated* signals can *also* facilitate a very similar effect. Therefore, it appears that the brain can learn to integrate spatially discrepant pieces of sensory information that would normally be segregated if repeated events with a consistent temporal structure are presented for a brief period of time. Additionally, the delta parameter for the visual likelihood showed an increase in the central bias in the post-test phase ( $t(30) = 3.385$ ,  $p = .002$ ), likely reflecting the same “tunnel vision” phenomenon as commented on previously.

### Summary of $P_{Common}$ Results

The results of the previous experiments demonstrate how subjects' tendencies to integrate audiovisual stimuli are plastic, and can be modified based on previous experience. A summary table of how spatial and temporal factors influence the tendency to bind is shown below.

Experiment	Time	Space	Change in $P_{Common}$
Experiment 1	None	None	<b>None</b>
Experiment 2	Uncorrelated	Incongruent	<b>None</b>
Experiment 3	Congruent	Congruent	<b>None</b>
Experiment 4	Uncorrelated	Congruent	<b>None</b>
Experiment 5	Congruent	Incongruent	<b>Significant Increase</b>
Experiment 6	Correlated	Incongruent	<b>Significant Increase</b>

**Table 10.** Summary of experimental changes in the  $P_{Common}$  parameter, based on the temporal and spatial characteristics of the exposure-task stimuli.

Surprisingly, all of our hypotheses regarding how the prior for integrating stimuli would change based on experience (except for the first experiment) were incorrect. We had hypothesized that repeated presentations of spatiotemporally congruent stimuli (experiment 3) would teach the brain an association between the auditory and visual stimuli, and thus increase the tendency to integrate in the post-test localization phase. Instead, this experiment produced a null result, as the  $P_{Common}$  parameter did not change between the two sessions. Instead, the only conditions which changed  $P_{Common}$  were conditions in which the visual and auditory exposure-task stimuli were (1) spatially discrepant, and (2) presented with a fixed temporal relationship (either synchronous or consistently staggered by 400ms). Interestingly, none of the exposure-task conditions resulted in a *decrease* in  $P_{Common}$ . Even in experiments #2 and #4, where the brain may plausibly have learned to segregate the signals based on the relationships between stimuli,  $P_{Common}$  values were consistent from the pre-test to the post-test phase, leaving the question of how the brain can learn to effectively *segregate* information unanswered.

## Discussion

As noted in a recent review, one of the major unsolved questions in a Bayesian account of perception is how priors change or update over time, as new sensory experience is acquired (Seriès & Seitz, 2013). Here, we demonstrate that our prior tendency to integrate multisensory stimuli can be increased through exposure to audiovisual stimuli with fixed spatial and temporal relationships. Specifically, this prior is increased through not only exposure to spatially disparate/temporally synchronous stimuli, but also spatially disparate/temporally *correlated* stimuli. Surprisingly, this prior is not significantly increased when observers are exposed to spatially congruent/temporally congruent stimuli.

It has been well-documented that perceptual priors in human observers can change (Knill, 2007; Körding & Wolpert, 2004; Miyazaki, Nozaki, & Nakajima, 2005; Sotiropoulos, Seitz, & Seriès, 2011). This principle even applies to priors that govern integration; for instance, subjects can learn to integrate artificially correlated signals that would not normally be integrated, such as luminance and stiffness, with repeated exposure to statistical co-occurrences (Ernst, 2007). Further demonstrating this principle is a recent investigation showing that discrepant visual and vestibular cues may be initially segregated, but with repeated exposure to their presentation, subjects can learn to integrate them (Kaliuzhna, Prsa, Gale, Lee, & Blanke, 2015). This demonstrates the brain's plastic capacity to learn to integrate discrepant pieces of information based on previous sensory experiences.

Additionally, the visual “light-from-above” prior can be modified and applied to novel visual contexts (Adams, Graf, & Ernst, 2004), and additional studies have revealed that this prior can be modified by not only crossmodal feedback, but within-modality feedback as well (Adams, Kerrigan, & Graf, 2010). The authors conclude that one of the driving influences behind



recalibration may be what they deem the “oops factor,” as sequential presentations of similar stimuli may be used to update and refine how an object is represented. In their paradigm, intermittent feedback served to update the object representation. In our paradigm, even though observers are given no feedback, our exposure task may be functioning in a similar manner: repeatedly presenting stimuli outside of the range used in the pre-test or post-test phases may cause the brain update the boundaries it uses to determine which stimuli should be integrated as a single object, and which stimuli should be segregated as separate objects. Temporal factors, including both synchrony and correlations, seem to be the most influential variables involved in updating these spatial boundaries, causing the brain to revise its prior probability of binding even when stimuli in the exposure task were presented from extremely discrepant locations.

These results are quite theoretically important when interpreted in light of recent ideas about how predictive coding may be implemented in the brain. As noted in a recent review, “One influential framework to explain the intricacies of multisensory processing is that of predictive coding. The predictive coding framework states that the brain produces a Bayesian estimate of the environment (Friston, 2010). According to this view, stochastic models of the environment exist somewhere in the brain, which are updated on the basis of processed sensory information. These stochastic models (see Klemen & Chambers, 2012, for a review) thus provide the brain areas lower in the sensory processing hierarchy with predictions (or in Bayesian terms “priors”) that can be used to adjust the processing of ongoing sensory input. A strong mismatch between the prediction and the actual sensory input will then result in a major update of the internal model. For example, when unexpected sensory input is present, our internal model may require updating to deal with this change in representations” (Talsma, 2015, p.2).

Our studies provide evidence for precisely this phenomenon: when repeatedly presented with pieces of sensory information that are normally segregated in one dimension (i.e. space), the brain can use information in another dimension (i.e. time) to learn about a relationship between the signals, and can then use this information to update its internal model of the world. Our studies demonstrate that this extremely plastic capacity of the brain can result in a change of its internal model of the causal structure of the world in an extremely short period of time (less than ten minutes), revealing an adaptive advantage of the system to learn about new associations between sensory signals in the environment, even in adulthood.

While our study of recalibration is based purely upon behavioral data, recent neurophysiological studies have provided some insights regarding how our findings might be implemented at the level of an individual neuron. For instance, in normally-reared adult cats, repeated exposures to audiovisual stimuli that are aligned in space but separated in time results in a merging of the distinct unisensory responses into a fused neural signature similar to what is seen for integrated stimuli, indicating plasticity that results in a new tendency to integrate across time (Yu, Stein, & Rowland, 2009). Another study found that in adult cats reared in darkness, previously naïve neurons could exhibit multisensory integration tendencies with as little as six hours of exposure to spatiotemporally synchronous audiovisual stimuli (Yu, Rowland, & Stein, 2010). According to the authors, “[the results] suggest that multisensory integration is subject to continual modification throughout life . . . though this possibility requires far more experiential support” (Yu et al., 2010, p. 4912). The finding that multisensory integration tendencies in neurons exhibit a high degree of plasticity into adulthood is congruent with the behavioral findings from our current study, as we were able to manipulate adult observers’ tendencies to integrate in a very small time window.

Our study reveals important insights regarding the spatial and temporal factors underlying audiovisual recalibration, and could inform protocols to improve sensory integration in individuals with sensory integration deficits. For instance, the finding that the tendency to integrate increases the most with exposure to temporally synchronous but spatially discrepant audiovisual stimuli indicates that to increase the brain's capacity to integrate, visual and auditory signals should be separated when repeatedly presented. Future studies will be needed to determine how these newly discovered principles can inform protocol to improve integration in clinical populations (e.g. autism, schizophrenia), and should also continue to probe whether these principles are specific to audiovisual integration, or represent general principles that apply to other modality combinations as well.

