**Title**
Own Body Perception as Bayesian Causal Inference



**Author**
Samad, Majed

**Publication Date**
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Own Body Perception as Bayesian Causal Inference

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychology

by

Majed Jamal Samad

2016

ABSTRACT OF THE DISSERTATION

Own Body Perception as Bayesian Causal Inference

by

Majed Jamal Samad

Doctor of Philosophy in Psychology

University of California Los Angeles, 2016

Professor Ladan Shams, Chair

This dissertation investigates the principles of multisensory integration that underlie the perception of ownership over one's body. To that end, three experimental approaches have been utilized: 1) an investigation of the rubber hand illusion from the perspective of Bayesian causal inference operating in peripersonal space has indicated that this phenomenon is governed by the same principles of statistical inference that govern perception of external objects, 2) an investigation of the same model formulated to operate in the somatotopic space – that which lines the surface of the body – revealed that the integration of visual and tactile representations is again governed by the same process of causal inference, and 3) an investigation of the malleability of the somatotopic space has revealed that brief exposure to synchronous visual-tactile pairs at different locations can cause a recalibration of that space.

In combination, these three investigations have made use of the Bayesian causal inference model that has been implemented in different ways, in order to model the respective spaces of relevance. Seeking to synthesize a complete account of body ownership, I then proceed to propose a unified account that makes use of the principles of Bayesian causal inference and performs a combined computation that operates in both somatotopic and peripersonal spaces in performing the inference as to which object is my body.

The dissertation of Majed Jamal Samad is approved.

Jamie Feusner

Martin Monti

Dario Ringach

Ladan Shams, Committee Chair

University of California, Los Angeles

2016

DEDICATION

This work is dedicated to the least among us, those that toil without rest, who will have no knowledge of what is contained herein, yet upon whose laboring backs all of this becomes possible.

$\smile\frown$ May all beings be happy $\frown\smile$

EPIGRAPH

*The foot feels the foot*

*when it feels the ground.*

—Buddha

TABLE OF CONTENTS

LIST OF FIGURES

## VITA

| | |
|---|---|
| 2006 | I.B. Diploma, American Community School, Beirut, Lebanon |
| 2010 | M.Eng. in Biochemical Engineering *with Bioprocess Management*, University College London, UK |
| 2012 | M.A. in Psychology *Cognitive Neuroscience*, University of California, Los Angeles |
| 2012-2016 | Graduate Teaching Assistant, University of California, Los Angeles |
| 2016 | Ph.D. Candidate in Psychology *Cognitive Neuroscience with Computational Cognition*, University of California, Los Angeles |

## PUBLICATIONS AND PRESENTATIONS

Samad, M., Shams, L. (submitted). Recalibrating the Body: Visuotactile Ventriloquism Aftereffect.

Samad, M., Shams, L. (2016 poster). Recalibrating the Body: Visuotactile Ventriloquism Aftereffect. Presented at the 2016 Society for Neuroscience annual meeting, San Diego, CA.

Samad, M., Shams, L. (2016). Visual-Somatotopic Interactions in Spatial Perception. *NeuroReport*.

Samad, M., Shams, L. (2015 talk). A Visuotactile Ventriloquist Illusion. Presented at the 2016 International Multisensory Research Forum, Pisa, Italy.

Samad, M., Chung, A., Shams, L. (2015). Perception of Body Ownership is Driven by Bayesian Sensory Inference. *PLoS ONE*.

Samad, M., Chung, A., Shams, L. (2013 poster). Towards a Computational Account of the Rubber Hand Illusion. Presented at the 2013 Society for Neuroscience annual meeting, San Diego, CA.

Samad, M., Chung, A., Shams, L. (2013 poster). The Rubber Hand Illusion as Bayesian Causal Inference. Presented at the 20th annual Joint Symposium on Neural Computation, California Institute of Technology, Pasadena, CA.

Samad, M., Shams, L. (2012 poster). Visuotactile Synchrony is not a Necessary Condition for the Rubber Hand Illusion. *Journal of Vision*.

# Chapter 1

# Background and Introduction

The unity of consciousness is the most immediate fact confronting us, all our experiences being staged coherently on a common stage, seamlessly woven together as a single piece of fabric is from many individual threads. While obvious and immediate it nevertheless remains a great mystery, especially as regards the circuitry of the brain that achieves this unification, despite the large strides we have taken in recent years to shed light on this and related questions. A subset of our perceptual world – arguably the most fundamental – is that which represents our own physical bodies and identifies with them as such. This forms the foundation of the sense of self-consciousness and serves as the nexus about which the rest of perceptual reality turns. It substantializes that ephemeral feeling of being a self and anchors all expressions of subjective experience in bodily phenomenology. Therefore, it would seem that the elucidation of the conditions that engender the appearance of this faculty, as well as those that result in its deviations from the norm, would be of pressing scientific concern.

This field of study situates itself within the broader realm of multisensory perception. When a ventriloquist's speech is perceived as emanating from the puppet's mouth – an

erroneous, though entertaining, percept – the perceptual system shifts its estimate of the location of the auditory source towards the visual signal because this is the inference that is most consistent with the totality of sensory data. In other words, the moving lips of the puppet are inferred to be more likely to have caused the sound than the ventriloquist's motionless lips. The scientific study of these phenomena concerns itself with observing the requirements for the occurrence of such illusions, such as proximity in spacetime or congruence along one of many possible dimensions of comparison.

The rubber hand illusion (RHI) is very well suited for use in the study of body representation because it is robust and easy to induce, as well as being so stark and vivid an experience that it often evokes in the naïve observer hysterical fits of laughter at the absurdity of what is perceived. In brief, the illusion involves positioning a lifelike model of a hand on a table in front of the subject, in a plausible location and posture congruent to that of the hidden real hand, and then applying synchronized brushstrokes to both. The sensory information impinging on the brain is then processed and the resultant inference can be approximated figuratively as: the time-locked brushstrokes are most likely to have originated in the same event, namely the stroking of a single hand, which is therefore the visible fake hand, which must therefore be my hand. This offers a unique window of insight into the mechanisms of multisensory integration that underlie the fundamental way we perceive ourselves. Recently, researchers have even started to propose computational models that can capture the observed behavior of human subjects in multisensory experimental tasks. While there is still considerable debate as to the correct family of mathematical frameworks, researchers are making good progress in identifying and comparing from amongst models that are achieving progressively better fits to collected data.

What follows is a body of work aimed at the characterization of the processes of multisensory integration that underlie several related aspects of body representation. I will begin with an investigation of the computational underpinnings of the rubber hand illusion (RHI) and attempt to explain it by recourse to a Bayesian causal inference model (Chapter 2). As the RHI involves a trimodal process operating in both space and time, I will need to simplify the paradigm to be able to study the phenomenon more systematically. To that end, I will restrict myself to the study of the malleability of visual and tactile spatial representations along the surface of the body in order to arrive at a more quantitative characterization of the Bayesian model (Chapter 3), and then proceed to examine whether these representations undergo recalibration following repeated presentations of synchronous though spatially discrepant stimuli (Chapter 4). At the end, this will comprise a body of work that elucidates the mathematical principles underlying the multisensory interactions the figure into the dynamic spatiotemporal representation of one's own body.

## 1.1 Multisensory Integration

Before we begin our specific discussion of the mechanisms that underlie body representation, it would be helpful to give a brief survey of the major findings from the literature on multisensory interactions. In general, the dominant impetus motivating the study of multisensory interactions derives from a concern regarding the question of how it is that parallel streams of information entering, or being processed by, the brain ever get combined and integrated into a coherent whole, a problem known as "the binding problem" (Revonsuo, 1999).

Robert Welch and David Warren (1980) laid the foundations for the principled behavioral investigation of what they called "intersensory bias", namely the behavioral effect of combining two discrepant signals leading to an estimate that is slightly different from one or both. They examined several competing theories to account for the effect, some of which are in fact still under debate at present, such as: (1) the modality appropriateness hypothesis, which posits that information from a modality that the nervous system deems most appropriate to the task at hand is overweighted regardless of the noise in the signal, and (2) the modality precision hypothesis, which states just the opposite, that signals are weighted entirely as a function of the amount of noise corrupting them. After conducting a thorough and exhaustive review of the literature at the time, they synthesized an account that occupies a sort of middle ground between these alternatives, bringing them both to bear on the inference process, and also suggests an additional factor, what they called the "unity assumption" that represents the subject's predisposition to regard the situation as representing a single or multiple event(s) (Welch and Warren, 1980). As we shall see in subsequent sections, this was exactly the insight that paved the way for the Bayesian causal inference models with which I will be most concerned.

As mentioned earlier, this field of study aims at uncovering the spatial, temporal, and other factors that influence whether or not stimuli are to be integrated or segregated. The seminal electrophysiological studies in this regard come from the heroic efforts of Meredith and Stein (1986) who recorded from the superior colliculus of the live anesthetized cat as it was presented with spatially coincident stimuli that were either unisensory or multisensory and observed a striking superadditivity of the response in the multisensory case (Meredith and Stein, 1986). Avillac et al. (2007) extended this work to the monkey ventral intraparietal

cortex and reported broadly analogous findings in terms of the nonlinearity of the combined neuronal responses. A seminal Nature Neuroscience review article emerged in 2008 that set out the principles that had been discovered from all the work that had been done all the way from single-unit recordings up to neuroimaging in humans. In brief, the authors summarized that for multisensory integration to occur, the stimuli must be coincident in space and time, these principles being known as the spatial rule and the temporal rule, respectively (Stein and Stanford, 2008). Additionally, another principle was expounded, namely the law of inverse effectiveness, which states that multisensory facilitation is greatest when the individual signals are least effective.

The field rapidly progressed to the study of these phenomena as they occur in human perception when experimenters began to study the famous old trick known as ventriloquism, which we will begin to discuss in more detail in the next section.

## Spatial Factors

Ventriloquism, the process by which auditory signals are perceived to emanate from different spatial source than they do in reality, relies on multisensory interactions that engender this interplay between the auditory and visual spaces. Several researchers have begun to elucidate many of the conditions required to elicit this strange phenomenon. Specifically, a spatial disparity between crossmodal stimuli may be overcome and the stimuli bound if and only if there is at least one other dimension along which the stimuli correspond (Hairston et al., 2003). Additionally, further research has demonstrated that this illusory effect is critically dependent on the perception of the unity of the signals, which the authors as-

sayed explicitly, and moreover that a contrasting repulsive effect occurs when such unity is not perceived (Wallace et al., 2004). Research has shown that this effect can be successfully modeled as a near-optimal computation involving maximum likelihood estimation on sources with signals being weighted by their reliabilities, a quantity that is equal to the inverse of their variance (Alais and Burr, 2004).

Researchers have naturally attempted to extend these paradigms to other combinations of modalities, and an audiotactile ventriloquist illusion was subsequently discovered wherein the tactile modality "captures" the auditory estimate of position along azimuth (Caclin et al., 2002; Bruns et al., 2011; Renzi et al., 2013). In the pioneering study that discovered this, Charles Spence's group demonstrated that auditory spatial localization judgments in **external** space along azimuth could be biased by vibrotactile stimulation of the index fingers when the arms were outstretched (Caclin et al., 2002). An extension of this work revealed that the direction of biasing reversed upon crossing the arms, indicating that the illusion was in fact operating in external, and not body-related or anatomical, spatial coordinates (Bruns et al., 2011). To the best of my knowledge, no study has yet examined similar ventriloquist-like spatial biasing along the surface of the body, which was therefore a question I myself set out to investigate, and will discuss in far greater detail in the chapters to follow (see Chapter 3).

## Temporal Factors

While vision is a very reliable modality with regards to spatial estimates, given that it has a spatially coded topography in the cortex and that its principle input is spatial in nature,

it does not have comparably good resolution for temporal information. Therefore, when experimental arrangements are produced so as to provide dual channels to similar information that is highly temporally constituted, the auditory modality will dominate, in keeping with the modality appropriateness hypothesis, and the visual modality will recalibrate. The prime example of just such a setup is what is known as the sound-induced flash illusion and its variants.

