From THE DEPARTMENT OF NEUROSCIENCE
Karolinska Institutet, Stockholm, Sweden

WHERE IMAGINATION MEETS SENSATION

MENTAL IMAGERY, PERCEPTION, AND MULTISENSORY INTEGRATION

Christopher C. Berger

Stockholm 2016

Where Imagination Meets Sensation
Mental Imagery, Perception, and Multisensory Integration

Thesis for Doctoral Degree (Ph.D.)

By
Christopher C. Berger
M.A.

Principal Supervisor:
Prof. Henrik Ehrsson
Department of Neuroscience
Karolinska Institutet

Co-supervisor:
Dr. Predrag Petrovic
Department of Clinical Neuroscience
Karolinska Institutet

Opponent:
Prof. Shinsuke Shimojo
Biology Division; Computation & Neural Systems
California Institute of Technology

Examination Board:
Dr. India Morrison
Department of Clinical & Experimental Medicine
Linköping University

Prof. Abdel El Manira
Department of Neuroscience
Karolinska Institutet

Prof. Gustaf Gredebäck
Department of Psychology
Uppsala University

"Isn't it pretty to think so."

Ernest Hemingway
The Sun Also Rises

ABSTRACT

What happens if we imagine seeing something while we are listening to something? Will it change what we hear? What happens if we imagine hearing something while we are looking at something? Will it change what we see? In everyday perception, our brains integrate the information provided to us by our different senses in order to form a coherent perception of the world around us in a process referred to as *multisensory integration*. However, sometimes the information provided to our senses arises from within, as is the case when we imagine a sensory stimulus; for example, when you picture in your mind the face of a loved one, or imagine how they sound when they say your name. The term *mental imagery* is used to refer to these willed simulations of sensory stimuli in our minds. Empirical research on mental imagery has demonstrated that there is a great deal of similarity in how we consciously experience, and in how our brains process the sensory stimuli we imagine and the sensory stimuli we perceive from the external world. However, whether our brains integrate stimuli that are imagined in one sense and perceived in the other has never before been explored. The main aim of this thesis was to investigate this possibility.

There were two main goals of the work comprising this thesis. First, to examine whether mental imagery is integrated with incoming sensory stimuli from a different sensory modality to change perception, and second, to examine the neural correlates of these mental imagery-induced changes in perception. Multisensory illusions have come to be a hallmark of multisensory integration as they are an easy and demonstrable way of measuring the integration of cross-modal sensory stimuli. Here, we have made use of classic multisensory illusions, and adapted them to investigate whether mental imagery in one sensory modality can integrate with veridical sensation in another sensory modality to produced fused multisensory percepts. We also used functional magnetic resonance imaging (fMRI) to examine whether brain areas related to multisensory integration of real stimuli were involved in multisensory integration of real and imagined stimuli.

By supplanting a real sensory stimulus with an imagined one in three different classic multisensory illusions, we found that imagined sensory stimuli were integrated with real sensory stimuli from a different sensory modality to change perception. Moreover, we found that these imagery-induced multisensory illusions followed the same spatial and temporal rules as classic multisensory illusions (**Study I**), as well as the unity-assumption rule of multisensory integration (**Study II**). Furthermore, we found that the neural correlates of a mental imagery-induced multisensory illusion were closely related to those known to be involved in integration of real multisensory stimuli (**Study III**). Lastly, consistent with what is known about adaptation to real multisensory stimuli, we found that repeated pairings of imagined and real stimuli from different sensory modalities lead to changes in future perception of the latter (**Study IV**). Together, these findings suggest that, indeed, what we imagine hearing can change what we see, and what we imagine seeing can change what we hear, affording mental imagery a larger role in multisensory perception than has ever before been considered.

LIST OF PUBLICATIONS

This thesis is based on the following publications, which are referred to in the text by their roman numerals (**Studies I-IV**)[1]:

I.   **Berger, C. C.** & Ehrsson, H. H. (2013). Mental Imagery Changes Multisensory Perception. *Current Biology 23*(14), 1367-72.

II.   **Berger, C. C.** & Ehrsson, H. H. What We Imagine Hearing Changes What We See. *Submitted Manuscript.*

III.   **Berger, C. C.** & Ehrsson, H. H. (2014). The Fusion of Mental Imagery and Sensation in the Temporal Association Cortex. *The Journal of Neuroscience 34(41)*, 13684-13692.

IV.   **Berger, C. C.** & Ehrsson, H. H. Rapidly Induced Auditory Plasticity by Imagined Visual Stimuli. *Manuscript.*

---

[1] For a complete list of publications by Christopher C. Berger visit:
http://130.237.111.254/ehrsson/Berger_Website/pub.html

CONTENTS

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
AV async.	Audiovisual Asynchronous, Real Visual Stimulus
AV sync.	Audiovisual Synchronous, Real Visual Stimulus
AV_i async.	Audiovisual Asynchronous, Imagined Visual Stimulus
AV_i sync.	Audiovisual Synchronous, Imagined Visual Stimulus
BOLD	Blood Oxygen Level Dependent
EEG	Electroencephalography
ERP	Event Related Potential
fMRI	Functional Magnetic Resonance Imaging
GLM	General Linear Model
L. STS	Left Superior Temporal Sulcus
LCD	Liquid-Crystal Display
MEI	Multisensory Enhancement Index
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MVPA	Multi-Voxel Pattern Analysis
PET	Positron Emission Tomography
PPI	Psycho-Physiological Interaction
PSE	Point of Subjective Equality
SEM	Standard Error of the Mean
STS	Superior Temporal Sulcus

1 INTRODUCTION

1.1 MENTAL IMAGERY

> **imag·i·na·tion** (i-ˌma-jə-ˈnā-shən) *n.* **1.** The act or power of forming a mental
> image of something not present to the senses or never before wholly perceived
> in reality ("Imagination", 2016)

As humans, we have an uncanny ability to imagine sensations in our minds at will. If I asked you to imagine seeing your nation's flag, or to imagine hearing the voice of a loved one, or the tactile sensation of a kiss on your cheek, you would probably have an easy enough time doing that. It is a fascinating, almost oxymoronic ability we have: to see what we do not see, to hear what we do not hear, to move what we do not move, to feel what we do not feel. Of course, despite how peculiar it seems on the surface, for most of us, it is a completely ordinary part of our lives. We do it when we daydream (read: *procrastinate*), when we imagine descriptions we read in a book (you are welcome E. L. James), and, perhaps most importantly, when we cannot remember what comes after the letter Q in the alphabet, and rehearse the 'alphabet song' in our mind to help us along. We imagine all sorts of things that we are not perceiving all the time, yet figuring out what exactly these mental images are, how they are generated by the brain, and their purpose, is—to put it figuratively for maximum imagery within this paragraph—one slippery bar of soap.

This act of imaging—or imagination as one might call it—has come to be called *mental imagery* by those who study it; mostly in order to avoid confusion with the more fanciful interpretations of the word imagination, and perhaps to create some distance from it in order to sound more scientific. Imagination is just one of those words that means too many things to too many people anyway, and so for our purposes, I will use the phrase mental imagery to refer to our ability to generate sensory experiences in our mind of things we are not actually perceiving: what you 'saw' when you imagined the flag, what you 'heard' when you imagined the voice of the loved one, and what you 'felt' when you imagined the kiss on your cheek in the examples above. Although philosophers and great thinkers have pondered the nature of mental imagery since antiquity, it was not until late into the 19th century, with the development of experimental psychology, that it really moved into the realm of scientific inquiry (Kosslyn, 1994).

As the field of experimental psychology really started to get going at the turn of the 20th century, a lively debate began to develop amongst experimental psychologists about the nature of mental imagery. At the real heart of the debate was a disagreement about how 'perception-like' mental imagery is. On one side, there were those who argued that there is no fundamental difference between the things we imagine in our mind an the things we perceive (James, 1890; Kosslyn, 1994; Perky, 1910). On the other side of the debate were those who argue that there is some other underlying mental process at play that is fundamentally different from perception (Pylyshyn, 1973, 2002). This argument would be

taken up in different ways throughout the 20th century, but the gist of the argument remains the same.

In addition to believing that mental imagery and perception did not differ in kind, proponents of the perception-like theories in the early part of the 20th century, believed that mental images were the most fundamental form of thought (Boring, 1942; Danziger, 1979; Titchener, 1902). That is, that all the thoughts we have, however complex, could be reduced to sensory-like impressions in our mind. Researchers on the other side of the debate, however, believed to have found empirical evidence of 'imageless thought'—thoughts that were completely devoid of sensory information (Angell, 1911; Boring, 1942). As was common practice in the early days of psychology, both sides conducted experiments that relied mostly on 'trained-observers' to introspect on the contents of their mind, and therefore both sides of the debate tended to produce results consistent with their theory (of course). This debate would ultimately contribute to a loss of faith in what can be learned by studying the contents of the mind at all, and the field of psychology would move towards studying behavior (i.e., behaviorism) almost exclusively for several decades (Hothersall, 2004).

In the later part of the 20th century, researchers would begin to investigate the contents of the mind once again, this time with new theories and new experimental techniques. And as the field began to revive itself, so to did the debate about mental imagery. On one side of the argument were proponents of the 'picture-like' (or analog) theory of mental imagery, arguing once again that the images we conjure up in our minds are very much like pictures; they are copies of sensory impressions from the external world that maintain essential features of the physical world (Halpern, 1988; Kosslyn, 1973, 1994). On the other side, were proponents of propositional-theory (also known as discrete theory), proposing that mental imagery is the product of a language-like more basic form of thought. That is, that mental imagery is the result of accessing and combining discrete symbols, which form an abstract (rather than analog) impression of the thing in mind, and that this is sufficient to account for our perception-like experiences when we imagine stimuli (Pylyshyn, 1973, 2002).

This debate in its original form, however, has died down considerably in recent years, with behavioral and neuroimaging studies having made great strides in our understanding of mental imagery and its relation to perception. Proponents of perception-like theories of imagery have come to modify their theories to accommodate mental processes that are decidedly non-sensory-like, while still maintaining that mental imagery and perception still rely on overlapping mechanisms (Borst, Kosslyn, & Denis, 2006; Kosslyn, 1994). Propositional theory on the other hand, has few hard-lined supporters today; mainly in light of its inability to account for mounting evidence that mental imagery and perception seem to rely on largely overlapping, rather than distinct neural representations (Ganis, Thompson, & Kosslyn, 2004; Kosslyn, Ganis, & Thompson, 2001; Kosslyn, 1994). Nevertheless, vestiges of this debate still exist, and there are those who are still opposed to the idea that mental imagery and perception rely on largely overlapping mechanisms. Many arguments today against perception-based theories are methodological in nature. The most common criticism

of behavioral experiments involving mental imagery is that tacit knowledge (i.e., that the participants implicitly know what the experiment is designed to study and respond accordingly, even if unintentionally) or by response bias (i.e., that the participants know what the experiment is aimed towards and consciously alter their behavior to conform to what they believe is expected of them) can account for results of studies demonstrating similarities between mental imagery and perception. In neuroimaging experiments, the criticisms are usually levied against the neuroimaging technique itself or other un-accounted for or inherent confounding variables. These concerns are not groundless, and experimental psychologists and cognitive neuroscientists have gone to great lengths to try to address these concerns. In this next section, I will give an introduction to some of the research that has shaped our current understanding of mental imagery that is important for this thesis. I will begin this section with the early behavioral studies on mental imagery and then move onto more recent behavioral and neuroimaging studies. Importantly, I will mainly focus on findings on visual and auditory mental imagery, as most of the most classic findings in the field concern these sensory modalities, as do the studies (i.e., **Studies I-IV**) in this thesis.

1.2   MENTAL IMAGERY AND PERCEPTION

Our everyday experience tells us that when we see something, and we imagine seeing something, that these two conscious experiences—one real and one imagined—are similar in some ways but different in others. For instance, when you imagine your nations flag, you see in your mind all the features of the flag, its colors, its overall shape, the distance between symbols or lines, etc. At the same time, however, you are certainly aware that you are not actually looking at a flag, and all the features of the flag described above, are not as clear, or as sharp as they would be if you were actually looking at a flag, are they? Why are these two experiences different? Is it because mental imagery is fundamentally different than perception—that is, mental imagery involves a different underlying process than perception—or is it that mental imagery fundamentally the same as perception, but differs only in strength or vividness? Much of the work in experimental psychology has attempted to establish the latter argument against those who claimed the former.

One of the earliest, and certainly one of the most fascinating, studies on mental imagery was an experiment conducted by a woman named Mary Cheves West Perky. Perky devised an ingenious experiment in which she asked participants to look at a screen in front of them and to imagine seeing specific objects on that screen. The objects were simple objects like a tomato, a book, a banana, an orange, a leaf, and a lemon. Perky then played a trick of sorts on her participants: she (well her assistants did, anyway) surreptitiously presented a patch of color (matched in size and shape to each of the objects) on the screen while the participants were imagining the objects (Perky, 1910). The fascinating part of the experiment is that none of the twenty-seven participants (well, except for three who were thrown out of the study because of some clumsy mishandling of the projection apparatus: *nice job assistants!*) noticed that the objects Perky displayed on the screen in front of them were real rather than imagined. The participants had confused their mental imagery with real percepts that they were actually

seeing! The participants were even surprised at how what they imagined sometimes looked different than they expected it to (e.g., the banana being upright when they meant to imagine it flat), and were even sometimes indignant when asked at the end of the experiment whether they were sure they had imagined the objects (Perky, 1910). This study was the first to provide strong empirical evidence for the notion that imagery and perception are very similar indeed.

Further research on the topic of visual mental imagery was concerned with whether the spatial properties of visual perception and visual imagery were comparable. To this end Stephen Kosslyn, and colleagues devised a series of experiments in which participants were shown fictional maps or drawings to memorize (**Figure 1**) and where subsequently quizzed on the details of these maps. Specifically, participants memorized a map such as the one presented in **Figure 1** and were then verbally given one object on the map, asked to close their eyes and form a mental image of the map, and then moments later verbally given a second item on the map. The participants were then asked to trace the distance between the first location and the second location in their mind.
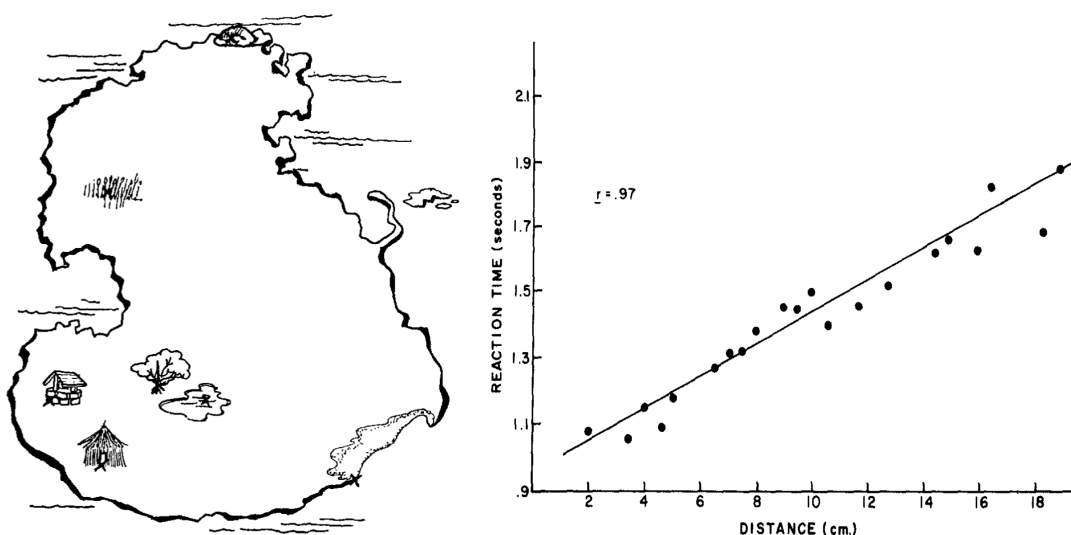

Figure 1. *How far from the hut to the well?* Example map (left) and results (right) from Kosslyn, Ball, & Reiser (1978). The plot on the right demonstrates the strong linear relationship between the distance between object pairs (e.g., the hut and the well) and their physical distance on the physical map that was memorized. This finding provided support for the notion that our visual mental imagery is picture-like by demonstrating that the time it took to scan between objects of a map in our mind scaled according to the true physical distance between the objects on that map.

The results from these experiments revealed a strong positive correlation between the time it took to scan the distance between two objects in the participants' minds and the physical distance on the maps they had memorized. Subsequent experiments demonstrated that the time it took to scan the distance between two objects also increased if the physical size of the same map was increased (in this case size *does* matter). That is, if the map in **Figure 1** was printed on a poster rather than in this thesis and you memorized the poster-sized map, and I memorized the thesis-sized map, it would require more time for you to trace the distance

between the hut and the well in your mind than it would for me. Kosslyn concluded that, combined, these results demonstrated that the spatial relationship between objects in the physical world appeared to be preserved, suggesting that mental imagery is quasi-pictorial in nature rather than descriptive (Kosslyn, Ball, & Reiser, 1978; Kosslyn, 1973).

It was argued however, that rather than demonstrating the preservation of spatial content in visual mental imagery, these findings could merely represent the participants' ability to simulate the physical laws of moving objects (Finke & Pinker, 1982) because the participants were specifically instructed to imagine an object moving from one location to another. And even more seriously, it was argued that these findings may merely be due to tacit knowledge about the purpose of the experiment in light of the fact that a similar relationship between the response times of participants and the physical distance between objects could be obtained when participants were asked to trace the distance in their mind between fictional locations for which they were only ever told the distances verbally (Mitchell & Richman, 1980; Richman, Mitchell, & Reznick, 1979). Follow-up experiments managed to provide support for Kosslyn's original interpretation, however. By using simple random dot patterns, asking participants to remember the locations of the dots, then taking them away, and placing an arrow that either pointed to a location where a dot was or not. In these experiments, the participants were to indicate whether the arrow pointed to a location where a dot was previously. The key manipulation here, is that when the arrow did point to the location where a dot was, the distance between the arrow and a dot varied from 4-12cm. Just as in Kosslyn's original study, and without the explicit instruction to imagine scanning the distance between the arrow and the dot, the participants' reaction times to determine whether the arrow pointed in the location of a dot or not *increased* as the distance between the arrow and the previous location of the dot increased (Finke & Pinker, 1982). This suggests that when the participants were asked to determine whether there was a dot in a specific location they implicitly formed a mental image and mentally scanned the image to determine whether there was a dot in a particular location (with the arrow being their starting point)(Borst et al., 2006; Borst & Kosslyn, 2008; Finke & Pinker, 1982). These studies re-affirmed the notion that the spatial content of visual stimuli is preserved in visual perception.

Similarly, studies on auditory imagery investigated whether the *temporal* features of auditory stimuli were preserved when imagining auditory stimuli and came to a similar conclusion. In experiments that I can only imagine must have felt to the participants a little like being a contestant on a game show, Andrea Halpern (1988) gave participants a task in which they were presented with two one-word song lyrics from familiar songs, and asked to determine whether the second song lyric was from the same song as the first (Halpern, 1988). The songs used were "Do re mi" (from The Sound of Music), "Hark the Herald Angles Sing", and "The Star Spangled Banner" (the American National Anthem). These songs were selected based on specific features critical to the experiment such as familiarity, whether they had lyrics that started on the odd-numbered beats from 1-13, and had unique words in the first phrase. Halpern found that the participants' reaction time increased as a function of the relative distance between the two lyrics within the song. This means that it took the same

amount of time for participants to determine whether the second lyric was in the song if the first lyric was form beat 1 and the second was from beat 5 (relative distance = 4 beats) as when the first lyric was form beat 7 and the second was from beat 11 (relative distance = 4 beats), but took *less time* when the first lyric was from beat 5 and the second was from beat 7 (relative distance = 2 beats). Moreover, this finding was consistent regardless of whether the participants were instructed to imagine hearing the lyrics in their mind or whether they were given notes instead of lyrics. The results form Halpern's experiments are consistent with Kosslyn's experiments on the preserved spatial content of visual images in visual imagery, and suggest that there is a preserved temporal format for auditory perception.

These findings from visual and auditory imagery are conceptually aligned with the widely popular and ingenious experiments conducted by Roger Shepard and Jacqueline Metzler at Stanford University around the same time. Shepard and Metzler (1971) came up with a unique experiment designed to measure how we solve a very simple problem that can be summarized in the following terms: Are these two objects the same or different? Shepard and Metzler asked participants to examine pictures of two three-dimensional objects side-by-side, and determine as quickly as possible whether they were the same of object or not (see **Figure 2**). The objects were either the same objects only rotated in the picture-plane (i.e., the picture of the object was rotated clockwise-or counterclockwise), the same objects only rotated in the depth-plane (i.e., rotated along their vertical axis within the picture), or they were different objects (see **Figure 2** for examples)(Shepard & Metzler, 1971). Shepard and Metzler (1971) found that the time it took for participants to determine whether the objects were the same (when they were in fact the same, only rotated) was correlated with the extent that the second object was rotated. That is, it took participants longer to determine that two objects were the same if they were rotated 80° than if they were rotated 60°. This suggests that participants were creating an internal mental image of the object and then rotating it in their mind to try to match the other object. These findings, along with others that would serve to rule out alternative explanations for this effect such as response bias or differences in eye movements (Cooper & Shepard, 1973), were some of the first and most convincing that mental imagery is perceptual rather than propositional.
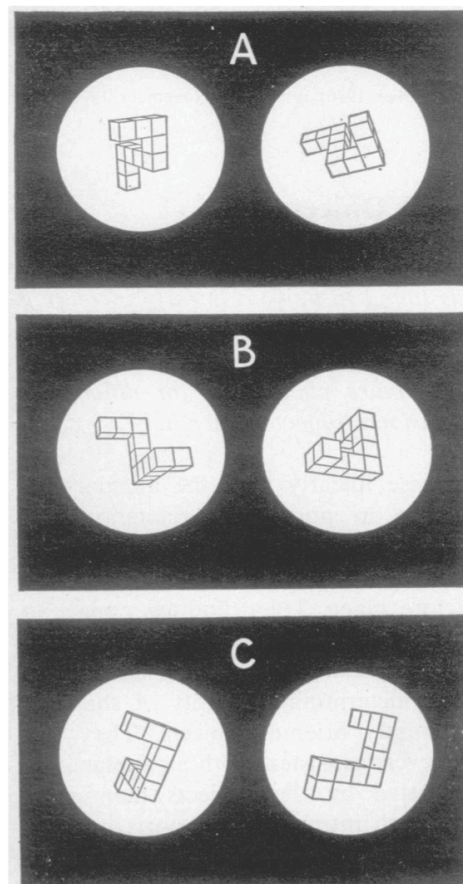
Figure 2. *Are the two objects the same or different?* Example stimuli used in experiments by Shepard & Metzler (1971) demonstrating the importance of mental imagery in human cognition. The objects were either the same objects rotated on their face (e.g., 80° clockwise) (**A**), the same objects rotated in depth (e.g., 80° in along the vertical axis)(**B**), or two objects that were different (**C**). Shepard and Metzler found that the time it took to determine whether the pairs were the same or not, was directly proportional to the degree of rotation no mater which plane the objects were rotated. These results suggest an important role of mental imagery in how we think[2].

Further studies on mental imagery examined the interaction of mental imagery and perception ongoing perception. Many of these early behavioral studies were interested in examining the hypothesis that mental imagery and perception rely on overlapping neural representations, and therefore imagining sensory stimuli in a given sensory modality should inhibit or enhance sensory perception within that sensory modality. In this line, Segal and Fusella (1970) made use of psychophysics—an experimental technique in which some aspect of a stimulus is systematically varied as a participant makes perceptual judgments about that stimulus—in a modern version of the Perky effect described above. In this modernized version of the Perky effect, Segal and Fusella (1970) varied how visible visual stimuli were (from not visible at all to completely visible) while the participants indicated whether there was a visual stimulus present or not, and recorded on a trial-by-trial basis how well participants were able to detect the visual stimuli when they were imagining a visual stimulus

---

[2] This figure was reproduced with permission from the American Association for the Advancement of Science (License number: 3851970431432)

compared to when they were not imagining anything. Consistent with Perky's original findings, they found that the participants were better at detecting the visual stimulus when they were *not* imagining anything compared to when they *were* imagining the visual stimulus. Moreover, they also found that participants were more likely to think they saw something that was not there when no visual stimulus was presented. However, Segal and Fusella (1970) also wanted to examine whether this hindrance of visual perception was merely the result of a shift in attention when imagining the visual stimuli or whether it was because the participants' visual imagery was interfering with their ability to perceive real stimuli because of overlapping neural representations. They reasoned that imagining an auditory stimulus should hinder visual perception in the same manner as imagining a visual stimulus if the effect is due to changes in attention; however, if the effect reflects competition for neural resources, auditory imagery should not hinder visual perception as much as visual imagery, and visual imagery should not hinder auditory perception as much as auditory imagery. The results from these experiments were in support of this idea: the participants' performances were worst when they imagined a stimulus in the same sensory modality. These results were strong behavioral evidence in support of the notion that imagery and perception rely on overlapping neural representations.

Another study conducted in the same vein by Martha Farah, from the Massachusetts Institute of Technology also made use of psychophysics to investigate the relationship of visual imagery and visual perception. Unlike the previous experiment, however, this study sought to examine whether there were instances in which visual imagery should facilitate visual perception, in order to try to dissociate purely propositional accounts of mental imagery (which purport that some independent non-perceptual process can explain effects found for mental imagery) from perception-based accounts of mental imagery (i.e., that the things we imagine are directly relatable to the things we perceive in the world). Farah argued that to make the case for the latter, it is important demonstrate that *both* the content and spatial format of the mental image were critical to alter visual perception. Thus, she examined whether imagining a specific letter facilitated visual processing of that letter, and when it was in the same location as the imagined letter. That is, if imagining the letter *H* facilitates processing of the letter *H* but not *T*, and this is due to a direct rather than abstract similarity in how we process imagined and perceived visual stimuli, it should only facilitate the perception of the letter *H* (and not *T*) and only when imagined in the same location. Indeed, this is exactly what Farah found: that if participants imagined a visual stimulus that was the same in content (e.g., imagine *H*; see *H*), and location, perception was significantly enhanced compared to when nothing was imagined or a different visual stimulus was imagined (Farah, 1985). Similar results were found for auditory imagery: Farah and Smith (1983) examined whether imagining hearing one tone facilitated perception of that tone vs. another tone in a detection task (a task in which participants judge whether a stimulus was present or not while the intensity of the stimuli are varied across trials). They found that imagining a tone selectively facilitated perceiving a tone of the same frequency, but not of a tone of a different frequency (Farah & Smith, 1983). Findings such these went a long way towards refuting

propositional theories of mental imagery, and provided strong behavioral evidence that certain aspects of mental imagery are perception-like rather than propositional.