## References

- Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the “light-from-above” prior. *Nature Neuroscience*, *7*(10), 1057–1058. <http://doi.org/10.1038/nn1312>
- Adams, W. J., Kerrigan, I. S., & Graf, E. W. (2010). Efficient Visual Recalibration from Either Visual or Haptic Feedback: The Importance of Being Wrong. *The Journal of Neuroscience*, *30*(44), 14745–14749. <http://doi.org/10.1523/JNEUROSCI.2749-10.2010>
- Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. <http://doi.org/10.1016/j.cub.2004.01.029>
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*(7), 1391–1397. <http://doi.org/10.1364/JOSAA.20.001391>
- Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*(5), 23. <http://doi.org/10.1167/9.5.23>
- Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482–489. <http://doi.org/10.3758/BF03208826>
- Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension: 1. Evidence from auditory–visual temporal order judgment. *International Journal of Psychophysiology*, *50*(1–2), 147–155. [http://doi.org/10.1016/S0167-8760\(03\)00130-2](http://doi.org/10.1016/S0167-8760(03)00130-2)
- Bertelson, P., Frissen, I., Vroomen, J., & de Gelder, B. (2006). The aftereffects of ventriloquism: patterns of spatial generalization. *Perception & Psychophysics*, *68*(3), 428–436.

- Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*(6), 578–584.  
<http://doi.org/10.3758/BF03207374>
- Canon, L. K. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, *84*(1), 141–147.  
<http://doi.org/10.1037/h0028925>
- Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, & Psychophysics*, *75*(5), 790–811. <http://doi.org/10.3758/s13414-013-0475-4>
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5), 7. <http://doi.org/10.1167/7.5.7>
- Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, *63*(4), 719–725. <http://doi.org/10.3758/BF03194432>
- 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), 1–14. <http://doi.org/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*(3), 315–327.  
[http://doi.org/10.1016/S0001-6918\(03\)00043-X](http://doi.org/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*(1–2), 93–100.  
<http://doi.org/10.1016/j.actpsy.2004.10.004>

- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. <http://doi.org/10.1038/nrn2787>
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778. <http://doi.org/10.1038/nn1268>
- Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual Localization Ability Influences Cross-Modal Bias. *Journal of Cognitive Neuroscience*, *15*(1), 20–29. <http://doi.org/10.1162/089892903321107792>
- Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Perceptual and Motor Skills*, *37*(3), 967–979.
- Kaliuzhna, M., Prsa, M., Gale, S., Lee, S. J., & Blanke, O. (2015). Learning to integrate contradictory multisensory self-motion cue pairings. *Journal of Vision*, *15*(1), 15.1.10. <http://doi.org/10.1167/15.1.10>
- Klemen, J., & Chambers, C. D. (2012). Current perspectives and methods in studying neural mechanisms of multisensory interactions. *Neuroscience & Biobehavioral Reviews*, *36*(1), 111–133. <http://doi.org/10.1016/j.neubiorev.2011.04.015>
- Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of Vision*, *7*(8), 13. <http://doi.org/10.1167/7.8.13>
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, *2*(9), e943. <http://doi.org/10.1371/journal.pone.0000943>
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. <http://doi.org/10.1038/nature02169>

- Lewald, J. (2002). Rapid Adaptation to Auditory-Visual Spatial Disparity. *Learning & Memory*, 9(5), 268–278. <http://doi.org/10.1101/lm.51402>
- Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, 16(3), 468–478. [http://doi.org/10.1016/S0926-6410\(03\)00074-0](http://doi.org/10.1016/S0926-6410(03)00074-0)
- Machulla, T.-K., Luca, M. D., Froehlich, E., & Ernst, M. O. (2012). Multisensory simultaneity recalibration: storage of the aftereffect in the absence of counterevidence. *Experimental Brain Research*, 217(1), 89–97. <http://doi.org/10.1007/s00221-011-2976-5>
- Miyazaki, M., Nozaki, D., & Nakajima, Y. (2005). Testing Bayesian Models of Human Coincidence Timing. *Journal of Neurophysiology*, 94(1), 395–399. <http://doi.org/10.1152/jn.01168.2004>
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, 17(1), 154–163. [http://doi.org/10.1016/S0926-6410\(03\)00089-2](http://doi.org/10.1016/S0926-6410(03)00089-2)
- Parise, C. V., Harrar, V., Ernst, M. O., & Spence, C. (2013). Cross-correlation between auditory and visual signals promotes multisensory integration. *Multisensory Research*, 26(3), 307–316.
- Parise, C. V., Spence, C., & Ernst, M. O. (2012). When Correlation Implies Causation in Multisensory Integration. *Current Biology*, 22(1), 46–49. <http://doi.org/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(2), 137–146. <http://doi.org/10.3758/BF03198746>

- Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. Thomas (1941) revisited. *Psychological Research*, 49(1), 17–22.
- Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences*, 95(3), 869–875.
- Roseboom, W., & Arnold, D. H. (2011). Twice Upon a Time Multiple Concurrent Temporal Recalibrations of Audiovisual Speech. *Psychological Science*, 22(7), 872–877.  
<http://doi.org/10.1177/0956797611413293>
- Sato, Y., & Aihara, K. (2009). Integrative Bayesian model on two opposite types of sensory adaptation. *Artificial Life and Robotics*, 14(2), 289–292. <http://doi.org/10.1007/s10015-009-0675-0>
- Sato, Y., Toyozumi, T., & Aihara, K. (2007). Bayesian Inference Explains Perception of Unity and Ventriloquism Aftereffect: Identification of Common Sources of Audiovisual Stimuli. *Neural Computation*, 19(12), 3335–3355.  
<http://doi.org/10.1162/neco.2007.19.12.3335>
- Seriès, P., & Seitz, A. R. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, 7, 668. <http://doi.org/10.3389/fnhum.2013.00668>
- Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, 14(9), 425–432. <http://doi.org/10.1016/j.tics.2010.07.001>
- Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, 12(1), 7–10.
- Sotiropoulos, G., Seitz, A. R., & Seriès, P. (2011). Changing expectations about speed alters perceived motion direction. *Current Biology*, 21(21), R883–R884.  
<http://doi.org/10.1016/j.cub.2011.09.013>



- Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, 19.  
<http://doi.org/10.3389/fnint.2015.00019>
- Van Wanrooij, M. M., Bremen, P., & John Van Opstal, A. (2010). Acquired prior knowledge modulates audiovisual integration. *European Journal of Neuroscience*, 31(10), 1763–1771. <http://doi.org/10.1111/j.1460-9568.2010.07198.x>
- Vroomen, J., & de Gelder, B. (2004). Temporal Ventriloquism: Sound Modulates the Flash-Lag Effect. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), 513–518. <http://doi.org/10.1037/0096-1523.30.3.513>
- Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, 22(1), 32–35. <http://doi.org/10.1016/j.cogbrainres.2004.07.003>
- Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2), 252–258. <http://doi.org/10.1007/s00221-004-1899-9>
- Wallace, M. T., & Stein, B. E. (2007). Early Experience Determines How the Senses Will Interact. *Journal of Neurophysiology*, 97(1), 921–926.  
<http://doi.org/10.1152/jn.00497.2006>
- Witten, I. B., & Knudsen, E. I. (2005). Why Seeing Is Believing: Merging Auditory and Visual Worlds. *Neuron*, 48(3), 489–496. <http://doi.org/10.1016/j.neuron.2005.10.020>
- Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability Matching as a Computational Strategy Used in Perception. *PLoS Comput Biol*, 6(8), e1000871.  
<http://doi.org/10.1371/journal.pcbi.1000871>

- Wozny, D. R., & Shams, L. (2011a). Computational Characterization of Visually Induced Auditory Spatial Adaptation. *Frontiers in Integrative Neuroscience*, 5.  
<http://doi.org/10.3389/fnint.2011.00075>
- Wozny, D. R., & Shams, L. (2011b). Recalibration of Auditory Space following Milliseconds of Cross-Modal Discrepancy. *The Journal of Neuroscience*, 31(12), 4607–4612.  
<http://doi.org/10.1523/JNEUROSCI.6079-10.2011>
- Xu, J., Yu, L., Rowland, B. A., Stanford, T. R., & Stein, B. E. (2012). Incorporating Cross-Modal Statistics in the Development and Maintenance of Multisensory Integration. *The Journal of Neuroscience*, 32(7), 2287–2298. <http://doi.org/10.1523/JNEUROSCI.4304-11.2012>
- Yamamoto, S., Miyazaki, M., Iwano, T., & Kitazawa, S. (2012). Bayesian calibration of simultaneity in audiovisual temporal order judgments. *PloS One*, 7(7), e40379.  
<http://doi.org/10.1371/journal.pone.0040379>
- Yu, L., Rowland, B. A., & Stein, B. E. (2010). Initiating the Development of Multisensory Integration by Manipulating Sensory Experience. *The Journal of Neuroscience*, 30(14), 4904–4913. <http://doi.org/10.1523/JNEUROSCI.5575-09.2010>
- Yu, L., Stein, B. E., & Rowland, B. A. (2009). Adult Plasticity in Multisensory Neurons: Short-Term Experience-Dependent Changes in the Superior Colliculus. *The Journal of Neuroscience*, 29(50), 15910–15922. <http://doi.org/10.1523/JNEUROSCI.4041-09.2009>