The sound-induced flash illusion, discovered in 2000 by researchers at the California Institute of Technology, is a well known instance of auditory dominance over a visual judgment (Shams et al., 2000). The paradigm is simple enough to implement: flashes and beeps are presented in various stimulus configurations and the subject is asked to judge the number of flashes that were seen. When the number of beeps exceeds the number of flashes, subjects' visual numerosity judgments are either dictated entirely by, or biased towards, the number of beeps. In an extension of this work, the authors also found that the temporal window of integration, i.e. the maximum stimulus onset asynchrony that permits the illusory experience was $\sim 100ms$, which was consistent with previous single cell recordings (Shams et al., 2002). This finding is in keeping with the modality appropriateness hypothesis insofar as the auditory modality is a much more reliable channel of information in the temporal domain than the visual modality. However, this effect only seems to occur when the beeps are more in number than the flashes, and this asymmetry has thus far resisted satisfactory explanation. Finally, a more recent study has even revealed that feedback training with repeat exposures to the illusion does nothing to restore veridical perceptions, and therefore that the illusion is robust to decision factors and high-level cognitive biases (Rosenthal et al., 2009).

Subsequent to the discovery of the sound-induced flash illusion came its extension to other

modality pairings, specifically what came to be known as the touch-induced flash illusion. In this version of the illusory experience the biasing of the numerosity judgment is occasioned by the greater temporal acuity of the somatosensory modality in comparison with that of the visual modality (Violentyev et al., 2005). Finally, an experiment was conducted where all three modalities were used in a parametrically manipulated design that was used to support the notion that human perception follows optimal statistical inference using a computational model of such that will be discussed in more detail below (Wozny et al., 2008).

## Recalibration

One of the interesting consequences of multisensory interactions is their tendency to engender an effect known as "recalibration". This occurs when a perceptual map shifts its alignment with perceptual maps from other modalities whenever misaligned information gets integrated. This can occur in either spatially or temporally encoded mappings, and these will be discussed in what follows.

### Spatial Recalibration

The ventriloquist aftereffect is a phenomenon observed subsequent to exposure to spatially disparate audiovisual stimuli, as in the cases of ventriloquism discussed above, where auditory localization judgments are shifted in the direction of the visual stimulus. The first psychophysical report of this effect comes from Gregg Recanzone's paper for a PNAS colloquium in 1998 in which he claimed to have found evidence that 20-30 minutes of exposure to 8° disparate audiovisual stimuli was sufficient to produce an enduring shift in the auditory representations (Recanzone, 1998). Noteworthy is that the temporal synchrony of the audio-

visual stimuli was a necessary condition for the occurrence of the aftereffect, as indeed it was seen to be for the occurrence of the ventriloquist illusion itself. This is interesting in that it indicates that the stimuli must correspond on some dimension for their successful binding to overcome a lack of correspondence on another dimension. In an extensive replication of this initial finding, Lewald sought to broaden its validity and examine it in greater detail. Apart from a verification of the finding as well as its extension to a larger set of spatial disparities and auditory tone frequencies, Lewald also found evidence that the aftereffect is specific to the frequency of presented tones and did not transfer to other frequencies (Lewald, 2002). Interestingly, a subsequent study found robust generalization across frequencies in stark contrast to these initial findings, with implications regarding the central or peripheral neural underpinnings of this recalibration (Frissen et al., 2003). Finally, and building on the findings of Wallace et al. (2004) that observed a dependence of the illusion on the perceived unity of the signals, this effect was explained using a Bayesian causal inference model that will be discussed in detail in later sections of this work wherein the auditory representations are encoded as Gaussian distributions with means shifted in the direction of visual representations (Wozny and Shams, 2011a).

Spatial recalibration need not be restricted to audiovisual remappings, and in fact has been investigated with respect to visuo-proprioceptive recalibration. Insofar as this has relevance to the interactions that figure into the generation of the body representation, these studies merit careful investigation. Using either prisms (Baily, 1972; Held and Hein, 1958) or psychophysical stimulus presentation techniques (Bedford, 1989), these studies have shown unequivocally that the visuo-proprioceptive maps that inform reach movements do indeed recalibrate upon extended exposure to conflicted pairings, and that this adaptation

is enhanced when movements in the new mapping are made, but that the adaptation does not generalize to the overall body representation and is limited to the limb and/or posture that has been adapted.

**Temporal Recalibration**

We have been discussing the intriguing effect that occurs when spatially encoded maps shift their alignment with respect to each other to overcome a spatial disparity. However, a corresponding effect for temporally encoded information has been recently found that is of equal import, though it has received less attention in the literature. Specifically, Fujisaki and colleagues reported in 2004 that subjects exposed to audiovisual stimuli with a constant lag between them exhibited a commensurately shifted point of subjective simultaneity (PSS), which, however, was only roughly 10% of the total lag that they were exposed to (Fujisaki et al., 2004). Importantly, this initial study could not disentangle which modality was specifically being recalibrated as the judgments were inherently crossmodal, being judgements regarding the simultaneity of crossmodal stimuli. A similar study from the same year reported broadly consistent findings, with the exception that the reported maximal exposure lag that produced the recalibration was $\sim 100ms$ whereas in the former it was larger: $\sim 235ms$ (Vroomen et al., 2004). It is important to note that congruence in the non-shifted dimension is essential for the realization of this effect, as it was in the case of spatial recalibration. In other words, the stimuli must be spatially coincident if their temporal non-coincidence is to drive recalibration.

A similarly intentioned study was conducted to examine whether such an effect occurs in audiotactile integration. It was observed that although there was indeed an aftereffect of

asynchronous presentation, this was not in terms of a shift in PSS but rather a widening of the just noticeable difference (JND), which the authors explain as being potentially due to the lack of lifelong exposure to the asynchronies in this modality pairing in contrast to the situation with audiovisual integration, which lead to an increase in the temporal window of integration (Navarra et al., 2007).

## Higher Order Multisensory Integration

Thus far we have been discussing the integration of information from low-level spatial and temporal maps, such as flashes and beeps. However, similar forms of multisensory integration have also been observed with higher level percepts, such as speech sounds and ownership over body parts. Two notable examples of this form of multisensory integration will be used for illustration: the McGurk effect whereby phonemes are distorted due to mismatched visual and auditory stimuli, and the rubber hand illusion whereby hand ownership judgment is distorted by mismatched visual and somatosensory stimuli.

The McGurk effect was first reported in 1976 in a Nature brief report detailing how mismatched audio (/ba/) and visual phonemes (/ga/) are combined to produce a percept of intermediate phonetic identity (/da/). This effect is explained as arising from the life-long experience with both modalities of speech perception and thereafter the expectation of matching phonemic identity, which leads to a perceptual averaging effect as is observed in simpler audiovisual integration paradigms (Mcgurk and Macdonald, 1976). Further, attention seems to play a crucial role in this perceptual averaging, shown by a recent dual-task study that had subjects perform a concurrent auditory or visual task that depleted at-

tentional resources and observed that subjects' perception of the McGurk phonemes was severely suppressed under such conditions, implying that it is not as automatic and early as was initially suspected (Alsius et al., 2005).

Another perceptual illusion that seems to depend on higher-level as well as spatiotemporal factors is the rubber hand illusion, first reported by Botvinick and Cohen in a Nature brief communication in 1998. Briefly, the authors described the induction of the illusion as being dependent on placing a lifelike model of the hand in a position congruent to the arm, deviated only in the azimuth dimension and with the real hand occluded from vision by a barrier, and the application of synchronized brushstrokes to both hidden real hand and visible rubber hand (Botvinick and Cohen, 1998). This produces the illusory feeling of ownership over the rubber hand, involving a remapping of visuoproprioceptive space similar to what occurs in ventriloquism with audiovisual maps. This seems to show that the perception of our own bodies relies on the same principles of multisensory integration that were thought to operate only at lower levels. Subsequent studies have revealed that this illusion is even accompanied by a physiological anxiety response when the owned rubber hand is threatened (Armel and Ramachandran, 2003).

## Body Representation

Before disembarking from the station of multisensory interactions, we shall discuss one last topic that has vexed cognitive researchers for decades, with recent work suggesting it may prove amenable to the same techniques that have been utilized for the study of multisensory interactions. I am referring to the study of the body representation, not to be confused with

the somatomotor homunculi that straddle the central sulcus, but rather the more integrated sense of inhabiting a physical, corporeal object that one feels both a sense of agency over, as well as presence within (see Herrera et al. (2006) for a discussion of the notions of agency and presence).

### Body Schema and Body Image

In the early twentieth century, researchers sought out to propose a conceptual classification scheme for the various kinds of body representations, seeming as they did to dissociate under certain circumstances and to serve sometimes dramatically different purposes. There are several such classification schemes, but by far the most influential is that which splits the body representations into a so-called "body image" and a "body schema", the former being the representation that is more visual (the way the body looks like from the outside) and the latter being that which is more sensorimotor (the way the body feels from the inside). This division is along broadly analogous lines to the more general action/perception dichotomy in mental representations in general. The clinical cases where this proposed taxonomy reveals its merits are deafferentation and numbsense. In the former, patients are stripped of their abilities to use body representations to guide actions, and as such are deficient in their body schemas. In numbsense, patients are no longer able to identify which part of their bodies has been touched, but can still make accurate movements, and are therefore impaired in their body image (de Vignemont, 2010; Head and Holmes, 1911).

### Embodied Cognition

In recent years, there has been a growing interest in using Embodied Cognition as a framework for understanding many – or, depending on who you ask, all – cognitive processes in the brain. This framework claims that the body and its needs and peculiarities are of fundamental importance in shaping our cognitive processes. One example of the type of process that is described here is the acquisition of action verb concepts, which are proposed to be derived from a mental simulation of the actual motor output that would be performed in said action verb. This theory remains controversial and does not enjoy universal support, though some of its claims are well verified, such as that "cognition is for action" by the established "where" stream in visual perception (Wilson, 2002).

### Peripersonal Space

While we are in the process of conducting this brief survey of the multisensory interactions that figure into the body representation, we will give an introductory treatment of the notion of peripersonal space, which has become so central to all discussions of the body representation of late. This topic will be further discussed at length in Section 1.5 below.

Peripersonal space refers to the idea that there exists a privileged region surrounding the body wherein information processing is prioritized, and which is the functional area within which the body is capable of defending itself and acting upon the world. This is a relatively new notion that has emerged with the findings from the initial studies that discovered bimodal neurons in monkey parietal and frontal cortices, which respond to visual and tactile stimulation, and whose visual receptive fields (RFs) extend out into space adjacent

to the tactile RFs (Rizzolatti et al., 1997). Moreover, these visual RFs are anchored to the tactile RFs, meaning that they are very rapidly and dynamically remapping to remain associated with their preferred part of the body, even as the body and/or the eyes move freely. Even more intriguingly, Graziano has observed that these neurons will spontaneously remap onto a visible dummy hand if the real hand is hidden and will have visual RFs that will remain anchored thereto (Graziano et al., 2000). In addition, tool use has been shown to be able to extend the peripersonal space, which has therefore been hypothesized to be a functional map of the body, demarcating the boundary within which it is possible to act upon the world (Ladavas, 2002). Avillac et al.'s study (2007), mentioned above in Section 1.1, recorded from neurons in monkey area VIP and also found evidence for these bimodal neurons in the parietal lobe, showing that they integrated visual and tactile targets in a similar fashion to those recorded by Graziano (Avillac et al., 2007). Taken together, these findings lend credence to the possibility of visuotactile interactions that resemble ventriloquism as the neurophysiology of these neurons concords with the audiovisual neurons Meredith and Stein recorded from in the superior colliculus, albeit operating in a different spatial reference frame – body-centered peripersonal space as opposed to external audiovisual space.

### Body Matrix

While many theories exist regarding the importance of, and delineating various functions of, the body representations, we shall not have to survey them all, the preceding discussions having given us a good deal of the flavor of the rest. However, it would be quite remiss of this survey if it did not include the modern approach to theorizing about body representations as being a result of multisensory integrative processes, with a stronger thrust to including

homeostatic mechanisms in the analysis, as well as interoceptive systems and their relationship to emotional processing. Lorimer Moseley and colleagues (2012) defined a theoretical construct that they called the "body matrix", which was purported to capture the notion of peripersonal space and explain it as arising out multisensory interactions involving proprioception, vision, and touch, and which is coded in body-centered coordinates. Moreover, it is said to include homeostatic regulatory functions and to coordinate between this and the cognitive representations of the body, thereby contrasting quite dramatically from the old body image/schema dichotomy (Moseley et al., 2012).