More recent behavioral experiments on the similarities between mental imagery and perception, however, have evolved along with our understanding of human perception. One technique that has been fruitful in furthering our understanding of visual perception is binocular rivalry. *Binocular rivalry* is a visual phenomenon in which when one object is presented to one eye and a completely different object is presented to the other eye, our brain fails to merge these two objects, and we are only consciously aware of one object at a time (Blake & Logothetis, 2002). This is conceptually similar to what happens when we look at ambiguous figures like the Necker cube and the duck/rabbit illusions (see **Figure 3** for examples). Faced with conflicting pieces of information, our perception fluctuates between the two percepts. Unlike, in these ambiguous figures however, binocular rivalry has long been considered outside of our conscious control: we cannot consciously force ourselves to see one stimulus over the other (Blake & Logothetis, 2002). For example, in binocular rivalry, a house might be presented in one eye, and a face in the other eye, and rather than a fused combination of a house and a face, we fluctuate back and forth between seeing a house or a face without much control over it[3]. For these reasons, binocular rivalry is an ideal paradigm to study conscious visual perception and has been used extensively to examine which aspects of visual perception are related to early processing of the stimuli by the visual cortex, and which aspects are related to late processing by higher level visual cortices. For instance it turns out that certain illusions of visual perception that rely on early processing of stimuli are still present when the stimuli are suppressed from conscious awareness in a rivalry paradigm. For example, it has been found that the visual motion aftereffect (i.e., a visual illusion in which a static visual stimulus is perceived as moving in the opposite direction of a previously moving stimulus[4]) occurs for suppressed (i.e., not consciously perceived) moving stimuli (Lehmkuhle & Fox, 1975; O'Shea & Crassini, 1981). The effects of visual priming (i.e., the facilitation of visual perception via brief prior exposure to that visual stimulus) on the other hand, seem to be extinguished when the participant does not consciously perceive the visual stimuli, suggesting that priming is a process that requires conscious awareness of the stimuli and relies on a later stage of visual processing (Cave, Blake, & McNamara, 1998). These findings in the field of perception would be useful in disentangling the similarities and difference between mental imagery and perception.

---

[3] However, it should be noted that visual attention can modulate the temporal dynamics (rate and duration of switching) of rivaling stimuli with practice. If observers focus on one stimulus that is already perceived (i.e., dominant) carefully it can increase the duration that stimulus is consciously perceived (i.e., dominance) at the expense of the other, but switching still occurs none-the-less, outside of the observer's control (Ooi & He, 1999). Furthermore, specific features of the rivaling stimuli (e.g., salience, contrast, context) will make one stimulus more dominant than the other (Blake & Logothetis, 2002).

[4] A classic example of this illusion is the 'waterfall effect', whereby if an observer looks at a waterfall for a period of time then looks to the trees or rocks to the side of the waterfall, they will appear to move upwards for a brief period.

One study made use of the motion aftereffect described above to examine whether visual mental imagery of motion would lead to a visual motion aftereffect (Winawer, Huk, & Boroditsky, 2010). In this study, Jonathan Winawer and colleagues (2010) investigated whether imagining a visual grating stimulus (i.e., a series of dark grey horizontal bars spanning the length of the visual display equally spaced from each other vertically) altered the perceived motion of moving dots presented afterwards. By psychophysically varying the extent that the moving dots were random, ranging from all the dots moving upwards to all the dots moving downwards, and completely random motion in the middle, Winawer et al. (2010) found that imagining a visual grating moving upwards caused the participants to see a given set of moving dots as moving downwards and imagining the visual grating moving downwards caused the participants to perceive the same set of dots as moving upwards. Interestingly, further experiments using this paradigm found that even listening to descriptions of literal visual motion (e.g., "the squirrels were teaming up the tree") and metaphorical motion (e.g., "stock prices were falling), lead to this motion aftereffect suggesting an inherent link between language comprehension and visual imagery (Dils & Boroditsky, 2010). In light of the studies on visual perception described above, these findings provided strong evidence in favor of overlapping representations of mental imagery and visual perception at a very early stage of visual processing.

Another study made use of the binocular rivalry paradigm described above to examine the effects of visual imagery on visual perception (Pearson, Clifford, & Tong, 2008a). If you recall, from the section above, I described experiments which found seemingly conflicting results: the Perky-effect studies found that visual perception was hindered by visual imagery, whereas the studies by Farah (in which the participants imagined *Hs & T*s) found that visual mental imagery of an object facilitated visual perception of that object. While both studies nicely demonstrate that visual mental imagery can alter perception within that sensory modality, they do so with opposite results; the problem being that there is no visual perception analog here, and so there is no way to compare these results with imagery to results with visual perception. It could simply be that imagining visual stimuli in these task changes the way in which the participants attend to the visual stimuli. By utilizing binocular rivalry, however, Pearson et al., (2008) were able to examine whether visual mental imagery facilitated or hindered subsequent visual perception. In this experiment, participants briefly (750 ms) viewed rivalry displays of a visual grating stimulus presented horizontally in one eye and a visual grating stimulus presented vertically in the other, then were either instructed to imagine the visual stimulus they just perceived (i.e., the dominant visual stimulus) or to imagine the visual stimulus they did not just perceive for 10 seconds. They found that when participants imagined the dominant visual stimulus, the stability of that visual stimulus was very high, whereas when they imagined the non-dominant visual stimulus, the stability of that visual stimulus dropped dramatically. That is, if the participant saw a vertical grating stimulus on the previous trial, and imagined seeing that stimulus for 10 seconds, they were more likely to see that stimulus again; however, if the participant saw a vertical grating stimulus and imagined a horizontal grating stimulus for 10 seconds, they were much less likely to see the

vertical stimulus again. A benefit of this paradigm, in contrast to earlier studies presented above on the effects of imagery on perception, is that real visual stimuli could also be presented in the periods between rivalry displays and the result on visual perception could be compared with those of mental imagery. In further experiments, Pearson et al. (2008) found that a weak visual stimulus presented in the 10 s periods between rivalry displays produced the same effect on visual perception stability suggesting that the observed changes in visual perception from visual imagery have the same underlying mechanisms as visual perception itself.

Another interesting experiment, which demonstrated a strong link between mental imagery and perception, measured pupil dilation in response to imagined visual stimuli (Laeng & Sulutvedt, 2013). In this study, the participants imagined the same visual stimulus (e.g., a triangle) at different levels of luminance while their pupil diameters were measured. Laeng and Sulutvedt (2013) found that the participants' pupils constricted when imagining a bright visual stimulus, but dilated when imagining dark visual stimulus. These changes in pupil dilation were very closely related to those of perceiving the same stimuli, and suggest that when one imagines a visual stimulus, it engages processes of visual perception at an early enough stage in processing that the pupils constrict to protect the eyes sensitive light-absorbing cells (Laeng & Sulutvedt, 2013). One particularly striking aspect of this paradigm is that it runs counter to what one would expect if the results were due to cognitive effort, as previous studies have found that increasing cognitive effort is associated with increased pupil dilation (Hess E. & Polt J., 1964; Kahneman, Tursky, Shapiro, & Crider, 1969).

Despite the strong case made from behavioral evidence that mental imagery is perception-like and relies on overlapping representations with perception, it has been argued that truly convincing evidence for either cannot come from behavioral experiments alone (Anderson, 1978, 1979). This is a 'black box problem' of sorts: when we can only control what goes in and what comes out of the black box, there will *always* be differing interpretations for what goes on inside (Anderson, 1978, 1979; Farah, 1989b). To truly determine how what goes in affects what comes out, we must look inside the black box. In this analogy, for our purposes, our brain is that black box, and what 'goes in' is a stimulus, real or imagined, and what comes out is some measurable change in behavior (e.g., response times) or perception, but to fully rule out whether these changes can be accounted for by perception-like representations, language-like propositional representations, both, or neither, we have to look inside the brain. In this next section, I will give a brief overview of the literature that has looked inside the brain and their findings concerning mental imagery and perception.
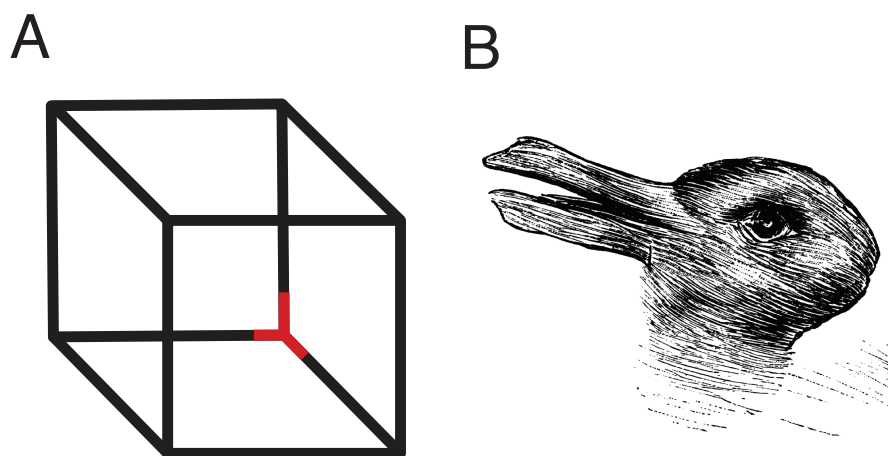
A

B

Figure 3. *Front or back; duck or rabbit?* Examples of the Necker cube (**A**) and duck/rabbit (**B**) ambiguous image illusions. The red corner in the Necker cube can either be perceived as being in the front of the cube (i.e., closest to you) or on the inside of the cube (i.e., furthest from you). The duck/rabbit ambiguous image drawing was originally used by Joseph Jastrow (1900) to illustrate the malleability of perception, and can either be perceived as, appropriately enough, a duck or a rabbit.

1.3   NEURAL FOUNDATIONS OF MENTAL IMAGERY

The neural basis for mental imagery is one of those things that are tricky to study in other animals. We cannot just ask a monkey or a mouse to imagine seeing or hearing something (even if we do so politely) while we record from electrodes in their brain[5], and so the bulk of our knowledge on this subject comes from studies on humans. Some of the earliest evidence for the neural basis for mental imagery would rely on human lesion studies. It was discovered, for example, that brain damaged patients who lost the ability to perceive color, also lost the ability to imagine color (De Renzi & Spinnler, 1967), that patients with damage to their parietal lobe who lost the ability to perceive visual stimuli on one side of their body (i.e., hemispatial neglect) also lost the ability to imagine stimuli on that side of their body (Bisiach & Luzzatti, 1978; Farah, 1989b). One interesting case study in this regard found that if you had the patient imagine themself in a particular location and imagine looking at the scene in a very familiar town square (Piazza del Duomo, Milan, Italy), they would accurately describe from their visual imagery, all the buildings on their right side, but not those from the left. If you then immediately had them imagine standing at the opposite vantage point in the square and describe the scene, they could now imagine and describe all the buildings on their right side that they could not just moments before. Moreover, they could no longer imagine or describe the buildings they had just imagined and described moments before that were now on their left (previously on their right) (Bisiach & Luzzatti, 1978). In the case of auditory imagery, studies of patients with left or right temporal lobe excision (i.e., removal of their

---

[5] Although a close approximation might be possible: some interesting work which has recently been taken underway by researchers using associative learning paradigms in monkeys are making some strides in this direction; see (Albright, 2012) and Discussion (section 5) below for more on this.

temporal lobes)—found that patients performed equally poorly on perceptual and imagery versions of a pitch comparison task. The imagery pitch comparison task in this study was similar to the pitch comparison task described above that was used by Halpern (1988) to investigate whether the temporal properties of auditory stimuli were preserved in imagined auditory stimuli. Participants were given two song lyrics of familiar songs and asked to compare whether the pitch of the first one was higher than the second, forcing them to imagine hearing the song. In the perceptual version of the task, the participants performed the same task, however, this time they actually heard the song while making their comparison. Zatore and Halpern (1993) compared lesion patients' and healthy controls' perception of familiar songs, and found that temporal lobectomy patients performed equally poorly on the imagery and perceptual versions of the task, compared to healthy controls who performed well on both tasks (Zatorre & Halpern, 1993). These findings were consistent with the idea that if imagery and perception involve shared representations, then we should see visual imagery impairments where we see impairments in visual perception. However, brain lesions are always large, involving damage to multiple areas, and interrupt the underlying white matter fiber tracts. Therefore, to determine whether mental imagery and perception rely on exactly the same cortical areas, neuroimaging experiments with good anatomical resolution—functional magnetic resonance imaging in particular—would be needed. Neuroimaging experiments would expand upon the findings from lesion studies considerably, largely converging on evidence in favor of shared representations of mental imagery and perception. At the time of Stephen Kosslyn's seminal review on the neural basis of mental imagery in 2001, over 50 neuroimaging studies had been conducted on visual imagery alone (Kosslyn et al., 2001). While I certainly will not cover all of these studies, I will try to survey some of the most influential neuroimaging studies on visual and auditory imagery that are relevant to the topic of this thesis.

Some of the earliest neuroimaging experiments examining the neural correlates of mental imagery made use of electroencephalography (EEG)—a neuroimaging technique in which electrodes placed on the scalp record fluctuations of electrical current generated by neurons in the brain (Nunez & Srinivasan, 2006). Using EEG, Martha Farah and her colleagues (1989) showed participants words, specifically either concrete nouns (e.g., *surfboard*) or abstract nouns[6] (e.g., *escapism*), and asked them to imagine the concrete nouns as vividly as possible and to simply view the abstract nouns while recording event related potentials (ERPs) (i.e., the time-locked response of an electrode to a stimulus)(Farah et al., 1989). They found that imagining the concrete nouns led to a significant positive difference in ERPs compared to the abstract noun (i.e., no-imagery) control for electrodes placed over areas involved in visual perception (i.e., occipital lobe). They found the same effect of imagery vs. no-imagery trials when participants were given the words verbally. Further experiments using EEG, made use

---

[6] Interestingly, abstract words were added to the experimental protocol as the no-imagery control because pilot studies revealed that participants found it too difficult *not* to imagine anything when presented with concrete words. This is consistent with research discussed above demonstrating the involvement of visual imagery in language comprehension (Dils & Boroditsky, 2010; Farah, Péronnet, Gonon, & Giard, 1988).

of perceptual facilitation effect of mental imagery on perception described earlier, in which imagining an *H* facilitated perception of an *H,* whereas imagining a *T* did not (Farah et al., 1989; Farah, 1985). In this study, ERPs were recorded while the participants were imagining either an *H* or a *T* while detecting whether an *H* or *T* was presented. They found that mental imagery of a letter facilitated visual detection of the same letter, as they found before (Farah, 1985), and that this perceptual facilitation was accompanied by changes in ERPs recorded from the occipital recording sites (i.e., over the visual cortex) compared to when they imagined a different letter than the one that was perceived (Farah et al., 1989). These early studies investigating the neural basis of mental imagery using EEG were consistent with the notion of shared representations of mental imagery and perception; however, the poor spatial resolution of EEG limits what can be said about how much perception and mental imagery truly involve overlapping mechanisms. Experiments using other neuroimaging techniques would prove more useful in this regard.

Researchers have also made use of other neuroimaging techniques such as single photon emission computed tomography (SPECT) and positron emission tomography study (PET) in the search for the neural correlates of mental imagery. SPECT and PET are neuroimaging techniques in which a radioactive tracer is injected into the bloodstream of the participant and then measured by the scanner as tracer-laced blood flows to different parts of the brain over the time course of the task (Gulyás & Sjöholm, 2007; Rahmim & Zaidi, 2008). The functional capacity of these techniques, from a neuroimaging standpoint, relies on the fact that increased blood flow to different parts of the brain is tightly coupled with neural activity (J. J. Chen, Wieckowska, Meyer, & Pike, 2008; Logothetis & Wandell, 2004). Goldenberg and colleagues (1988) conducted an experiment using (SPECT) that was conceptually similar to the EEG experiment described above in which participants were given lists of concrete words to imagine (Farah, 1989b; Goldenberg et al., 1989). Goldenberg et al. (1988) gave participants concrete words and asked participants to either imagine seeing the objects they referred to or to memorize the words while in the SPECT scanner and found increased blood flow (indicative of neural activity) to parts of the brain responsible for visual perception (i.e., occipital and temporal cortices). These findings were consistent with those from a (PET) study in which Roland and Friberg (1985) found increased blood flow to areas critical for visual processing (i.e., occipital, posterior parietal, and posterior inferior temporal areas) while participants imagined leaving their home and navigating through their neighborhood (Farah, 1989b; Roland & Friberg, 1985). They were also consistent with another PET study in which participants observed grids of different patterns of darkened and non-darkened squares, and were subsequently asked to indicate whether a darkened square was in a given location previously—a task which requires mental imagery (Kosslyn et al., 1994; Kosslyn, Thompson, & Alpert, 1997). This study found increased blood flow to areas of the visual cortex for this imagery task compare to control conditions in which the participants merely viewed the same stimuli (Kosslyn et al., 1997). In the case of auditory imagery, converging evidence for shared representations of imagery and perception came from another study conducted using PET which found that imagining familiar tunes was associated with

increased blood flow to areas of the brain (i.e., the auditory cortex) commonly associated with processing auditory stimuli (Halpern & Zatorre, 1999). Because SPECT and PET involve the injection of radioactive tracer, they are rather invasive techniques and so the vast majority of the neuroimaging research conducted on mental imagery would be done using less expensive and non-invasive techniques such as functional magnetic resonance imaging (fMRI)[7].

Like PET or SPECT, fMRI is a neuroimaging technique that makes use of the changes in blood flow related to neural activity, however, unlike PET/SPECT, fMRI does not require the use of radioactive tracers. Instead, fMRI makes use of the blood oxygen level dependent (BOLD) response; capitalizing on the fact that increased neural firing consumes oxygen (and glucose), which is then replenished by oxygen rich blood. Due to the difference in the magnetic properties of oxygen-rich blood (i.e., oxyhemoglobin is diamagnetic) and non-oxygen-rich blood (i.e., deoxyhemoglobin is paramagnetic), this causes distortions in the magnetic field that are then measured by a radiofrequency coil around the participants head. Early fMRI studies investigating the neural correlates of mental imagery largely sought to elaborate upon the findings from earlier lesion and PET studies with new experimental designs and questions concerning the extent of the involvement of early sensory cortices in mental imagery. The first fMRI study on mental imagery examined BOLD responses to imagining objects (e.g., a tree) compared to baseline BOLD activity, and perceiving those objects compared to baseline BOLD activity. This study revealed largely overlapping activations for visual imagery and visual perception in portions of the frontal, parietal, and occipital-temporal cortices (Ganis et al., 2004). In the first study on auditory imagery using fMRI, BOLD responses were compared between conditions in which the participants viewed silent videos of everyday scenarios and imagined hearing the associated sounds (e.g., someone blow-drying their hair) and a control condition in which the same videos were digitally scrambled. They also compared BOLD responses between viewing the same videos but while hearing the associated sounds and the scrambled-video control condition (Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005). They found that the hearing the sounds led to increased activity in the primary auditory and secondary auditory cortices, and imagining the sounds lead to increased activity in the secondary auditory cortex (Bunzeck et al., 2005). These studies converged with those form lesion, EEG, and PET/SPECT studies, in showing that imagery and perception seemed to rely on mostly overlapping rather than distinct neural representations. So, it is settled then. Mental imagery and perception share overlapping and not distinct neural representations, right? Well, not quite. It remained to be gleaned from the studies presented above whether imagining sensory stimuli could activate *primary* sensory cortices. Such a finding would stand as a strong refutation of propositional theories of mental

---

[7] It is important to note that as fMRI became more widely available, PET/SPECT neuroimaging studies measuring regional blood flow alone have become largely obsolete due to their invasiveness and cost. Many of the current neuroimaging studies using these techniques today make use of different radioisotopes catered to questions concerning the underlying neurobiology of regional brain activation, which is something fMRI cannot assess (Small et al., 2008).

imagery (i.e., non-perception-based theories), and so many subsequent fMRI (and PET studies) would serve to establish whether mental imagery could activate early sensory cortices.

The vast majority of neuroimaging studies on visual imagery would come to affirm the notion that visual mental imagery can activate the primary visual cortex (Kosslyn et al., 2001); and although early studies on auditory imagery have been more mixed about the involvement of the primary auditory cortex when it comes to imaging auditory stimuli, recent evidence seems to converge on a consensus that auditory imagery also activates the primary auditory cortex (Oh, Kwon, Yang, & Jeong, 2013). Concerning visual imagery, some of the most convincing evidence of activation of the primary visual cortex comes from a study which made use of retinotopic mapping—mapping the visual input from the eyes on the early visual areas of the cortex (Van Essen, Newsome, & Maunsell, 2002)–to compare imagining and perceiving visual objects of different sizes and found that imagined visual stimuli activated the earliest portions of the visual cortex (i.e., area 17) and that the locus of activation in this area shifted with the size of the imagined visual stimulus in the same manner as real visual stimuli (Tootell, Hadjikhani, Mendola, Marrett, & Dale, 1998). Similarly, although studies have found activation of the primary auditory cortex for imagining non-speech auditory stimuli (e.g., simple tones or music without lyrics)(Kraemer, Macrae, Green, & Kelley, 2005; Yoo, Lee, & Choi, 2001), the most convincing evidence for shared representations of perceived and imagined auditory stimuli comes from a study which made use of tonotopic mapping—mapping the frequency of sounds presented to the ears to the early areas of the auditory cortex (Saenz & Langers, 2014)–and found that imagining tones of different frequencies selectively activates portions of the primary auditory cortex (i.e., Heschl's gyrus) responsible for perceiving those frequencies (Oh et al., 2013). Together, these studies, and others like them, went a long way towards refuting propositional accounts of mental imagery, and suggested that early sensory cortices are involved in imagining the low-level features of sensory stimuli.

Further neuroimaging studies on mental imagery have approached the problem of whether imagery and perception involve overlapping neural representations by examining whether the *content* (rather than specific *features* such as the size of the stimuli) of imagined sensory stimuli led to same content specific brain activation of perceived stimuli. To get to this question, however, one must first have an understanding of which areas are responsible for content specific visual perception. Neuroimaging and electrophysiology studies making use of binocular rivalry (described in section 1.2 above) have pointed to a distinction between early visual areas (i.e., striate cortex) and later visual areas (i.e., extrastriate areas), with the former being involved in representing low-level spatial features of visual stimuli provided by the retina, and the latter being involved in the conscious perception of visual stimuli (Tong, Nakayama, Vaughan, & Kanwisher, 1998; Tootell et al., 1998). Although it was believed early on that conscious visual perception of stimuli is resolved by the primary visual cortex (Blake, 1989), electrophysiology recordings in monkeys found very little selectivity for the consciously perceived visual stimulus within the primary visual cortex during binocular

16

rivalry (Logothetis & Schall, 1989). This suggested that conscious visual perception is resolved at a later stage of visual processing. The question is where? Tong and colleagues (1998) addressed this question in a neuroimaging experiment using binocular rivalry on humans. They had participants view binocularly rivaling stimuli (i.e., a house and a face) and asked them to report their conscious visual perception while recording BOLD responses from regions just outside the primary visual cortex (i.e., extrastriate areas) which are known to respond selectively to house and face stimuli (Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997), and found that the participants' conscious visual percept selectively correlated with activity in these areas as if the visual stimuli were presented one at a time (i.e., monocularly) (Tong et al., 1998). This finding suggests that conscious visual perception of rivaling stimuli is still resolved in visual areas of the brain, but ones that lie outside the primary visual cortex. Thus, when it comes to examining the similarities between mental imagery and perception, perhaps it is more functionally relevant to consider the similarities between mental imagery and the perception in areas of the brain involved in the conscious perception of visual stimuli.

To this end, in one study, O'craven and Kanwisher (2000) investigated whether mental imagery of houses and faces involved content specific activation in extrastriate areas found to be selectively responsible for conscious perception of those stimuli. They found that imagining faces and houses lead to content specific activation within the ventral ocipito-temporal cortex called the fusiform face area and the ventromedial cortical region called the parahippocampal place area, respectively. These findings dovetailed nicely with the known selectivity for visual stimuli of these kinds within those regions, and served as strong evidence that mental imagery recruits the same neural machinery of conscious visual perception in a content specific manner. Another fMRI study conducted in the same vein utilized multi-voxel pattern analysis (MVPA) of the BOLD signal to examine whether the brain encodes the content and location of imagined visual stimuli in the same manner as perceived visual stimuli. Multi-voxel pattern analysis examines the fine-grain patterns of activation in the brain over time associated with a given stimulus or condition to create classifiers based on those stimuli or conditions (Norman, Polyn, Detre, & Haxby, 2006). These classifiers are then applied to new BOLD data from a different set of conditions to determine whether they can significantly predict the pattern of activation above chance. If they can, then the patterns of activation observed in the first set of data are reflected in the patterns of activation in the second (Norman et al., 2006). Thus, by having participants imagine seeing objects of different categories and in different spatial locations, and training classifiers on the brain activity from perceiving those objects from different spatial locations, Cichy et al. (2011) were able to determine whether the patterns of activity involved in the latter, could be used to predict brain activity of the former. They found that the classifiers trained on patterns of brain activity in response to visual stimuli of different categories (e.g., faces vs. houses) could significantly predict the patterns of activity in visual areas associated with perceiving those categories, and that the visual areas responsible for perceiving those categories maintained their category selectivity when imagining the stimuli (i.e., patterns of

activity in the occipital face are were significantly predicted by imagining faces, but not by imagining houses). Furthermore, they found that the patterns of activity associated with the spatial location of imagined visual stimuli could be predicted in early visual areas. Together, these neuroimaging experiments suggest that the content and spatial location of imagining visual stimuli and perceiving visual stimuli make use of overlapping neural mechanisms.