## IX.

### **The effects of selective and divided attention on sensory integration**

#### **Abstract**

In our daily lives, our capacity to selectively attend to stimuli within or across sensory modalities enables enhanced perception of the surrounding world. While previous research on selective attention has studied this phenomenon extensively, two important questions still remain unanswered: (1) how selective attention to a single modality impacts sensory integration processes, and (2) the mechanism by which selective attention improves perception. We explored how selective attention impacts performance in both a spatial task and a numerosity judgment task, and employed a Bayesian Causal Inference model to investigate the computational mechanism(s) impacted by selective attention. Here, we report three interesting findings: in the spatial domain, selective attention improves precision of the visual sensory representations, but not the auditory sensory representations; in the temporal domain, selective attention improves the sensory representations in both modalities; in both tasks, selective attention exerts minimal influence over the tendency to integrate sensory stimuli. We conclude with a discussion of how these results relate to recent theoretical considerations regarding the topic of selective attention.

## **Introduction**

In our daily lives, our capacity to selectively attend to information from a single sensory channel is extremely important as we attempt to accurately process information from the surrounding world. For instance, in order to effectively read and comprehend passages in a book, one needs to allocate attentional resources exclusively towards processing the visual information on the page. However, if we want to listen to a lecture in podcast format and fully comprehend what is being discussed, we will need to exclusively attend to the auditory information, at the expense of sensory stimuli in other modalities. This process of selectively attending to a single sensory modality is critical for being able to quickly and effectively navigate a busy world in which important information could come from different sensory channels at any given time.

Previous research indicates that selective attention improves processing in the attended modality. Behaviorally, selective attention to a single sensory modality has been shown to improve sensory discriminations in the attended modality (Bonnell & Haftser, 1998), decrease reaction time to targets (Spence & Driver, 1997), and improve spatial discrimination (left vs. right) judgments (Spence, Nicholls, & Driver, 2001). Neuroimaging studies indicate that selective attention to either visual or auditory stimuli in multisensory environments can increase activity in the corresponding sensory cortices, while dividing attention across those two modalities results in only a slight, simultaneous activation of both brain regions (Jennifer Adrienne Johnson & Zatorre, 2006; Jennifer A. Johnson & Zatorre, 2005; Laurienti et al., 2002; Loose, Kaufmann, Auer, & Lange, 2003; Shomstein & Yantis, 2004). This general idea is consistent with several ERP studies indicating that the effect of selective attention to one type of sensory input is to enhance activity in the applicable cortical area (Alho, Woods, & Algazi, 1994;

Eimer & Schröger, 1998; Hötting, Rösler, & Röder, 2003). Thus, studies indicate that processing is improved for the attended modality, but the *mechanism* involved remains unclear.

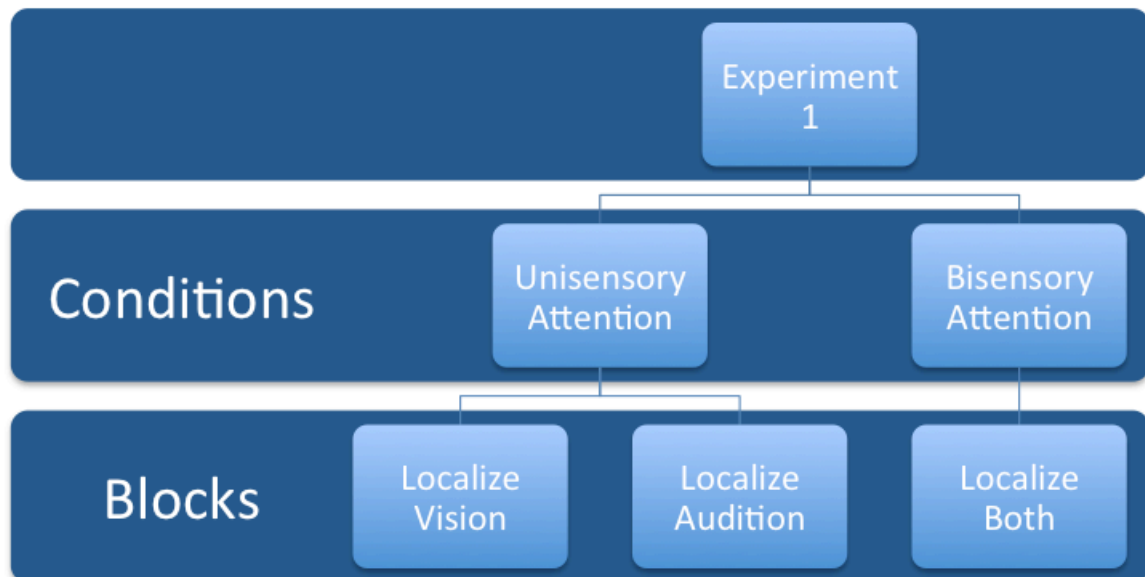
Computationally, models assuming optimal Bayesian integration of sensory cues have accurately captured human observer's performance on a number of multisensory tasks (Alais & Burr, 2004; Ernst & Banks, 2002; Körding et al., 2007). However, as noted in a recent review, "the success with which Bayesian . . . models can describe the interaction between attention and multisensory integration remains yet to be answered" (Talsma, 2015, p.3). Therefore, we aim to provide insight into *how* selective attention exerts its beneficial affects in a Bayesian framework by employing a Bayesian Causal Inference model (Beierholm, Quartz, & Shams, 2009; Körding et al., 2007; Rohe & Noppeney, 2015; Samad, Chung, & Shams, 2015; Wozny, Beierholm, & Shams, 2010; Wozny & Shams, 2011) and comparing conditions of selective and divided attention. Because the effect of attention could potentially differ in separate modalities, tasks, or domains, we explore these questions systematically by implementing both a spatial task and a numerosity task, and testing how attention to the visual or auditory modality alone differs from conditions where attention is allocated to both modalities at the same time.

Most previous studies investigating selective attention indicate that it improves processing of an attended feature (Ball & Sekuler, 1980; Corbetta & Shulman, 2002; Doshier & Lu, 2000; Motter, 1994; Posner, Snyder, & Davidson, 1980). However, this could be due to improving the sensory representations (reducing noise), or due to improving expectations about when and where things will occur in the environment. Using our computational model, we aim to establish whether selective attention exerts effects on the sensory representations or *a priori* expectations by quantitatively estimating both of these components in each observer in each task.

Finally, while the question of attention's impact on integration has been explored extensively by previous research and thoroughly discussed in several recent reviews (Koelewijn, Bronkhorst, & Theeuwes, 2010; Talsma, 2015; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010), studies investigating the question of how (or if) attention can influence the integration of sensory signals have yielded heterogeneous results. For instance, depending on the paradigm, it has been shown that selective attention does not influence integration (Bertelson, Vroomen, Gelder, & Driver, 2000; Shore & Simic, 2005; Vroomen, Bertelson, & Gelder, 2001), increases integration (Alsius, Navarra, Campbell, & Soto-Faraco, 2005), or even reduces integration (Hugenschmidt, Mozolic, & Laurienti, 2009; Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2007; Werkhoven, Erp, & Philippi, 2009). One of the main problems with some of the previous studies examining this question is that the measure of integration is confounded with unisensory processing; therefore, a change in unisensory processing (improved reliability, for example) could result in a change in interaction between the two modalities and be misinterpreted as a change in integration. Our Bayesian model provides a measure of integration tendency that is not confounded by unisensory precision (or noise), and therefore can provide a clearer picture of whether attention influences unisensory precision, multisensory integration or both. Therefore, utilizing the causal inference model, we quantitatively estimated the integration tendency for each individual subject in both selective and divided attention conditions, and in both spatial and temporal tasks to address this question more rigorously.