**The Importance of Interoception**

Related to the foregoing discussion, is the proliferation of theories of awareness that implicate the body representations as being of fundamental import. The most popular of these is Bud Craig's theory of the "sentient self", claiming that the anterior insula contains a re-representation of the interoceptive sensations from the internal milieu upon which is overlaid to-be-integrated information arriving from many other brain areas. In this way, he claims that this representation provides the basis for the subjective feeling of what he calls the "global emotional moment", and that if this were to be modular, would account for the so-called stream of consciousness in its trajectory through serial repetitions of the module (Craig, 2010). A very related account has been proposed by another group of researchers and differs only its specification of the processes computed in the insula as being interoceptive inference, and the further elaboration on the arising of the separate subjective notions of agency and presence (Seth et al., 2012).

## 1.2 Rubber Hand Illusion

Having thus surveyed the prominent experimental paradigms of multisensory interactions, we can extend our foray into the intersection of this field with that concerning the representation of one's own body. The primary tool for the study of this phenomenon is the rubber hand illusion. For perception scientists, illusory experiences represent great assets in unveiling the mechanisms of normative experience insofar as they function like microelectrodes that can patch clamp onto a phenomenology and record traces of its modulation by a variety of experimental conditions.

### Initial findings

Botvinick and Cohen reported two measures of the illusion: drift in the proprioceptively localized position of the hidden arm in the direction of the seen rubber hand, as well as positive scores on questionnaires that directly assess the degree of the feeling of ownership over the rubber hand (Botvinick and Cohen, 1998). Subsequent to this, as mentioned above, Armel and Ramachandran discovered that the illusion was accompanied by a physiological response measured by increases in skin conductance in response to threatening stimuli directed at the rubber hand, but only when it was owned and not otherwise (Armel and Ramachandran, 2003). Moseley and colleagues reported in 2008 that the skin temperature of the stimulated hidden hand dropped significantly upon self-attribution of the rubber hand as compared to the unstimulated hand, a surprising finding suggesting a disownership of the real hand as a result of the illusion, though we have heard anecdotal reports that this particular finding is difficult to replicate (Moseley et al., 2008). Another study examining the

influence of anatomical and postural factors on the illusion, showed that the rubber hand must be in a position that is both anatomically plausible and congruent with the real hand's posture in order for the illusion to occur (Tsakiris and Haggard, 2005). The illusion has been also reported to be sensitive to spatial mismatches in hand-centered space, whether these mismatches occurred in tactile stimulation vectors or in hand position (Costantini and Haggard, 2007). Finally, a study that addressed the question of the spatial limits of the RHI observed that the illusory reports dropped off significantly after a distance $27.5cm$ intervened between the real and rubber hands, and was at a minimal level after that, indicating a spatial zone with a non-linear boundary surrounding the body where objects to-be-embodied are preferentially processed, in support of the peripersonal space hypothesis (Lloyd, 2007).

## Neuroimaging and Stimulation

By now there have been several neuroimaging studies of the rubber hand illusion that have reported broadly concordant findings. The first of these reported that the ventral premotor cortex shows activity in fMRI scans that correlates with the experience of the illusion, as well as revealing a network of regions involving the intraparietal sulcus and cerebellum that are involved in the recalibration period leading up to the illusion (Ehrsson et al., 2004). A subsequent study by the same group found similarly that the ventral premotor cortex was bilaterally involved, and moreover expanded upon these studies by indicating the involvement of the insula and anterior cingulate cortices in the generation of the anxiety response to threat of the hand, modulated by the extent of ownership (Ehrsson et al., 2005). Additionally, Tsakiris et al. (2007) conducted a PET study that manipulated the synchrony

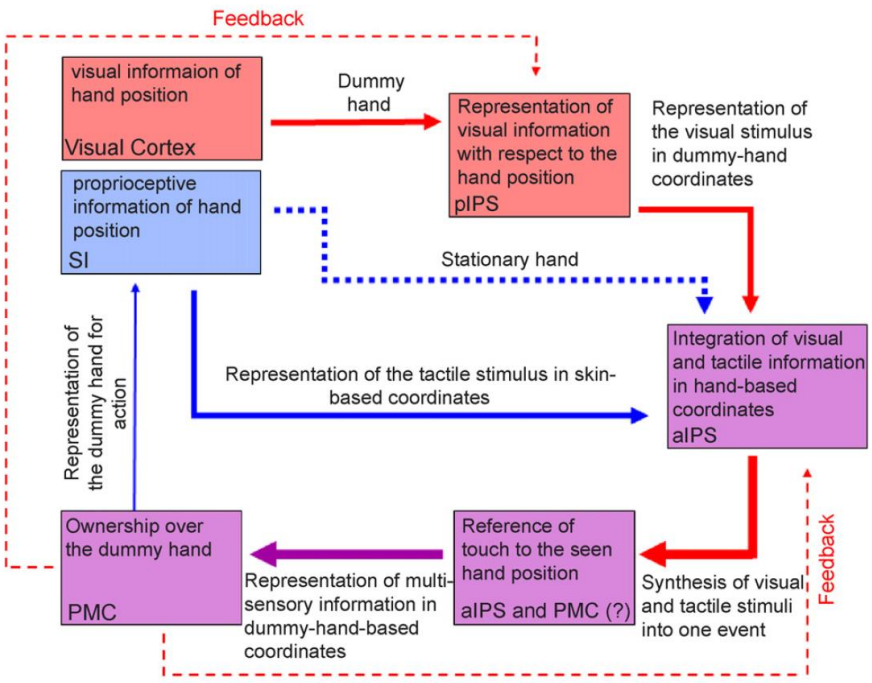


Figure 1.1: Makin et al. (2008)'s peripersonal space model of the RHI

of visuotactile stimulation of a rubber hand and showed that activity in the right posterior insula was correlated with proprioceptive drift and, moreover, that failure to illicit the illusion was associated with activity in somatosensory cortices. An EEG study examining the effects of synchronous visuotactile stimulation on somatosensory ERPs was conducted and found an enhanced N140 in this condition compared to one where visuotactile stimulation was uncorrelated (Press et al., 2008). Finally, a recent exhaustive fMRI study of the neuronal correlations of body-related multisensory integration and disintegration across the visual, tactile, and proprioceptive modalities confirmed the previous findings that a network of areas is responsible for the integration that occurs, namely a network spanning premotor cortices, intraparietal cortices, and cerebellar regions (Gentile et al., 2013).

## Cognitive Models

With regard to the underlying framework accounting for the illusion specifically, and body representation more generally, several qualitative neurocognitive models have been proposed. The first of these emerged from a sweeping and thorough review of the literature and can be outlined as follows (see Figure 1.1): a mechanism of peripersonal space remapping that depends upon an initial visuo-proprioceptive integration step is purported to occur somewhere in the posterior parietal cortex, followed by a referral of all visual stimuli occurring near the dummy hand to the dummy-centered coordinates thereby computed and the eventual feeling of ownership over the dummy hand that this produces, purported to occur in the premotor areas (Makin et al., 2008).

A subsequently proposed model built upon Makin et al.'s by including a role for top-down constraints on objects to be embodied (see Figure 1.2) – proposed to be occurring in the temporoparietal junction – and proposing that the recalibration of visual and tactile maps occurs in posterior parietal and ventral premotor cortices, as in Makin et al., and finally that the right posterior insula is the seat of the subjective experience of ownership (Tsakiris, 2010). From a slightly different approach, Graziano and Botvinick (2002) wrote a review that attempted to bridge the often disparate neuropsychological and psychological conceptualizations of the body representation, many of the themes of which have been discussed in Section 1.1 above. They propose a model that is very similar to the peripersonal space model described above, in that it describes the body representation as being composed of the integration of low-level joint position information with crossmodal information about the body and a top-down representation of the body, namely the the body schema, resulting

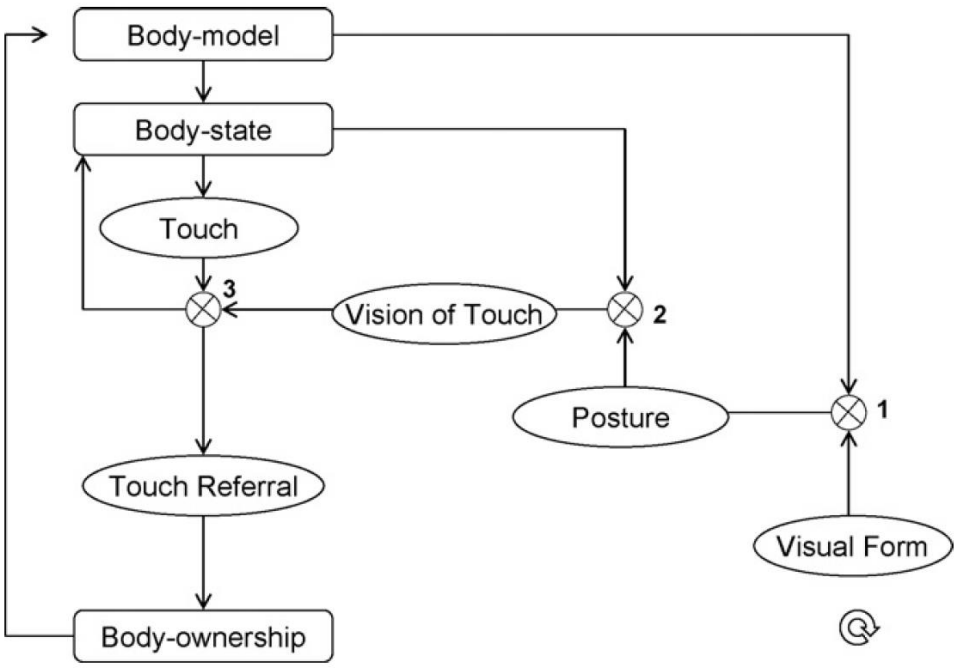


Figure 1.2: Tsakiris (2010)'s neurocognitive model of body ownership

in a coherent zone surrounding and representing the feeling of inhabiting the body. On a related note, Macaluso and Maravita (2010) presented a sweeping survey of the literature and emphasized the dynamic nature of the body representation, looking at specific examples of how tool use remaps peripersonal space, and how dummy hands very rapidly remap the visual-tactile correspondences.

## Recent findings

More recently there has been a flurry of interest in the continued study of the illusion and many fascinating and often conflicting reports have emerged from this latter wave of research. One such study from 2009 revealed the temporal binding window for the multi-sensory integration that underlies the RHI (Shimada et al., 2009). The authors used video

techniques to introduce various levels of visual feedback delay and assess the effect that this had on ownership ratings and proprioceptive drift, and reported that $300ms$ was the largest delay that permitted the illusion, and that longer delays caused significant reduction in both measures. Holle et al. (2011) showed two years later that it was possible to dissociate proprioceptive drifts from ownership reports, a surprising finding that cast doubt on previous studies that assumed these to always be assaying the same underlying neural process. Specifically, the researchers rotated the rubber hand by 180° and found significant proprioceptive drift when stroking was synchronous, but no ownership reports. A very important study was conducted in 2012 to assess whether the susceptibility to the RHI – a very poorly understood aspect of the illusion – is a trait-like or a state-like variable by studying its long-term stability (Bekrater-Bodmann et al., 2012). This study found firstly that the subjective ratings were stronger when the illusion was induced in a vertical setup as compared to a horizontal one (i.e. when the spatial disparity between the real and rubber hands was vertical instead of azimuthal), and that these were correlated across two sessions conducted six months apart. In addition, activity in the ventral premotor cortex was correlated with the subjective strength of the illusion and across-sessions. However, the proprioceptive drift was not correlated across sessions, suggesting yet again that this may not be the ideal behavioral proxy of the RHI. One more piece of evidence supporting the dissociation of these two measures comes from another study from 2012 that attempted to induce supernumerary limb rubber hand illusions – i.e. using two rubber hands – and found that while the subjective feeling of ownership was present for both dummy limbs when they were stroked in synchrony with the real hand, proprioceptive drift was abolished by this highly non-ecological experimental setup (Folegatti et al., 2012).