In summary, the neuroimaging evidence seems to suggest that mental imagery and perception rely on largely overlapping neural mechanisms. Moreover, it appears that imagining sensory stimuli can activate early sensory cortices, refuting the notion that mental imagery is propositional in nature, and supporting perception-based theories of imagery. Interestingly, however, all the research on the similarities between imagery and perception thus far have relied on demonstrating similarities within a single sensory modality (i.e., similarities between visual imagery and visual perception, or between auditory imagery and auditory perception) despite the fact that perception routinely involves the integration of sensory stimuli from several sensory modalities (Ghazanfar & Schroeder, 2006; Stein & Stanford, 2008). Thus, to truly examine whether mental imagery is perception-like, one must consider its relation to converging sensory information from other sensory modalities. This is the aim of this thesis. In this next section I will discuss what is known about the multisensory nature of perception before outlining the aims of this thesis in more detail.

1.4   PERCEPTION AND MULTISENSORY INTEGRATION

As you walk the streets of a busy city during the day, you will see buildings, streets, cars, bicyclists, and fellow pedestrians; you will hear the cacophony of passing cars, bells of bicyclists, and the indistinct chatter of people around you; you will feel the wind, the sun, or the rain on your face, and perhaps the bump of a passing stranger. How is all of this information put together into a coherent perception of the world? How do we know which sights go with which sounds, or whether the bump we just felt was from a passing pedestrian, bicyclist, or car? Without some form of integration of these pieces of information, we would not be able to determine which sounds go with which objects we see, and the result would be, perceptually speaking, incomprehensible.

The proposed mechanism for how the brain solves this information-processing problem has been appropriately dubbed 'multisensory integration'. Multisensory integration can be most generally defined as the effect that sensory stimuli from different sensory modalities have on an organism when they are combined rather than independent (Stein & Stanford, 2008). In the case of human perception, this can take the form of an enhancement or hindrance of perception in one or more sensory modalities, as is the case in multisensory interactions (i.e., when one sensory modality effects the processing of another sensory modality). Imagine the following: you are approached by a stranger on the bus while you are looking down at your feet (I don't know why, maybe you just bought a pair of new shoes and you're admiring how awesome your feet look in your new kicks—it's just an example, so roll with it) minding your own business, when the stranger says something to you. You're not quite sure what they said, but you think you heard them say, "are you going to take a dump?", but that can't be right. No. Even for public transportation, that is just too weird of a question (at least in most places: *I'm looking at you San Francisco*). So you politely look up from your amazing feet, and ask, "what's that?" Then they begin to repeat themselves, but this time, as your gaze makes its way up to their face, a shiny button on their shirt distracts you, and you hear them repeat their question about your bowel movements. Still not convinced that a stranger could be asking you such a question, you look them directly in the face now and ask more directly, "what?!" They repeat the question, this time, with you looking directly at their lips as they speak, and now you *hear* it clearly: "Are you going to vote for Trump?" and all at once you realize that the button was a 'Vote for Trump' button, and you sincerely wish they asked you the other question. In this colorful example, your auditory perception was hindered every time you heard the question because of the noisy environment of the bus; however, once you looked at their mouth, *violà*! Your visual perception made what they said completely clear. In this case, your visual perception facilitated your auditory perception, but this could have been the other way around; your auditory perception could have facilitated your visual perception. Moreover, it could have been the case that your auditory perception was hindered by visual perception. Nevertheless, it is the interaction of the senses in this manner that we are discussing when we discuss multisensory interactions. In the case of multisensory integration, however, we are talking about the integration of sensory stimuli from two or more sensory modalities into a fused multisensory percept. That is, rather than

one sensory stimulus interacting or affecting the processing of stimuli in another sensory modality, the combination of two stimuli produces a fused, multisensory percept. Such would be the case in the example above if you heard someone say "bad" while you were watching the lip movements of the Trump supporter saying "gad" (whom, for the purpose of this example, you cannot hear at all), and you clearly hear "dad", a fused multisensory percept of the heard and seen stimuli as a result of their integration. Below I will try to outline some of the key features within the large field of multisensory integration that are critical to this thesis. Because the studies contained within this thesis primarily examine multisensory integration and interactions of audiovisual stimuli, you will notice that I will mostly focus on the literature involving audiovisual stimuli. This is not to be taken as an indication that there is little research on the integration and interaction of different sensory modalities. Much to the contrary in fact, there is a vast and rich literature of the integration and interaction of other combinations; they just happen to fall outside the scope of this thesis.

Research on multisensory integration really starts to get going with findings by Alex Meredith and Barry Stein in the field of Neuroscience. Meredith and Stein (1983) were recording from individual neurons in the superior colliculus—a structure in the brain known to be involved in orientation towards sensory stimuli—of some of your favorite childhood pets (cats and hamsters) when they found cells with unique response properties. They found cells that not only responded to inputs from different senses (i.e., visual and auditory), but that responded more to the combination of visual and auditory stimuli than stimuli presented alone (Meredith & Stein, 1983). Although previous electrophysiology studies had found evidence of cells in the superior colliculus of mammals (Stein & Arigbede, 1972) and non-mammals (Bastian, 1982) that respond to more than one kind of sensory stimulus this was the first evidence of what happens to the firing property of a cell in a more ecological context with more than one sensory input at a time (cf., Newman & Hartline, 1981) and seemed to provide a ostensible solution to the binding problem of perception (i.e., how it is that stimuli are combined from different senses to form perception). These findings would subsequently be expanded upon, and further experiments would uncover neurons with multisensory response properties in the cortex (i.e., outside the superior colliculus) of the cats (Wallace, Meredith, & Stein, 1992; Wallace, Stein, & Virginia, 1994), nonhuman primates (Bruce, Desimone, & Gross, 1981; Dahl, Logothetis, & Kayser, 2009; Perrodin, Kayser, Logothetis, & Petkov, 2014; Schroeder & Foxe, 2002) and humans (Calvert, Hansen, Iversen, & Brammer, 2001; Gentile, Petkova, & Ehrsson, 2011; Werner & Noppeney, 2010b).

In characterizing the response properties of multisensory neurons, researchers have come to develop several basic principles of the integration of multisensory stimuli (Stein & Stanford, 2008). The first is the *temporal rule* of multisensory integration, which states that cross-modal sensory stimuli are more optimally integrated if they occur at the same time. The second is the *spatial rule*, which, states that cross-modal stimuli are more optimally integrated if they occur in the same location. The third is the principle of *inverse effectiveness*, which states that the strength of the multisensory response is inversely proportional to the strength of the individual stimuli. That is, the weaker the stimuli are on

their own, the stronger the magnitude of the multisensory response when they are presented together. Lastly, another guiding principle is the so-called, '*unity assumption*' rule, which states that in addition to the low-level spatial and temporal factors above, the strength of the multisensory response is related to the relationship of high-level features of those stimuli through prior experience (L. Chen & Vroomen, 2013; Vatakis & Spence, 2007). For example, hearing a dogs bark while seeing a cats mouth move would be less optimally integrated than hearing a dogs bark and seeing a dogs mouth move. These principles play an important role in a wide range of studies geared towards understating how we combine the onslaught of stimuli bombarding our senses at any given moment to form a coherent perception of the world.

In humans, the majority of studies characterizing the neural mechanisms of multisensory integration have been conducted using fMRI. Using various techniques, which manipulate the basic principles of multisensory integration described above, fMRI studies investigating the convergence of multisensory stimuli have identified the superior temporal sulcus (STS) as a key cortical region involved in the integration of auditory and visual stimuli (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Bonath et al., 2007; Bushara et al., 2003; Calvert, Campbell, & Brammer, 2000; Calvert, 2001; Marchant, Ruff, & Driver, 2011; Noesselt et al., 2007; Stevenson & James, 2009). Studies investigating how we integrate seen and heard speech (Beauchamp, Nath, & Pasalar, 2010; Gibert, Fordyce, Nath & Beauchamp, 2012; Peelle, Eason, Schmitter, Schwarzbauer, & Davis, 2010; Stevenson & James, 2009; Wyk et al., 2010) as well as how we integrate seen and heard natural objects (e.g., a hammer hitting a nail) (Beauchamp, Lee, Argall, & Martin, 2004; Noesselt et al., 2007), and simple shapes and sounds on a computer screen (Driver & Noesselt, 2008; Marchant et al., 2011; Noesselt et al., 2007), have all converged on the STS as a critical cortical region responsible for integrating these stimuli into a coherent percept. These findings have dovetailed nicely with electrophysiology and anatomical studies which have found neurons in the STS involved in multisensory integration in non-human primates (Bruce, Desimone, & Gross, 1981; Dahl, Logothetis, & Kayser, 2009; Perrodin, Kayser, Logothetis, & Petkov, 2014; Schroeder & Foxe, 2002). This is also consistent with anatomical studies on non-human primates which have found that this region receives direct projections from the auditory and visual cortices, making this the ideal candidate region for the integration of audiovisual stimuli (Kaas & Collins, 2004; Lyon & Kaas, 2002; Schroeder & Foxe, 2002; Seltzer & Pandya, 1994). Thus, the STS plays a pivotal role in the integration of auditory and visual stimuli, and will be of particular interest for our investigation of the neural mechanism involved in the integration of imagined visual and perceived auditory stimuli.

How does the integration of sensory stimuli effect perception? Some of the most dramatic examples of the perceptual consequences of multisensory integration can be readily observed in multisensory illusions. Multisensory illusions are perceptual illusions that arise through the integration of two or more sensory stimuli. One such illusion is the cross-bounce illusion in which two moving visual objects which are normally perceived as passing-by one another, are perceived as colliding and bouncing off one-another when an auditory stimulus is

presented at the moment they meet (Sekuler, Sekuler, & Lau, 1997). Consistent with the temporal principle of multisensory integration, the auditory stimuli presented at the moment the objects meet promotes the bounce percept much more than when auditory stimuli are presented before or after the objects meet. Furthermore, a comparison of trials in which the objects were perceived as bouncing and trials in which the objects were perceived as crossing revealed increased BOLD responses in multisensory areas of the cortex, as well as in the superior colliculus in one fMRI experiment (Bushara et al., 2003). Interestingly, this illusion can also be produced by presenting a tactile stimulus at the moment the objects meet or by briefly flashing visual stimulus at the moment the objects meet, leading some to speculate that this effect might merely be due to alterations in spatial attention (Shimojo & Shams, 2001a). Further studies, however, have revealed that this illusion cannot be accounted for by attentional mechanisms alone (Grassi & Casco, 2009, 2012). For example, it has been found that the content of the auditory stimulus affects whether we perceive the objects to bounce-off or cross-by one another, such that an auditory stimulus that reflects the typical acoustics of colliding objects (i.e., a damped auditory stimulus) produces the illusion, whereas the same sound played backwards (i.e., a ramped auditory stimulus) does not, even though the latter is perceived as louder and more perceptually salient than the former (Grassi & Casco, 2009). Furthermore, a signal detection analysis revealed that these auditory stimuli (i.e., damped and ramped) have the same effect on visual sensitivity in a visual detection task involving identical moving stimuli, suggesting that the effect cannot be explained by alterations in spatial attention alone (Grassi & Casco, 2009).

Another multisensory illusion that nicely demonstrates the perceptual consequences of multisensory integration is the ventriloquist effect. The ventriloquist effect gets its name from the classic act of stagecraft in which a puppeteer moves the mouth of a puppet while keeping his or her own mouth still, thereby shifting the perceived location of the puppeteers voice to the mouth of the puppet. In the study of multisensory integration, however, the ventriloquist effect refers to the translocation of auditory stimuli (usually simple tones) towards the perceived location of visual stimuli (usually simple flashes of light or briefly presented shapes) (Howard & Templeton, 1966; Wallace et al., 2004). Psychophysical behavioral experiments have revealed that the ventriloquism effect is the result of near optimal binding of audiovisual stimuli (Alais & Burr, 2004), and neuroimaging experiments investigating the neural substrates of the ventriloquism effect in humans have revealed the involvement of the STS in the ventriloquist effect (Bischoff et al., 2007), as well as dynamic changes in activity in the auditory cortex (Bonath et al., 2007); findings which correspond nicely with electrophysiology studies on non-human primates (Bruce et al., 1981; Kayser & Logothetis, 2009; Seltzer & Pandya, 1994). Furthermore, researchers have found that repeated exposure to spatially disparate audiovisual stimuli can lead to plasticity of the auditory perceptual system causing observers to misperceive auditory stimuli in the direction of the previously related visual stimulus—a so called, ventriloquism aftereffect (Frissen, Vroomen, & de Gelder, 2012; Frissen, Vroomen, de Gelder, & Bertelson, 2005; Recanzone, 1998; Woods &

Recanzone, 2004; Wozny & Shams, 2011). Thus, the ventriloquism illusion is a powerful tool for investigating the perceptual consequences of audiovisual integration.

The McGurk illusion is another important multisensory illusion that has been useful in furthering our understanding of multisensory integration. The McGurk illusion is an audiovisual speech illusion in which an auditory stimulus (e.g., /ba/) dubbed over videos of people silently articulating an incongruent speech stimulus (e.g., /ga/) leads to an illusory fused auditory percept (e.g., /da/) (similar to the example described above). This illusory fused percept is the result of the compromise between the seen articulation in the video and the heard sound (McGurk & MacDonald, 1976). The benefit of this illusion for investigating multisensory integration, is that while normally the successful integration of what we hear and see during everyday speech perception goes unnoticed, in the McGurk illusion this same successful integration of what we hear and what we see leads to a very tangible change in auditory perception. Researchers can then manipulate various features of the auditory or visual stimuli to help us understand the underlying mechanisms of normal multisensory speech perception. Furthermore, consistent with other studies implicating the STS as a key region involved in integrating both speech and non-speech stimuli (Beauchamp, Argall, et al., 2004; Calvert et al., 2000; Marchant et al., 2011; Noppeney, Josephs, Hocking, Price, & Friston, 2008; Perrodin et al., 2014; Stevenson & James, 2009; Szycik, Stadler, Tempelmann, & Münte, 2012; Werner & Noppeney, 2010a), neuroimaging experiments have found that the STS is critically involved in the McGurk illusion (Nath & Beauchamp, 2012; Szycik et al., 2012), and that creating a temporary virtual lesion to this region using transcranial magnetic stimulation (TMS) severely disrupts the illusion (Beauchamp et al., 2010).

In the studies of this thesis, I have made use of the three multisensory illusions to determine whether (i) mental imagery can lead to changes in multisensory perception consistent with multisensory integration, (ii) whether these multisensory effects from imagined stimuli on perception follow the same temporal, spatial, and unity assumption principles as the real-stimulus versions of the illusions, and (iii) whether mental-imagery induced multisensory illusions are associated with the same neural processing as the integration of real audiovisual stimuli. This line of research is uncharted territory. Despite the abundance of research outlining the similarities between mental imagery and perception, whether mental imagery integrates with stimuli from a different sensory modality to change perception has never before been examined. Given how important the integration of sensory stimuli from our different sensory modalities is for perception, to understand mental imagery and its relationship to perception, we must consider its relationship to perception not just within a given sensory modality but also between them. The following sections will outline our modest attempt to do just that.

2 AIMS

The general aim of this thesis was to investigate what role, if any, mental imagery has in multisensory perception. More specifically, we sought to examine whether mental imagery is integrated with real sensory stimuli from a different sensory modality to change perception using behavioral and psychophysics techniques. We also aimed to establish whether the integration of real and imagined cross-modal audiovisual stimuli is associated with activation in cortical areas related to the integration of real audiovisual stimuli using fMRI.

2.1 STUDY I AIM

- To examine whether mental imagery is capable of integrating with sensory stimuli from a different sensory modality to change multisensory perception in three different multisensory illusions.

2.2 STUDY II AIM

- To examine whether multisenory integation of real and imagiend sensory stimuli adheres to the 'unity-assumption' rule of multisensory integration, and also rule out possible alternative explanations for the imagery-induced cross-bounce illusion.

2.3 STUDY III AIM

- To investigate whether the cortical activation and connectivity patterns associated with the integration of imagined visual and real auditory stimuli in the imagery-induced ventriloquist effect is the same as for the classical ventriloquist effect (i.e., with real auditory and real visual stimuli).

2.4 STUDY IV AIM

- To examine whether visual imagery can lead to short term cross-modal plasticity of auditory perception and change how we perceive auditory stimuli in the future.

3 METHODS

3.1 PARTICIPANTS

All participants (*n* = 288) were recruited from the Stockholm student population, were screened for psychiatric or neurologic disorders, as well for any contradictions to the magnetic scanner in **Study III**, had normal hearing and vision (or corrected to normal with lenses). All participants gave their informed written consent at the outset of the experiment, and received monetary compensation in exchange for their participation at the conclusion of the experiment. All experiments were approved by the Regional Ethical Review Board of Stockholm and were conducted in accordance with the declaration of Helsinki.

3.2 EXPERIMENTAL STIMULI

One of the great challenges of the studies contained within this thesis was setting up the experimental stimuli and apparatuses. Although the main theme of all the studies in this thesis revolved around whether mental imagery can integrate with cross-modal sensory stimuli, most of the experiments in each study had their own unique question to answer and therefore required their own unique stimuli specifically catered to each question. I will try to concisely summarize the key stimuli used for each experiment in **Studies I-IV** in this section.

3.2.1 Stimuli for Study I

In **Experiments 1A and 1B** of **Study I**, we modified the original cross-bounce illusion setup (Sekuler et al., 1997) for our purposes. The visual stimuli consisted of two blue moving discs on a grey background on a 13-inch (33.02 cm) MacBook laptop screen. The discs (radius = 7.5 cm) moved diagonally 45° across the screen at a rate of 125 mm per second, crossing at the central fixation and disappearing off the screen in the opposite corner they came onto the screen. The auditory stimulus was a 200 ms beep sound, and the motor/tactile stimulus was a quick ≈200 ms finger tap with their index finger on the desk (with their palm facing downwards), or a ≈200 ms finger lift (the same movement as the finger tap, but with their palm facing upwards), which the participant imagined during the experiment, but heard or performed before the experiment began (although they actually heard the sound or performed the finger movement in the second half of **Experiment 1B of Study I**). The finger lift condition was strategically chosen in order to compare with the finger lift condition in **Experiment 1A**. Previous research has demonstrated that in addition to sounds, tactile stimuli can produce the cross-bounce illusion, thus we tested whether mental imagery of a motor movement which includes a strong tactile component (i.e., tapping the tip of the index finger against a surface) would also produce this effect when imagined at the moment of coincidence (Shimojo & Shams, 2001a). However, in order to control for the possibility that any increase perceived bounce was due to non-specific effects of mental imagery (e.g., modulation of attention or response bias), we included a condition with motor imagery almost identical to the finger tap but without the critical tactile event from tapping the tip of the

index finger (i.e., the finger lift condition) in **Experiment 1B**. All the stimuli that were not imagined were controlled using PsycoPy (Peirce, 2007, 2008).

In **Experiments 2A and 2B** of **Study I**, we made use of the ventriloquist illusion (Howard & Templeton, 1966). In this illusion auditory stimuli are translocated towards the simultaneously presented but spatially disparate visual stimuli. Thus, we presented auditory stimuli (a mixed 3000 Hz and 4000 Hz sine-wave tone; amplitudes = .5 N/m^2) from 12 different locations in **Experiment 2A**, and asked participants to localize the auditory stimuli as accurately as possible by clicking in the location where they heard a sound on a horizontal line spanned length of the visual display and appeared in the same vertical plane as the sounds. The auditory stimuli were pre-recorded binaurally using a dummy head (KU 100 dummy head audio system; Neumann artificial head stereo) placed in the exact location the participants' head would be during the experiment, and presented to the participants via headphones. Visual stimuli (white disc; radius = 9 cm) were imagined in four different locations (33.4 cm to the left or right of fixation and 79 cm to the left or right of fixation). Together, this resulted in the following stimulus combinations: auditory stimulus alone, visual imagery of a disc in the same location as the auditory stimulus, visual imagery of a disc in a location 15° away from the auditory stimulus, and visual imagery of a disc at a location 30° away from the auditory stimulus. Auditory stimuli presented in the auditory alone condition were presented from the same four locations as where the visual stimuli were imagined in the other conditions. In a separate experiment conducted at the conclusion of the imagery experiment, a real-stimulus version of the experiment was conducted that was identical to the imagery version of the experiment except that the visual stimuli were actually presented rather than imagined. All visual stimuli that were not imagined were projected onto the wall 79 cm in front of the participants from an overhead projector and controlled using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993).

In **Experiment 2B** of **Study I**, we made use of a psychophysical staircasing procedure (Vroomen, Bertelson, & de Gelder, 2001). Auditory stimuli were therefore presented from one of 41 possible locations (20 per hemifield and one at fixation) with the most distant sounds at 48°. As in **Experiment 2A,** all the auditory stimuli were pre-recorded binaurally using a dummy head placed in the exact location the participants would be tested during the experiment. The imagined visual stimulus was the same white disc as in **Experiment 2A,** and was imagined at fixation at the end of a countdown. Auditory stimuli were presented using the psychophysical staircase procedure during two counterbalanced conditions, and participants made binary judgments about the location of the stimuli (i.e., left or right). In one condition participants simply localized the auditory stimuli, in the other condition the participants imagined a visual stimulus flashing at the center of the screen at the same time they heard the auditory stimuli. All non-imagined stimuli for this experiment were controlled using PsychoPy software (Peirce, 2007, 2008).

In **Experiment 3** of **Study I**, we made use of the McGurk illusion in which audiovisual stimuli combine to change auditory perception. In one classic version of the illusion, the

auditory speech stimulus /ba/ dubbed over video of someone (silently) articulating /ga/ leads to an illusory auditory speech percept /da/. Furthermore, congruent sounds (speech sounds which have a similar articulation) such as /ka/ are accurately perceived. However, for our purposes, we modified this experiment so that participants imagined hearing the auditory stimuli while viewing the visual stimuli and asked them to report what they believed the person in the video was saying (i.e., da or ga). The visual stimuli consisted of six different videos of different people articulating /ga/ three times at 40 bpm. The participants were trained, and then subsequently instructed to imagine hearing the auditory stimuli at the same rate following the end of a countdown. The imagined auditory stimuli were the auditory speech stimuli /ba/ or /ka/. At the end of the experiment, the participant was tested for whether they perceived the classic McGurk illusion in a free response task that made use of a standard McGurk illusion video from the internet (example: https://www.youtube.com/watch?v=aFPtc8BVdJk). All stimuli during the experiment were controlled using PsyScope software (Cohen et al., 1993) on a 24-inch iMac computer.

3.2.2 Stimuli for Study II

In **Study II** we made use of the cross-bounce illusion again; however, we made some changes to the visual and auditory stimuli compared to **Experiment 1A** and **1B** in **Study I.** The visual stimuli consisted of two black discs (visual angle = .702°) moving diagonally at a 45° angle (rate = 10.081°/s) from the top left and right corners of a white square on the screen (height = 15.095°, width = 15.095°). We also manipulated the extent that the discs overlapped: the discs either overlapped 100% at the center of the screen (as in **Study I**), overlapped by 80%, or overlapped by 60% in the center of the screen. The different overlap conditions were achieved by removing the respective frames in which the discs overlapped by more than 80% or 60%. Importantly, the removed frames were added to the beginning and end of the to moving stimuli (not seen on-screen) in order to keep the speed and overall trial duration constant across all 3 overlap conditions. The auditory stimuli in this experiment consisted of a 200 ms damped auditory stimulus (i.e., harmonic complex tone, consisting of harmonics 1-10 of a 250 Hz fundamental for which the maximum amplitude was linearly decreased from ± 1 to 0 across its duration) or a ramped auditory stimulus (i.e., the same as the damped auditory stimulus but played in reverse). All visual and auditory stimuli that were not imagined were controlled using PsychoPy software (Peirce, 2007, 2008) on a 21.5-inch iMac computer at a viewing distance of 100 cm.

3.2.3 Stimuli for Study III

Study III required special consideration concerning the stimuli and the procedures in order to adapt the ventriloquism paradigm to the unique environment of the scanner. The visual stimuli consisted of a white disc (20 mm) on a black background (as in **Experiments 2A** and **2B**, in **Study I**). During the main experiment the discs were imagined either 20° to the left or 20° to the right of the central fixation (0°); during functional localizer scans they were actually presented. The visual stimuli were presented via MR-compatible LCD video goggles (NordicNeuroLab). The auditory stimulus was a slightly shorter version (100 ms

duration) of the mixed tone used in **Experiments 2A** and **2B** in **Study I**, played in mono-mode via MR-compatible stereo headphones. Thus, the auditory stimuli were generally perceived as spatially centered. Importantly, we chose to present the auditory stimuli in this manner for two reasons: First, we wanted to maximize the number of repetitions of all of the necessary conditions for the experiment within a reasonable amount of time (i.e., under 90 minutes) and so we minimized the number of spatial locations of the auditory stimuli, and second we feared that the loud and noisy environment of the scanner would make precise localization of the auditory stimuli in different locations rather difficult, and therefore, behavioral measurements of the effect on auditory stimuli difficult to obtain across a small number of trials. Concerning the latter reason, there were two different methods of stimulus presentation in the literature for examining the neural correlates of ventriloquism, and so we chose the method of stimulus presentation in which auditory stimuli were presented centrally, and visual stimuli to the left or the right of fixation, because it fit best with our time and behavioral constraints (Bischoff et al., 2007; Bonath et al., 2007). Finally, in addressing our question concerning multisensory integration of imagined visual and perceived auditory stimuli, our key manipulation was the temporal relationship between the stimuli (i.e., auditory stimuli at the same or different time as the imagined visual stimuli) consistent with the temporal rule of multisensory integration (Stein & Stanford, 2008). For all critical comparisons and controls, there were seven different stimulus combinations in this experiment: (1) imagined visual stimulus to the left at the same time as the auditory stimulus, (2) imagined visual stimulus to the right at the same time as the auditory stimulus, (3) imagined visual stimulus to the left at a different time as the auditory stimulus, (4) imagined visual stimulus to the right at a different time as the auditory stimulus, (5) imagine visual stimulus on the left without an auditory stimulus, (6) imagine visual stimulus on the right without an auditory stimulus, (7) hear auditory stimulus without imagining a visual stimulus. Each stimulus combination was presented six times (in 12 second blocks each). In functional localizer scans, the visual stimuli were actually presented to the participant rather than imagined in the same stimulus combinations described above. All the stimuli that were not imagined were controlled using PsychoPy software (Peirce, 2007, 2008) on a 13-inch Macbook computer within a control room.