## Experiment #1: Single Task vs. Dual Task

The conditions in experiment 1 consisted of blocks requiring selective attention (i.e., “unisensory attention”), and blocks requiring divided attention (i.e., “bisensory attention”). While the audiovisual environment was the same for all blocks, the unisensory attention condition involved blocks in which participants localized either *only* the auditory stimulus, or *only* the visual stimulus. The bisensory attention blocks required them to localize both. Thus, our unisensory attention blocks were single-task, requiring only one response, and our bisensory attention blocks were dual-task, requiring responses to both stimuli.



**Fig. 12.** Design schematic depicting the conditions and blocks for experiment 1. As shown above, while the bisensory attention condition only included one block type that required localization of both auditory and visual stimuli, the unisensory attention condition was made up of two different block types, with each block requiring only the localization of stimuli from one sensory modality.

## Methods

### *Participants and Apparatus*

Twenty-five research volunteers at the University of California-Los Angeles (eight males) completed the experiment; one subject was excluded due to negligence with the response device during the task. Potential participants were screened before the experiment and ruled

ineligible to participate if they had a history of any neurological conditions (e.g. seizures, epilepsy, or stroke), had uncorrected vision or hearing impairments, or had experienced significant head trauma.

Eligible participants sat at a desk in a dimly lit room with their chins positioned on a chinrest 52 cm from a projection screen. The screen was a black, acoustically transparent cloth subtending much of the visual field ( $134^\circ$  width  $\times$   $60^\circ$  height). Behind the screen were 5 free-field speakers (5 x 8 cm, extended range paper cone), positioned along azimuth  $6.5^\circ$  apart,  $7^\circ$  below fixation. The middle speaker was positioned below the fixation point, and two speakers were positioned to the right and two to the left of fixation. The visual stimuli were presented overhead from a ceiling mounted projector set to a resolution of 1280 x 1024 pixels with a refresh rate of 75 Hz.

### *Stimuli*

The visual stimulus was a white-noise disk (.41 cd/m<sup>2</sup>) with a Gaussian envelope of  $1.5^\circ$  FWHM, presented  $7^\circ$  below the fixation point on a black background (.07cd/m<sup>2</sup>), for 35 ms. The center of visual stimuli overlapped the center of one of the five speakers behind the screen positioned at  $-13^\circ$ ,  $-6.5^\circ$ ,  $0^\circ$ ,  $6.5^\circ$ , and  $13^\circ$ . 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.

### *Eyetracker and Response Mechanism*

Prior to the presentation of stimuli, participants were required to have their gaze centered on the fixation point. To ensure that participants' gaze for each trial was starting from the same location, gaze position and fixation time were recorded at 60Hz with a ViewPoint eye tracker



(Arrington Research, Scottsdale, AZ) and PC-60 software (version 2.8.5,000). Stimuli were not displayed until the recorded gaze angle was within  $1.5^\circ$  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.

### *Procedure*

In order to familiarize participants with the task, each session in experiment 1 started with three different practice periods of auditory-only trials, visual-only trials, and interleaved auditory-only, visual-only, and bisensory trials. For bisensory trials, subjects were told that “the sound and light could come from the same location, or they could come from different locations.” As a reminder, a red ‘S’ or a blue ‘L’ was placed inside the cursor to remind subjects to respond to the sound or light, respectively. Each trial started with the fixation cross, followed after 750 ms (if the subject was fixating properly) by the presentation of the stimuli. 450 ms after the stimuli, fixation was removed and a cursor appeared on the screen vertically just above the horizontal line where the stimuli were presented 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. 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.

The practice session was followed by 1425 test trials lasted for approximately 90 minutes. The stimulus conditions included five unisensory auditory locations, five unisensory visual locations, and all 25 combinations of auditory and visual locations (bisensory conditions). Three different blocks were implemented three times each in the experiment in a Latin-square design, and in a given block, participants were given one of three possible instructions: localize only the auditory stimulus, localize only the visual stimulus, or localize *both* the auditory and visual stimulus. In bisensory attention blocks, subjects were asked to report both the location of the auditory stimulus and the location of the visual stimulus in sequential order. The order of these two responses was consistent throughout the session, and was counter-balanced across subjects. After the completion of all 1425 trials, analysis using the Bayesian Causal Inference model was performed.

## **Model**

For the spatial tasks, we used a Bayesian causal observer model with eight free parameters to model participants' data from both the unisensory and bisensory attention conditions (see Chapters 1 and 2). The parameters included the prior probability of a common cause ( $p_c$ ), the uncertainty of vision ( $\sigma_v$ ), the uncertainty of audition ( $\sigma_A$ ), the bias in the visual sensory likelihood ( $\Delta x_v$ ), the bias in the auditory sensory likelihood ( $\Delta x_A$ ), how the visual likelihood variance changes with eccentricity ( $\Delta\sigma_v$ ), and the mean and variance of the participant's prior bias for localizing stimuli towards the central location ( $x_p, \sigma_p$ ). Model comparisons between this eight-parameter model and other versions with fewer parameters (i.e. Körding et al., 2007) have shown that this model is superior both in terms of log-likelihood and BIC values (see Chapter 1).

For the numerosity task (experiment 3), we used a simpler version of the model with only four free parameters ( $p_c$ ,  $\sigma_v$ ,  $\sigma_A$ ,  $\sigma_p$ ), since the spatial biases motivating additional parameters in this Bayesian model have not been demonstrated in the numerical domain. Utilizing this model, we assumed that the bias for localizing stimuli towards the central location was normally distributed with a mean at  $\mu_p$  (1.38, based on results from Wozny, Beierholm, and Shams, 2008), and a standard deviation,  $\sigma_p$ , that captured the strength of this prior. We modeled the likelihood terms with Gaussians centered at  $\mu_v$  and  $\mu_a$  (the true stimulus locations on a given trial), and standard deviations that captured the strength of the likelihoods ( $\sigma_v$  and  $\sigma_a$ ).

In all experiments, each participant's data was modeled using the best-fitting decision strategy (see Chapter 2). For each subject, free parameters (either eight free parameters for the spatial task, or four free parameters for the numerosity task) were fit to the data. This was done simulating 10000 trials using MATLAB's *fminsearchbnd* function (Mathworks, 2006), maximizing the likelihood of the parameters of the model. Participants' overall datasets were divided into trials that from blocks that required *unisensory* attention (i.e. trials where a response was only required from one modality) and trials that required *bisensory* attention (where a response was required for both modalities), and separate parameter fits were obtained for each of these datasets.

## Results

After the optimized parameters were fit to individual subjects' data for the unisensory attention and bisensory attention datasets, we compared the model parameters between the two attention conditions. Two-tailed paired-samples t-tests were performed to determine whether there were any significant differences in parameters between the unisensory attention condition and the bisensory attention condition. The results for all parameters are summarized in Table 11.

	$P_c$	$\sigma_v^{***}$	$\sigma_A$	$\Delta x_v$	$\Delta x_A$	$\Delta \sigma_v$	$x_p$	$\sigma_p$
Unisensory Attention	.41	.91	11.68	-.50	3.65	.58	0.01	25.25
Bisensory Attention	.35	1.23	11.65	-.58	3.15	.60	-1.96	26.03