Another interesting recent study demonstrated that RHI induction on the left hand reduces the pseudoneglect that is observed in non-pathological individuals when asked to bisect a straight line (Ocklenburg et al., 2012). Finally, a very recent paper reported the interesting effect that merely expecting a tactile event to occur on a rubber hand that is positioned appropriately is enough to induce an illusion as indexed by skin conductance responses, and that merely looking at a rubber hand does not suffice, but that a stimulus must fall within peripersonal space, remapped or not (Ferri et al., 2013). In a concerning conflict of results, I (see Chapter 2) have recently obtained evidence for precisely this latter effect, namely that the RHI can occur in the absence of vision (Samad et al., 2015).

## Motor-Induced Variant

Since the first demonstration of the RHI, researchers have examined variants of it to attempt to shed light on its generalizability. One notable variant is that which utilizes synchronized actions made by the subject rather than passive touches. Briefly, this makes use of visual-kinesthetic congruence to induce multisensory integration as opposed to visual-tactile congruence. The first such experiment revealed that active movements resulted in proprioceptive drifts that encompassed the entire hand despite the movement being restricted to one finger, whereas when the movements were passive or tactile stimulation was instead applied to a single finger, the proprioceptive drift thereby produced was fragmentary and restricted to the stimulated finger (Tsakiris et al., 2006). A later study showed that subjective reports of ownership were as high in this motor-induced variant as in the more conventional condition (Dummer et al., 2009). Subsequent investigations using the motor-induced RHI

variant found a dissociation between the feeling of agency and ownership over the rubber hand by manipulating whether the movements were active or passive (Kalckert and Ehrsson, 2012), and that synchronized motor actions produced both of the typical RHI measures (ownership reports and proprioceptive drift) and that these were correlated (Sanchez-Vives et al., 2010).

## Self-Touch Variant

Apart from the motor-induced variant described above, there is also an interesting variant that relies on only touch and proprioception. The discovery of this variant reported that when subjects made experimenter-guided taps to a hidden rubber hand that were synchronized with another (visible) experimenter's taps to the subject's visible hand, subjects perceived the visible experimenter's hand as their own and that they were touching their own hand, indexed by questionnaires and proprioceptive drifts (Davies et al., 2010). In addition, Ehrsson et al. (2005) discovered yet another variant that does not seem to require vision at all. There, blindfolded subjects administer experimenter-guided touches to a rubber hand in synchrony with experimental touches to their contralateral hand, resulting in an illusory experience akin to the RHI, but where it seems the two signals to be bound are both proprioceptive in nature, one being kinesthetic from the active hand and the other passive from the static hand (Ehrsson et al., 2005).

## Other findings

As of this writing there have been over 100 published articles that examined various aspects of the rubber hand illusion. Of these, many are of only tangential interest to my express aims, but it will do us no harm to briefly mention some notable findings. First, it was reported that the RHI experience correlates with empathic and schizotypy questionnaire items (Asai et al., 2011). Intriguingly, a paper from 2011 reported that heartbeat counting accuracy – an established interoceptive sensitivity task – was negatively correlated with RHI susceptibility, suggesting that interoceptive awareness plays an important role in the integration of body-related signals (Tsakiris et al., 2011). Following on this, a recent paper demonstrated for the first time that it is possible to induce the illusion using synchronized visual feedback of heartbeats flashing over the dummy hand, with no tactile stimulation required (Suzuki et al., 2013). These latter two results support the theories of conscious selfhood by interoceptive predictive coding described above. Another notable finding relates to the oft-cited dissociation between action and perception subsystems and posits a corresponding dissociation in body representations, supported by experiments that showed that ballistic reaches towards their hands were accurate despite having experienced the illusion and exhibiting the expected proprioceptive drift when assessed using perceptual judgments (Kammers et al., 2009). Other curious findings relate to the fact that skin tone and texture have no effect on the induction of the illusion (Haans et al., 2008). Another very influential study investigated the possibility of using a psychometric approach to further specify and obtain greater insight into the questionnaire that is conventionally utilized, and observed that the subjective experience can be grouped into three major subcomponents: ownership,

location, and agency (Longo et al., 2008).

### The Mirror Illusion

Finally, I would like to briefly mention the large literature that has emerged around the so-called mirror illusion. This refers to an experimental paradigm that is closely related to the rubber hand illusion and is very well suited for the study of visuo-proprioceptive integration. The method consists of positioning a mirror such that it reflects the image of the subject's hand so as to appear to vary in location with respect to the real hand which is hidden behind the mirror. The effects of this visuo-proprioceptive disparity on subsequent reaches is often measured and is taken to be analogous to proprioceptive drift in the rubber hand illusion (Holmes and Spence, 2005; Holmes et al., 2004). Recently, these paradigms have been used to test van Beers et al. (1998, 2002) direction-dependent reliability model by causing visuo-proprioceptive discrepancies either in a horizontal plane (Snijders et al., 2007) or in a parasagittal plane (Tajima et al., 2015), and have found evidence that the visuo-proprioceptive integration occurs when the reflected object is a hand, and that this depends on a spatial window that is consistent with the predictions of the model.

## 1.3 Full-body Generalizations of the Rubber Hand Illusion

Now we shall shift gear and give a brief survey of a few interesting experimental attempts to study the general representation of the entire body, by what would at first appear to be

two slightly different though related generalizations of the rubber hand illusion.

## Full-Body Illusion

In 2007, two rival groups in Europe simultaneously discovered a novel type of illusion named the full body illusion (FBI), which can be thought of as a generalization of the rubber hand illusion, though by different means of its induction, which can cause a significantly different type of illusion to be perceived. I consider the Ehrsson (2007) variant to be more faithful to the original RHI paradigm and more naturally adapted to the whole body, though this is by no means uncontroversial. In short, this illusion makes use of stereo cameras placed behind a seated subject that are connected to a head-mounted display that the subject is wearing, which gives them a view of themselves from behind. The experimenter then supplies synchronized taps to the subject's chest and the corresponding location under the cameras (see Figure 1.3), analogous to the synchronized brushstrokes in the RHI. This illusion was measured using ownership questionnaires and skin conductance responses to threats applied to the camera between conditions where the stroking was synchronous compared to when it was asynchronous (Ehrsson, 2007). The Lenggenhager et al. (2007) variant of this illusion was discovered nearly simultaneously in France and was similar in most respects except that it utilized a different stroking strategy, applying taps to subjects' back which rendered them visible in the camera's field of view, and which could be synchronous (no-delay) or asynchronous (video delay). In this illusion, the authors utilized questionnaires as well as an adaptation of the proprioceptive localization measure to the full body case wherein subjects were passively moved while standing and blindfolded, and then asked to return

Figure 1.3: Henrik Ehrsson administering the taps that induce the full-body illusion

to their location of origin, and found that their estimates were more biased towards the perceived body when the stroking was synchronous (Lenggenhager et al., 2007). I favor the former variant because it is more similar to the RHI in its requirement of binding a visionless tactile signal with a touchless visual signal. The latter, in contrast, involves the same signal containing visual and tactile aspects, both of which are brought into conflict with an egocentric spatial representation, and as such seems to investigate a different question than that posed by the RHI and Ehrsson's FBI.

## Body Swapping and The Barbie Effect

In an extension of this work, Petkova and Ehrsson (2008) showed that it was possible for subjects to feel as though they had switched bodies with the experimenter. They

achieved this startling effect by having the cameras affixed to the experimenter's head and aimed at a metronome-synchronized handshake with the subject, who thereby perceived this from the experimenter's point of view. The validity of the subjective phenomenology was demonstrated by galvanic skin responses recorded from the subject being higher when the experimenter's wrist was threatened compared to when the subject's hand was (Petkova and Ehrsson, 2008).

Several experiments were conducted by the Ehrsson group at the Karolinska Institute designed to assess the effect that body transfer illusions of the type described above have on the perceived size of objects and the world. Among these, subjects experienced a full body illusion in which ownership over a miniature doll's body or a giant's body was induced via similar mechanisms to those previously discussed. When the body they transferred into was tiny, they reported that objects were smaller and farther away, and vice versa for the giant body, these effects being reported across ten experiments that utilized a variety of test measures from questionnaire items, to physiological measures, to verbal distance estimation. This surprising finding is interpreted within the embodied cognition framework and taken to be indicative of a general process whereby the body that one inhabits is a foundation for the perception of all of the space that surrounds it (van der Hoort et al., 2011).

## Peripersonal Space and Cardiac-Induced FBI

The Lenggenhager et al. (2007) method for inducing the FBI has been utilized to assess the extent of peripersonal space, and in an intriguing study that investigated this using modulations of tactile detection reaction times by looming or receding auditory stimuli found that

the these boundaries correspondingly shift forward when subjects experience their bodies at the location shown in front of them (Noel et al., 2015).

Finally, one last noteworthy study involved the use of interoceptive signals to induce a full body illusion. This is analogous to the cardiac-induced RHI discussed above in Section 1.2. In this case, the authors recorded subjects' heartbeats and displayed these as halos overlaid on the image of their bodies seen from behind that flashed either in real time or with a video delay introduced, and observed the induction of the Lenggenhager et al. (2007) variant of the full body illusion measured by shifts in the self-location metric and greater cross-modal congruency effects (Aspell et al., 2013).

## Effect of other modalities in virtual worlds

The preceding set of experiments discussed naturally leads on to the consideration of the sorts of effects one expects to obtain in studies that utilize a completely artificial environment, what is called virtual reality, as opposed to the merely altered reality of the experiments surveyed above. To this end, a recent study examined the extent to which multisensory stimulation in virtual reality strengthens the feeling of presence by manipulating the multisensory cues that were available (auditory, tactile, and olfactory) as well as the level of visual fidelity in a massively factorial design (2x2x2x2) with 322 subjects. The measure of presence was quantified by four sets of questionnaires assessing presence, spatial layout, and object location. The study found main effects for the addition of cues from all non-visual modalities in terms of increasing presence, but surprisingly visual fidelity showed no effect (Dinh et al., 1999). In addition, research from Joseph O'Doherty's and Miguel Nikolelis's

groups has been hard at work at improving the ability of brain-computer interfaces to provide sensory feedback to monkeys through stimulating somatosensory cortex using implanted electrodes, and have had some recent success in providing virtual tactile and proprioceptive stimuli that enabled the monkey to successfully interact with virtual objects (Dadarlat et al., 2015; O'Doherty et al., 2011).

## 1.4 Proprioception

At this point, we should take pause to investigate one of the lesser understood modalities of the brain, and in particular discuss its framing in the context of sensory-perceptual functions. Proprioception has traditionally been studied in the context of motor control, despite its being inherently sensory in nature that nevertheless serves its highest function in the feedback loops that modulate motor output and correct erroneous trajectories. However, its role in passive body location judgments has been somewhat understudied, although there are several seminal studies that do report such experimental investigations. If the goal is to understand the body representations that the brain constructs, however, the elucidation of proprioceptive function would appear to have an important place in a body of work such as what is being presented here.

### Localization Accuracy and Precision

The classic series of studies that investigated the crucial question of how proprioception and vision are integrated in the formation of a visuoproprioceptive estimate was performed by Robert van Beers and colleagues. In the first of these, a paradigm was designed where

subjects localized targets that were either visual or proprioceptive (location of fingertip under experimental tabletop) or both, and then used statistical models to disentangle the contributions of the two modalities, in terms of the variances of their individual probability distributions. They found that proprioception was more precise in the radial direction with respect to the shoulder and that vision was more precise in the azimuthal direction, and moreover account for this finding by referring to the geometry of the arm and the noise in joint angle estimates (van Beers et al., 1998). The very next year, Beers et al. (1999) proposed a model for the integration of the information from the two modalities based on these results, and then performed a study to test it. They found that the integrative step performs direction-dependent estimations based on the directionally-weighted modality, and that the estimate therefore lies off the straight line connecting the (biased) unisensory estimates. Finally, in a subsequent study, van Beers et al. (2002) obtained further confirmation of this model from an adaptation paradigm where a displacement in the visual stimuli was gradually introduced and its effects on localization error assessed, and found a double dissociation between blocks where the displacement was in azimuth and those where it was in depth. These findings further supported their model, thus providing convergent evidence for it across a series of experiments and approaches, strengthening the case for both the directionally-specific precisions and the integration scheme.