3.2.4 Stimuli for Study IV

In **Study IV** we examined whether imagined visual stimuli can lead to changes in future auditory perception after repeated exposure to synchronous but spatially disparate imagined visual and real auditory stimuli. Thus, this experiment makes use of the imagery-induced ventriloquism effect to investigate whether there is an aftereffect on auditory perception (i.e., a 'ventriloquism aftereffect'). Thus, the visual stimuli consisted of a white disc on a black background. The auditory stimuli consisted of 50 ms white noise bursts. White noise bursts were chosen over other kinds of sounds (such as the tones used in **Study I** and **Study III**) in light of previous studies demonstrating very rapid recalibration of auditory stimuli from visual stimuli (Wozny & Shams, 2011). The auditory stimuli in **Experiment 1A** and **Experiment 1B** of this study were presented from one of five locations (±8°, ±16°, and 0°)

from self-amplified monitors arranged in a semicircular orientation along the azimuth on a shelf behind acoustically transparent fabric. The fabric was visually opaque, however, and the visual stimuli were projected onto the fabric from an overhead projector. In the test phase, auditory stimuli were presented either 8° to the left of fixation, 8° to the right of fixation, or at fixation in counterbalanced blocks. In all blocks, the participants imagined the visual stimulus appearing at 0° at the same time as the auditory stimulus. During the test phases, the participants heard the auditory stimuli come randomly form one of the 5 locations described above and pressed whether they heard the sound come from the left or the right. In **Experiments 2A** and **2B,** however, we wanted to examine whether any observed effects on auditory perception were frequency specific. The rationale for these experiments were in line with previous research on the ventriloquism aftereffect which has found that the aftereffect does not transfer across disparate frequencies (Frissen et al., 2005; Recanzone, 1998; Woods & Recanzone, 2004). Thus, the auditory stimulus presented in the exposure phase for **Experiments 2A** and **2B** of **Study IV** was a 50 ms 4 kHz sine-wave auditory stimulus. The auditory stimulus used in the test phase of these experiments was the same as in **Experiments 1A** and **1B** of **Study IV.** In **Experiments 1B** and **2B**, the visual stimuli were actually presented (rather than imagined) for comparison with any imagery-induced effects. All stimuli in **Study IV** were controlled using PsychoPy software (Peirce, 2007, 2008) on a 24-inch iMac computer.

3.3   IMAGERY INSTRUCTIONS

Another very important methodological consideration for these experiments is how the participants were instructed to imagine the stimuli during these experiments. Because our main question concerning the integration of mental imagery and perception in these studies hinged on the participants being able to imagine specific auditory (**Study I & II**), visual (**Study I, III, & IV**), or tactile/motor (**Study I**) stimuli at precise moments in time and space, careful consideration was given to how participants were instructed to imagine these stimuli. At the outset of each experiment (**Studies I-IV**), the participants were shown (if visual), heard (if auditory), or performed (if tactile/motor) the stimulus or stimuli they were to imagine during that experiment. This was repeated as needed, until the participant indicated that they felt they were familiar enough with the stimulus that they could remember it and form a clear mental image of it in their mind. Because it was critical that participants imagine the stimuli at a precise time and/or location, temporal and/or spatial cues were used during the experiments.

3.3.1   Imagery Instructions for Experiments in Study I

In **Experiments 1A** and **1B,** and **Experiments 1A and 2** in **Study I & II,** respectively on the cross-bounce illusion, the participants made use of the moving discs to time their mental imagery of the specified stimulus. In **Experiments 1A & 1B** of **Study I** a cue prior to each trial instructed the participants which stimulus to imagine (i.e., auditory or motor/tactile) and when during that trial (500 ms before the discs met, at the moment they met, 500 ms after the discs met) or instructed them to simply view the discs. Specifically, the participants saw

instructions appear on the screen for 2 s prior to the start of each trial indicating what to imagine and where. The instructions were as follows: "imagine sound here", "imagine moving finger here", "view only", for conditions in which the participants imagined the sound, the finger movements, or simply viewed the visual stimuli, respectively. The auditory and motor/tactile stimuli were first heard and performed several times before the start of the experiment. Importantly, for the motor/tactile stimuli the participants had their hand on the table, performed the action several times before the start of the experiment, and were instructed to imagine moving their finger in that exact manner when cued to "imagine moving finger here". For all imagery conditions, the location of the text indicated at which point during the motion of the discs the participants should imagine the stimulus. That is if the "imagine sound here" was placed above the central fixation prior to a trial, the participant knew that they would imagine hearing the auditory stimulus 500 ms *before* the discs met in the center of the screen, whereas if the "imagine sound here" was placed at the fixation prior to the trial, the participant would imagine hearing the auditory stimulus at the moment the discs met.

In the ventriloquism **Experiments 2A and 2B** in **Study I**, the timing *and* the location of the imagined visual stimulus was very important. For this reason, the participants saw the visual stimulus they should imagine on a particular trial in the location where they should imagine it, prior to the trial. At the outset of the experiment, the participants were instructed that they should imagine this visual stimulus as vividly as possible, and in the exact location they saw it appear, at the end of a countdown. Next, a countdown from 3 began at the central fixation. At the end of the countdown, the participant imagined the visual stimulus in the appropriate location as vividly as possible, but while maintaining fixation in the center of the screen. On trials for which the participants did not need to imagine a visual stimulus no cue was presented prior to the countdown.

In **Experiment 3** in **Study I** the experiment was split into blocks of trials in which the participants were instructed to imagine hearing either /ba/ or /ka/ three times at a specific pace (40 bpm) at the end of a countdown. A metronome was used to allow participants to practice their timing before the start of the experiment. The countdown was also presented at the rate of 40 bpm, and so the participant was instructed to imagine the auditory stimulus in continuation of the beat of the countdown. At the same time, they imagined the auditory stimuli, a video of someone silently articulating /ga/ appeared on the screen with their mouth movements synchronized to the beat of the countdown. At the conclusion of the video, the participant was then asked to indicate their perception of what they thought the person in the video was articulating.

3.3.2   Imagery Instructions for Study II

In **Study II** the participants always imagined the auditory stimulus at the moment the discs met but were instructed which auditory stimulus (i.e., damped, ramped, or no auditory stimulus) to imagine before each block of trials. The block order was counterbalanced across participants. The participants heard both auditory stimuli at the outset of the experiment in

order to establish that they could hear the difference between them. There were two reasons for exposing the participants to both stimuli in the beginning of the experiment: (1) classic versions of the experiment did not have this block design, and so participants would have been exposed to the ramped and damped sound on a trial-by-trial basis, and (2) we were concerned that the stimuli were too similar and wanted to verify that the participants heard the difference between them (and would therefore, be able to imagine the correct stimulus during the experiment). Furthermore, we were confident that if this early exposure to the stimuli were to bias the participants' responses during the experiment it would make itself clear in as a lack of variation in the perception of bounce across the three overlap conditions (i.e., we would see no difference between the three overlap conditions for a given sound). We chose a block design because we feared that participants would be confused about which sound was which if we had instructed them to imagine the stimuli on a trial-by-trial basis. Before each block, the sound the participants were to imagine during that block was played repeatedly until they had a good memory for the sound and felt they could accurately and vividly imagine it throughout the experiment.

3.3.3 Imagery Instructions for Study III

In **Study III** in fMRI-adapted version of the imagery-induced ventriloquism illusion, the participants imagined the visual stimuli while in the scanner. They were first given general instructions as to what the stimuli would look like, and what they would need to do. The participants were given a practice session with one trial of each of the possible stimulus combinations during the calibration scans. This was done in the scanner in order to make sure the participant was familiar with the task while in the scanner, and so that they could acclimate themselves to the task with noise of the scanner. For this experiment, the participants were familiarized with the visual stimulus (white disc on a black ba) and the auditory stimulus (100 ms sine-wave mixed tone presented in mono-mode over headphones; perceived spatially as coming from 0°) they would have to imagine during the functional localizer scans. However, in order to cue the participant to the location and the timing that they should imagine the stimulus (60 bpm) during the imagery portions of the experiment, they saw an instruction appear below the central fixation instructing them whether they should imagine the visual stimulus or the auditory stimulus, "imagine circle" or "imagine sound", respectively. A countdown from three appeared at the same time as the imagery instruction, in the location that the participant should imagine the visual stimulus on that trial (i.e., 20° to the left of fixation, or 20° to the right of fixation). The countdown appeared in the central fixation before the imagine auditory stimulus trials. Following the countdown, in the same beat as the countdown, the participants imagined the respective stimulus as vividly as possible. The stimulus combinations were constructed around this basic setup, such that the imagery task was always to imagine the respective stimulus, in the specified location, at the specified timing (following the countdown), and the auditory stimuli were presented in synchrony or asynchrony with the participants imagery (during imagine circle trials; no visual or auditory stimulus was presented when they imagined the auditory stimulus).

3.3.4 Imagery Instructions for Study IV

In **Study IV** we simplified the imagery task for the participants considerably. Much like **Study III**, this experiment required that the participants imagine a visual stimulus repeatedly at a specific rate for an extended period of time. In Study IV, this was because we wanted to examine whether auditory adaptation occurred in response to repeated exposure to synchronous but spatially discrepant auditory and imagined visual stimuli. Thus, the participants had to imagine the visual stimuli once per second (60 bpm) for 30 s exposure periods). In order to cue the participant to the location and timing of the to-be-imagined visual stimulus (although the location always remained the same), the participant saw a countdown from 5 appear just above the central fixation. After 4 in the countdown, from 3 to 1, the participant saw the white disc they were to imagine flash on the screen at the central fixation (although the fixation cross disappeared at 4 in the countdown) at the same pace as the countdown and the same rate that the participant should imagine the visual stimulus. The participant then imagined the visual stimulus flashing as vividly as possible once per second for 100 ms in the same manner they saw it flashing from 3-1 in the countdown for the duration of the exposure period. At the end of the exposure period a red fixation-cross appeared on the screen informing the participant that they should stop imagining the visual stimulus. Thus, for our key manipulation the auditory stimuli were presented to the left of, the right of, or at fixation at the same time the participants were imagining the visual stimuli (following the rhythm, and at the same rate as the countdown) during the exposure phase.

3.4 PSYCHOPHYSICAL TECHNIQUES

In **Study I, Study II, and Study IV** we systematically varied some aspect of the stimuli that the participants were reporting in an attempt to strengthen our conclusions by examining changes in perception of the stimuli that would be very difficult (if not impossible) to explain by response bias or tacit knowledge.

In **Experiment 2B** of **Study I** following the methodology of Bertelson and Ascherleben (1998) we made use of a psychophysical staircaising procedure in which the location of the auditory stimulus was systematically varied as participants made simple judgments about whether the sound came from the left of the right. The 'staircase' procedure involves starting at one extreme of the stimulus extremes (i.e., 48° to the left or right) and gradually moving the location of the stimulus towards the center one step at a time as participants make correct responses. Sounds starting on the left and moving towards the center are referred to as the left staircase, and sounds starting at the right and moving towards the center are referred to as the right staircase. During the experiment, sounds from the left and right staircase are randomly selected. Eventually, as the participant continues to make correct responses, the sounds move closer to the center, and at a certain point the participant can no longer distinguish between sounds on the left or the right. When participants make a mistake, the auditory stimulus moves back one step in the staircase (i.e., in the opposite direction as the previous auditory stimulus in that staircase), if they make a correct response, the auditory stimulus moves one step closer to the center (and even past the central fixation and into the

other hemifield if necessary). From the first mistake, however, we begin counting the number of response reversals—a response that was different than the previous response—until the participant has made 8 such reversals. If auditory stimuli are translocated towards imagined visual stimuli, then we should see imagined visual stimuli cause the reversals to happen earlier in the staircase (i.e., when the auditory stimuli were further away from the center) and the staircases should converge towards the center more slowly. Thus, we employed the staircasing procedure under two conditions: one in which the participants imagined a visual stimulus in the center of the screen at the same moment that they heard each sound, and one in which they did not imagine the visual stimulus. An estimate of the region of uncertainty (i.e., the area in which the participant could not distinguish between left and right sounds) can then be calculated as the average distance between the left and right staircases on the 8 reversals (i.e., responses that were different than the previous response in that staircase). In this paradigm we predict that the area of uncertainty should be significantly larger when participants imagined the visual stimulus than when they did not on the basis of previous studies using this paradigm for real visual stimuli (Bertelson & Aschersleben, 1998; Bertelson, 2000). Importantly, this procedure obviates concerns that the participants are adjusting their responses to please the experimenter (response bias) or basing their responses (even without their knowledge) on tacit knowledge about the purpose of the experiment, as there is no clear response pattern from the vantage point (explicit or implicit) of the participant which would ideally conform to either the hypothesis nor alternative hypothesis without careful consideration on the part of the participant. Particularly when one considers that the experiment did not end until the participant made 8 responses one each staircase that was different than the previous. Without the participant knowing which staircase they were on (left or right) and which response they had given on that staircase previously, it would have been impossible to systematically or tacitly alter their response for or against the hypothesis without the experimenter's knowledge.

In **Experiments 1A, 1B and 2**, in **Study II** we manipulated the extent that the moving discs overlap (100%, 80%, or 60%) at the center of the screen, and thus, how much they appeared to bounce-off one another, in the cross-bounce illusion. The extent that the discs overlap alters the perception of whether they bounce or not because truly bouncing objects never completely overlap. Thus, the 60% overlap condition, should be the most 'bounce-like' and the 100% overlap the least 'bounce-like'. This manipulation as chosen on the basis of previous work on the cross-bounce illusion (Grassi & Casco, 2009, 2012). The most important feature of this manipulation is that it would indirectly serve to rule out any effect of response bias for our imagined sound manipulation (i.e., damped and ramped sounds) by making it unclear to the participant what the ideal response strategy would be across trials in order to conform to the experimental hypothesis. That is, the results of **Experiment 1A** or **1B** could rule out such implicit or explicit response biases. For instance, if the results indicated that the damped sound increased the perception of bounce approximately the same amount across all overlap conditions, or that the ramped sound decreased or did not change across all levels of the overlap conditions, then this would reveal a response bias on the part

of the participant. That is, they thought that when they imagined the damped sound, they should say they saw the discs bounce rather than cross, and when they heard the ramped sound, they should say the sounds crossed rather than bounced on average. However, if the effects were truly perceptual in nature, we should see that the proportion of bounce responses in the ramped and damped sound conditions scales with the overlap conditions (which is indeed what we find). Furthermore using this paradigm, we were able to use signal detection analysis to explicitly test whether there was a systematic change in sensitivity to the visual stimuli (i.e., attentional mechanism), or response bias pattern associated with imagining the auditory stimuli at the moment the discs met in the center, in a separate experiment. To this end, in accordance with previous studies (Grassi & Casco, 2009), we had participants report whether the discs overlapped or not on each trial, keeping all other aspects of the experiment identical. In this way, if one or both of the visual stimuli caused a change or shift in attention (as measured by the sensitivity index d' here) or response bias (captured by c here) we would be able to measure it. In order to calculate the sensitivity index d' and response bias $c,$ the participants' responses were first coded as hits (i.e., partially overlapping response was given for partially overlapping display), misses (i.e., fully overlapping response given on partially overlapping display), correct rejections (i.e., fully overlapping response was given for a fully overlapping display), and false alarms (i.e., partially overlapping response was given for a fully overlapping display. We were then able to calculate the Hit and False alarm rates as follows: Hit rate = Hits/(Hits+Misses), and False-Alarm (FA) rate = FA/(FA+Correct-Rejections), repetively, for each participant, for the 60% and 80% overlap displays for each imagery condition (i.e., imagine ramped sound or imagined damped sound). The sensitivity index d' [$d' = \Phi^{-1}$ (Hit rate) $- \Phi^{-1}$ (FA rate)] and the response bias index c [$c = -(\Phi^{-1}$ (Hit rate) $+ \Phi^{-1}$ (FA rate))/2] were then calculated. In this formulation, positive perceptual sensitivity values would indicate better performance on the visual discrimination task, and negative values on the response bias index c should indicate a bias towards the 'partially overlapping' response. We correct for infinite values—a common occurrence in signal detection analyses—using the 1/(2N) rule (Hautus, 1995; Macmillan & Kaplan, 1985). Thus, these values would reveal whether imagining a damped compared to ramped sound led to systematic changes in visual attention or response bias that could perhaps explain the observed effects of our main experiment on whether damped compared to ramped sounds altered the perceived motion of visual stimuli. We predicted that imagining a ramped and damped sounds could reduce sensitivity compared to the view only condition, due to attention load (Alais, van Boxtel, Parker, & van Ee, 2010; Watanabe & Shimojo, 1998), but predicted that there should be no significant differences between damped and ramped sound on sensitivity or response bias indices. Such differences would be predicted by alternative, response bias or attentional mechanisms. Thus, together, this study was uniquely designed to confirm or rule out alternative accounts of this imagery-induced multisensory illusion.

In **Study IV** we were interested in whether there was a shift in auditory perception following exposure to spatially disparate imagined visual and real auditory stimuli. Thus, we systematically varied the spatial location of the auditory following exposure periods, in order

to obtain an estimate of each participant's point of subjective equality (PSE) for the different adaptation locations (i.e., imagined visual stimulus to the left of the auditory stimulus, imagined visual stimulus to the right of the auditory stimulus, or imagined visual stimulus in the same location as the auditory stimulus). The PSE refers to the location at which the participants can no longer distinguish between auditory stimuli from the left or the right in this study. The participants made binary (left or right) responses to indicate the perceived location of the auditory stimuli following the exposure periods. One key feature of this paradigm is that the visual stimulus was always imagined in the center of the screen, with the auditory stimuli presented to the left, right or same location. This meant that from the outset, the effect we were interested in was in the opposite direction of what lay intuition, tacit knowledge, or response bias would predict. That is, if the participant noticed the discrepancy between the visual and auditory stimulus during the exposure periods, they would surely think that they should respond more in the direction of that they heard the auditory stimulus when making their responses in the test phase. However, the predicted adaptation effect is in the opposite direction: we predicted that the participants should perceive sounds to come more from the left following exposure periods where the visual stimulus was imagined in the center and the auditory stimulus was heard on the right (and vise versa for auditory stimuli presented on the left of the imagined visual stimuli in the center). This stimulus setup also had the added benefit of making it easier for the participants to imagine the visual stimuli consistently throughout the experiment. The PSEs were calculated by first fitting a logistic regression curve to the participants' responses across the different auditory locations, and then calculating spatial location that corresponded to the 50% probability of a 'rightward' response. The PSEs for each adaptation condition (leftward adaptation, rightward adaptation, same location adaptation) were calculated for each participant individually and then compared across participants at the group level.

3.5   FMRI: METHODOLOGICAL CONSIDERATIONS

Functional Magnetic Resonance Imaging (fMRI) is a highly versatile neuroimaging technique with vast clinical and research applications. The benefits of using fMRI for research applications are that it enables us to measure activity of the entire brain non-invasively in live and awake humans engaged in a variety of behavioral and cognitive tasks. The activity measured in fMRI is the blood oxygen level dependent (BOLD) response, which takes advantage of the hemodynamics associated with neural activation. When neurons fire in a particular region in the brain, glucose and oxygen are metabolized in that region, leading to increased blood flow, and with it, oxygenated hemoglobin (Brown, Perthen, Liu, & Buxton, 2007; Logothetis & Wandell, 2004; Logothetis, 2008). A radio-frequency coil around the participants head then measures the subsequent change in the otherwise uniform magnetic field that results from the influx of oxygenated hemoglobin (Brown et al., 2007). This imaging technique has been extremely useful in investigating the neural correlates of a wide variety of behavioral, cognitive, and emotional processes. However, fMRI also has its drawbacks. In this section, I will briefly touch upon some of the methodological drawbacks of fMRI and the things we have done to try to get the most out of this technique despite them.

One methodological consideration in fMRI is head movement. Unfortunately, both the quality of the scans and the functional inferences that are to be made on them are highly susceptible to head movement (Khanna, Altmeyer, Zhuo, & Steven, 2015; Yendiki, Koldewyn, Kakunoori, Kanwisher, & Fischl, 2014). For this reason, the participants were instructed and reminded throughout the experiment to keep their head as still as possible. Further precautions were taken to keep their head in place during the scanning by carefully packing their head into the radio frequency (RF) head coil with padding. Some movement, however, is unavoidable despite all of these precautions; therefore, further offline corrections for head movement are made. The offline-processing and statistical analysis of the fMRI data was handled using SPM 8 (Statistical Parametric Mapping, Wellcome Center for Neuroimaging) and Matlab (The MathWorks Inc., Natick, MA, USA). The functional images are first visually inspected for movement artifacts. If artifacts are found, then that participant or that run can be thrown out (no such artifacts were discovered in our data set, however). Further inspection of the data for head motion was then conducted using the Artifact Detection Toolbox (ART)(Massachusetts Institutet of Technology, Cambridge, MA, USA), which searches for signal intensity and movement artifacts in the fMRI images. Next, the images are re-aligned to compensate for any head-movement during each run of the experiment. Following image re-alignment (and other pre-processing of the data, such as normalization to MNI space, reslicing, and smoothing) the detected artifacts from ART can then be used as nuisance regressors in the general linear model (GLM) in addition to the x, y, and z movement parameters to covary out movement related confounders.

In using fMRI we to investigate the neural correlates of the imagery-induced ventriloquism, much of what we are interested in, is *where* in the brain multisensory specific activation occurs. However, each person's brain is slightly different in size and shape, and so making anatomical generalizations at the group level requires that we place all the brains into the same coordinate space. For this reason, we take a high-resolution anatomical scan of each participant's brain and then co-register the functional images to that participant's anatomical scan. Furthermore, so that we can make inferences across participants, the anatomical and functional images are warped into Montreal Neurological Institutet (MNI) standard space. This standard coordinate space also allows us to make comparisons across experiments.

Our key experimental question was centered on whether the imagined visual and real auditory stimuli led to multisensory integration, and therefore, whether the imagery-induced ventriloquism effect was associated with activity in key multisensory regions. However, there are many ways to characterize multisensory integration in neuroimaging. Common analysis techniques for characterizing multisensory responses in neuroimaging include the criterion of the mean, whereby the observed activation is greater than the average of the responses to the sensory stimuli alone; the criterion of the maximum, whereby the observed activation is greater than the maximum response to a single sensory stimulus; or the criterion of superadditivity, whereby the activation observed is greater than the sum of the responses to all sensory stimuli presented alone (Calvert & Thesen, 2004; Laurienti, Perrault, Stanford, Wallace, & Stein, 2005; Stein & Stanford, 2008; Wallace et al., 2004). However, one

problem the lesser stringent criteria (criterion of the mean or criterion of the maximum) is that, in contrast to electrophysiology, the spatial resolution of fMRI (i.e., 2 x 2 x 2 mm voxel size, smoothed with an 8 mm full-width-at-half-maximum isotropic Gaussian kernel in our experiment) cannot rule out that the responses to multisensory stimuli merely reflects the activation of unisensory neurons in the same area, rather than reflecting the response of multisensory neurons which serve to integrate cross-modal stimuli. The problem with the more stringent criterion (the criterion of superadditivity) is that (a) it this analysis may reveal superadditive responses by summing positive and negative activations, such that rather than finding regions responsive to stimuli from different sensory modalities, superadditivity is found because an activation occurs in the presence of one stimulus and is suppressed below baseline in another, and (b) even if these concerns are accounted for, this criterion may be too stringent as there are multisensory neurons that do have linear subadditive response profiles, and therefore, lead to an increase in false negatives (Calvert & Thesen, 2004; Laurienti et al., 2005). However, in these concerns can be obviated by systematically manipulating the conditions under which two sensory stimuli are likely to be combined (Calvert & Thesen, 2004). This method has the added benefit of controlling for any possible effects of attention as the same stimuli are presented in both conditions; only their relationship has changed. In our experiment, we have made use of the temporal rule of multisensory integration to avoid these concerns, by manipulated the temporal relationship between the imagined visual and real auditory stimulus (i.e., presented synchronously or asynchronously). Any resulting differences in activation should thusly reflect the multisensory integration of these stimuli.

Lastly, further consideration for inferences about fMRI concerns what we can and cannot infer from the fMRI signal. For all its benefits, fMRI is unavoidably correlative in its relation to neural activity. As stated above, the activation observed in fMRI is the BOLD response, and is related to the hemodynamics associated with neural activation, not neural activity itself. Work examining the extent to which the BOLD response is indicative of neural signals using simultaneous fMRI and intracortical electrophysiology recordings have found that the hemodynamic response does indeed seem to closely reflect neural activity, and seems to most closely reflect local field potentials (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Logothetis & Wandell, 2004). Nevertheless, one must keep in mind the correlative nature of the technique, and the many levels removed we are from measuring what we truly wish to measure (i.e., neural activity). As for our part, we have done our best to keep in mind the correlative nature of the BOLD signal in interpreting our results, and where possible we have couched our findings within a framework that takes into account the underlying neurophysiology.

4 OVERVIEW OF STUDIES

4.1 STUDY I: MENTAL IMAGERY CHANGES MULTISENSORY PERCEPTION

4.1.1 Study I Background and Rationale

In light of previous research on the similarities between mental imagery and perception described above, **Study I** was really a series of experiments designed as a proof of concept for the multisensory integration of real an imagined stimuli. Thus, we made use of classic versions of multisensory illusions, adapted for this purpose by having participants imagine a stimulus in one sensory modality and measuring their perception of the other sensory modality.