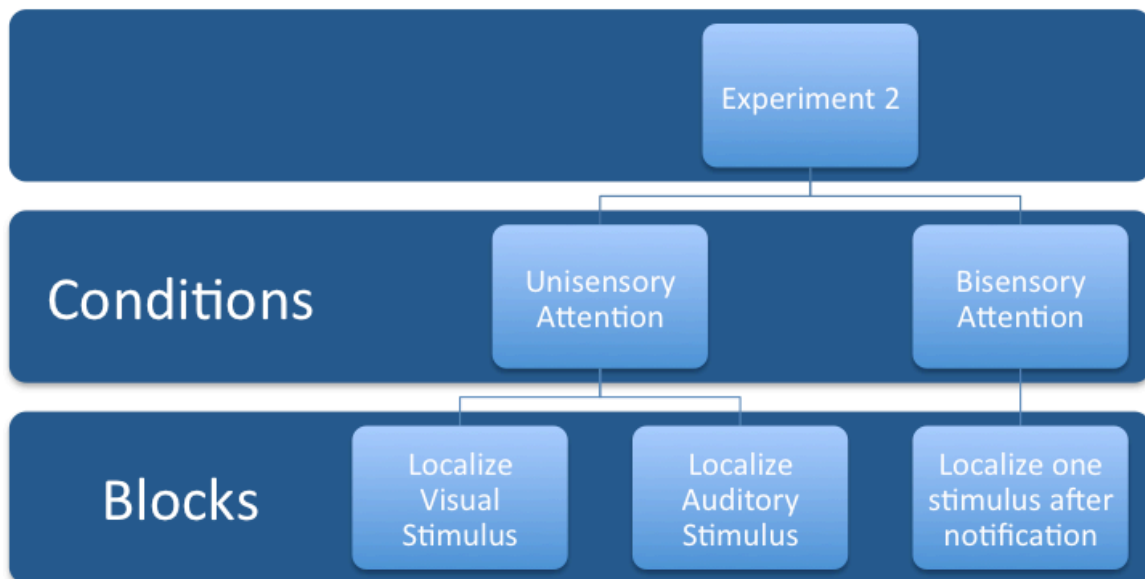
**Table 11.** Optimized parameter values from experiment 1. \*\*\* indicates  $p < .001$ .

The prior probability for a common cause in the unisensory attention condition was found to be slightly larger than in the bisensory condition; however, this difference was not significant ( $t(24) = 1.518, p > .05$ ), indicating that the tendency to bind is not significantly altered due to different allocations of attention. In regards to the question of whether attention can affect the precision of sensory representations, the standard deviation of the visual likelihood ( $\sigma_v$ ) in the unisensory attention condition was found to be significantly smaller than that of the bisensory attention condition ( $t(24) = -4.161, p < .001$ , two-tailed paired t-test). However, a similar trend was not found in the auditory domain, as the standard deviation of the auditory likelihood ( $\sigma_a$ ) in the unisensory attention condition was nearly identical to that of the bisensory attention condition (11.68 compared to 11.65, respectively,  $t(24) = .061, p > .05$ ).

In summary, experiment 1 indicated that while selective attention to the visual modality significantly increased precision in the visual domain, selective attention to the auditory modality did *not* increase sensory precision in the auditory domain. Results also showed that under conditions requiring attention to a single sensory modality, the tendency to integrate stimuli was not significantly altered, compared to a condition requiring attention to two sensory modalities.

## Experiment #2: Single Response Only

While the findings from experiment 1 provided preliminary evidence for selective attention improving the precision in visual sensory representations, there remained the possibility that requiring the localization of *one* stimulus in the unisensory attention condition and *two* stimuli in the bisensory attention condition could be driving the observed effects in the visual likelihood distribution, due to differences in the demands required by single and dual tasks. Thus, a second experiment was conducted to eliminate these differences in memory requirements by making a small change in the bisensory attention condition: participants would now be required to pay attention to both stimuli as they occur, but would be notified immediately afterwards as to which stimulus (visual or auditory) they were to localize. Therefore, in both attention conditions in experiment 2, participants were only required to localize *one* stimulus in any given block.



**Fig. 13.** Design schematic for experiment 2. The conditions and blocks for experiment 2 were quite similar to the first experiment, but now in the bisensory attention condition, participants were only required to localize *one* stimulus, and were notified of which stimulus they were to report when the cursor popped up on the screen.

## Methods

Twenty-eight students and research volunteers at the University of California-Los Angeles completed the experiment, and twenty-six participants were included in the data analysis (two participants had extremely inaccurate responses indicating negligence while performing the task). All participants were naïve to the experimental tasks, and the setup, design, and implementation was the same as in experiment 1, with one exception: since the bimodal blocks now only required *one* response per trial, additional trials were included to gather sufficient responses to both visual and auditory stimuli in that condition. Thus, 1950 trials were included in the overall experiment.

## Model

As in experiment 1, a Bayesian causal observer model was implemented to model participant data, and changes in parameter values were compared between the unisensory and bisensory attention conditions.

## Results

Parameters were fit to the individual subject's data separately for the unisensory attention and bisensory attention datasets. T-tests for the parameter fits for both the unisensory and bisensory data are shown below.

	$P_c$	$\sigma_v^{**}$	$\sigma_A$	$\Delta x_v$	$\Delta x_A$	$\Delta \sigma_v$	$x_p$	$\sigma_p$
Unisensory Attention	0.45	1.15	12.26	-0.55	2.60	0.48	-0.13	26.33
Bisensory Attention	0.44	1.32	11.06	-0.68	3.39	0.55	-0.27	17.73

**Table 12.** Optimized parameter values from experiment 2. \*\* indicates  $p < .01$ .

The standard deviation of the visual likelihood ( $\sigma_v$ ) in the unisensory attention condition was again found to be significantly smaller than that of the bisensory attention condition ( $t(25) = -3.16, p = p.005$ ). This finding was strong enough to survive a Bonferroni Correction for the

eight t-tests performed on this dataset. A similar trend was not found in the auditory domain, as the standard deviation of the auditory likelihood ( $\sigma_a$ ) in the unisensory attention condition was slightly larger to that of the bisensory attention condition (12.26 compared to 11.06, respectively); this difference was not significant, due to the large variability in subjects' auditory likelihood standard deviation fits. Thus, it appears that vision, a sense that is already highly adept at processing spatial information, receives an extra boost from selective attention, but audition, which is much less precise in terms of spatial information, receives no additional benefit.

A few other trends emerged that, while interesting, did not reach significance at the required Bonferroni-corrected p-value. Visual localizations often show a localization towards the center of visual space (e.g. Adam, Davelaar, Gouw, & Willems, 2008; Fortenbaugh & Robertson, 2011), and our estimates of the mean of the visual likelihood distribution show that the amount of this bias *decreases* with selective visual attention (Unisensory attention:  $-0.55^\circ$ , Bisensory attention:  $-0.67^\circ$ ), indicating that selective attention may make the visual sensory representation more accurate. (A similar trend existed in the auditory domain, but the magnitude of the effect was even smaller). Additionally, bisensory attention appeared to increase reliance on the central prior, indicating that as attentional resources are spread across modalities, the brain may rely in preexisting information to influence the final sensory estimates.

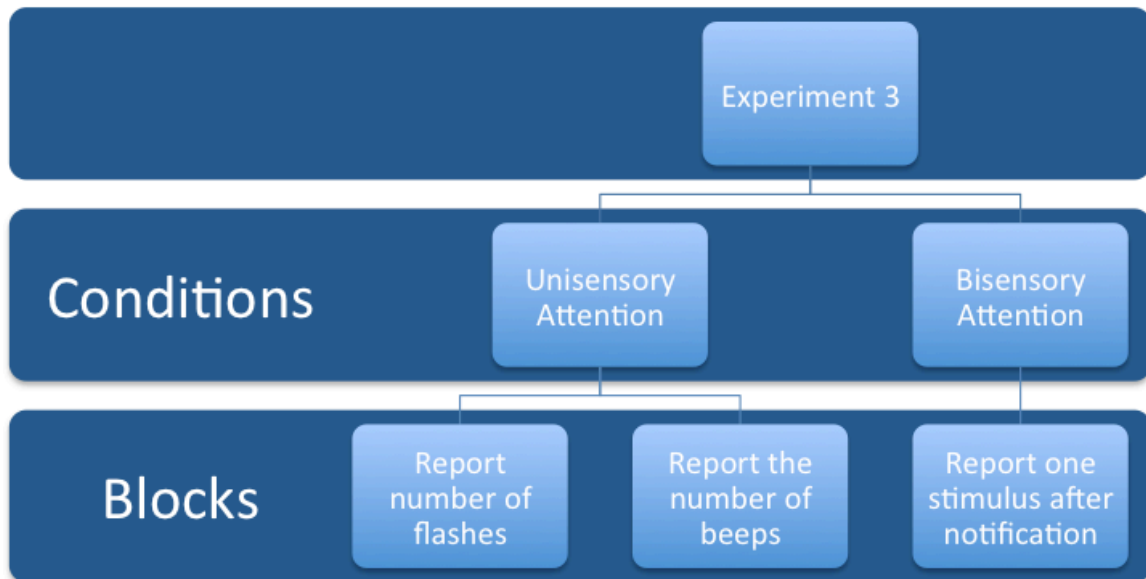
Finally, it is important to note that the prior for integrating sensory information, known as the “prior probability of a common cause” or “ $p_c$ ”, was not significantly different between the two conditions. The values between the two conditions were nearly identical, indicating that selective attention to a single sensory modality does not impact the tendency to integrate multisensory information.