## Drift in absence of vision

A discordant finding that generated vigorous debate in the literature concerns the nature of unisensory proprioceptive representations in the absence of visual input. Several early

studies reported that the proprioceptive estimate drifts towards the body midline, and that this drift develops gradually the longer that the limb has been occluded from vision, and furthermore is stronger if the limb has been passively positioned (Paillard and Brouchon, 1968). However, this result has been contested by other research that appeared to show that proprioception does not in fact drift in the absence of vision across several experiments that manipulated whether the initial hand position or its trajectory during reaches was visible (Desmurget et al., 2000). Finally, one intriguing idea was put to the test by a research study in 2003 that examined the hypothesis that it was movement speed that generated the drifting estimates and verified that this was indeed the case and not a fading of proprioceptive representations, namely that faster movements incurred greater amounts of drift than slower movements (Brown et al., 2003). Recently, however, a group of researchers working under Eli Brenner have proposed a model for the optimal combination of visual and proprioceptive information that can account for these drifts in a manner that depends on the number of reaches that have been performed without vision, and not necessarily the duration of occlusion (Smeets et al., 2006). There is, thus, clearly a need for more research in order to ascertain more clearly the cause for these unisensory biases and the nature of the relationship to movement.

## Imaging

An fMRI study was performed in 2007 to investigate the cortical contributions to visuoproprioceptive integration and revealed that the posterior intraparietal sulcus was the ultimate destination for the processing of information regarding a visual stimulus near the

proprioceptively felt hand (Makin et al., 2007).

## 1.5 Visual-Tactile Interactions

In contrast to the difficulties in research on proprioception, the other half of the somatosensory field of representations has received considerably more attention and elucidation, namely the tactile modality. This modality offers several advantages over its proprioceptive cousin. First, the primary somatosensory cortex is principally a somatotopic map of the surface of the body, and as such occupies the lowest rung in the hierarchy of somatosensory processing. This is advantageous because it enables a relatively pure investigation into somatosensory processes, so that its interaction with other modalities can be cleanly investigated. Secondly, it is far easier to create and manipulate tactile stimuli as compared to proprioceptive stimuli, which necessarily imply movement of the body part involved and thus make systematic studies laborious and inherently limited (e.g. it is difficult if not impossible to create a unimodal condition lacking proprioceptive information).

### Bimodal Neurons

Let us begin with a survey of the single-cell recording experiments that investigated neurons with both visual and tactile receptive fields. We must of course begin with the series of studies by Michael Graziano that report results of recording from a variety of sensorimotor areas in the macaque brain (Graziano and Botvinick, 2002; Graziano et al., 2000; Graziano and Gross, 1998). Across these studies, the central underlying finding is that there exists a network of brain regions, principally ventral premotor cortex, that contain

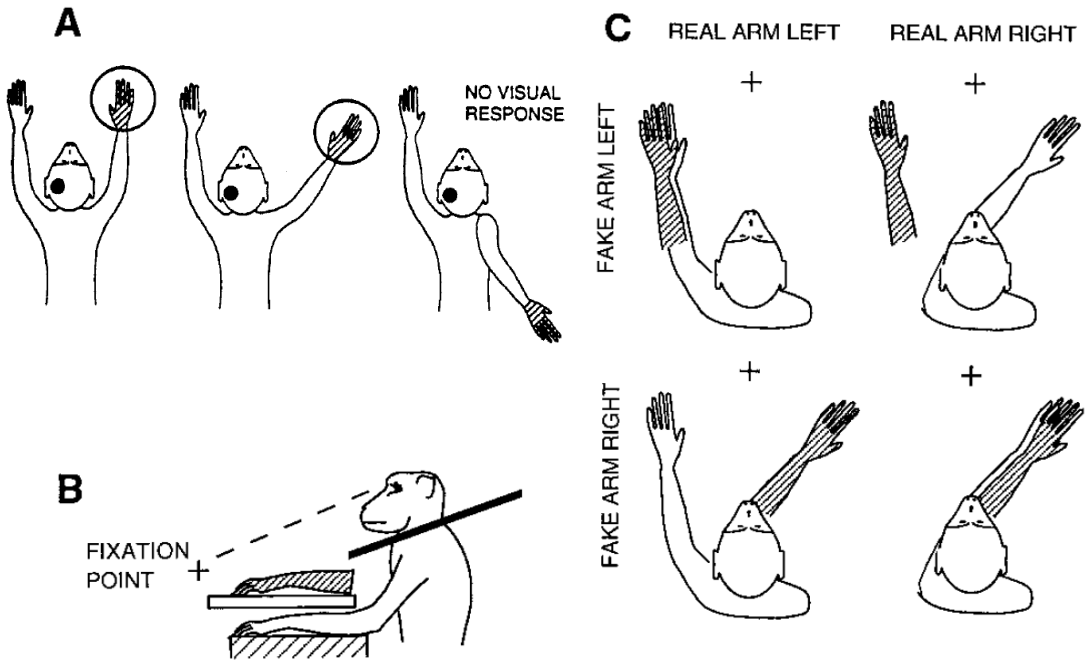


Figure 1.4: **A.** Figure reproduced from Graziano et al. (1994) showing the anchoring of the visual RF to the hand. **B and C.** Figure reproduced from Graziano et al. (2000) showing the response of bimodal neurons to visual stimulation near a fake arm.

so-called bimodal neurons responsive to both visual and tactile stimuli, with the curious property that visual receptive fields of these neurons are said to be anchored to the tactile receptive fields (see Figure 1.4A) (Graziano et al., 1994; Graziano and Botvinick, 2002). In addition, what is intriguing about these neurons is that they respond to the position of a realistic fake monkey arm when their real arm is hidden but with a congruent posture implicating the role of visuo-proprioceptive integration (see Figure 1.4B-C) (Graziano et al., 2000), and that eye movements do not alter their body-centered visual receptive field coding that remain anchored to the relevant part of the body (Graziano and Gross, 1998).

## Neuropsychological Findings

Complementary to the neurophysiological evidence discussed in the previous section are clinical observations from patients with right hemispheric damage that results in the clinical condition known as extinction (Ladavas, 2002). In particular, an effect that has been named cross-modal visuotactile extinction, a visual stimulus in the right hemifield and near the right hand (i.e., on the same side of the body as the lesion) inhibits the processing of a tactile stimulus on the left hand (i.e., on the opposite side of the lesion, and thus, the impaired side), and that this extinguishing effect was as strong as was obtained by using a tactile stimulus on the right hand (Ladavas et al., 1998). What is most interesting about these findings, however, is the fact that this extinguishing effect relied on the visual stimulus being near the body, and was greatly attenuated by its positioning in far space (Ladavas et al., 1998). This result concords quite nicely with the single-cell recording studies that similarly observe a boundary surrounding the body beyond which no, or very weak, visuotactile interactions can be observed (Graziano et al., 1994).

## Behavioral Findings

Finally, it would be remiss of this review if it did not briefly remark upon the various findings involving visuotactile interactions from behavioral paradigms. A very well established effect is what is commonly referred to as the visual enhancement of touch effect (VET), first discovered by Patrick Haggard's research group at UCL in 2001 (Kennett et al., 2001), which itself built off of earlier findings that reported a speeding effect of vision of the hand on tactile detection rates (Tipper et al., 1998). They observed that merely looking at one's

hand improved tactile two-point discrimination even though the visual stimulus contained only task-irrelevant information (Kennett et al., 2001). Subsequently, this effect has been further investigated using different methodologies, and found an visual modulation of tactile ERPs (Taylor-Clarke et al., 2002), and that it exhibits the characteristic inverse effectiveness that is typical of multisensory facilitation and moreover is stronger when the hand is perceived to belong to oneself (Longo et al., 2008). Another notable visuotactile interaction was reviewed in an article by Spence et al. (2004) and is referred to as the crossmodal congruency effect. Briefly, it entails a speed advantage for elevation discrimination responses regarding vibrotactile pulses to the index finger or thumb – while subjects grasped a cube embedded with tactors and LEDs – crucially, when the light distractors were congruent with the tactile stimuli (Spence et al., 2004). Finally, one last effect worthy of note is what is known as the cutaneous rabbit illusion, and occurs when trains of taps at discrete locations along the forearm are perceived as consecutively marching along the arm in a semi-continuous fashion, and feel like a rabbit's footsteps – hence the name (Geldard and Sherrick, 1972).

Taken together, these findings imply the existence of a tight coupling between tactile stimuli on the surface of the body and visual stimuli that are nearby. Evidence comes from single channel recordings – reviewed above in Section 1.1 – all the way up the heirarchy to behavioral effects and together serves to motivate the further investigation of the conditions that enable stimuli from these two modalities to be integrated. Graziano and Cooke (2006) have attempted to synthesize a coherent account of the function of this tightly coupled network of brain regions in the parietal and frontal lobes, and have speculated on its importance for defensive behaviors that seek to protect the body. Their overarching thesis is that peripersonal space ought to be understood under the larger framework of an ethological

understanding of the function of this network, such that the behavior of defending the body surface is the underlying evolutionary force driving its development.

## 1.6 Computational Modeling

In the last section of this survey, I will discuss briefly the computational modeling of human perception and behavior as it relates to the topic of current study. To operate at David Marr's "computational level" is crucial for the understanding the information processing task that the brain is attempting to perform (Marr, 1982). As brain processes typically require very large circuits of neuronal assemblies that are thus very difficult to model using neurobiological simulations, this approach has multifaceted advantages. We therefore, for the time being, rely on normative statistical inference models that explain how signals embedded in noise can be extracted and processed by a system that operates upon the same information that brains are sensitive to.

### Maximum Likelihood Estimation (MLE)

Michael Landy and colleagues provided a nice summary of the sorts of models of multisensory integration that were available towards the end of the twentieth century and offered their views as to the correct family of models that should be further investigated. They delineated the spectrum of possible models into those which they termed Weak Observer (also called Weak Fusion) models, wherein the cues to the system are completely modular and get combined in a simple fashion, and those they termed Strong Observer (also called Strong Fusion) models, where the inference proceeds on all the data simultaneously and maximizes

the log likelihood of the data given the model, along the lines of Bayesian inference systems. They propose to strike a middle ground between these two camps and offer a theory they call Modified Weak Fusion, which would nowadays be better known as Maximum Likelihood Estimation (MLE), where the various coexisting cues do indeed get combined, as in Weak Fusion, but not without the intermediary step of their being weighted according to the inverse of their variances, a quantity known as a signal's reliability (Landy et al., 1995), as:

$$\hat{x} = \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}} \tag{1.1}$$

This style of thinking about the mathematization of inferential brain processes has proven to be extremely influential in the intervening years, and is now a hallmark of initial basic attempts to model behavioral data. Let us cursorily begin to take stock of what has been achieved in this regard.

I should remark in passing at this point that this class of models is the ultimate realization in mathematical form of the "modality precision" hypothesis described by Welch and Warren in their seminal paper from 1980, and that while it has its merits, it does not fully formalize the complete theory of "intersensory bias" that they championed (Welch and Warren, 1980). We shall have more to say on this topic when we examine the Bayesian family of models in what follows.

### Visual and Haptic Integration

I will restrict this review to the types of models that are implemented when combining cues across modalities, which is the topic of specific relevance to my research, and gloss over the vast literature on cue combination within the same modality as it is not qualitatively

different from our current purview. The key thought behind this form of modeling is that subjects will rely more heavily on the more reliable modality. Experimenters often systematically introduce noise to the unisensory estimates and fit each subject's psychometric function in order to determine the just noticeable difference (JND) that can then be used to determine relative weightings when a multisensory estimate is produced. This has been done using a setup that allows the manipulation of haptic and visual estimates of the height of a virtual object and results across several experiments concurred with MLE model predictions (Ernst and Banks, 2002; Gepshtein et al., 2005).

### Visual and Proprioceptive Integration

The MLE approach is a very general one, and as such, can be applied to any combination of modalities. Of perhaps more relevance to this body of work is integration that involves the proprioceptive modality. Relating to this, two papers reported attempting to model arm trajectory perception using this framework and were successful at doing so (Reuschel et al., 2010; Serwe et al., 2009). However, where Reuschel et al. (2010) found the expected reduction in the bisensory estimate variance due to the benefit of integration, Serwe et al. (2009) did not. The reasons for this are likely to do with the different nature of the task that subjects performed, which in the latter paper involved stimulus pairs that would not be likely to be integrated due to their being of a highly artificial nature.