In the first two experiments (**Experiments 1A & 1B**) we made use of the *cross-bounce illusion* in which an auditory stimulus presented at the moment two passing stimuli meet promotes the illusory perception that they bounce-off, rather than cross-by one another. For our purposes, the participants imagined the auditory stimulus. Consistent with classic findings on the cross-bounce illusion (Sekuler et al., 1997), as well as the temporal rule of multisensory integration (Stein & Stanford, 2008), we predicted that, imagining an auditory stimulus at the moment two discs meet would promote this illusory bounce percept compared to when the auditory stimulus was imagined before, or after the objects met, or not imagined at all. Furthermore, we also included an additional finger-tap condition at the moment the discs met, in which the participants imagined tapping their finger at the moment the discs met. Consistent with previous studies demonstrating that the cross-bounce illusion can be elicited by tactile events we predicted that imagining a finger tap at the moment of coincidence would also produce the illusion because of the dynamic tactile feedback from the fingertip. In a control experiment (**Experiment 1B**) we included condition in which the participants imagined a finger-lift—the same movement as in the finger tap condition, but importantly, without the tactile component—at the moment of coincidence. In contrast to the finger tap condition of **Experiment 1A**, we predicted that the finger lift condition should not produce the illusion, despite the fact that both conditions demand the same amount of attention and involve very similar kinds of motor imagery.

In two additional experiments (**Experiments 2A & 2B**) we made use of a different multisensory illusion—*the ventriloquist effect*—to examine whether imagined visual stimuli could integrate with real auditory stimuli to change auditory perception. In the classic ventriloquism paradigm, a visual stimulus presented at the same time, but different location than an auditory stimulus leads to a translocation of auditory perception towards the visual stimulus. Thus, for our purposes, we had participants imagine a visual stimulus at the same time as a spatially disparate auditory stimulus and measured their auditory perception. We predicted that imagining a visual stimulus in a different location than an auditory stimulus would lead to a translocation of auditory stimuli away from their veridical source, and that imagining a visual stimulus in the same location as an auditory stimulus would improve auditory localization compared to when auditory stimuli were presented alone. This result

would be consistent with studies using real audiovisual stimuli in the classic ventriloquism illusion (Alais & Burr, 2004; Howard & Templeton, 1966). In **Experiment 2B** we made use of a psychophysical staircase procedure in an experiment designed to rule out non-perceptual accounts if the imagery-induced ventriloquism illusion.

Finally, in a third experiment (**Experiment 3**), we made use of yet another classic multisensory illusion—the *McGurk illusion*—to examine whether auditory imagery could lead to cross-modal changes in speech perception. In the classic McGurk illusion, certain incongruent auditory speech stimuli (e.g., /ba/) dubbed over visual speech stimuli (e.g., /ga/) lead to an illusory auditory speech percept (e.g., /da/) compared to congruent auditory speech stimuli (e.g., /ka/)(McGurk & MacDonald, 1976). Thus, in this experiment, we predicted that imagining hearing /ba/ at the moment participants were seeing someone silently articulate /ga/ would promote the illusory perception that the person articulating /ga/ was articulating /da/, compared to when they imagined hearing the congruent auditory stimulus /ka/ or simply viewed the person articulating /ga/. This finding would be further evidence of the cross-modal influence of auditory imagery on visual perception, and would be consistent with the so-called 'reverse-McGurk' illusion whereby real auditory speech stimuli have been shown to change visual shape perception (Charles Spence & Deroy, 2012; Sweeny, Guzman-Martinez, Ortega, Grabowecky, & Suzuki, 2012).

4.1.2 Study I Results and Conclusions

Consistent with our hypothesis, in **Experiment 1A** of **Study I** we found that imagining an auditory stimulus at the moment the two discs met significantly promoted the illusory perception that the discs bounced-off rather than crossed-by one another, compared to when the participants imagined an auditory stimulus before or after the discs met, or when they were simply viewing the discs (**Figure 4A**). As expected, we found that imaging the finger tap at the moment the discs met also promoted the illusory bounce percept. Importantly, however, in **Experiment 1B** of **Study I** we found that imagining a finger lift—i.e., a movement like the finger tap movement in **Experiment 1A** but without the dynamic tactile percept from the tip of the finger—at the moment the discs met did not promote the illusory bounce percept compared to simply viewing the discs. **Experiment 1B** also replicated the effect for imagining a sound at the moment the discs met on the perception of the bounce percept. These results were consistent with those from a subsequent 'real-stimulus' version of the experiment in which the participants actually moved their finger or heard a sound at the moment the discs met (see **Figure 4B**).
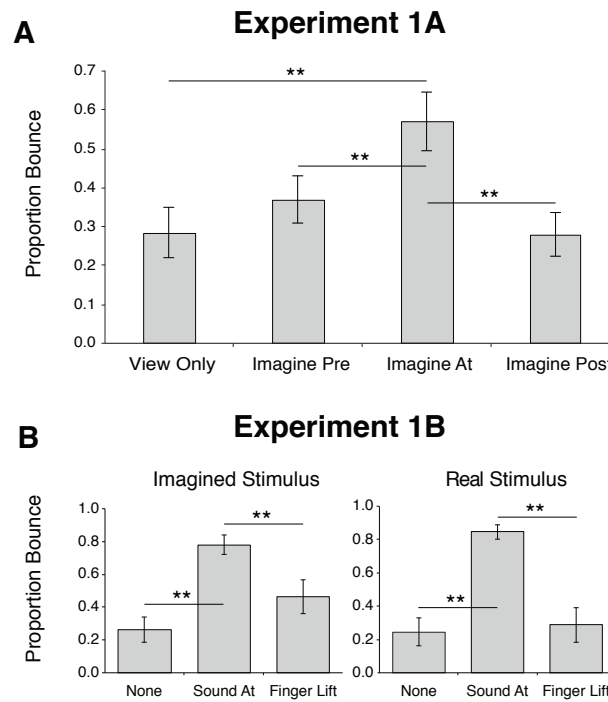
Figure 4. *Visual Bounce-Illusion from imagining sounds.* The results from Experiment 1A (**A**) and Experiment 1B (**B**) from **Study I.** Experiment 1A (**A**) demonstrated that imagining a sound at the moment the discs met, compared to before or after, or not imagining a sound at all, promoted the illusory perception that the discs bounced-off rather than crossed-by one another. The results from Experiment 1B (**B**) demonstrate that this effect was not a general effect that resulted form imagining anything at the moment of coincidence, as imagining a finger lift did not promote the illusory bounce percept (left). The same result was found in a 'real-stimulus' version of the experiment in which participants heard a real sound and actually moved their finger at the moment of coincidence (right). Asterisks indicate significant *p*s < .01. Error bars denote ± SEM.

The results from **Experiment 2A** & **2B** of **Study I** revealed that imagined visual stimuli altered the perceived location of auditory stimuli. In **Experiment 2A** we calculated a percent visual bias (% visual bias)—the mean localization error when they heard an auditory stimulus and did not imagine a visual stimulus subtracted form the mean localization error when the participants imagined a visual stimulus at the same time as an auditory stimulus divided by the actual spatial disparity between the imagined visual and the real auditory stimulus, then multiplied by 100—for conditions where the imagined visual and real auditory stimulus came from different locations. We found that imagining a visual stimulus at the same time but in a different location as an auditory stimulus led to an illusory translocation of the auditory stimulus towards the imagined visual stimulus. We also found a larger % visual bias when the disparity between the visual and auditory stimulus was greater (i.e., 15° vs. 30°) (**Figure 5A**). Furthermore, we calculated a multisensory enhancement index (MEI)—the mean localization error when the auditory stimulus was presented alone, subtracted from the mean localization error when the participants imagined a visual stimulus at the same time as an auditory stimulus, divided by the mean localization error when the auditory stimulus was presented alone—and found that imaging a visual stimulus at the same time and same location as an auditory stimulus significantly enhanced auditory localization ability (see **Figure 5B**). That is, participants were better at localizing an auditory stimulus when they imagined an auditory

stimulus in the same location compared to when they heard the auditory stimulus alone. Both the % visual bias and MEI results were in line with the results from a 'real-stimulus' version of the experiment conducted at the conclusion of the imagery version of the experiment (see **Figure 5AB**).

Experiment 2B in **Study I** made use of a psychophysical staircasing procedure to examine whether visual imagery could bias visual perception in conditions that would be difficult to explain by response bias or tacit knowledge about the experiment. In this procedure, participants made binary (left or right) judgments about the location of sounds presented one at a time from the left or the right. The sounds began at the far left and right extreme edges of the visual display and moved gradually towards the center as participants made correct responses (see Methods above or **Experiment 2B** in **Study I** for more details). When participants made a wrong response, the sound location moved back one step (i.e., away from fixation) in the staircase. A region of uncertainty was then calculated for each subject as the average distance between the left and right staircases at which the first eight response reversals (i.e., a response reversal is defined as a response that was different than a previous response) occurred. We found that the region of uncertainty significantly increased when participants imagined seeing a visual stimulus in the center of the screen at the same time as they heard the auditory stimuli compared to when they were localizing the auditory stimuli alone (see **Figure 5CD**). This result is consistent with the hypothesis that auditory stimuli were translocated towards the imagined visual stimuli as a result of multisensory integration rather than due to response bias or tacit knowledge on the part of the participant.

In **Experiment 3** in **Study I** we found that imagining an auditory speech stimulus could change visual speech perception. Making use of the McGurk effect, we found that imagining hearing /ba/ while participants watched silent videos of someone articulating /ga/ significantly promoted the illusory perception that the person in the video was articulating /da/ compared to when participants were imagining /ka/ or were simply viewing the videos. Moreover, we found that this effect was specific to individuals who perceived the classic version of the illusion in which /ba/ dubbed over someone articulating /ga/ leads to an illusory auditory perception of /da/ in a post-experiment free-response test for the effect (see **Figure 6**). These findings suggest that auditory speech imagery can alter visual speech perception and are consistent with the reverse-McGurk effect (Charles Spence & Deroy, 2012; Sweeny et al., 2012).
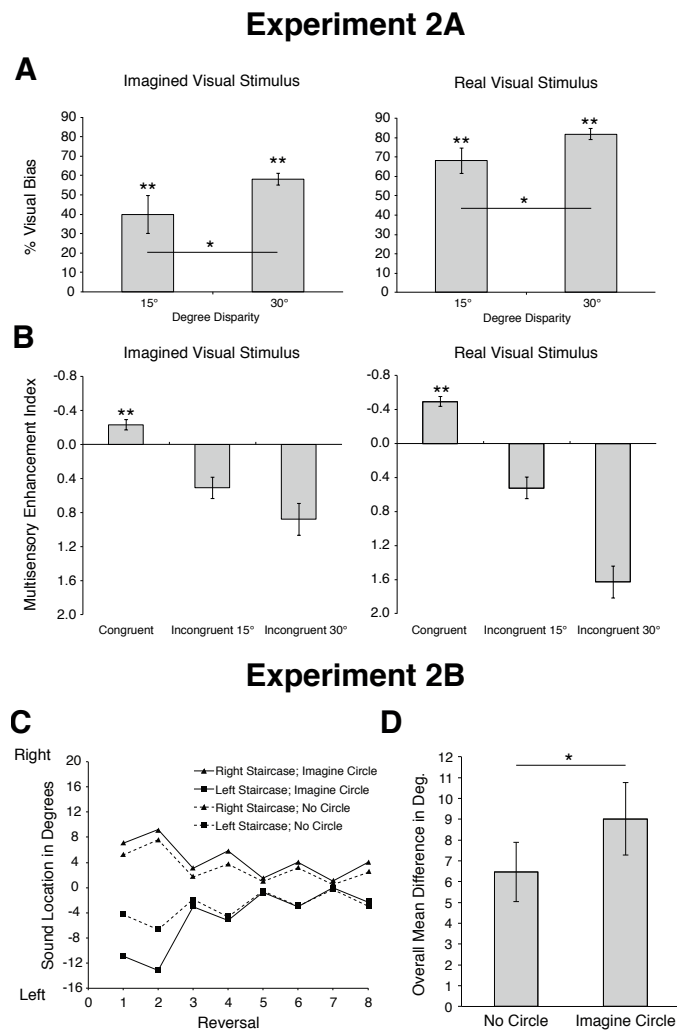
Experiment 2A

A

Imagined Visual Stimulus

Real Visual Stimulus

B

Imagined Visual Stimulus

Real Visual Stimulus

Experiment 2B

C

D

Figure 5. *Ventriloquist Illusion from imagining visual stimuli.* The results from Experiment 2A (**A-B**) and Experiment 2B (**C-D**) from **Study I.** Imagining a visual stimulus at the same time, but different location than an auditory stimulus significantly biased participants' auditory perception (**A**, left) in the same manner as real visual stimuli (**A**, right). Imagining a visual stimulus at the same time and same location as an auditory stimulus significantly improved participants' ability to identify the source of an auditory stimulus (**B**, left). The same was found for real visual stimuli (**B**, right). Imagining a visual stimulus at the center of the screen at the same time participants heard sounds from two converging staircases led to a translocation of auditory stimuli towards the imagined visual stimulus, as indicated by slower convergence of the auditory stimuli as participants made response reversals (C) and the increased region of uncertainty (i.e., the mean distance between the left and right staircases at response reversals 1-8)(D). Single asterisks indicate significant $ps < .05$, significant $ps < .01$. Error bars denote ± SEM.
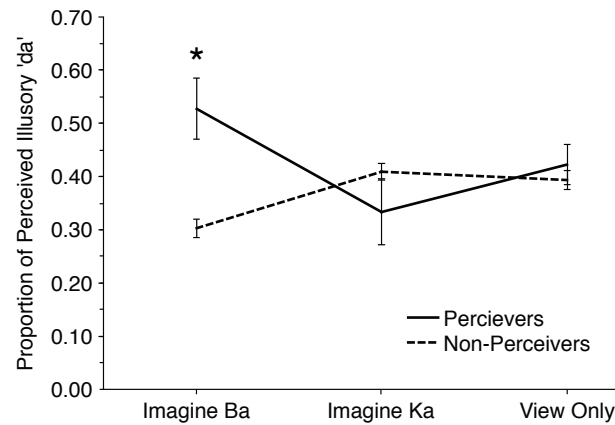
Figure 6. *What I imagine hearing can change what I think you said.* Auditory imagery of /ba/ significantly promoted the illusory percept that people articulated /da/ rather than /ga/ compared to auditory imagery of /ka/ or simply viewing the visual stimuli. This imagery-induced effect on speech perception was specific to participants who perceive the classic McGurk illusion. Asterisk indicates a significant $p < .05$. Error bars denote ± SEM.

Together, the findings from **Experiments 1-3 of Study I** demonstrate that mental imagery in one sensory modality can change perception of a different sensory modality. Specifically, we found that imagined auditory and tactile stimuli can alter visual motion perception in **Experiments 1A & 1B**, that imagined visual stimuli can alter the perceived location of auditory stimuli in **Experiments 2A & 2B**, and that imagined auditory stimuli can alter visual speech perception in **Experiment 3**. These behavioral results are consistent with the findings presented here, and those from previous research documenting multisensory integration of real sensory stimuli. Further, they provide support for perception-based theories of mental imagery, and suggest that mental imagery can integrate with incoming sensory stimuli from a different sensory modality to shape perception.

4.2 STUDY II: WHAT WE IMAGINE HEARING CHANGES WHAT WE SEE

4.2.1 Study II Background and Rationale

In **Study II** we sought to explore whether the imagery-induced *cross-bounce illusion* was a general effect of auditory imagery on visual perception, or whether it was specific to the content of the imagined auditory stimulus. Previous studies on the classic cross-bounce illusion have found that damped sounds, which simulate the acoustics of a collision, promote the illusory bounce percept, whereas ramped sounds (the same sounds played in reverse) do not, even though the ramped sounds are perceived as being louder, and more perceptually salient (Grassi & Casco, 2009, 2012). To this end, we reasoned that if the auditory imagery-induced cross-bounce illusion relies on the same multisensory mechanisms as the veridical version of the illusion (Bushara et al., 2003), then we should expect imagined damped and not imagined ramped sounds to produce the cross-bounce illusion. Thus in **Experiment 1A** of **Study II** we sought to examine whether imagined damped, but not imagined ramped sounds led to the cross-bounce illusion. Further, we sought to rule out alternative explanations of the illusion such as response bias or tacit knowledge by subtly manipulating the extent the moving discs overlapped, and are therefore perceived as crossing. We predicted that if participants were responding on the basis of tacit knowledge or response bias, we would see uniform responses (of bounce or cross) for a given imagined sound across the different overlap conditions; however, if participants are reporting their genuine perception, then the probability of a bounce percept should scale according to the extent that the discs overlap. **Experiment 1B** of **Study II** was a 'real-stimulus' version of the experiment, in which the auditory stimuli were actually heard (rather than imagined) at the moment of coincidence, to compare with the imagery version of the experiment.

Furthermore, in a third experiment (**Experiment 2**, **Study II**) we specifically sought to examine whether the effect of auditory imagery on visual perception could be attributable to purely attentional mechanisms by asking participants to attend to whether the visual stimuli overlapped or not, and measuring their performance (which requires visual attention) on this task, while imagining the ramped, damped, or simply viewing the stimuli. We predicted that performance on the visual attention task would decrease when participants imagined the auditory stimuli, compared to when they did not imagine anything; however, we predicted that if the effect were multisensory rather than attentional, there should be no difference in performance (i.e., sensitivity index = d') when imagining the damped vs. imagining ramped sounds. Moreover, in this same experiment we could measure the extent to which imagining a particular auditory stimulus, or not imagining an auditory stimulus, led to a particular response strategy (i.e., response bias index = c).

4.2.2 Study II Results and Conclusions

In **Experiment 1A** we found that imagining a damped sound significantly promoted the illusion that the discs bounced-off rather than crossed-by one-another, whereas the imagining the same sound played backwards (i.e., a ramped sound) did not, compared to simply viewing

the stimuli (see **Figure 7a**). Moreover, we found that the increased bounce-percept scaled according to the extent that the discs overlapped, suggesting that the effect represents a genuine change in perception rather than response bias or tacit knowledge. These observations were consistent with findings of the 'real-stimulus' version of the experiment conducted on an independent group of subjects in **Experiment 1B** (see **Figure 7b**). Furthermore, we also calculated the strength of the imagery-induced cross-bounce illusion as the mean difference between the proportion of perceived bounce when the participants imagined hearing the damped vs. ramped sound and found that it was significantly correlated with how vividly the participants were able to imagine the auditory stimuli (see F**igure 8**).
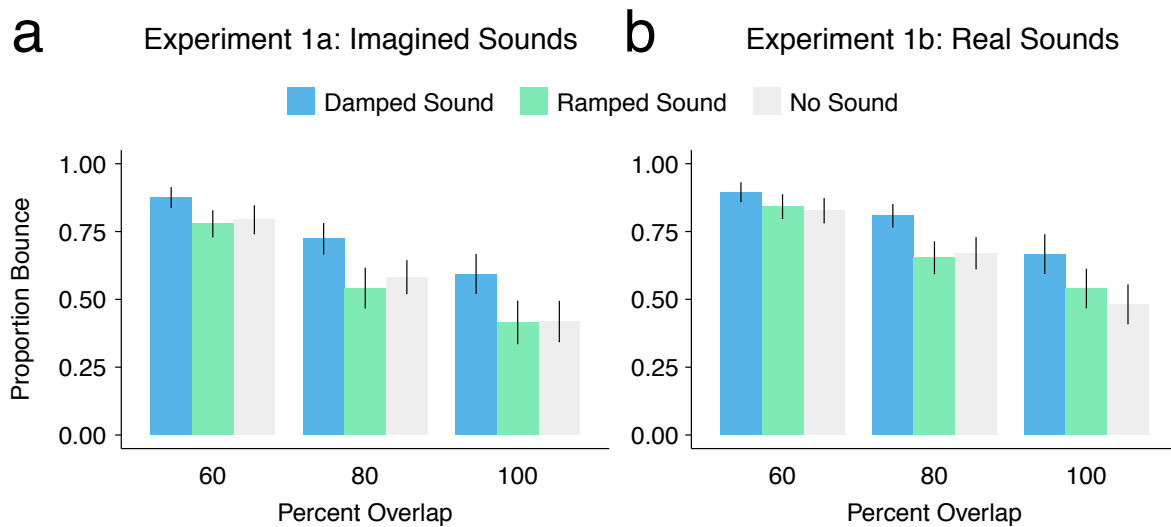
Figure 7. *What one imagines hearing changes what one sees.* The results from Experiment 1a (**a**) and Experiment 1b (**b**) on the perceived bounce of two passing objects as a function of the type of sound imagined (i.e., damped, ramped, or none), or real sound heard in Experiment 1b, across the three different overlap conditions (i.e., the extent that the discs overlapped in the center of the screen, which is directly related to how 'bounce-like' they are perceived be). Error bars denote ± SEM.

The results from **Experiment 2** revealed that although the participants' sensitivity (d') changed across the degree to which the discs overlapped (60 vs. 80), there was no change in the participants' sensitivities when they imagined a damped, ramped, or simply viewed the stimuli (see **Figure 9a**). This finding suggests that imagining the auditory stimuli at the moment the discs met did not lead to systematic changes in visual attention that might account for the increased perception of bounce. That is, this result suggests that the effect cannot be due to the possibility that the participants' attention was drawn away from the visual stimuli when they imagined the damped sound and thereby increased their propensity to respond that the discs bounce more than cross. Moreover, we also analyzed whether imagining a damped auditory stimulus lead to a biased response pattern compared to when they imagined a ramped auditory stimulus or simply viewed the visual stimuli. The results confirmed that imagining a damped sound did not lead to a particular response strategy compared to imagining a ramped sound or not imagining a sound. Once again, the only difference in response strategy was observed between the overlap conditions (see **Figure 9b**).
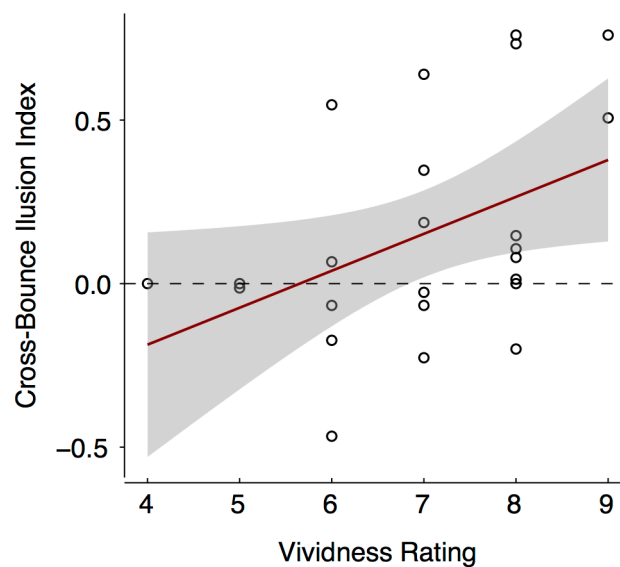
Figure 8. *How well one imagines sounds predicts how much they see the discs bounce.* Plot (and 95% confidence interval bands) of the significant relationship between the participants' self-reported vividness of the imagined sounds and the strength of the cross-bounce illusion. Dotted line denotes the divide between the participants who perceived the illusion and those who did not.
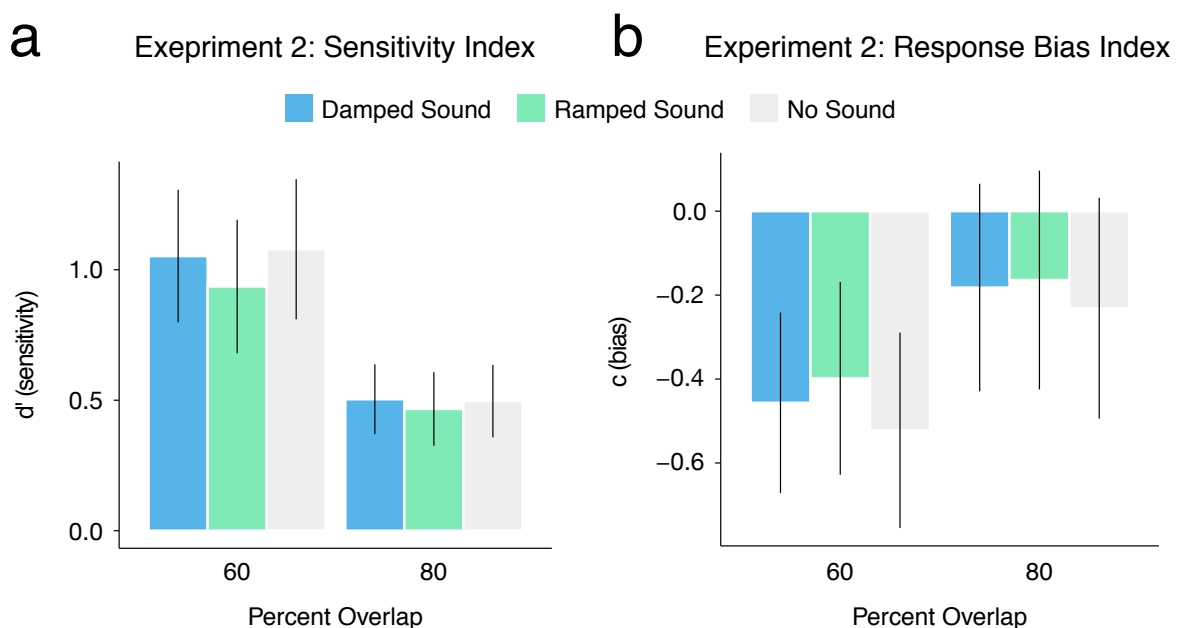

Figure 9. *What one imagines does not alter one's attention nor how one responds, generally.* Plots of the sensitivity indices (i.e., d')(**a**) and response biases (c)(**b**) as a function of the overlap and imagery conditions indicating. These results indicated that imagery of sounds had no influence of sensitivity compared to not imagining a sound, nor did imagining a sound lead to a particular response strategy compared to not imagining a sound. Only the overlap conditions (60% and 80%) had a significant effect on sensitivities and response biases.