In summary, experiment 2 replicated the findings from experiment 1, and ruled out the possibility that the initial findings were due to differences in the demands caused by localizing either one or two objects. In both experiments, selective attention significantly increased precision in the visual domain, did not increase sensory precision in the auditory domain, and did not significantly change the prior probability of integrating stimuli across space.



### Experiment #3: Beep-Flash Numerosity Judgment Task

While experiments 1 and 2 examined the effects of selective attention on sensory integration along a *spatial* dimension, experiment 3 was conducted to determine whether or not selective attention to a single sensory modality could affect integration along a *temporal* dimension. Similar to experiment 2, subjects participated in conditions involving unisensory and bisensory attention; in the unisensory attention condition, participants knew ahead of time which modality they needed to report. In the bisensory attention condition, participants did not know which modality to report until after the stimuli were presented, and thus had to pay attention to both modalities throughout the task.



**Figure 14.** Design schematic for experiment 3. Similar to experiment 2, participants in the numerosity judgment paradigm were either reporting the *same stimulus* for every trial in a block (the unisensory attention condition), or had to wait until after stimuli were presented before they knew which stimulus had to be reported (the bisensory attention condition).

### Methods

Twenty-five subjects from the University of California-Los Angeles completed the experiment, but only twenty-four were included in data analysis (one had extremely variable responses indicating negligence during the task). Participants were screened before the

experiment according to the eligibility criteria listed in experiment 1. Eligible subjects were seated in a chinrest 57cm away from a CRT monitor, which was flanked on each side by two speakers. Before each trial in the experiment, subjects were required to have their eyes fixated on a white cross, displayed at the center of the computer monitor, and the same eyetracking method as described in previous experiments was used. Visual stimuli consisted of a white disc flashed for one frame, which was approximately 10 ms. The flash was presented 7° below the fixation point on an otherwise dark monitor. Auditory stimuli were ramped noise bursts of similar duration.

In order to familiarize participants with the task, each session started with practice trials requiring unisensory visual attention, unisensory auditory attention, and bisensory attention to both modalities. In the unisensory visual attention trials, subjects were presented with the instruction, “Report the number of flashes you see.” Following the presentation of these instructions for 1800 milliseconds, a fixation cross appeared on the screen, and once participants fixated on the cross for longer than 250 milliseconds, stimuli were displayed. After the stimuli were presented, subjects had two seconds to enter their response on keyboard placed on the table in front of them, responding with a number from 1-4. It is important to note that in the unisensory attention blocks, participants could be presented with either unimodal or bimodal stimuli, but were always consistently reporting one modality throughout a block. In the unisensory auditory attention trials, the protocol was identical to the unisensory visual attention trials, but the instruction presented before each trial read, “Report the number of beeps you hear.” The bisensory attention trials proceeded as follows: first, the fixation cross was presented at the center of the screen. Once participants had been fixated on the cross for more than 250 milliseconds, stimuli were displayed. Following the stimulus presentation, instructions appeared

on the screen, requiring participants to report either the number of flashes, or the number of beeps. From the time instructions appeared, participants had 2 seconds to make their response. To cut down on time lost while reading the instructions, the modality to be reported was listed in capital letters, so after the first few presentations of the instructions, most participants simply looked for the capitalized word.

The practice session was followed by 1200 trials that consisted of two unisensory attention blocks (one visual, one auditory), and two bisensory attention blocks, where reporting flashes and reporting beeps was interleaved. The presentation of the blocks followed a Latin-square design, with each subject receiving a unique ordering of the blocks. It is important to note that in the unisensory visual block, there was always at least 1 flash presented on every trial, and in the unisensory auditory block, at least one beep was always presented. In the bisensory block, participants were never asked to provide a report for a modality that was not presented with any relevant stimuli.

## **Model**

As in experiments 1 and 2, a Bayesian causal observer model was fit to participant data. Two important changes to the model were implemented: first, the mean for the prior  $\mu_p$  was now fixed at 1.38, based upon the optimal value found by Wozny, Beierholm and Shams (2008) in a numerosity judgment task. Next, the number of parameters was reduced from eight to four, as the parameters that were included to capture systematic biases in spatial localization ( $\Delta x_v$ ,  $\Delta x_A$ ,  $\Delta \sigma_v$ ) were omitted.

## Results

As in previous experiments, parameters were fit to the individual subjects' data separately for the unisensory attention and bisensory attention datasets.

	$P_c$	$\sigma_v^{**}$	$\sigma_A^{**}$	$\sigma_p$
Unisensory Attention	0.52	0.85	0.38	0.93
Bisensory Attention	0.48	1.01	0.44	0.95

**Table 13.** Optimized parameter values from experiment 3. \*\* indicates  $p < .01$ .

The trend from previous experiments for the visual likelihood to improve with unisensory attention was preserved, with the unisensory attention condition exhibiting an increase in sensory precision compared to the bisensory condition:  $t(23) = -2.954$ ,  $p < .01$ . Interestingly, the auditory likelihood distribution also exhibited a significant increase in precision in the unisensory attention condition, yielding  $t(23) = -3.378$ ,  $p < .01$ . The slight increase in the  $p_c$  parameter was not significant, with the paired-samples t-test finding  $t(23) = 0.894$ ,  $p > .05$ . Finally, the difference in  $\sigma_p$  between the two conditions was also insignificant:  $t(23) = -.269$ ,  $p > .05$ .

Taken together with the results from experiments 1 and 2, these results highlight some surprising findings: selective attention to the visual modality increases sensory precision in both the spatial and temporal domains, but selective attention to the auditory modality only sharpens sensory representations in the temporal domain.

## Discussion

Many studies have investigated the role of attention in multisensory integration; the results have been mixed, no clear picture has emerged and as a result hypotheses involving complex relationships have been proposed (Koelewijn, Brokhorst, & Theeuwes, 2010; Martens, Kandula, & Duncan, 2010; Talsma, Senkowski, Soto-Faraco, Woldorff, & 2010). Therefore, the question of how selective attention influences sensory integration has yet to be illuminated. Even more, studies have investigated the role of attention on visual and auditory perception, but again a clear computational characterization of the effect of attention on perceptual processing has proved elusive.

In our first two experiments, we tested the effects of selective attention to a particular sensory modality using a localization task. Experiment 1 assessed the differences between unisensory attention and bisensory attention by asking participants for localization of a single stimulus in the unisensory condition, or localization of stimuli in two modalities in the bisensory condition. In experiment 2, we always required participants to localize only one stimulus; in the unisensory attention condition they knew beforehand which modality was relevant, and in the bisensory attention condition they did not, and thus needed to pay attention to both.

Consistent with the idea of attention as a mechanism that reduces noise and spatial uncertainty (Serences, 2011; Whiteley & Sahani, 2012) and consistent with previous findings showing benefits for an attended sensory modality (Bonnell & Haftser, 1998; Spence & Driver, 1997; Spence et al., 2001), results from the first two experiments showed significant increases in the precision of the visual likelihood representations in the unisensory attention condition.

Interestingly, selective attention did *not* enhance auditory representations, which opens the question of how effectively the brain can select auditory inputs for enhanced spatial processing

when visual information is still accessible (i.e. when the eyes are open, when a cursor is visible, etc.).

The first two experiments also found that in a localization task, selective attention did *not* change the tendency to integrate stimuli; a weak trend towards increased integration was found in the first study, but nearly identical priors for integration were recorded in both attention conditions in the second experiment. Pre-existing literature on this topic is quite mixed, as previous studies investigating whether attention influences multisensory integration have found conflicting results. Some investigations have found that attention has no effect on multisensory integration (Bertelson, et al., 2000; Shore & Simic, 2005). In studies investigating the integration of facial and vocal information, a few findings indicate that attention does influence integration. For instance, Alsius, Navarra, Campbell, & Soto-Faraco (2005) found that performing a secondary visual or auditory task reduced the prevalence of the McGurk effect, indicating that integration processes break down under conditions that require attention to multiple sources of information at once. Additionally, Massaro (1998) investigated facial expressions with both visual and auditory components and found increased influence of the sensory modality to which attention was directed.