### Visual and Vestibular Integration

Another area of relevance is that concerning the integration of exteroceptive information with vestibular information. One study of note found a statistically optimal model that could

predict the reduction in variance in a heading discrimination task with visual and vestibular signals, initially using a pure MLE model. However, the vestibular signal was always weighted higher than the visual in cases of conflict, regardless of the reliability differences, which they address by introducing a prior on the vestibular signal to overweight it, in essence making the model a true Bayesian one, which will be discussed further below (Butler et al., 2010).

## Bayes Decision Theory

In building upon the reliability-weighted cue combination schemes described above, Bayesian models make a minor but important modification by introducing a prior term that incorporates the influence of past experience. In fact, the MLE models are special cases of Bayes Decision Theory in which the prior is a uniform distribution. By way of introduction, let us revisit a quote from the father of visual psychophysiology, the great Hermann von Holmholtz:

> "The general rule according to which visual representations determine themselves is that we always find present in the visual field such objects as would have to exist in order for them to produce the same impression on the neural apparatus under the usual normal conditions of the use of our eyes" – Helmholtz (1867).

In his notion of "unconscious inference", we can see the germ of what modern investigators formalize as Bayesian inference, namely the fact that perception is an active process that represents the observer's best guess as to the cause of the sensations that are registered on the sense organs. The transition from the MLE approach to the fully characterized Bayesian theory is broadly analogous to Welch and Warren's discussion of how the "modality precision" theory can be expanded upon to become the theory that they termed "intersensory bias", a key component of which states that past knowledge is brought to bear on the current inference to be made (Welch and Warren, 1980).

This influence of past knowledge, what is termed the prior, can be readily seen in the most basic form of Bayes Rule, shown below:

$$p(H|e) = \frac{p(e|H)p(H)}{p(e)} \tag{1.2}$$

where $H$ denotes a hypothesis under consideration, and $e$ the evidence that has been gathered. The influence of prior knowledge is introduced in the term $p(H)$, which denotes how likely a given hypothesis is believed to be **prior** to gathering any evidence – hence the name.

### A Brief Survey of Bayesian Models

Without spending too much time on the discussions of Bayesian models in general – models which are by now very well established and utilized across many disciplines – I would like to introduce and comment upon the way in which they have been formulated as models of perception, and primarily vision. This will hopefully put the subsequent section that deals with the Bayesian model of multisensory integration in its appropriate context. For instance, in the seminal Nature Neuroscience paper by Weiss et al., they formulated and tested a normative Bayesian model of motion perception that was equipped with a prior that assumes that motion is generally slow, and found that it was able to reproduce typically observed human biases and motion illusions (Weiss et al., 2002). From another area of computational vision modeling, Mamassian and Landy have characterized models based on Bayes Decision theory aimed at accounting for the constraints utilized in inferring shape from shading (Mamassian and Landy, 2001), and shape-from-contour (Mamassian and Landy, 1998). These highly sophisticated models demonstrate how the nervous system combines cues from different parts of the visual system, and moreover how biases are treated

as additional cues that serve to constrain the interpretations of a scene, of which there are clearly infinite when the problem is as ill-posed as that of perceiving the three-dimensional world from sensing a two-dimensional retinal image (Mamassian and Landy, 2001). Finally, a highly influential model has been gaining a lot of traction in the field in recent years and is called the free energy hypothesis (Feldman and Friston, 2010). At every moment in time, the brain is claimed to be actively predicting sensory events, which represents what is typically referred to as top-down processing, while the sensory information is simultaneously being compared to these predictions, and hence generating a prediction error that flows up the hierarchy, which is referred to as bottom-up processing (Feldman and Friston, 2010). In Karl Friston's proposal, the claim is that the brain attempts to minimize a free energy quantity, which denotes the Bayesian surprise brought about by the prediction error, through a process of modifying the internal model of the world (Feldman and Friston, 2010).

## Bayesian Causal Inference

Of particular interest to my proposed program is the form of Bayesian modeling that implements what is called causal inference. Briefly, this mathematical framework performs a hierarchy of inferences, the first of which infers the most likely causal structure, and the second of which uses that causal structure to estimate properties of the stimuli. As there are two levels of inference, there are thus two levels of prior as well: one which corresponds to the expected causal structure that is often referred to as $p_{common}$ or $p(C = 1)$, and another which corresponds to the properties to be estimated. This model has been used quite successfully to account for a wide variety of multisensory phenomena as it can very

parsimoniously determine when to integrate signals and when to segregate them (Kording et al., 2007; Beierholm et al., 2009b; Wozny et al., 2010; Wozny and Shams, 2011a; Samad et al., 2015).

### Spatial information

A remarkable series of papers has established the role of the Bayesian causal inference as the appropriate statistical framework to use as a model of the processes involved in audiovisual localization, and as it is a general model, spatial integration by generalization. Briefly, the model performs a hierarchical pair of inferences, the first establishing the weight on causal structure for the signals (i.e., whether they came from the same or different sources), and the second performing inference on the optimal location for each of the signals, with the degree of integration or segregation of the signals determined by the causal structure deemed most probable (Beierholm et al., 2009b; Kording et al., 2007; Wozny et al., 2010; Wozny and Shams, 2011a).

### Temporal information

Being a general and normative model, this has been adapted to the temporal domain and has been successfully utilized to model the flash-beep illusion, which depends on numerosity judgments (Wozny et al., 2008).

### Recalibration

Finally, the same Bayesian model has been utilized to account for the recalibration which takes place when spatially discrepant stimuli are presented to a subject for an extended period

of time (Wozny and Shams, 2011a). With these three pieces together, the Bayesian model seems very well suited to model the RHI, and by extension body perception in general. However, to do so it requires its extension to both spatial and temporal domains simultaneously, and as such becomes a significantly larger model than these previous versions. In Chapter 2, I will describe in detail my attempts to investigate its behavior and predictions as well as its success at accounting for the illusion and various other empirical effects associated with it (Samad et al., 2015). In Chapter 3, I will present data showing the applicability of the model to visual-tactile integration along the surface of the skin in somatotopic coordinates (Samad and Shams, 2016).

## 1.7 Aims of the Dissertation

As the burgeoning field that studies the cognitive and perceptual aspects of body representation has been surveyed in the foregoing conversation, we will now proceed to document the experimental work that I have undertaken in the effort to contribute to this field and address the areas in which it is lacking. We have discussed the use of perceptual illusions such as the rubber hand illusion and the full body illusion as well as the use of computational modeling frameworks to model multisensory interactions. As the emphasis in body representations has begun to swing in favor of regarding them as the result of such computations, it appears that the next step would be to continue investigating the multisensory principles that govern the generation of the body representations and to attempt to discover the computational frameworks that can best instantiate these processes.

In Chapter 2, I will describe an experiment that was conducted with the aim of fur-

thering our understanding of the computational mechanisms that underlie the rubber hand illusion. I applied the Bayesian causal inference model to the phenomenon and tested and confirmed some of its predictions, lending it much credibility as the framework for explaining the illusion. In Chapter 3, I will describe an experiment that aimed at uncovering a new multisensory interaction that has never been observed before, namely visuotactile ventriloquism. Briefly, this involves the use of visual and tactile stimuli along the surface of the arm, where tactile localization estimates are shifted towards simultaneously presented nearby visual stimuli. This work also made use of the Bayesian causal inference model to quantitatively account for the reported visuotactile ventriloquist effect, such that it produced parameter estimates to fit the observed data. In Chapter 4, I extend the investigation into this visuotactile interaction to a study of how prolonged exposure to spatially discrepant stimulus pairs can induce a visuotactile recalibration, akin to the audiovisual ventriloquist aftereffect. These experimental sections will be tied together in Chapter 5 that will synthesize a novel theoretical and computational account I have called the Bayesian body hypothesis from the body of work contained in between the covers of this dissertation, as a generalization of the causal inference model to the case of body-related representations. Finally, the appendix documents the development of a new tool I have created to aid future researchers with the use of the Bayesian causal inference model that I have devoted much of the research described in this dissertation to.

Thus, the overarching **aims of the dissertation** are to propose and test out a computational model of body ownership that can be utilized to account for body ownership illusions, which require characterization of the peripersonal space, as well as illusions of the tactile modality with representations characterized by their somatotopic space. The predominant

hypothesis that spans all the work depicted here is that the computations that engender body ownership are no different from those that are involved in perception of objects in the external world, and therefore that a normative model based on fundamental principles of statistical inference ought to be well-supported by the data.

# Chapter 2

# Perception of Body Ownership is Driven by Bayesian Sensory Inference

## 2.1 Abstract

Recent studies have shown that human perception of body ownership is highly malleable. A well-known example is the rubber hand illusion (RHI) wherein ownership over a dummy hand is experienced, and is generally believed to require synchronized stroking of real and dummy hands. Our goal was to elucidate the computational principles governing this phenomenon. We adopted the Bayesian causal inference model of multisensory perception and applied it to visual, proprioceptive, and tactile stimuli. The model reproduced the RHI, predicted that it can occur without tactile stimulation, and that synchronous stroking would enhance it. Various measures of ownership across two experiments confirmed the predictions: a large percentage of individuals experienced the illusion in the absence of any tactile stimulation, and synchronous stroking strengthened the illusion. Altogether, these findings

suggest that perception of body ownership is governed by Bayesian causal inference – i.e., the same rule that appears to govern the perception of outside world.

## 2.2 Introduction

Intuitively, our sense of ownership of our body and body parts appears inherent, stable, and immutable. However, recent research has shown an incredible degree of malleability in our sense of body ownership and perception. For example, using simple and brief manipulation of sensory input, the subject may experience ownership over another person's body and disownership of one's body (Petkova and Ehrsson, 2008), may experience the body in another location (Ehrsson, 2007; Lenggenhager et al., 2007), or may adopt ownership of artificial bodies (van der Hoort et al., 2011) or body parts (Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005). While the protocols and brain regions involved in these alterations of body ownership have been investigated by recent studies, the governing rules and computational mechanisms of body ownership remain poorly understood (Ehrsson et al., 2004; Gentile et al., 2013; Tsakiris et al., 2007). The goal of this study was to gain insight into the principles that govern body ownership in humans. To this end, we used a well-established and extensively studied body-ownership illusion known as Rubber Hand Illusion (RHI).

In the RHI (Armel and Ramachandran, 2003; Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005) a dummy hand is misattributed to oneself when positioned in an anatomically and posturally plausible location near the occluded real hand and stroked synchronously with that of the occluded real hand. The original paradigm used for the study of the illusion consisted of occluding a participant's arm and placing a visible rubber hand medial to it, and

stroking the index fingers of both with paintbrushes either synchronously or asynchronously. Such experiments led to the conclusion that the synchrony of the stroking is a critical condition for the illusory experience. For instance, Manos Tsakiris and Patrick Haggard state in one of the classic RHI studies that "the necessary condition for the inducement of the illusion is the presence of synchronized and spatially congruent visual and tactile stimulation" (Tsakiris and Haggard, 2005). It has also been reported that the rubber hand must be in a position that is both anatomically plausible and congruent with the real hand's posture in order for the illusion to occur (Tsakiris and Haggard, 2005).

In the original demonstration of this effect and several subsequent studies, the illusion was assessed by two measures: ratings on a questionnaire that assessed degree of ownership for the fake hand, and change in the localization of the hidden hand after exposure to the rubber hand ("proprioceptive drift"). The two measures were found to be correlated and only subjects receiving synchronous stroking (and not those subjected to asynchronous stroking) experience the illusion and exhibit the aforementioned proprioceptive drift. In addition, the RHI can cause an increase in skin conductivity – a physiological measure of anxiety or arousal – in response to a threat to the rubber hand (Armel and Ramachandran, 2003).

While several qualitative neural models have been proposed to describe the brain areas that may be involved in this intriguing phenomenon, as well as their hypothesized processing and communication (Makin et al., 2008; Tsakiris, 2010), computational theories have yet to emerge. The Rubber Hand Illusion obviously involves interactions among visual, tactile and proprioceptive modalities. Furthermore, the perception of the illusion can be characterized as inference of a common cause for proprioceptive, tactile and visual sensations, whereas the absence of illusion can be characterized as perception of independent sources for the

visual (rubber hand), and proprioceptive and tactile (real hand) sensations. Therefore, the perception of the RHI appears to depend on a process of causal inference operating on three sensory stimuli.