The findings from **Study II** suggest that the cross-modal effect of imagined auditory stimuli on visual motion perception are specific to the kinds of sounds that produce the

classic version of the illusion.   Moreover, these findings firmly establish that this effect represents a genuine phenomenon of perception and go further to rule out that the effects we observed for the increased perception of bounce can be explained by differences in attention or response strategies for the different kinds of stimuli. These results of **Experiment 2** were even stronger evidence that the effect cannot be explained by attentional mechanisms or response bias, as we did not observe significant difference between imagined auditory stimuli and not imagining any stimuli. This finding stands in slight contrast to those from the classic version of the experiment which have found that a real auditory stimulus can effect sensitivity to the stimuli compared to no auditory stimulus (Grassi & Casco, 2009). Thus, when real auditory stimuli are presented at the moment of coincidence, attention *alone* cannot account for the perception of bounce, and the results here demonstrate that attention cannot account for the perception of bounce *at all*. Furthermore, the finding that imagined damped and not imagined ramped sounds produced the cross-bounce illusion is consistent with the 'unity assumption' principle of multisensory integration which proposes that only meaningful combinations of multisensory stimuli are optimally integrated (De Gelder & Bertelson, 2003; Vatakis & Spence, 2007; Welch & Warren, 1980). That is, because a damped sound simulates the acoustics typical of a collision, but the ramped sounds do not, imagery of the former and not the latter is optimally integrated to change perception. Finally, these results provide additional evidence that the integration of real and imagined sensory stimuli across the senses relies on the same integrative mechanisms as the integration of real cross-modal stimuli.

4.3 STUDY III: THE FUSION OF SENSATION AND MENTAL IMAGERY IN THE TEMPORAL ASSOCIATION CORTEX

4.3.1 Study III Background and Rationale

Previous studies have found that that mental imagery within a specific sensory modality involves largely overlapping neural representations with perceiving those sensory stimuli (Bunzeck et al., 2005; Cichy et al., 2011; Ehrsson, Geyer, & Naito, 2003; Kosslyn et al., 2001; O'Craven & Kanwisher, 2000; Oh et al., 2013; Roth et al., 1996) Therefore, we hypothesized that if mental imagery and perception rely on overlapping neural representations, then the integration of imagined and real sensory stimuli across sensory modalities should involve overlapping mechanisms as well. To investigate this possibility we made use of functional magnetic resonance imaging fMRI and the imagery-induced ventriloquist effect from **Study I.** Specifically, we examined whether synchronous, but spatially disparate imagined visual but real auditory stimuli lead to increased activation in the multisensory temporal association cortex. We also examined whether synchronous, but spatially disparate imagined visual but real auditory stimuli lead to increased effective connectivity between the temporal association cortex and the auditory cortex. Moreover, we examined the relationship between the strength of the imagery-induced ventriloquism effect and activity in the multisensory temporal association cortex and the strength of the effective connectivity between the temporal association cortex and the auditory cortex. In order to limit our search to brain areas responsive to multisensory stimuli a 'real-stimulus' version of the ventriloquism paradigm served as a functional localizer for this study.

4.3.2 Study III Results and Conclusions

We first assessed the behavioral results obtained from the scanner by calculating a ventriloquism index for each subject. The ventriloquism index was calculated by coding the responses for left, center, and right, as -1, 0, and 1, respectively, then averaging by condition, and subtracting each participants average bias in the auditory only condition (from functional localizer blocks) from the synchronous and asynchronous audiovisual conditions. This analysis controls for false positives (i.e., instances where the participants indicate that the auditory stimulus comes from same direction as the imagined visual stimulus because of a baseline perceptual bias in a particular direction). The results form this analysis revealed that imagining a visual stimulus at the same time (synchronously; AV_i sync.) but different location than an auditory stimulus lead to a significant translocation of auditory stimuli towards the imagined visual stimuli compared to when the visual stimuli was imagined at a different time (asynchronously; AV_i async.) and location than auditory stimuli (see **Figure 10A**). The same result was found during functional localizer runs in which real visual stimuli were presented synchronously (AV sync.) or asynchronously (AV async.) in spatially disparate locations rather than imagined (see **Figure 10B**).
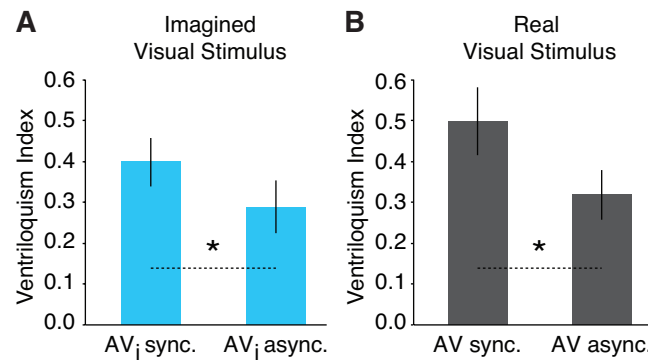
Figure 10. *What one imagines seeing in the scanner changes where one hears sounds in the scanner.* The behavioral results from Study III revealing a significantly stronger ventriloquism effect (i.e., translocation of auditory stimuli towards visual stimuli) for imagining visual stimuli at the same time as, but different location than, real auditory stimuli (AV$_i$ sync.) compared to imagining visual stimuli at a different time as, and different location than, real auditory stimuli (AV$_i$ async.) in the main experiment (**A**). The same comparison between synchronous (AV sync.) vs. asynchronously presented (AV async.) spatially disparate audiovisual stimuli from functional localizer blocks (**B**). Asterisks denote significant *p*s < .05. Error bars denote ± SEM.

The main fMRI analysis compared the BOLD response for AV$_i$ sync. compared to AV$_i$ async. stimuli and found significant activation within the left superior temporal sulcus (L. STS) (**Figure 11A-B**). We also calculated the strength of the imagery ventriloquism effect for each participant (i.e., the difference of the AV$_i$ sync. and AV$_i$ async. ventriloquism indices) and examined whether there was any activity that was related to the strength of the imagery-induced ventriloquism effect in a whole brain multiple regression analysis. This analysis revealed that activity in the L. STS could be predicted by the strength of the mental imagery-induced ventriloquist effect (see **Figure 11C**).

A psychophysical interaction (PPI) analysis was conducted in order to examine whether AV$_i$ sync. was associated with increased effective connectivity between the STS and the auditory cortex compared to AVi async. The benefit of a PPI analysis is that it allows you to examine changes in effective connectivity between one region and remote brain areas as a function of your experimental manipulation. The PPI analysis revealed increased connectivity between the L. STS and the auditory cortex during the AV$_i$ sync. condition compared to the AV$_i$ async. condition (see **Figure 11D-E**). Furthermore, we assessed the relationship between the strength of the mental imagery-induced ventriloquism effect and the effective connectivity between the L. STS and the auditory cortex across participants, and found that a stronger imagery induced ventriloquism effect was associated with stronger effective connectivity between the L. STS and the auditory cortex (see **Figure 11F**).
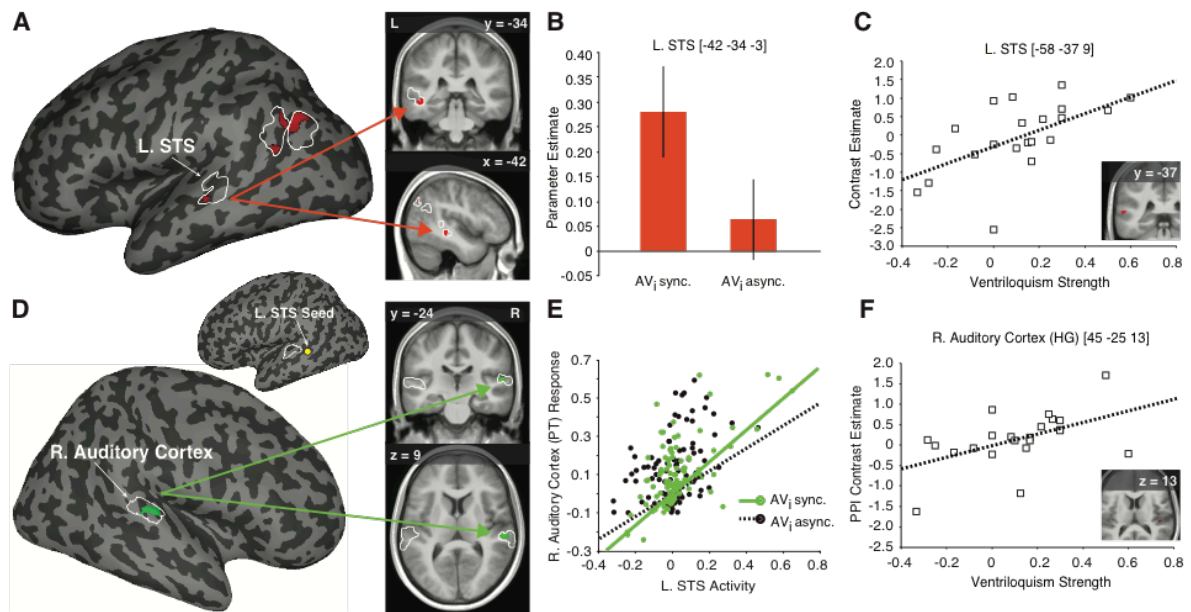
Figure 11. *The neural correlates of the imagery-induced ventriloquist effect.* Imagining visual stimuli at the same time, but different location as real auditory stimuli (AV$_i$ sync.) lead to a significant increase in the BOLD response in the left superior temporal sulcus (L. STS) compared to imagining visual stimuli at a different time and location (AV$_i$ async.)(**A**). Parameter estimates for this comparison reveal this was due to a significant increase of activation in AV$_i$ sync. vs. AV$_i$ async., rather then de-activation (**B**). The strength of the mental imagery-induced ventriloquism effect was significantly predicted by the strength of the activity in the L. STS across participants for the above comparison (**C**). The AV$_i$ sync. condition was also associated with increased effective connectivity between the L. STS and auditory cortex compared to the AV$_i$ async. conidion (**D**). Plot of the increased effective connectivity for one representative subject in the form of a steeper regression slope for the AV$_i$ sync. condition compared to the AV$_i$ async. condition (**E**). The strength of the mental imagery-induced ventriloquism effect was also significantly predicted by the strength of the effective connectivity between the L. STS and auditory cortex for the above contrast (AV$_i$ sync. vs. AV$_i$ async.) across participants (**F**). Error bars denote ± SEM.

Together, the findings from **Study III** suggest the integration of real and imagined audiovisual stimuli involves the same neural mechanisms as the integration of real audiovisual stimuli. Moreover, our findings suggest that the imagery-induced ventriloquism effect is associated with activity in the multisensory association cortex, and with increased connectivity between the multisensory temporal association cortex and the auditory cortex. This finding suggests that the temporal association cortex may play a modulatory role in auditory perception when integrating spatially disparate imagined visual and real auditory stimuli. These findings are also consistent with previous neuroimaging studies on the classic version of the ventriloquist effect (Bischoff et al., 2007; Bonath et al., 2007). The involvement of the L. STS is consistent with previous neuroimaging studies implication the STS in the integration of audiovisual stimuli across a range of stimulus types (Beauchamp, Lee, et al., 2004; Beauchamp et al., 2010; Bischoff et al., 2007; Driver & Noesselt, 2008; Nath & Beauchamp, 2011; Noesselt et al., 2007; Werner & Noppeney, 2010b). Furthermore, the results from this study and previous neuroimaging studies in humans are also consistent with electrophysiological recordings in nonhuman primates which have found that the STS

contains cells that integrate auditory and visual stimuli at the level of the single neuron (Bruce et al., 1981; Dahl et al., 2009; Perrodin et al., 2014; Schroeder & Foxe, 2002). These findings also provide strong support for 'perception-like' theories of mental imagery, and demonstrate that the overlap in neural representation of real and imagined sensory stimuli extend to the more perceptually ecological context of multisensory perception.

4.4 STUDY IV: RAPIDLY INDUCED AUDITORY PLASTICITY BY IMAGINED VISUAL STIMULI

4.4.1 Study IV Background and Rationale

In **Study IV** we investigated whether repeated exposure to the imagery-induced ventriloquist illusion could lead to plasticity of our auditory system. Research on the classic ventriloquist effect has found that repeated exposure to synchronously presented spatially disparate audiovisual stimuli leads to a recalibration of auditory system in the direction of the visual stimulus even when the visual stimulus is no longer present—i.e., a ventriloquism aftereffect (Lewald, 2002; Recanzone, 1998; Wozny & Shams, 2011). In light of the behavioral and neuroimaging evidence presented above suggesting that the integration of imagined and real multisensory stimuli seems to rely on similar mechanisms as the integration of real multisensory stimuli, we predicted that there should be a measureable change in future auditory perception following repeated exposure to synchronous, but spatially disparate imagined visual and real auditory stimuli—i.e., an imagery-induced ventriloquist aftereffect. Thus, in **Study IV** we took advantage of the imagery-induced ventriloquist effect to examine whether it can lead to changes in *future* auditory perception. In this experiment we had participants imagine the visual stimulus for 100 ms, once per second, in the center of the screen for 30 s. In separate blocks, an auditory stimulus was either presented in the same location (i.e., same adaption) as the imagined visual stimulus, 8° to the left of the imagined visual stimulus (i.e., rightward adaptation), or 8° to the right (i.e., leftward adaptation)[8] of the imagined visual stimulus. Following this exposure phase, the participants heard sounds come from five different locations (i.e., ±16°, ±8°, or 0°) and indicated whether they heard the sound come from the left or the right. From these data we measured the participants point of subjective equivalence (PSE)—the point at which the participants could no longer perceive an auditory stimulus from the left or the right. We predicted that the PSE would be shifted to the right following leftward adaption (i.e., auditory stimuli presented to the right of the imagined visual stimulus), but shifted to the left following rightward adaption (i.e., auditory stimuli presented to the left of the imagined visual stimulus), compared to a non-adaptation control condition (i.e., auditory stimuli presented in the same location as the imagined visual stimulus).

4.4.2 Study IV Results and Conclusions

The results from **Experiment 1a** of **Study IV** revealed a significant leftward shift in the PSEs for sounds localized following rightward adaption compared to sounds localized following same location adaptation. Moreover, we found a significant rightward shift in PSEs

---

[8] I know that may seem like a mistake because a sound on the left is rightward adaptation and a sound on the right is rightward adaption, but rest assured it is not. It is leftward adaptation when the auditory stimulus was presented on the right because the auditory stimulus should be translocated towards the visual stimulus, which in this case is on the left; hence, leftward adaption. The opposite, of course is true for rightward adaption, in which an auditory stimulus presented on the left leads to a translocation of the auditory stimulus towards the visual stimulus on its right, and is therefore, rightward adaptation.

for sounds localized following leftward adaption, compared to same location adaption (see **Figure 12AB**). These results were consistent with a 'real-stimulus' version of the experiment (**Experiment 1b**) conducted on a separate group of subjects (see **Figure 12CD**). These results represent the first evidence of a shift in the perceived location of auditory stimuli following cross-modal adaptation of real and imagined stimuli.

Additional experiments examined whether this imagery induced-ventriloquism aftereffect was selective to the specific type of sound being played. This would strengthen the conclusion that the same neural mechanisms involved in the plasticity of the auditory system were engaged as with real stimuli in the classic ventriloquism aftereffect which has been shown to be specific to the type of sounds being presented during adaptation (Recanzone, 1998; Woods & Recanzone, 2004). Furthermore, these experiments were conducted to rule out that the effect was merely the result of being presented with auditory stimuli in a spatial location to the left or right repeatedly. Note, that this is highly unlikely, given that this would mean that the participants responded counter to the location the auditory stimulus was presented during the adaptation periods when making their responses in the test phases. Nevertheless, in **Experiment 2a** we examined whether adaptation to an imagined visual stimulus at the same time as, but different location than a sine-wave auditory stimulus would change future auditory perception of a white-noise auditory stimulus (used in both adaptation and test phases in **Experiments 1a & 1b** above), which we predicted would not be the case. In line with previous studies on the ventriloquist aftereffect that have demonstrated that the ventriloquism aftereffect does not transfer across sounds with different frequencies (Recanzone, 1998; Woods & Recanzone, 2004), we found no transference of adaption from the sine-wave auditory stimulus to the white noise stimulus from **Experiments 1a & 1b** (see **Figure 13AB**). This was consistent with the results obtained from a 'real-stimulus' version of the experiment conducted on a separate group of participants, in which the participants actually saw rather than imagined the visual stimuli (see **Figure 13CD**).

Together, the results from these experiments suggest that a systematic spatial discrepancy between visual mental imagery and a real auditory stimulus can lead to multisensory recalibration and plasticity in our auditory perceptual system. Moreover, we found that this imagery-induced ventriloquism aftereffect was specific to the auditory stimuli to which the adaption occurred. This is consistent with previous studies on the ventriloquism aftereffect (Recanzone, 1998; Woods & Recanzone, 2004). Furthermore, these findings cannot be explained by changes in attention, response bias, or tacit knowledge, once again providing strong evidence in support of perception-based theories of mental imagery. Indeed, it is hard to see how the adaptation aftereffect—and the underlying plastic neuronal changes—would arise unless the imagery-induced ventriloquist effect engaged the same basic multisensory integration mechanisms as the classic ventriloquist effect (i.e., with real audiovisual stimuli).
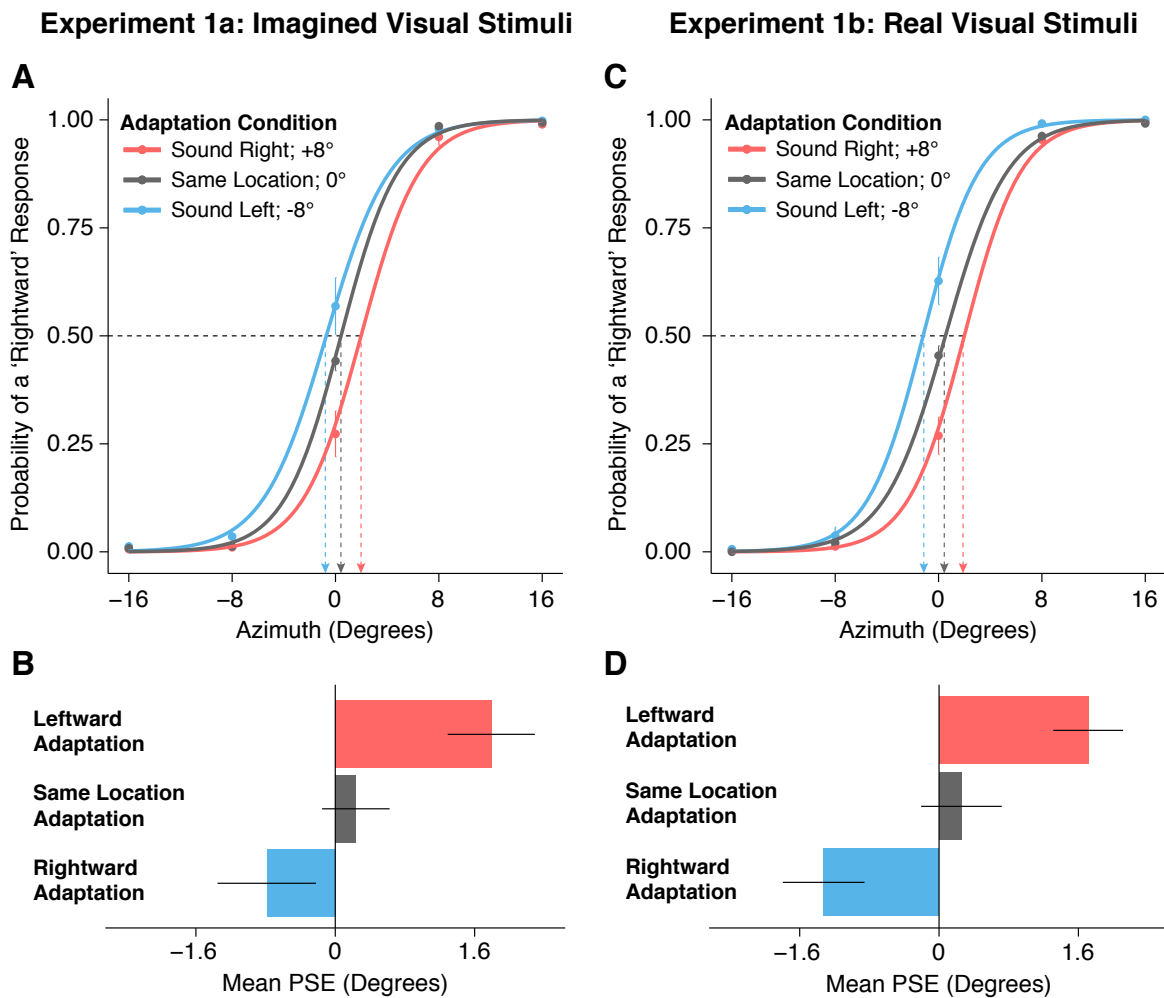
Figure 12. *What I imagine seeing changes where I hear sounds in the future.* The results from **Study IV** demonstrating the imagery-induced aftereffect. Logistic regression curves fitted to the group localization data for the leftward, rightward, and same location adaptation conditions for imagined visual stimuli (**A**), and the bar plot of the mean of the PSEs across participants (**B**). Logistic regression curves fitted to the group localization data for the leftward, rightward, and same location adaptation conditions for the 'real-stimulus' version of the experiment in which visual stimuli were actually presented rather than imagined (**C**), and the bar plot of the mean PSEs across participants for this experiment (**D**). Dotted lines pointing to the x-axis in **A** and **C** indicate the PSE across participants. Error bars denote ± SEM.
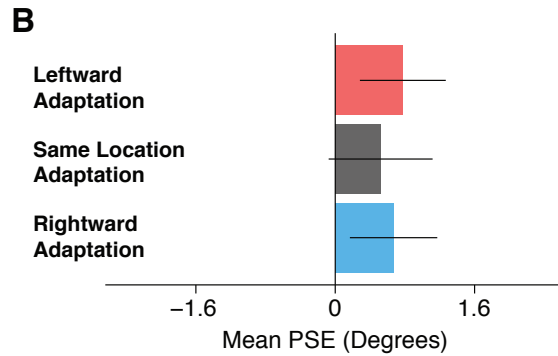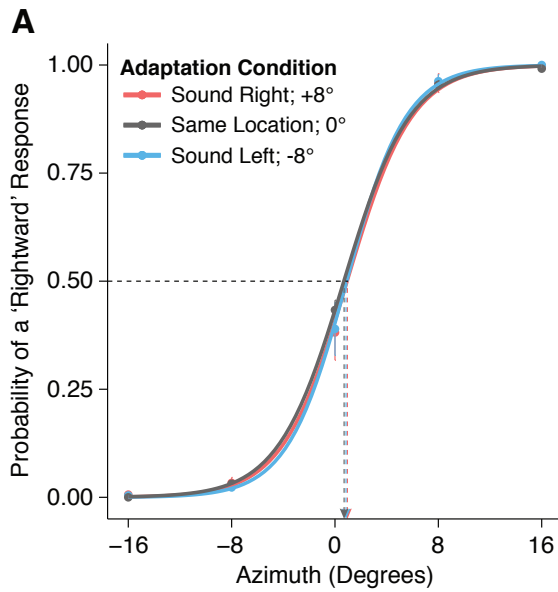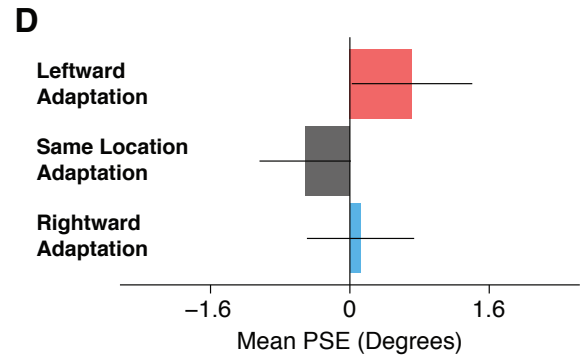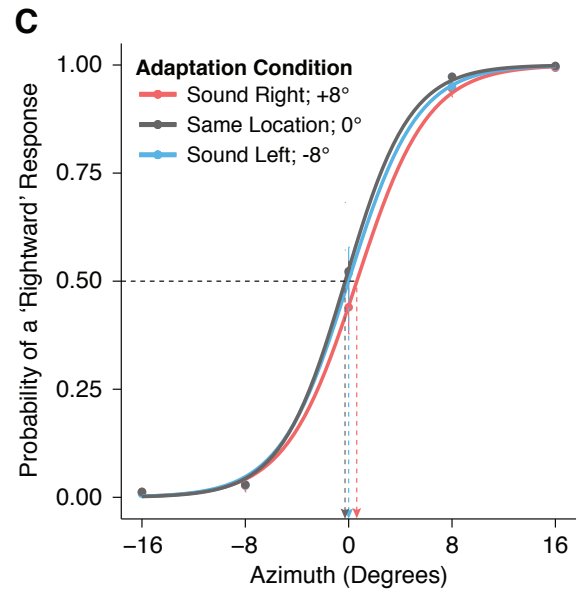
**Experiment 2a: White Noise Aftereffect
4kHz Training with Imagined Visual Stim.**

**Experiment 2b: White Noise Aftereffect
4kHz Training with Real Visual Stim.**

Figure 13. *What I imagine seeing does not change where I hear different sounds in the future.* The results from the control experiments of **Study IV** demonstrating that the imagery-induced aftereffect does not transfer across sound types. Logistic regression curves fitted to the group localization data for the leftward, rightward, and same location adaptation conditions for imagined visual stimuli (**A**), and the bar plot of the mean of the PSEs across participants (**B**). Logistic regression curves fitted to the group localization data for the leftward, rightward, and same location adaptation conditions for the 'real-stimulus' version of the experiment in which visual stimuli were actually presented rather than imagined (**C**), and the bar plot of the mean PSEs across participants for this experiment (**D**). Dotted lines pointing to the x-axis in **A** and **C** indicate the PSE across participants. Error bars denote ± SEM.

5 DISCUSSION

Can what you imagine in one sensory modality change what you perceive in another? Yes, it can! In the experiments within this thesis we have found evidence that imagined sensory stimuli can fuse with incoming sensory signals to change perception of the world around us. In **Studies I** and **Study II,** we found that imagining a sound at the moment two passing objects meet promoted the illusory perception that they bounce off one another rather than cross-by one another. In **Study II** we found that this was only for sounds that simulate the acoustics of a collision. In **Study I**, we also found evidence that imagining an auditory speech stimulus can change visual speech perception. In **Study I** and **Study III** we found that imagining a visual stimulus can change the perceived location of an auditory stimulus. Furthermore, In **Study III** we found that this visual imagery induced change in auditory localization was associated with activity in the superior temporal sulcus (STS), a key region involved in the integration of real audiovisual stimuli. We also found that the visual imagery-induced ventriloquist effect was associated with increased effective connectivity between the STS and the auditory cortex. These activation and connectivity patterns were closely related to those found for the influence of real visual stimuli on auditory perception. Lastly, in **Study IV**, we found that repeated exposure to the imagery-induced ventriloquist effect can change where we hear sounds in the future. Together, these findings make a strong case that what we imagine in one sensory modality can change what we perceive in another.