Contrary to these findings, several previous studies have indicated that the effect of selective attention to a particular modality is to *reduce* multisensory integration, when integration is assessed by analyzing reaction times (Hugenschmidt, Mozolic, & Laurienti, 2009; Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2007; Talsma, Doty, & Woldorff, 2006). In these studies, responses to multisensory trials were compared to responses to unisensory trials as predicted by the independent race model (Miller, 1982, 1986). Other studies have used ERP analyses, where ERP waves produced in response to audiovisual objects are compared to the sum

of the waves when the auditory and visual components were presented alone. We find that, at least in the spatial domain, the brain's pre-existing prior for integration is *not* significantly impacted by the allocation of modality-specific attention. While this finding does not exclude the possibility that other types of feature-based attention could impact attention (i.e. object-based or spatial attention), it does provide some reliable evidence that integration for low-level sensory features may operate independently of attention, similar to what has been reported in the spatial domain previously.

From a theoretical perspective, our results inform current ideas about attention in a Bayesian framework (Talsma, 2015; Talsma et al., 2010). Previous accounts of attention using Bayesian modeling have focused on attention serving the role as an additional prior to enhance selective perception (Dayan & Zemel, 1999; Yu & Dayan, 2004). A more recent account has focused on how attention may impact perception in environments, where inference problems may be computationally intractable. In this framework, rather than simply serving as a prior, attention serve as a mechanism to refine perceptual accuracy by approximating the inference to be performed (Whiteley & Sahani, 2012). Here, we used a simple task to determine which element of a Bayesian model is impacted by modality-specific attention in a spatial and a temporal task. We did not find evidence that attention impacts the integration process in any way, but we did find that it enhances the precision of sensory representations of the visual modality in both the spatial and temporal domains, and the auditory modality exclusively in the temporal domain. Additionally, we found that auditory spatial representations were *not* enhanced, which introduces the question of how and when auditory spatial processes may be impacted by attention, if at all.

Finally, future research will need to investigate the question of additional factors influencing integration and segregation under conditions of selective and divided attention. For instance, semantic congruency likely plays a role in how information is either integrated or segregated in the brain. As we read a book, we may find it relatively easy to filter out distracting auditory information that is incongruent with the words on the page, but listening to a lecture on headphones while encountering an environment with large amounts of visual information may make it more difficult to process what we hear at the same moment. Additional experiments should continue to investigate how and when attention can influence integration or segregation of higher-level features in more realistic settings, and also probe how its allocation and deployment can be more effectively cultivated and implemented.



## References

- Adam, J. J., Davelaar, E. J., Gouw, A. van der, & Willems, P. (2008). Evidence for attentional processing in spatial localization. *Psychological Research*, *72*(4), 433–442.  
<http://doi.org/10.1007/s00426-007-0126-2>
- Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. <http://doi.org/10.1016/j.cub.2004.01.029>
- Alho, K., Woods, D. L., & Algazi, A. (1994). Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology*, *31*(5), 469–479. <http://doi.org/10.1111/j.1469-8986.1994.tb01050.x>
- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual Integration of Speech Falters under High Attention Demands. *Current Biology*, *15*(9), 839–843.  
<http://doi.org/10.1016/j.cub.2005.03.046>
- Ball, K., & Sekuler, R. (1980). Models of Stimulus Uncertainty in Motion Perception. *Psychological Review*, *87*(5). Retrieved from  
<http://search.proquest.com/docview/1290901538?accountid=14512>
- Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*(5), 23.  
<http://doi.org/10.1167/9.5.23>
- Bertelson, P., Vroomen, J., Gelder, B. D., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*(2), 321–332. <http://doi.org/10.3758/BF03205552>

- Bonnel, A.-M., & Haftser, E. R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception & Psychophysics*, *60*(2), 179–190.  
<http://doi.org/10.3758/BF03206027>
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. <http://doi.org/10.1038/nrn755>
- Dayan, P., & Zemel, R. S. (1999). Statistical models and sensory attention. In *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470)* (Vol. 2, pp. 1017–1022 vol.2). <http://doi.org/10.1049/cp:19991246>
- Doshier, B. A., & Lu, Z.-L. (2000). Noise Exclusion in Spatial Attention. *Psychological Science*, *11*(2), 139–146. <http://doi.org/10.1111/1467-9280.00229>
- Eimer, M., & Schröger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology*, *35*(03), 313–327.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. <http://doi.org/10.1038/415429a>
- Fortenbaugh, F. C., & Robertson, L. C. (2011). When here becomes there: attentional distribution modulates foveal bias in peripheral localization. *Attention, Perception, & Psychophysics*, *73*(3), 809–828. <http://doi.org/10.3758/s13414-010-0075-5>
- Hötting, K., Rösler, F., & Röder, B. (2003). Crossmodal and intermodal attention modulate event-related brain potentials to tactile and auditory stimuli. *Experimental Brain Research*, *148*(1), 26–37. <http://doi.org/10.1007/s00221-002-1261-z>
- Hugenschmidt, C. E., Mozolic, J. L., & Laurienti, P. J. (2009). Suppression of multisensory integration by modality-specific attention in aging. *Neuroreport*, *20*(4), 349–353.  
<http://doi.org/10.1097/WNR.0b013e328323ab07>

- Johnson, J. A., & Zatorre, R. J. (2005). Attention to Simultaneous Unrelated Auditory and Visual Events: Behavioral and Neural Correlates. *Cerebral Cortex*, *15*(10), 1609–1620.  
<http://doi.org/10.1093/cercor/bhi039>
- Johnson, J. A., & Zatorre, R. J. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *NeuroImage*, *31*(4), 1673–1681.  
<http://doi.org/10.1016/j.neuroimage.2006.02.026>
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372–384. <http://doi.org/10.1016/j.actpsy.2010.03.010>
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, *2*(9), e943.  
<http://doi.org/10.1371/journal.pone.0000943>
- Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y.-F., Field, A. S., & Stein, B. E. (2002). Deactivation of Sensory-Specific Cortex by Cross-Modal Stimuli. *Journal of Cognitive Neuroscience*, *14*(3), 420–429. <http://doi.org/10.1162/089892902317361930>
- Loose, R., Kaufmann, C., Auer, D. P., & Lange, K. W. (2003). Human prefrontal and sensory cortical activity during divided attention tasks. *Human Brain Mapping*, *18*(4), 249–259.  
<http://doi.org/10.1002/hbm.10082>
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. [http://doi.org/10.1016/0010-0285\(82\)90010-X](http://doi.org/10.1016/0010-0285(82)90010-X)
- Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, *40*(5), 331–343. <http://doi.org/10.3758/BF03203025>

- Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *The Journal of Neuroscience*, *14*(4), 2178–2189.
- Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2007). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, *184*(1), 39–52. <http://doi.org/10.1007/s00221-007-1080-3>
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. <http://doi.org/http://dx.doi.org/10.1037/0096-3445.109.2.160>
- Rohe, T., & Noppeney, U. (2015). Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLoS Biol*, *13*(2), e1002073. <http://doi.org/10.1371/journal.pbio.1002073>
- Samad, M., Chung, A. J., & Shams, L. (2015). Perception of Body Ownership Is Driven by Bayesian Sensory Inference. *PLoS ONE*, *10*(2), e0117178. <http://doi.org/10.1371/journal.pone.0117178>
- Serences, J. T. (2011). Mechanisms of Selective Attention: Response Enhancement, Noise Reduction, and Efficient Pooling of Sensory Responses. *Neuron*, *72*(5), 685–687. <http://doi.org/10.1016/j.neuron.2011.11.005>
- Shomstein, S., & Yantis, S. (2004). Control of Attention Shifts between Vision and Audition in Human Cortex. *The Journal of Neuroscience*, *24*(47), 10702–10706. <http://doi.org/10.1523/JNEUROSCI.2939-04.2004>
- Shore, D. I., & Simic, N. (2005). Integration of visual and tactile stimuli: top-down influences require time. *Experimental Brain Research*, *166*(3-4), 509–517. <http://doi.org/10.1007/s00221-005-2391-x>