A Bayesian causal inference model (Beierholm et al., 2009b; Kording et al., 2007; Magnotti et al., 2013; Shams et al., 2005; Wozny et al., 2008, 2010) has been shown to successfully account for a variety of human multisensory perceptual phenomena, and a recent human fMRI study has provided further support for the brain carrying out this type of computation (Rohe and Noppeney, 2015). This model makes an inference about the causal structure of the sensations, namely whether they have a common cause or independent causes, based on the similarity of the sensory signals and the prior probability of a common cause. The stimulus properties (location, time, etc.) will then be estimated according to the inferred causal structure, entailing integration of senses only if warranted by the inferred causal origin. Therefore, both the causal inference and integration problems are solved in a coherent and unified fashion. Of interest, this model has accounted for multisensory integration of spatial information (Beierholm et al., 2009b; Kording et al., 2007; Wozny et al., 2010), as well as temporal information (Wozny et al., 2008), and crossmodal sensory recalibration (Wozny and Shams, 2011a). As the RHI involves all of these aspects, namely, spatial and temporal crossmodal interactions, crossmodal recalibration, and causal inference, the Bayesian causal inference framework appears to be the ideal framework for a computational understanding of the RHI. Therefore, in this study, we adopted this framework and examined whether Bayesian causal inference can account for the RHI.

## 2.3 Bayesian Causal Inference Model

### Method

The Bayesian causal inference framework adopted here to model the RHI operates on both temporal and spatial information in order to infer the causal structure that is most likely to have produced the sensory signals (Figure 2.1).

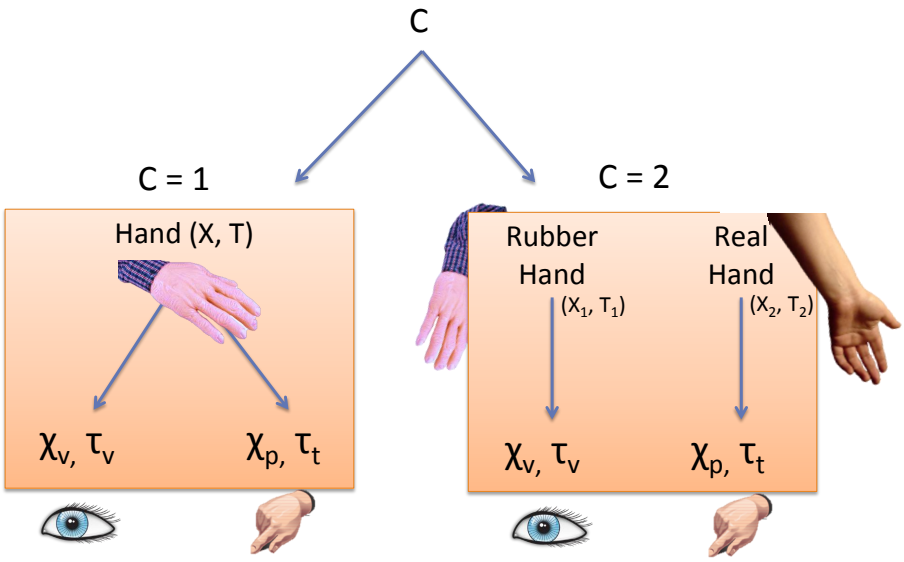


Figure 2.1: **Rubber Hand Illusion as Causal Inference.** Spatial signals ($\chi$) and temporal signals ($\tau$) coming from the visual ($\chi_v$, $\tau_v$) and somatosensory modalities (proprioception: $\chi_p$, tactile: $\tau_t$) are either integrated or segregated depending on whether the brain infers a common cause or independent causes for the sensations.

A visual cue to the location of the rubber hand and a proprioceptive cue to the location of the real hand provide spatial information, while a visual cue to timing of the seen stroking of the rubber hand and a tactile cue to the timing of the felt stroking of the real hand provide temporal information (Figure 2.1).

When stroking of the fingers occurs, both spatial and temporal information is available for the inference process. We modeled the spatiotemporal sensory input as bivariate Gaussians (see Figure 2.2). The assumption of a Gaussian distribution for proprioception is supported by distributions of proprioceptive localization judgments reported below in experiment 1. We tested the normality of these distributions and found that between 70-80% of subjects' data passed the Shapiro-Wilk, Anderson-Darling, Jarque-Bera, and Lilliefors' tests of normality. In addition, we make the assumption that the spatial ($\chi$) and temporal ($\tau$) signals are statistically independent, which allows us to derive an analytic solution to the combined likelihoods in the equations. Although tactile-proprioceptive interactions have been observed whereby a touch reduces the magnitude of errors in proprioceptive localization judgments without altering their pattern (Rincon-Gonzalez et al., 2011), this effect was small and confined to the right hand independently of handedness. Given that our experiments involved proprioceptive localization of the left hand only, and for the sake of parsimony, we assumed independence of tactile and proprioceptive signals.

The posterior probability of a causal structure given the visual (v), tactile (t), and proprioceptive (p) sensory signals is computed using Bayes Rule as follows:

$$p(C|\chi_v, \chi_p, \tau_v, \tau_t) = \frac{p(\chi_v, \chi_p, \tau_v, \tau_t|C)p(C)}{p(\chi_v, \chi_p, \tau_v, \tau_t)} \tag{2.1}$$

where C is a binary variable denoting the causal structure (1 vs. 2 causes); $\chi_v$ and $\chi_p$ denote the visual and proprioceptive sensations of location, respectively; and $\tau_v$ and $\tau_t$ denote the visual and tactile sensations of timing, respectively. Therefore, the posterior probability of

the signals having a single cause in the environment is computed as:

$$p(C=1|\chi_v,\chi_p,\tau_v,\tau_t)=\frac{p(\chi_v,\chi_p,\tau_v,\tau_t|C=1)p(C=1)}{p(\chi_v,\chi_p,\tau_v,\tau_t|C=1)p(C=1)+p(\chi_v,\chi_p,\tau_v,\tau_t|C=2)(1-p(C=1))} \tag{2.2}$$

where the likelihood probability is:

$$p(\chi_v,\chi_p,\tau_v,\tau_t|C=1)=\iint p(\chi_v,\chi_p,\tau_v,\tau_t|X,T)p(X,T)dXdT \tag{2.3}$$

and $p(C=1)$ is the prior probability of a common cause. $X$ and $T$ denote spatial and temporal attributes of the stimuli, respectively, which give rise to the visual $(\chi_v,\tau_v)$ and/or somatosensory $(\chi_p,\tau_t)$ neural representations. They are modeled as continuous random variables ($X$ ranges across the azimuthal space with zero indicating body midline; $T$ spans the duration of a trial with zero indicating the start of a trial), and have the following priors: $\mathcal{N}(\mu_X,\sigma_X)$ and $\mathcal{N}(\mu_T,\sigma_T)$, where $\mathcal{N}(\mu,\sigma)$ stands for a normal distribution with mean $\mu$ and standard deviation $\sigma$. Equation 2 shows that two factors contribute to the inference of a common cause: the likelihood (the first term in the numerator) and the prior (the second term in the numerator). A high likelihood (Equation 3) occurs if the spatiotemporal sensory signals are similar, such that greater similarity of spatial $(\chi_v,\chi_p)$ and/or temporal $(\tau_v,\tau_t)$ signals results in a greater likelihood that they are generated by a common cause (Equation 3). The prior probability of a common cause, $p(C=1)$, on the other hand, is independent of the present sensations, and depends on the observer's prior experience.

We assume that the nervous system tries to minimize the mean squared error in the spatiotemporal estimates of the events:

$$\text{Cost}=(\hat{X}-X)^2+(\hat{T}-T)^2 \tag{2.4}$$

Therefore, the optimal estimates under this quadratic error will be weighted averages of the two causal models, which is called model averaging. This implies that the optimal estimates will in most cases include influences of both causal models, except in the most extreme cases where the evidence fully supports one or the other. The optimal estimate of the position of the observer's arm, $\hat{X}_p$, calculated according to Bayes rule, will thus be:

$$\hat{X}_p = p(C=1|\chi_v, \chi_p, \tau_v, \tau_t)\hat{X}_{p,C=1} + (1 - p(C=1|\chi_v, \chi_p, \tau_v, \tau_t))\hat{X}_{p,C=2} \tag{2.5}$$

where $\hat{X}_{p,C=1}$ represents the best estimate of proprioceptive stimulus location under the assumption of common cause, which is thus equivalent to $\hat{X}_{v,C=1}$, the best estimate of visual stimulus location, both of which are computed according to Bayes Rule as:

$$\hat{X}_{v,C=1} = \hat{X}_{p,C=1} = \frac{\frac{\chi_v}{\sigma_v^2} + \frac{\chi_p}{\sigma_p^2} + \frac{\mu_X}{\sigma_X^2}}{\frac{1}{\sigma_v^2} + \frac{1}{\sigma_p^2} + \frac{1}{\sigma_X^2}} \tag{2.6}$$

and where $\hat{X}_{v,C=2}$ and $\hat{X}_{p,C=2}$ represent the best visual and proprioceptive estimates under the assumption of independent causes, computed according to Bayes Rule as:

$$\hat{X}_{v,C=2} = \frac{\frac{\chi_v}{\sigma_v^2} + \frac{\mu_X}{\sigma_X^2}}{\frac{1}{\sigma_v^2} + \frac{1}{\sigma_X^2}} \qquad \text{and} \qquad \hat{X}_{p,C=2} = \frac{\frac{\chi_p}{\sigma_p^2} + \frac{\mu_X}{\sigma_X^2}}{\frac{1}{\sigma_p^2} + \frac{1}{\sigma_X^2}} \tag{2.7}$$

Note that this model also produces temporal estimates ($\hat{T}_v$, $\hat{T}_t$: estimated timing of visual stimulus and tactile stimulus), which have not been described, but would be computed in an entirely analogous way to the spatial estimates above.

To simulate the spatiotemporal perceptions produced by this model in different tactile stimulation conditions (synchronous and asynchronous), we performed 100,000 trials of Monte Carlo simulations. We chose realistic values for the parameters. Means for sensory likelihoods corresponded to typically utilized distances/durations between stimuli (rubber

hand ($\chi_v$): $16cm$ from midline, real hand ($\chi_p$): $32cm$ from midline, temporal latency between stimuli during asynchronous stroking ($|\tau_v - \tau_t|$: 0.5-1 seconds). In addition, we simulated the effect of increasing distance between the real and rubber hands by moving the simulated position of the rubber hand from $16cm$ to $36cm$ away from the real hand in intervals of $2cm$, while holding all other parameters constant. The standard deviation of proprioception ($\sigma_p^2$) was set to $15mm$ (Jones et al., 2010; van Beers et al., 1998). Vision is known to have a superb spatial acuity, and a previous study with similar experimental conditions estimated this variability to be around 0.36 degrees (van Beers et al., 1998). In our set up, with an eye to rubber finger distance of $\sim$ 35-45$cm$, this translates to a standard deviation of a couple of millimeters. Therefore, the standard deviation of visual likelihood ($\sigma_v^2$) was set to $1mm$. The results are robust with respect to the exact value of this parameter. Temporal standard deviations were set to $20ms$ for both visual and tactile modalities based on research showing similar JNDs in a temporal task (Hirsh and Sherrick, 1961). For the sake of parsimony, $X$ and $T$ were assumed to be statistically independent and their priors to be uninformative. Therefore, the standard deviation of the spatial prior, $\sigma_v$, and the standard deviation of the temporal prior, $\sigma_t$, were set to large numbers to approximate uniform distributions. For the sake of parsimony, the prior probability of common cause, $p(C = 1)$, was set equal to 0.5.

## Results

Figures 2.2-2.4 show the simulation results. When the tactile signal is temporally congruent with the visual signal, i.e., when the stroking is synchronous, the inferred probability of a common cause is high, and the illusion is experienced (Figure 2.2a,b). When there is

a temporal delay between the two signals, i.e., in the asynchronous stroking condition, the inferred probability of a common cause is low, and the model favors independent causes, and thus, the rubber hand and real hand are estimated to be at distinct locations (Figure 2.2c,d). Therefore, the model can account for the RHI.

In addition, when assessing the effect of distance between the rubber hand and the real hand on the illusion, the inference of a common cause becomes increasingly less probable, and thus the illusion becomes increasingly weaker, as the distance between the two is increased, and the illusion starts to vanish as the distance approaches $30cm$ (Figure 2.3). These results closely match empirical findings which had shown the illusion deteriorates as a function of distance, and had found the spatial limits on the experience of the RHI was $27.5cm$ (Lloyd, 2007).