The idea that mental imagery and perception are similar in kind and rely on overlapping neural representations is not a new idea, and is supported en masse by previous work (Borst & Kosslyn, 2008; Ehrsson et al., 2003; Farah et al., 1989; Farah, 1989b; James, 1890; Kosslyn et al., 2001; Kosslyn, 1994; O'Craven & Kanwisher, 2000; Pearson, Clifford, & Tong, 2008b; Segal & Fusella, 1970). However, this idea has been met with a great deal of criticism since it was first proposed and there are those who have argued that mental imagery is merely the result of some other more basic level of conscious processing (Pylyshyn, 1973, 2002). Although this argument has its original roots in philosophy, and non-perception based theories of mental imagery have few hard-lined supporters today, skepticism about the results from studies on mental imagery has persisted and made its way into the modern psychological and cognitive neuroscience literatures. How do our results fit into this debate? Our results are in line with the behavioral and neuroimaging evidence that suggest that mental imagery and perception involve shared representations (Halpern & Zatorre, 1999; Huang & Sereno, 2013; Kosslyn et al., 2001; Oh et al., 2013); and that because of this, mental imagery has a functional impact on ongoing and future perception (Dils & Boroditsky, 2010; Farah & Smith, 1983; Farah, 1989a; Pearson et al., 2008b; Segal & Fusella, 1970; Winawer et al., 2010). However, in contrast to studies on within-modality sensory interactions, we have been able to demonstrate that these similarities extend to a more ecological, multisensory context. Everyday perception is ordinarily a multisensory process, so much so that some have come to question whether the sole function of the neocortex is to integrate the information from our different senses, and further, whether it was ever reasonable to assume

that the senses operate independently during everyday cognition (Ghazanfar & Schroeder, 2006). Thus, in order to understand the nature of mental imagery and its functional impact on perception clearly, one must take into account the multisensory context in which those mental images are generated and whether they integrated and interacted with our senses in the same manner our senses do normally. In this way, our findings are fundamentally different from previous studies investigating the similarities between mental imagery and perception, as well as the functional impact of mental imagery on perception; and by the same token, fundamentally important. However, vestiges of the classic mental imagery debate (i.e., how perception-like mental imagery is) can still be found today in methodological concerns about whether the results from experiments on mental imagery can be explained by tacit knowledge or response bias.

One of the most common criticisms of behavioral experiments investigating the similarities between mental imagery and perception is that the effects can be explained by response bias, i.e., that the participant knew or guessed the hypothesis of the experiment, and had an explicit strategy to change their response to conform to what we were looking for; or tacit knowledge, i.e., that the participants implicitly knew what the experiment was looking for and inadvertently responded according to the hypothesis because of an implicit bias. These arguments were originally championed by proponents propositional theories of mental imagery to explain away effects which seemed to confirm that mental imagery and perception rely on overlapping representations (Anderson, 1978, 1979; Pylyshyn, 1973, 2002), but have since been co-opted by anyone skeptical of results obtained by experiments on mental imagery or in behavioral experiments in psychology at large. Furthermore, in research on multisensory integration, a common concern when measuring the effect of cross-modal stimuli on perception is the extent to which the circumstances under which the resultant change in perception can be explained by attentional mechanisms alone, rather than by the integration of the cross-modal stimuli (Grassi & Casco, 2009; C Spence & Driver, 1997; Watanabe & Shimojo, 1998). Of course, attention plays an important role in the successful integration of multisensory stimuli (Bachmann, 2011; Gondan, Blurton, Hughes, & Greenlee, 2011; Mishra & Gazzaley, 2012; C Spence & Driver, 1996; Zvyagintsev, Nikolaev, Sachs, & Mathiak, 2011) and so it is not possible to completely preclude its role in studies on human perception. Therefore, it was important to ensure that attention, or explicit or implicit response biases alone cannot explain the behavioral and neuroimaging effects we observe in our studies if we are to make meaningful claims about the integration of real and imagined cross-modal stimuli.

In all the studies in this thesis (**Studies I-IV**), we have taken special care to avoid concerns regarding implicit or explicit response biases, and to minimize the role of attention as an alternative explanation for our results. For studies on the cross-bounce illusion, it was important to rule out (a) that the increased bounce effect was not due to tacit knowledge or response bias, and (b) that the effect could be due to changes in exogenous spatial attention rather than due to the integration of the imagined auditory stimuli with the moving visual stimuli. Thus, in order to compliment the results of **Experiment1A** of **Study I** in which we

demonstrated that imagining a sound at the moment two objects meet promotes an illusory bounce percept, we designed a control experiment to examine whether imagining anything at the moment of coincidence would produce the same illusion. In **Experiment 1B** of **Study 1** we found that imagining the act of lifting the index finger did not produce the illusory effect, suggesting that the effect cannot be explained by general attention or the cognitive task set of performing any kind of imagery. This finding suggested that the resultant change in motion perception found in **Experiment 1A** was due to the cross-modal influence of the sound on visual perception when imagined at the moment of coincidence; however, it is possible that imagining the finger lift and imagining the sound recruit attentional resources differently, as imagining a sound at the moment of coincidence may divert attention to an exogenous sensory stimulus at the moment of coincidence differently than an imagined finger lift, and the resulting change in perception could be due to unavoidable differences in attention when imagining these two very different types of stimuli. Note, that this interpretation of **Experiments 1A & 1B** of **Study I** is unlikely given that imagining a finger tap at the moment of coincidence significantly promoted the illusory bounce percept, whereas the finger lift did not; two almost identical imagined stimuli, except that the imagined finger tap contains a discrete sensory event which is akin to the vibration associated with a collision, and is known to induce the cross-bounce illusion (Shimojo & Shams, 2001b; Watanabe & Shimojo, 1998), and the imagined finger lift does not. However, an additional criticism could be that the participants were responding not based on their genuine change in perception but due to implicit or explicit response strategies, i.e., that the participants knew that we expected them to press bounce when they imagined the auditory stimulus at the moment the discs met and responded accordingly. Concerning explicit response strategies, a funneled debriefing following the methodology detailed by Bargh and Chartrand (2000) revealed that none of the participants had discerned the hypothesis of the experiments. However, in light of these concerns, we wanted to experimentally explore the possibility that these results are contaminated by attentional or response bias.

Previous studies on the cross-bounce illusion have manipulated the content of the auditory stimuli to rule out that the classic version of the cross-bounce illusion could be due to attention alone, and further have manipulated the extent to which the moving discs appear to be bouncing to rule out that the cross-bounce illusion is merely due to a change in response strategies of participants rather than a genuine change in perception (Grassi & Casco, 2009, 2012). Specifically, Grassi & Casco (2009) found that a damped auditory stimulus (which simulates the acoustics of naturally colliding objects) produces the illusion reliably and robustly, but that the same sound played backwards (i.e., a ramped auditory stimulus) does not, despite the fact that the latter is perceived as more perceptually salient (Grassi & Casco, 2009). Furthermore, by manipulating the extent that the discs overlap (and therefore are perceived as bouncing or not) and examining participants' performance in determining whether the discs overlapped or not (i.e., a visual discrimination task requiring attention) Grassi and Casco (2012) were able to examine whether the presentation of a sound led to specific change in attention which could explain the cross-bounce effect, as well as whether

the auditory stimulus produced a specific response strategy, by calculating sensitivity (*d*) and response bias indices (*c*), respectively. Thus in **Study II**, we made use of these techniques to examine whether imagining a damped sound vs. a ramped sound led to content specific changes in visual motion perception, and whether the changes in visual motion perception could be accounted for by attention or response biases alone, by calculating the sensitivity (*d′*) and response bias (*c*) indices from participants' ability to identify whether the discs overlapped or not. We found that imagining a damped, but not a ramped stimulus induced the cross-bounce illusion, and that the extent to which the imagined damped auditory stimulus changed visual motion perception changed as a function of the extent to which the discs overlapped (suggesting that the participants did not base their response on the simple heuristic, "press bounce when imagining the damped sound"). These results were consistent with a real-auditory stimulus version of the experiment conducted in a separate group of participants. Moreover, we found that imagining the damped vs. ramped auditory stimulus did not lead to significant changes in attentional sensitivity (*d′*) nor response biases (*c*) compared simply viewing the stimuli. This was strong behavioral evidence that the imagery-induced motion bounce illusion reflects the integration of the imagined auditory stimulus and the perceived visual stimulus rather than changes in attention or response biases. Taken together, we found that the cross-modal effects of auditory mental imagery on visual motion perception could not be accounted for by changes in visual attention or response biases (Experiment **1A & 2**, **Study II**), and that they conformed to the same temporal (**Experiment 1A, Study I**) and unity assumption (**Experiment 1A, Study II**) principles as multisensory integration of real sensory stimuli.

The classic McGurk illusion is usually immune to concerns over response bias or attention, because the resultant change in perception is a novel percept that is spontaneously reported by the participant rather than shift or bias in perception. However, in order to adapt this experiment to our experimental question, we asked participants to imagine hearing the auditory stimuli and to report what they thought the person in the video was saying (**Study I**, **Experiment 3**); thus, this is an effect of auditory stimuli (imagined in our case) on visual perception—a *reverse McGurk effect* of sorts (Charles Spence & Deroy, 2012; Sweeny et al., 2012). Because the visual stimulus is quite ambiguous, it was necessary to restrict the participants' responses in order to obtain responses consistent enough to examine statistically between the conditions of our experimental manipulation (this is much like restricting the responses in the cross-bounce illusion to 'bounce' or 'cross' as free response could lead to a wide array of descriptions about the motion that are not interesting or relevant, and would make comparing the responses statistically difficult). For this reason, we chose to give the participants the option between the non-illusory percept (i.e., /ga/), and the illusory percept (i.e., /da/). We found that auditory imagery of the incongruent auditory stimulus /ba/ significantly promoted the illusory percept that people in silent videos articulating /ga/ were articulating /da/ compared to when the participants imagined hearing /ka/, an auditory stimulus with a similar articulation. One could argue that this effect may only be due to a propensity to respond /ba/ when participants imagined hearing /ba/ that is independent of the

fusion of the imagined auditory and real visual stimuli (i.e., tacit knowledge); however, if this were the case, then we would see the same response tendency for those who perceived the classic illusion and those who do not (the participants were categorized as McGurk perceivers and non-perceivers on the basis of a post-experiment free-response test for the McGurk illusion). Instead we observed a dissociation between those who perceived the classic illusion and those who did not on their susceptibilities to auditory imagery-induced illusion. This suggests that the participants were responding on the basis of their genuine perception, and one that corresponded to their experience of the fusion of those same audiovisual stimuli when tested with real audiovisual stimuli. Interestingly, the categorization of McGurk perceivers and non-perceivers is a relatively new area of study, as it had previously been generally assumed that this was a universal effect (Gurler, Doyle, Walker, Magnotti, & Beauchamp, 2015; Magnotti et al., 2015; Nath & Beauchamp, 2012). These individual differences in the perceptibility of the McGurk effect have recently been found to be associated with where the participants look when observing others speak—individuals that focus on the mouth tend to perceive the illusion, whereas individuals whose gaze moves around the face tend not to perceive the illusion (Gurler et al., 2015)—above other suspected factors such as one's native language (Magnotti et al., 2015).

In **Experiment 2A** in **Study I,** we found that visual mental imagery led to a significant shift in the perceived location of auditory stimuli when imagined in a disparate location (as measured by the % visual bias), and led to a significant enhancement in auditory localization (as measured by the MEI)—i.e., ventriloquism effect—when imagined in the same location as an auditory stimulus. One possible concern with these findings is that the direction of the perceived shift in auditory localization is always in the direction of the imagined visual stimulus, and therefore, could just be the result of the following heuristic: "wherever I imagine the visual stimulus, click in that location." Thus, the effect could simply be the result of response bias or a shift in spatial attention towards the location of the visual stimulus rather than the perceptual shift due to the integration of the real auditory and imagined visual stimuli. This is a problem that also exists for classic experiments on the ventriloquist effect in which the participants indicate the perceived location of the auditory stimuli by pointing (Bertelson, Frissen, Vroomen, & de Gelder, 2006; Howard & Templeton, 1966; Thurlow & Jack, 1973; Vroomen & Stekelenburg, 2014), and so to address this concern, researchers have employed various psychophysical techniques to examine the ventriloquism effect under conditions in which the hypothesized shift in auditory perception cannot be explained by changes in attention or response bias (Alais & Burr, 2004; Vroomen & Stekelenburg, 2014).

One such technique is a psychophysical staircasing procedure in which auditory stimuli are presented one at a time from two descending staircases (randomly selected). In this procedure, the auditory stimuli are first presented from the far left or right, and gradually converge towards the center as participants correctly report whether they heard the sound come from the left or the right (Vroomen et al., 2001). When the participant makes an error, the next sound presented is from one step backwards in that staircase (i.e., further away from fixation), and this procedure continues until the participant has made a certain number of

response reversals (i.e., responses that were different than the previous response for that staircase). In order to determine whether a ventriloquism effect occurs, a visual stimulus is presented at the same moment as the auditory stimuli but in the center of the screen. The prediction is that because auditory stimuli are translocated towards the visual stimuli, the participants become uncertain about the location of the auditory stimuli earlier on in the staircases, and therefore start to make errors about whether an auditory stimulus came from the left or the right earlier on in the staircase. Importantly, this change in perception is measured from the left/right judgments about the sounds, and the final comparison between conditions is performed on the average distance between left and right staircases. In this way, the task, as well as the final analysis of the data to assess our hypothesis, is completely un-intuitive from the perspective of the participant. Even if they could somehow guess the hypothesis of the experiment, the participants (a) had no idea under what conditions the experiment will end (as the experiment ends once the participant has reached a pre-determined number of response reversals) and (b) had no idea what the other conditions of the experiment are before partaking in that respective condition (i.e., the participant was unaware that they would perform the auditory localization task again and under what circumstances, which they would need to know in order to adjust their response in a manner consistent with our hypothesis). Thus, **Experiment 2B** in **Study I** we employed this technique and found that imagined visual stimuli at the same time as the auditory stimuli, but in the center of the screen, led to earlier response reversals and slower convergence of the left and right staircases—consistent with the notion that the auditory stimuli were translocated towards the imagined visual stimuli—compared to when participants did not imagine the visual stimuli. In this setup, however, we cannot fully rule out that visual attention played a role, as one could argue that imagining the visual stimulus compared to not imagining the stimulus allocates visual attention resources to the center of the screen, and therefore leads to response reversals earlier on because attentional resources have been allocated to the center of the screen and away from the auditory stimuli presented on the left or the right. This is very unlikely, however, because the participants were fixating on the center of the screen in both the imagery and non-imagery portions of the experiment. Moreover, the observed increase in the region of uncertainty in the imagery condition is rather close to the center of the screen (see **Figure 5C**). If the effect were due to increased attentional resource allocation to the center of the screen when imagining the stimuli, one would predict that the effect would occur for sounds presented much further away from the center of the screen rather than the fine-tuned convergence we see in our results. Furthermore, one might also just as well predicted that if the imagined visual stimulus enhanced visual attention to the center of the screen, the participants' left-right discrimination performance should be better in the no-imagery trials than in the imagery trials. Lastly, the behavioral results obtained from the scanner in **Study III** also serve to rule out that spatial attention alone can produce these results, as we found that synchronously imagined visual (and real auditory) stimuli lead to a significantly larger ventriloquist-effect than asynchronously imagined visual stimuli (i.e., two conditions which are matched in attention). Importantly, this result is also consistent with the temporal rule of multisensory integration.

In **Study IV**, we once again made use of the ventriloquism effect but to a new end. In this experiment, we examined whether the translocation observed in the previous imagery-induced ventriloquism studies could produce the same ventriloquism aftereffects observed for real audiovisual stimuli (Frissen et al., 2012, 2005; Jain, Sally, & Papathomas, 2008; Recanzone, 1998; Woods & Recanzone, 2004). Once again, we wanted to rule out the possibility that the participants' responses could be explained by response biases or attentional mechanisms alone. Thus, we designed the experiment in such a way that the predicted shift in perception, would be in the opposite direction of directed attention. That is, the participants always imagined the visual stimulus in the center of the screen, and the auditory stimuli were presented to the left, the right, or the in the same location as the imagined visual stimulus (and to the left, right, or same location as fixation) during the exposure phases (i.e., repeated pairings of spatially discrepant imagined visual and real auditory stimuli). In the test phases, if one had noticed that the auditory stimulus came from the left or the right (although most participants did not because the auditory stimuli were translocated towards the imagined visual stimulus in the center), then the most obvious response strategy would be for them to respond more favorably to the left, and right, respectively when presented with sounds from different locations during the test phase. However, our predicted (and observed) aftereffect was in the opposite direction (i.e., the participants should respond that they heard auditory stimuli in the test phases in the opposite direction of the location of the auditory stimuli in the exposure phases (in left-sound, and right-sound adaptation conditions). These findings rule out that the participants were responding on the basis of implicit or explicit response strategies. Furthermore, in line with classic ventriloquism aftereffect studies which have demonstrated that the ventriloquism aftereffect does not transfer across auditory stimuli of disparate frequencies (Recanzone, 1998; Woods & Recanzone, 2004), we examined whether the imagery-induced aftereffect also showed the same auditory stimulus specificity. We found that the imagery-induced ventriloquism aftereffect did not transfer across different sounds-types, suggesting that, (a) these results cannot be explained by some sort of 'ironic spatial attention mechanism' whereby hearing auditory stimuli to the left or right repeatedly produces a response tendency in the opposite direction independent of the imagined visual stimulus, and (b) that this imagery-induced ventriloquism aftereffect adheres to the multisensory integration principles of real audiovisual stimuli.

Furthermore, we also made use of the ventriloquist effect in **Study III**, but to examine whether the imagery-induced ventriloquism effect was associated with activation in key areas of the cortex involved in audiovisual integration; namely the STS and its neural connections with the auditory cortex. One important consideration of early fMRI experiments on multisensory integration was how to assess whether BOLD activation in response to audiovisual stimuli reflected multisensory integration (i.e., responses that reflect the integration of audiovisual stimuli as were first characterized by electrophysiology recordings in mammals) or merely reflected the activation of two separate populations of unisensory neurons intermingled in the same area (i.e., voxel). In light of the response properties of some

multisensory neurons observed in electrophysiology studies, which responded superadditively to cross-modal stimuli (i.e., more to the presentation of cross-modal stimuli than to the sum of the responses to unisensory stimuli presented alone), it was proposed that this strict threshold be used to dissociate unisensory from multisensory activation in BOLD responses observed in human neuroimaging experiments (Calvert, 2001). Adhering to this criterion, neuroimaging experiments in humans have observed superadditive BOLD responses in the STS in response to audiovisual speech and non-speech stimuli (Calvert et al., 2000, 2001; Stevenson, Geoghegan, & James, 2007; Werner & Noppeney, 2010b). It has been argued, however, that this criterion is too strict and will lead to too many false negatives, and further it is also not entirely immune to mis-characterizations (i.e., superadditivity can be observed as the result in deactivation of a region compared to resting baseline in response to a unimodal stimulus). Thus, when possible, a better characterization of multisensory responses can be obtained in neuroimaging experiments by manipulating the relationship between the audiovisual stimuli along some dimension which would lead to predicted differences in their integration in accordance with the basic principles of multisensory integration (i.e., spatial, temporal, inverse effectiveness) (Calvert & Thesen, 2004; Calvert, 2001). In our experiment, we first manipulated the temporal relationship between real visual and real auditory stimuli, and examined the BOLD responses to synchronous compared to asynchronous audiovisual stimuli in the entire brain. Consistent with previous studies, we observed increased BOLD responses in the STS for synchronously compared to asynchronously presented stimuli. We then performed the same comparison for imagined visual stimuli, and also found increased activity in the STS. Moreover, we performed a PPI analysis and observed increased effective connectivity between the STS and auditory cortex for synchronously compared to asynchronously presented real audiovisual stimuli, and found the same increased connectivity for synchronously imagined visual and real auditory stimuli compared to asynchronously imagined visual and real auditory stimuli. Thus, the same BOLD responses and connectivity patterns indicative of multisensory integration of real audiovisual stimuli were observed for imagined visual and real auditory stimuli as well. Furthermore, in whole-brain regression analyses we found that the behavioral index of the strength of the imagery-induced ventriloquist effect significantly predicted the strength of the BOLD response in the STS and the strength of the connectivity between the STS and auditory cortex for synchronously vs. asynchronously imagined visual and real auditory stimuli. These findings provide strong support for the notion that simultaneously imagined and real cross-modal stimuli lead to multisensory integration, and serve to provide further support for the behavioral evidence of multisensory integration obtained from the other studies in this thesis. Together, the findings from **Studies I-V** provide strong support for perception-based theories of mental imagery, and suggest that the perceptual experience of imagined sensory stimuli is supported by a largely overlapping neural representations of real and imagined sensory stimuli in the human cortex.

What are the limitations of mental imagery-perception interactions? Although we have been able to establish a place for mental imagery in multisensory perception, there are also

certain limitations to cross-modal interactions from imagined sensory stimuli. For instance, in pilot experiments, we examined whether auditory imagery could induce the sound-induced flash illusion (Shams, Kamitani, & Shimojo, 2000) without much success. In the sound-induced flash illusion the presentation of two or more beeps presented at the same time as a single visual stimulus produces the illusory perception that there was more than one visual stimulus presented (Shams et al., 2000). One possible explanation for why imagined auditory stimuli were unable to produce this illusion, is that mental imagery requires more processing time than its bottom up counterpart. This is consistent with behavioral evidence suggesting that it can take between 400 ms and 500 ms for participants to successfully generate a clear mental image from cue until image generation (Weber & Castleman, 1970). In the sound induced flash illusion the auditory stimuli which produce the illusion are presented for only 7 ms and only 57 ms apart (Andersen, Tiippana, & Sams, 2004; Shams et al., 2000). Thus, these rapid stimulus presentations may very well lie outside the scope of what is possible to simulate endogenously. Furthermore, matching the timing of the imagined auditory stimulus in this illusion is extremely difficult as the visual stimuli are only presented very briefly (< 20 ms), and the sound-induced flash illusion is much less forgiving about temporal disparities between the audiovisual stimuli than the other multisensory illusions we have managed to find cross-modal effects of imagery for (Andersen et al., 2004). For example, the sound induced flash illusion is abolished with temporal disparities between the auditory and visual stimulus as small as 100 ms (Shams et al., 2000), whereas the McGurk illusion persists for temporal disparities between the auditory and visual stimuli of up to 200 ms (van Wassenhove, Grant, & Poeppel, 2007).

Can mental imagery be studied in animals? Most of what we know about the neural foundations of mental imagery comes form behavioral and neuroimaging techniques in humans. This is, of course, because of the nature of mental imagery: you cannot just politely ask monkeys, cats, mice, hamsters, or zebrafish (etc…) to imagine things for us. Furthermore, even if we could, it is unclear whether they have the same phenomenological experiences humans do when doing so. However, it has been argued that mental imagery can be evoked clearly and vividly without the conscious intention to do so. Conceptually, this is nicely demonstrated by Wegner's studies on ironic processing—one cannot help but to imagine a white bear, even when explicitly instructed not to (Wegner, 1994). In Wegner's example, whether the result is mental image or conceptual was not of particular interest; however, if I ask you *not* to imagine a spider, you likely just did so, and if spiders bother you, I would wager that it is not the 'idea' of the spider but your quasi-sensorial experience (i.e., your mental image) of it that you find disturbing. One clear example of the automatic elicitation of visual mental images comes from experiments which have found that understanding motion language can produce a visual motion aftereffect (i.e., an aftereffect of visual perception in which static objects are perceived as moving in the opposite direction of previously moving visual stimuli)—suggesting that the linguistic descriptions of visual motion produce sufficiently vivid mental images that a visual motion aftereffect could be induced as a result even without the participants' conscious intention to imagine the visual

stimuli (Dils & Boroditsky, 2010; Winawer et al., 2010). Following this same logic, Schlack and Albright (2007) trained monkeys to associate visual motion of specific directions with specific static objects and then examined whether visual motion-sensitive neurons in the medial temporal cortex (i.e., visual area MT) demonstrated motion selectivity for the previously paired static objects. They found that MT neurons now displayed direction selectivity for the static objects compared to pre-training presentations of those same static objects, suggesting that following training, the static object automatically elicited a visual mental image of the moving stimulus (Albright, 2012; Schlack & Albright, 2007). Of course, whether the monkey experienced this visual image in a similar way (phenomenologically speaking) as the human participants in Dils and Boroditsky's (2010) experiment cannot be verified, and therefore, such findings should be interpreted with caution. Nevertheless, these studies open up an exciting new avenue for examining the neurophysiological basis for mental imagery.

How can our findings be useful in a clinical setting? One possible clinical application of our findings could be for research and development on the use of motor imagery to control robotic prosthetic limbs. If imagined motor stimuli are capable of integrating with real sensory feedback provided by a prosthesis, this could have important implications for the development of neurally controlled prosthetics. Furthermore, the fact that we have found that the integration of real and imagined sensory stimuli can lead to cross-modal plasticity suggests that our findings may be useful in a physical medicine and rehabilitation setting. For example, although research has found that mental imagery of motor movements may not necessarily be more effective than control tasks in restoring motor function in stroke patients (Ietswaart et al., 2011), it may be the case that mental imagery in a different sensory modality (e.g., visual imagery) is easier and more effective in restoring motor function. Furthermore, our findings may be useful in the study of dissociative disorders in which a breakdown in reality monitoring and hallucinations are prevalent; scenarios in which patients fail to distinguish between exogenous and endogenous sensory percepts (Johnson, 2006; Plaze et al., 2011). In schizophrenia, for example, it has been found that auditory hallucinations are the result of patients' inability to distinguish between self-generated speech and external speech (Bentall, 1990; Johns et al., 2001). One prevalent theory regarding the source of positive symptoms in schizophrenia, such as hallucinations, delusions, and confused thoughts and speech, is that they are the result of abnormal interactions or integration between different cortical areas of the brain (Friston & Frith, 1995). Thus, investigating whether schizophrenic patients demonstrate an increased propensity for mental imagery induced multisensory illusions could provide further support for this possibility.