- Spence, C., & Driver, J. (1997). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, *59*(3), 389–403.  
<http://doi.org/10.3758/BF03211906>
- Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*(2), 330–336.  
<http://doi.org/10.3758/BF03194473>
- Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, *19*.  
<http://doi.org/10.3389/fnint.2015.00019>
- Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. <http://doi.org/10.1016/j.tics.2010.06.008>
- Vroomen, J., Bertelson, P., & Gelder, B. D. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*(4), 651–659.  
<http://doi.org/10.3758/BF03194427>
- Werkhoven, P. J., Erp, J. B. F. van, & Philipp, T. G. (2009). Counting visual and tactile events: The effect of attention on multisensory integration. *Attention, Perception, & Psychophysics*, *71*(8), 1854–1861. <http://doi.org/10.3758/APP.71.8.1854>
- Whiteley, L., & Sahani, M. (2012). Attention in a Bayesian Framework. *Frontiers in Human Neuroscience*, *6*. <http://doi.org/10.3389/fnhum.2012.00100>
- Wozny, D. R., Beierholm, U. R., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision*, *8*(3). <http://doi.org/10.1167/8.3.24>

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability Matching as a Computational Strategy Used in Perception. *PLoS Comput Biol*, 6(8), e1000871.

<http://doi.org/10.1371/journal.pcbi.1000871>

Wozny, D. R., & Shams, L. (2011). Computational Characterization of Visually Induced Auditory Spatial Adaptation. *Frontiers in Integrative Neuroscience*, 5.

<http://doi.org/10.3389/fnint.2011.00075>

Yu, A. J., & Dayan, P. (2004). Inference, Attention, and Decision in a Bayesian Neural Architecture. *Advanced Neural Information Processing Systems*, 16, 1577–1584.

## X.

### **General Discussion and Conclusions**

#### **Summary**

This dissertation implemented a new Bayesian computational model to probe the nature of biases in spatial perception, recalibration of the tendency to integrate sensory stimuli, and the effects of attention on multisensory perception. These investigations aimed to shed light on how these factors can influence both likelihoods and priors in this computational framework.

The first chapter analyzed an extremely large spatial localization dataset, and showed that, on average, subjects exhibited central biases when localizing visual stimuli, peripheral biases when localizing auditory stimuli, and for spatially congruent multisensory stimuli, the emerging biases were determined at least partially by the inference about a common cause. Models were tested to determine which combination of parameters fit subjects' data the best, and these model comparisons revealed that the biases exhibited by subjects were best accounted for by a Bayesian model that allowed for systematic, symmetric shifts in the means of the sensory representations, as well as a prior over space.

The second chapter of the dissertation explored how brief sensory experiences may change how multisensory information is processed in the brain. Brief, ten-minute exposure tasks were employed, displaying audiovisual stimuli that varied in their spatial and temporal relationships. Counterintuitive principles were revealed regarding how/when the brain updates its tendency to bind sensory information, as well as the conditions in which the auditory representation of space is changed.

The final chapter of the dissertation investigated the effects of selective attention on sensory integration and unisensory processing. It was shown that modality-specific selective attention benefits both the precision of sensory representations for vision and audition in time, benefits only vision (and not audition) in terms of spatial processing, and does not impact the tendency to integrate multisensory stimuli. Taken together, findings from all three of these chapters reveal surprising inaccuracies in our perceptual systems, remarkable plasticity in the brain's tendency to integrate, interesting spatiotemporal principles regarding how the brain learns to change its tendency to integrate, and fascinating findings about how attention can selectively benefit unisensory processing, but minimally impact the tendency to bind sensory information.

### **Future Directions**

Additional projects implementing the refined model described in the first chapter of this dissertation are already underway. The first project focuses on the generalization and stability of the prior probability of a common cause; currently, it is unknown (1) whether the tendency to integrate multisensory stimuli is consistent across different tasks, and (2) whether the tendency to integrate multisensory stimuli is stable across time within a task. By testing the same group of subjects in a spatial localization task (Chapters 1, 2, and 3) and a numerosity judgment task (Chapter 3), we hope to determine whether this prior probability ( $P_{\text{Common}}$ ) reflects a *general* tendency to bind sensory information in different dimensions, or whether the estimated value for this parameter is *domain-specific* in either the spatial or temporal dimension. Additionally, by testing the same group of subjects on these tasks one week apart, we hope to determine for each specific task the degree to which this parameter may vary on a day-to-day basis. Preliminary results from this study indicate that while the inferred  $P_{\text{Common}}$  values are quite consistent within



a task across time, they do not generalize across tasks, indicating a domain-specific nature for this binding parameter.

The second project implementing the eight-parameter Bayesian model focuses on investigating the frame of reference in multisensory integration. Since visual information is encoded with respect to the eyes, and auditory information is encoded with respect to the head, the question of which frame of reference is used by the brain when integrating multisensory information is quite relevant. Previous work has shown that audiovisual integration aftereffects are encoded in a hybrid reference frame (Kopco, Lin, Shinn-Cunningham, & Groh, 2009); our spatial localization paradigm and Bayesian model present themselves as unique tools to probe this question further. Specifically, we are interested in the effects that eye and head position may have on the various parameters in the model, as these variables could influence any of the relevant parameters, including the integration prior, the sensory representations, or the prior over space. Preliminary evidence from these studies indicates that the spatial prior is at least partially dependent on eye position (i.e. may be encoded in a hybrid reference frame between eye-and-head position), and that when eye and head position is aligned, spatial localization using the visual response cursor is especially precise. Building on these insights, future investigations will aim to investigate the role of one more important factor in our spatial task: eye movement. In all of our localization studies up to this point, we have allowed the eyes to move as subjects make their response. By comparing our initial results to conditions where we force the eyes to fixate *throughout* the localization response, we can more effectively parse out the effects that eye and head position may have on the estimated parameters in our Bayesian model.

Finally, another preliminary study is already underway to determine the degree to which parameters in our model may correlate with various psychological disorders and abnormalities.

It is already quite well known that individuals with dyslexia, autism, schizophrenia, ADHD, synesthesia, and other conditions can exhibit abnormalities in sensory processing and sensory binding. By running a large number of control subjects through our various tasks and modeling their data, we seek to investigate whether any of the elements of our model correlate with scores from questionnaires that quantitatively estimate traits associated with these specific disorders. For instance, it has been shown that people that exhibit schizophrenic traits show abnormalities in their ability to integrate multisensory speech information (Stekelenburg, Maes, Van Gool, Sitskoorn, & Vroomen, 2013; Szycik et al., 2009). Therefore, it seems reasonable to hypothesize that people that score high on a scale that assesses schizophrenic traits (the PQ-B scale, Loewy, Pearson, Vinogradov, Bearden, & Cannon, 2011) may also exhibit a lower  $P_{\text{Common}}$  value in our model. These types of investigations could not only be informative to highlight where the deficit is in processing multisensory information in these individuals, but could also possibly yield our model's utility as a diagnostic tool in identifying psychological disorders. Future research will be needed to investigate this possibility further.

Finally, additional work will be necessary to highlight how neural circuits actually implement the computations discussed in the preceding chapters. As noted by previous researchers, much is still unknown about how neural circuits can implement Bayesian inference (Pouget, Beck, Ma, & Latham, 2013), although a recent investigation has provided substantial evidence that the unisensory estimates produced by the Bayesian Causal Inference model used in this dissertation are carried out by early-level visual and auditory cortices, and the integrated estimates are carried out primarily by regions in the intraparietal sulcus (Rohe & Noppeney, 2015). It is only through a combination of computational, physiological, and neuroimaging data that we will ultimately come to a more complete understanding of how the brain implements

Bayesian principles, and with investigations such as this dissertation shedding light on factors that can influence both likelihoods and priors in a Bayesian framework, a more complete and coherent understanding appears possible as future endeavors seek to reveal how the brain uses past experience and new evidence to enable perception of the surrounding world.