Further research on mental imagery induced multisensory illusions may be useful in examining exactly how it is that we are capable of distinguishing between real and imagined sensory stimuli during everyday perception. Perhaps, for example, the perceived difference in our perception of real and imagined stimuli relies on fine-grain patterns of activity within sensory and association cortices. Alternatively, it may also be the case that mental imagery and perception rely on the same fine-grain patterns of activity in these regions, and that the

differences instead lie with connectivity of the pre-frontal cortex to sensory and association cortices (such a finding would be in line with the proposed mechanism for failures to distinguish between endogenous and exogenous sensory percepts during hallucinations in schizophrenic patients, for example). Future neuroimaging experiments, which make use of additional analysis techniques such as MVPA, may be useful in examining these questions. Future experiments will also be useful in examining the neural basis of the imagery induced ventriloquism aftereffect.

In conclusion, it is clear that mental imagery has an important role to play in perception. More importantly, it also has an important role to play in multisensory perception. Multisensory integration is critical for our perception of the world around us and for our ability to act within it. The fact that mental imagery has a role to play in that process at all, let alone to the extent that we have observed in the studies within this thesis is both curious and fascinating. The studies I have presented here have only begun to scratch the surface of the complicated relationship between what we imagine in our minds and what we perceive in the external world, and I hope that future research in this line will continue to further our understanding of this complicated relationship. Future research will also serve to explore the boundaries between imagination and multisensory perception in order to determine how we are able to distinguish between these otherwise very similar processes.

6 ACKNOWLEDGMENTS

I would like to express my gratitude to the many people who contributed towards making this work possible. First and foremost, I would like to thank my main supervisor **Henrik Ehrsson**, and co-supervisor **Predrag Petrovic**. **Henrik Ehrsson**, special thanks for taking a chance on this crazy American showing up at your doorstep from California asking you (read: *begging you*) to join the lab. You have been the best supervisor a PhD student could ask for. Thank you for not only allowing me, but also encouraging me to explore new and sometimes strange ideas, and for your invaluable guidance in these endeavors.

I would like to express my gratitude to the staff at the MR-Research Center in Solna, especially to **Rouslan Sitnikov** and **Jonathan Berrebi,** for technical support with the fMRI experiments.

Thank you to the many members, past and present, of the Brain, Body, & Self Lab with whom I have had the pleasure to overlap with; **Giovanni Gentile**, **Arvid Guterstam**, **Björn van der Hoort**, **Zakaryah Abdulkarim**, **Malin Björnsdotter**, **Andreas Kalkert**, **Pawel Tacikowski, Valeria Petkova**, **Claudio Brozzoli**, **Loretxu Bergouignan**, **Hiske van Duinen**, **Laura Schmalzl**, **Catherine Preston**, **Konstantina Kilteni**, **Nora Preuss**, and to my summer students, **Zeynep Senveli** and **Josefin Ivarsson**. Thanks to **Martti Mercurio**, for the technical assistance with particularly tricky bits of python code. And thank you to all, for tolerating all the McGurk videos, my rants about the weather, for listening patiently about new data in lab meetings, and for your support and friendship.

To my friends at the Department of Neuroscience, **Giovanni Gentile** and **April Johnston**, special thanks for coming up with creative solutions to opening wine bottles and Thanksgiving dinners in Sweden, and for letting Tanaz and I escape the darkness and sleep on your air mattress in CA when we needed to. **Björn van der Hoort**, thanks for all the inspiring, but sometimes depressing, deep, philosophical, and existential conversations, for brightening my mornings by showing up as an abstract painting of yourself, for all the beers (the many, many, beers), and for figuring out SWISH. Thanks to **Arvid Guterstam** and **Hugo Zeberg** for all the laughs and for tolerating all the bacon-wrapped dates, and to **Zak Abdulkarim** for all the protein bars: the many, many, protein bars.

Thank you to my friends and family here in Sweden at large, especially to all the members of the Molapour/Rastin family, with special thanks to **Ario Molapour**, for your consistent support since moving here, for being a good roommate, friend, younger-brother-from-a-Persian-mother, and travel-buddy, and thanks in advance for being the only one I mention here who gives me a hard time about their place in this list. **Malin Lundahl**, thanks for being there for Tanaz and I over the past five years, for all the conversations about what's next, and for all the adventures past, present, and future. Thanks to **Ivo Toderov** (and Obi and Ines) for Maebee support, and for conversations about all the motorcycles I never purchased (despite your best efforts). Thanks, to all my friends from the Emotion Lab for the good times—with

7 REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology : CB*, *14*(3), 257–62.

Alais, D., van Boxtel, J. J., Parker, A., & van Ee, R. (2010). Attending to auditory signals slows visual alternations in binocular rivalry. *Vision Research*, *50*(10), 929–935.

Albright, T. D. (2012). On the Perception of Probable Things: Neural Substrates of Associative Memory, Imagery, and Perception. *Neuron*, *74*(2), 227–245.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, *21*(3), 301–308.

Anderson, J. R. (1978). Arguments concering representations for mental imagery. *Psychological Review*, *85*(4), 249–277.

Anderson, J. R. (1979). Further arguments concerning representations for mental imagery: A response to Hayes-Roth and Pylyshyn. *Psychological Review*, *86*(4), 395–406.

Angell, J. R. (1911). Imageless thought. *Psychological Review*, *18*(5), 295–323.

Bachmann, T. (2011). Attention as a Process of Selection, Perception as a Process of Representation, and Phenomenal Experience as the Resulting Process of Perception Being Modulated by a Dedicated Consciousness Mechanism. *Frontiers in Psychology*, *2*(December), 1–7.

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. *Handbook of Research Methods in Social and Personality Psychology.*

Bastian, J. (1982). Vision and Electroreception: Integration of Sensory Information in the Optic tectum of the Weakly Electric Fish Apteronotus albifrons. *Journal of Comparative Physiology*, *147*, 287–297.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*, *7*(11), 1190–2.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809–23.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(7), 2414–7.

Bentall, R. P. (1990). The illusion of reality: a review and integration of psychological research on hallucinations. *Psychological Bulletin*, *107*(1), 82–95.

Bertelson, P. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception*, *62*(2), 321–332.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of percieved auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482–489.

Bertelson, P., Frissen, I., Vroomen, J., & de Gelder, B. (2006). The aftereffects of ventriloquism: patterns of spatial generalization. *Perception & Psychophysics*, *68*(3),

428–36.

Bischoff, M., Walter, B., Blecker, C. R., Morgen, K., Vaitl, D., & Sammer, G. (2007). Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia*, *45*(3), 578–86.

Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, *14*(1), 129–133.

Blake, R. (1989). A neural theory of binocular rivalry. *Psychological Review*, *96*(1), 145–167.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews. Neuroscience*, *3*(January), 13–21.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard, S. a. (2007). Neural basis of the ventriloquist illusion. *Current Biology : CB*, *17*(19), 1697–703.

Boring, E. G. (1942). *Sensation and Perception in the History of Experimental Psychology*. New York: D. Appleton-Century Company, Incorporated.

Borst, G., & Kosslyn, S. M. (2008). Visual mental imagery and visual perception: Structural equivalence revealed by scanning processes. *Memory & Cognition*, *36*(4), 849–862.

Borst, G., Kosslyn, S. M., & Denis, M. (2006). Different cognitive processes in two image-scanning paradigms. *Memory & Cognition*, *34*(3), 475–90.

Brown, G. G., Perthen, J. E., Liu, T. T., & Buxton, R. B. (2007). A primer on functional magnetic resonance imaging. *Neuropsychology Review*.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*(2), 369–84.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H.-J., & Jancke, L. (2005). Scanning silence: mental imagery of complex sounds. *NeuroImage*, *26*(4), 1119–27.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience*, *6*(2), 190–5.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex (New York, N.Y. : 1991)*, *11*(12), 1110–23.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology : CB*, *10*(11), 649–57.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, *14*(2), 427–38.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology, Paris*, *98*(1-3), 191–205.

Cave, C. B., Blake, R., & McNamara, T. P. (1998). Binocular Rivalry Disrupts Visual Priming. *Psychological Science*, *9*(4), 299–302.

Chen, J. J., Wieckowska, M., Meyer, E., & Pike, G. B. (2008). Cerebral blood flow

measurement using fMRI and PET: A cross-validation study. *International Journal of Biomedical Imaging*, *2008*(1).

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. *Attention, Perception & Psychophysics*, *75*(5), 790–811.

Cichy, R. M., Heinzle, J., & Haynes, J.-D. (2011). Imagery and Perception Share Cortical Representations of Content and Location. *Cerebral Cortex (New York, N.Y. : 1991)*, (February).

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods*, 257–271.

Cooper, L. a., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. *Visual Information Processing: Proceedings*, *Vi*, 555.

Dahl, C. D., Logothetis, N. K., & Kayser, C. (2009). Spatial organization of multisensory responses in temporal association cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(38), 11924–32.

Danziger, K. (1979). The positivist repudiation of Wundt. *Journal of the History of the Behavioral Sciences*, *15*(3), 205–230.

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*(10), 460–467.

De Renzi, E., & Spinnler, H. (1967). Impaired Performance on Color Tasks in Patients with Hemispheric Damage. *Cortex*, *3*(2), 194–217.

Dils, A. T., & Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 16396–16400.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. *Neuron*, *57*(1), 11–23.

Ehrsson, H. H., Geyer, S., & Naito, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. *Journal of Neurophysiology*, *90*(5), 3304–16.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Farah, M. J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology. General*, *114*(1), 91–103.

Farah, M. J. (1989a). Mechanisms of imagery-perception interaction. *Journal of Experimental Psychology. Human Perception and Performance*, *15*(2), 203–11.

Farah, M. J. (1989b). The neural basis of mental imagery. *Trends in Neurosciences*, *12*(10), 395–399.

Farah, M. J., Péronnet, F., Gonon, M. a, & Giard, M. H. (1988). Electrophysiological evidence for a shared representational medium for visual images and visual percepts. *Journal of Experimental Psychology. General*, *117*(3), 248–57.

Farah, M. J., & Smith, A. F. (1983). Perceptual interference and facilitation with auditory imagery. *Perception & Psychophysics*, *33*(5), 475–8.

Farah, M. J., Weisberg, L. L., Monheit, M., & Peronnet, F. (1989). Brain Activity Underlying Mental Imagery: Event-related Potentials During Mental Image Generation. *Journal of Cognitive Neuroscience*, *1*(4), 302–316.

Finke, R. a, & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *8*(2), 142–147.

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.

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2005). The aftereffects of ventriloquism: Generalization across sound-frequencies. *Acta Psychologica*, *118*(1-2 SPEC. ISS.), 93–100.

Friston, K. J., & Frith, C. D. (1995). Schizophrenia: a disconnection syndrome? *Clinical Neuroscience (New York, N.Y.)*.

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Brain Research. Cognitive Brain Research*, *20*(2), 226–41.

Gentile, G., Petkova, V. I., & Ehrsson, H. H. (2011). Integration of visual and tactile signals from the hand in the human brain: an FMRI study. *Journal of Neurophysiology*, *105*(2), 910–22.

Ghazanfar, A. a, & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–85.

Gibert, G., Fordyce, A., Stevens, C. J., Laboratories, M. A., & Sydney, W. (n.d.). Role of form and motion information in auditory-visual speech perception of McGurk combinations and fusions. *Ethics*.

Goldenberg, G., Podreka, I., Steiner, M., Willmes, K., Suess, E., & Deecke, L. (1989). Regional cerebral blood flow patterns in visual imagery. *Neuropsychologia*, *27*, 641–664.

Gondan, M., Blurton, S. P., Hughes, F., & Greenlee, M. W. (2011). Effects of spatial and selective attention on basic multisensory integration. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(6), 1887–97.

Grassi, M., & Casco, C. (2009). Audiovisual bounce-inducing effect: attention alone does not explain why the discs are bouncing. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 235–243.

Grassi, M., & Casco, C. (2012). Revealing the Origin of the Audiovisual Bounce-Inducing Effect. *Seeing and Perceiving*, *25*(2), 223–233.

Gulyás, B., & Sjöholm, N. (2007). Principles of Positron Emission Tomography Imaging. In F. G. Hillary & J. DeLuca (Eds.), *Functional Neuroimaging in Clinical Populations*. New York: The Guilford Press.

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception & Psychophysics*, *77*(4), 1333–41.

Halpern, a R. (1988). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(3), 434–43.

Halpern, a R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex (New York, N.Y. : 1991)*, *9*(7), 697–704.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51.

Hess E., & Polt J. (1964). Pupil size in relation to mental activity in simple problem solving. *Science*, *143*, 1190–1192.

Hothersall, D. (2004). *History of Psychology* (4th ed.). New York: McGraw-Hill.

Howard, I. P., & Templeton, W. B. (1966). *Human Spatial Orientation*. London: Wiley.

Huang, R.-S., & Sereno, M. I. (2013). Bottom-up Retinotopic Organization Supports Top-down Mental Imagery. *The Open Neuroimaging Journal*, *7*(858), 58–67.

Ietswaart, M., Johnston, M., Dijkerman, H. C., Joice, S., Scott, C. L., MacWalter, R. S., & Hamilton, S. J. C. (2011). Mental practice with motor imagery in stroke recovery: Randomized controlled trial of efficacy. *Brain*, *134*(5), 1373–1386.

Imagination [Def. 1]. (n.d.). In *Merriam Webster Online,* Retreived April 24, 2016, from http://www.merriam-webster.com/dictionary/imagination

Jain, A., Sally, S. L., & Papathomas, T. V. (2008). Audiovisual short-term influences and aftereffects in motion: examination across three sets of directional pairings. *Journal of Vision*, *8*, 7.1–13.

James, W. (1890). *Principles of Psychology*. New York: Holt.

Jastrow, J. (1900). *Fact and Fable in Psychology*. Houghton, Mifflin and Company.

Johns, L. C., Rossell, S., Frith, C., Ahmad, F., Hemsley, D., Kuipers, E., & McGuire, P. K. (2001). Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychological Medicine*, *31*, 705–715.

Johnson, M. K. (2006). Memory and Reality, (November).

Kaas, J., & Collins, C. (2004). The resurrection of multisensory cortex in primates: Connection Patterns that Integrate Modalities. In G. A. Calvert, C. Spence, & B. Stein (Eds.), *The handbook of multisensory processes* (pp. 285–294). Cambridge, MA: Bradford.

Kahneman, D., Tursky, B., Shapiro, D., & Crider, A. (1969). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology*, *79*(1), 164–167.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(11), 4302–11.

Kayser, C., & Logothetis, N. K. (2009). Directed interactions between auditory and superior temporal cortices and their role in sensory integration. *Frontiers in Integrative Neuroscience*, *3*(May), 1–11.

Khanna, N., Altmeyer, W., Zhuo, J., & Steven, A. (2015). Functional Neuroimaging:

Fundamental Principles and Clinical Applications. *The Neuroradiology Journal*, *28*(2), 87–96.

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics*, *14*(1), 90–94.

Kosslyn, S. M. (1994). *Image and brain :the resolution of the imagery debate*. Cambridge, MA: The MIT Press.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Chabris, C. F., Rauch, S. L., & Anderson, A. K. (1994). Identifying objects seen from different viewpoints A PET investigation. *Brain*, *117*(5), 1055.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology. Human Perception and Performance*, *4*(1), 47–60.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews. Neuroscience*, *2*(9), 635–42.

Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: a positron emission tomography study. *NeuroImage*, *6*(4), 320–334.

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. *Nature*, *434*(7030), 158.

Laeng, B., & Sulutvedt, U. (2013). The Eye Pupil Adjusts to Imaginary Light. *Psychological Science*, (November).

Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research*, *166*(3-4), 289–297.

Lehmkuhle, S. W., & Fox, R. (1975). Effect of binocular rivalry suppression on the motion aftereffect. *Vision Research*, *15*(7), 855–859.

Lewald, J. (2002). Rapid Adaptation to Auditory-Visual Spatial Disparity. *Learning & Memory*, *9*(5), 268–278.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–78.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–7.

Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science (New York, N.Y.)*, *245*(4919), 761–763.

Logothetis, N. K., & Wandell, B. a. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, *66*, 735–69.

Lyon, D. C., & Kaas, J. H. (2002). Evidence for a modified V3 with dorsal and ventral halves in macaque monkeys. *Neuron*, *33*(3), 453–61.

Macmillan, N. a, & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185–199.

Magnotti, J. F., Basu Mallick, D., Feng, G., Zhou, B., Zhou, W., & Beauchamp, M. S. (2015). Similar frequency of the McGurk effect in large samples of native Mandarin Chinese and American English speakers. *Experimental Brain Research*, *233*(9), 2581–2586.

Marchant, J. L., Ruff, C. C., & Driver, J. (2011). Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target-detection performance for both modalities. *Human Brain Mapping*, *000*(January), 1–13.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(23), 746–748.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science (New York, N.Y.)*, *221*(4608), 389–91.

Mishra, J., & Gazzaley, a. (2012). Attention Distributed across Sensory Modalities Enhances Perceptual Performance. *Journal of Neuroscience*, *32*(35), 12294–12302.

Mitchell, D. B., & Richman, C. L. (1980). Confirmed reservations: mental travel. *Journal of Experimental Psychology. Human Perception and Performance*, *6*(1), 58–66.

Nath, A. R., & Beauchamp, M. S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(5), 1704–14.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage*, *59*(1), 781–787.

Newman, E. A., & Hartline, P. H. (1981). Integration of visual and infrared information in bimodal neurons in the rattlesnake optic tectum. *Science (New York, N.Y.)*, *213*(4509), 789–91.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(42), 11431–41.

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(3), 598–609.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430.

Nunez, P. L., & Srinivasan, R. (2006). *Electric Fields of the Brain: The neurophysics of EEG. Electric Fields of the Brain: The neurophysics of EEG* (2nd ed.). Oxford Univ Press.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–23.

O'Shea, R. P., & Crassini, B. (1981). Interocular transfer of the motion after-effect is not reduced by binocular rivalry. *Vision Research*, *21*(6), 801–804.

Oh, J., Kwon, J. H., Yang, P. S., & Jeong, J. (2013). Auditory imagery modulates frequency-specific areas in the human auditory cortex. *Journal of Cognitive Neuroscience*, *25*(2), 175–87.

Ooi, T. L., & He, Z. J. (1999). Binocular rivalry and visual awareness: The role of attention. *Perception*, *28*(5), 551–574.

Pearson, J., Clifford, C. W. G., & Tong, F. (2008a). The Functional Impact of Mental Imagery on Conscious Perception. *Current Biology*, *18*(13), 982–986.

Pearson, J., Clifford, C. W. G., & Tong, F. (2008b). The functional impact of mental imagery on conscious perception. *Current Biology : CB*, *18*(13), 982–6.

Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *NeuroImage*, *52*(4), 1410–9.

Peirce, J. W. (2007). PsychoPy--Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13.

Peirce, J. W. (2008). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10.

Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology*, *21*(3), 422–452.

Perrodin, C., Kayser, C., Logothetis, N. K., & Petkov, C. I. (2014). Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(7), 2524–37.

Plaze, M., Paillère-Martinot, M.-L., Penttilä, J., Januel, D., de Beaurepaire, R., Bellivier, F., … Cachia, A. (2011). "Where do auditory hallucinations come from?"--a brain morphometry study of schizophrenia patients with inner or outer space hallucinations. *Schizophrenia Bulletin*, *37*(1), 212–21.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A Critique of mental imagery. *Psychological Bulletin*, *80*(1), 1–12.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *The Behavioral and Brain Sciences*, *25*, 157–182.

Rahmim, A., & Zaidi, H. (2008). PET versus SPECT: strengths, limitations and challenges. *Nuclear Medicine Communications*, *29*(3), 193–207.

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(February), 869–875.

Richman, C. L., Mitchell, D. B., & Reznick, J. S. (1979). Mental Travel: Some Reservations. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(1), 13–18.

Roland, P. E., & Friberg, L. (1985). Localization of cortical areas activated by thinking. *Journal of Neurophysiology*, *53*(5), 1219–1243.

Roth, M., Decety, J., Raybaudi, M., Massarelli, R., Delon-Martin, C., Segebarth, C., … Jeannerod, M. (1996). Possible involvement of primary motor cortes in mentally simulated movement: a functional magnetic resonance imaging study. *Neuroreport*, *7*(7), 1280.

Saenz, M., & Langers, D. R. M. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research*, *307*, 42–52.

Schlack, A., & Albright, T. D. (2007). Remembering visual motion: neural correlates of associative plasticity and motion recall in cortical area MT. *Neuron*, *53*(6), 881–90.

Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Research. Cognitive Brain Research*, *14*(1), 187–98.

Segal, S. J., & Fusella, V. (1970). Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology*, *83*(3), 458–64.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*(6614), 308.

Seltzer, B., & Pandya, D. N. (1994). Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *The Journal of Comparative Neurology*, *343*(3), 445–63.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*(December), 2000–2000.

Shepard, R., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects, *171*(3972), 701–703.

Shimojo, S., & Shams, L. (2001a). Sensory modalities are not separate modalities : plasticity and interactions. *Current Opinion in Neurobiology*, 505–509.

Shimojo, S., & Shams, L. (2001b). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, *11*(4), 505–9.

Small, G. W., Bookheimer, S. Y., Thompson, P. M., Cole, G. M., Huang, S. C., Kepe, V., & Barrio, J. R. (2008). Current and future uses of neuroimaging for cognitively impaired patients. *The Lancet Neurology*, *7*(2), 161–172.

Spence, C., & Deroy, O. (2012). Hearing mouth shapes: Sound symbolism and the reverse McGurk effect. *I-Perception*, *3*(8), 550–2.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology. Human Perception and Performance*, *22*(4), 1005–30.

Spence, C., & Driver, J. (1997). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, *59*(3), 389–403.

Stein, B. E., & Arigbede, M. O. (1972). Unimodal and multimodal response properties of neurons in the cat's superior colliculus. *Experimental Neurology*, *36*(1), 179–196.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience*, *9*(4), 255–66.

Stevenson, R. a, Geoghegan, M. L., & James, T. W. (2007). Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Experimental Brain Research*, *179*(1), 85–95.

Stevenson, R. a, & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–23.

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2012). Sounds exaggerate visual shape. *Cognition*, *124*(2), 194–200.

Szycik, G. R., Stadler, J., Tempelmann, C., & Münte, T. F. (2012). Examining the McGurk illusion using high-field 7 Tesla functional MRI. *Frontiers in Human Neuroscience*, *6*(April), 95.

Thurlow, W., & Jack, C. (1973). Certain determinants of the "ventriloquism effect". *Perceptual and Motor Skills*, *36*(3), 1171.

Titchener, E. B. (1902). *Experimental psychology: A manual of laboratory practice*. New York: MacMillan & Co., Ltd.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular Rivalry and Visual Awareness\nin Human Extrastriate Cortex. *Neuron*, *21*(753-759), 753–759.

Tootell, R. B. H., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From retinotopy to recognition. *Trends in Cognitive Sciences*, *2*(5), 174–183.

Van Essen, D. C., Newsome, W. T., & Maunsell, J. H. (2002). The visual field representation in striate cortex of the macaque monkey asymmetries, anisotropies, and individual variability. *Vision Research*, *24*(5), 429–448.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, *69*(5), 744–756.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*(4), 651–9.

Vroomen, J., & Stekelenburg, J. J. (2014). A bias-free two-alternative forced choice procedure to examine intersensory illusions applied to the ventriloquist effect by flashes and averted eye-gazes. *The European Journal of Neuroscience*, *39*(9), 1491–8.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research.*, *91*(3), 484–488.

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. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *158*(2), 252–8.

Wallace, M. T., Stein, B. E., & Virginia, I. (1994). Cross-modal synthesis in the midbrain depends on input from cortex. *Journal of Neurophysiology*, *71*(1), 429–32.

Watanabe, K., & Shimojo, S. (1998). Attentional modulation in perception of visual motion events. *Perception*, *27*(9), 1041–54.

Weber, R. J., & Castleman, J. (1970). The time it takes to imagine. *Perception & Pschophysic*, *8*(3), 165–168.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*(1), 34–52.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667.

Werner, S., & Noppeney, U. (2010a). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(7), 2662–75.

Werner, S., & Noppeney, U. (2010b). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(8), 1829–42.

Winawer, J., Huk, A. C., & Boroditsky, L. (2010). A motion aftereffect from visual imagery of motion. *Cognition*, *114*(2), 276–284.

Woods, T. M., & Recanzone, G. H. (2004). Visually Induced Plasticity of Auditory Spatial Perception in Macaques. *Current Biology*, *14*, 1559–1564.

Wozny, D. R., & Shams, L. (2011). Recalibration of auditory space following milliseconds of cross-modal discrepancy. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(12), 4607–12.

Wyk, B. C. Vander, Ramsay, G. J., Hudac, C. M., Jones, W., Lin, D., Klin, A., … Pelphrey, K. a. (2010). Cortical integration of audio-visual speech and non-speech stimuli. *Brain and Cognition*, *74*(2), 97–106.

Yendiki, A., Koldewyn, K., Kakunoori, S., Kanwisher, N., & Fischl, B. (2014). Spurious group differences due to head motion in a diffusion MRI study. *NeuroImage*, *88*, 79–90.

Yoo, S. S., Lee, C. U., & Choi, B. G. (2001). Human brain mapping of auditory imagery: event-related functional MRI study. *NeuroReport*, *12*(14), 3045–3049.

Zatorre, R. J., & Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia*, *31*(3), 221–232.

Zvyagintsev, M., Nikolaev, A. R., Sachs, O., & Mathiak, K. (2011). Early attention modulates perceptual interpretation of multisensory stimuli. *Neuroreport*, *22*(12), 586–91.