In order to further examine the validity of the model as the computation underlying the RHI, we explored additional predictions of the model that can be tested empirically. It should be noted that in the absence of any tactile stimulation the input is purely spatial. Depending on the exact degree of sensory noise/precision and the distance between the real hand and rubber hand, the illusion may or may not occur based on spatial information alone. If the precision of spatial proprioceptive representations is not very high and/or the distance between the rubber and real hands is not very large, the inferred probability of a common cause would be large. In such a case, vision (location of the rubber hand) would capture proprioception (location of the real hand) and the rubber hand illusion would be perceived (ownership of the rubber hand would be experienced). The model simulations for this situation are illustrated in Figure 2.4. Here, in the absence of tactile signals (and temporal information) the visual and proprioceptive spatial signals are integrated, as shown

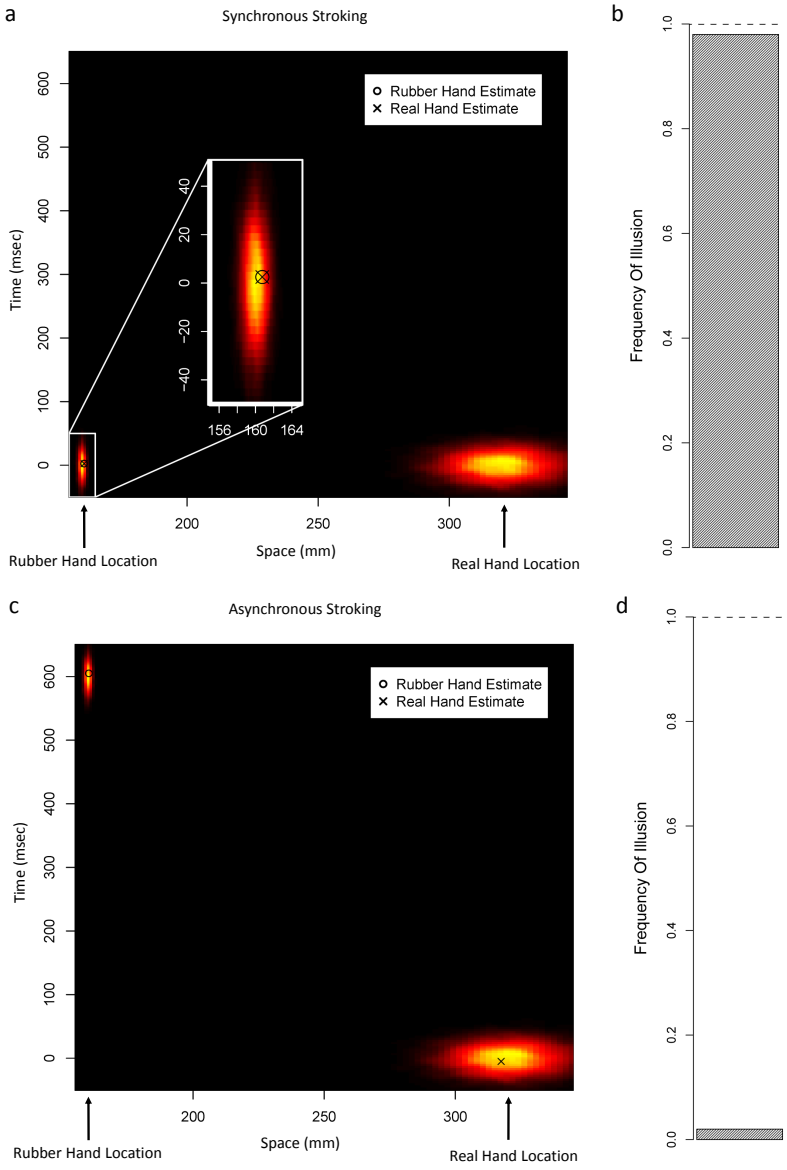


Figure 2.2: **Simulation Results** a) Synchronous Stroking: Distributions are the likelihoods representing the objective stimulus locations/timings. Marked points are the model estimates (MAP) of stimulus location/timing. b) Synchronous Stroking: The frequency of simulation runs in which a common cause is inferred is shown in the shaded bar. c) Asynchronous Stroking: Distributions are the likelihoods representing the objective stimulus locations/timings. Marked points are the model estimates (MAP) of stimulus location/timing. d) Asynchronous Stroking: The frequency of simulation runs in which a common cause is inferred is shown in the shaded bar.

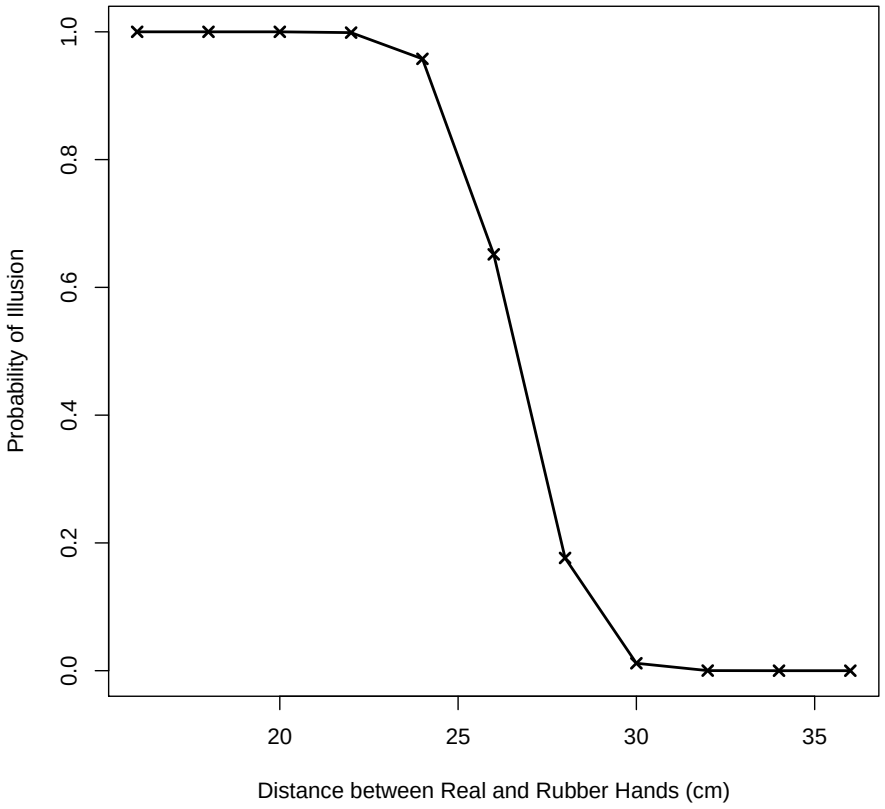


Figure 2.3: **Simulation Results: Spatial Extent.** The probability of experiencing the illusion is plotted as a function of the distance (in centimeters) between the rubber hand and the real hand. As the distance between the two increases, the illusion becomes weaker and eventually fails to occur. These results are qualitatively and quantitatively consistent with empirical findings from human participants (Lloyd, 2007).

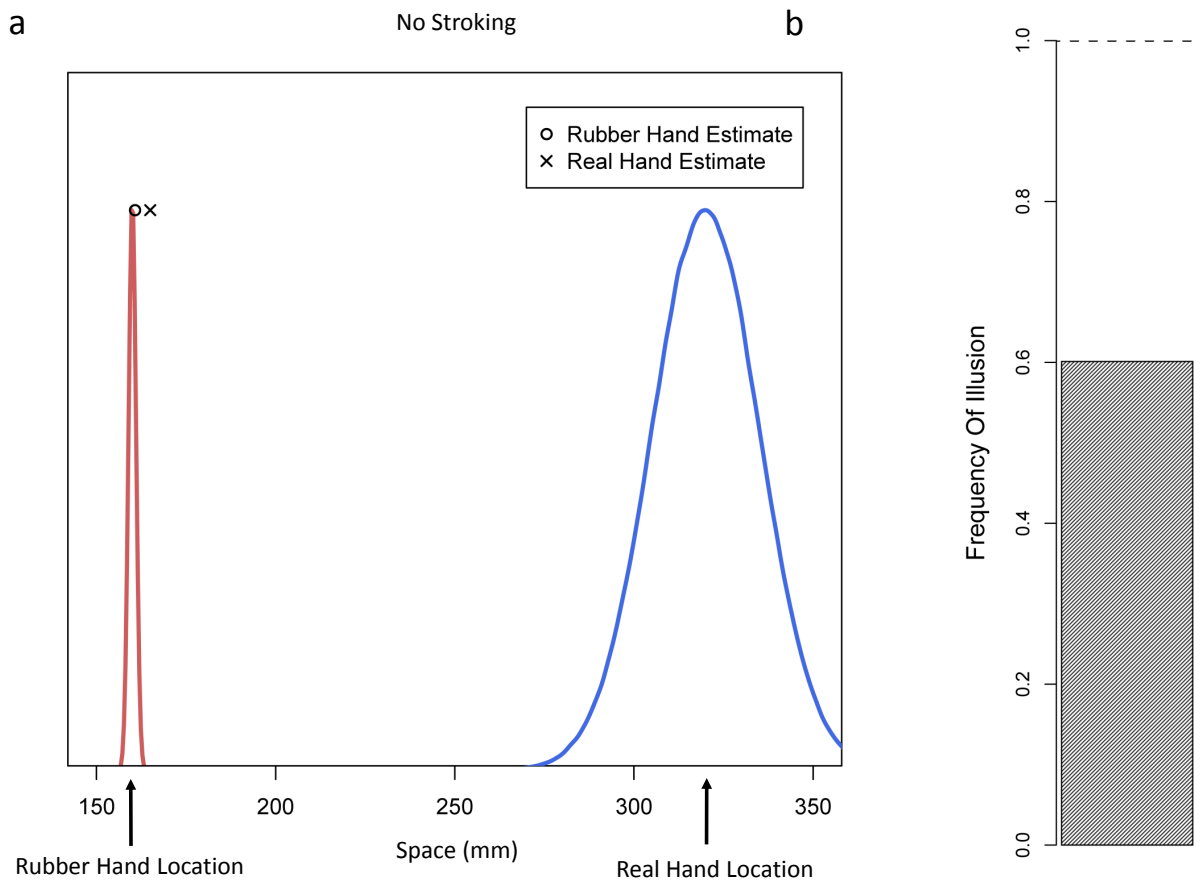


Figure 2.4: **Simulation Results: No Stroking.** a) Removing the temporal dimension from the model retains the illusory effect of overlapping spatial estimates. Marked points represent model estimate (MAP) of hand location. b) The frequency of simulation runs in which a common cause is inferred is shown in the shaded bar.

by the very close proximity of the spatial estimates. Therefore, the model predicts that if the distance between the real hand and rubber hand is not very large, the illusion should be perceived without any stroking, at least for those individuals who do not have very precise proprioceptive representations. This suggests the possibility of inducing the rubber hand illusion prior to the application of brush strokes. We tested this prediction experimentally as described below.

## 2.4 Experiment 1

The goal of this experiment was to test the hypothesis that tactile stimulation is not necessary for the induction of the rubber hand illusion.

### Method

#### Design

As in standard Rubber Hand Illusion studies, the left arm of the participants was hidden from their view, and a visible rubber hand was positioned in front of the observer in an anatomically plausible position. Unlike the standard RHI studies that probe the ownership of the rubber hand and the drift in proprioception of the real hand only after stroking of the hands, here, the subjective report of ownership and proprioception of the hand were examined before the application of tactile stimulation. This experiment consisted of four conditions: 'sync', 'async', 'no-stroke', and 'no-hand' (see Figure 2.5).

As the model predicts that synchronous tactile stimulation should strengthen the illusion, we measured subjective assessment of rubber hand ownership and the drift in proprioception in a group of subjects who received synchronous tactile stimulation ('sync' condition). To examine the role of synchronization of tactile stimulation, another group of participants received asynchronous stimulation ('async' condition). In addition, we had a group of subjects who never received tactile stimulation ('no-stroke' condition). To obtain a baseline for both reports of ownership and proprioceptive perception, a fourth group of participants underwent the exact same procedures but was not presented with any rubber hands ('no hand' condition).

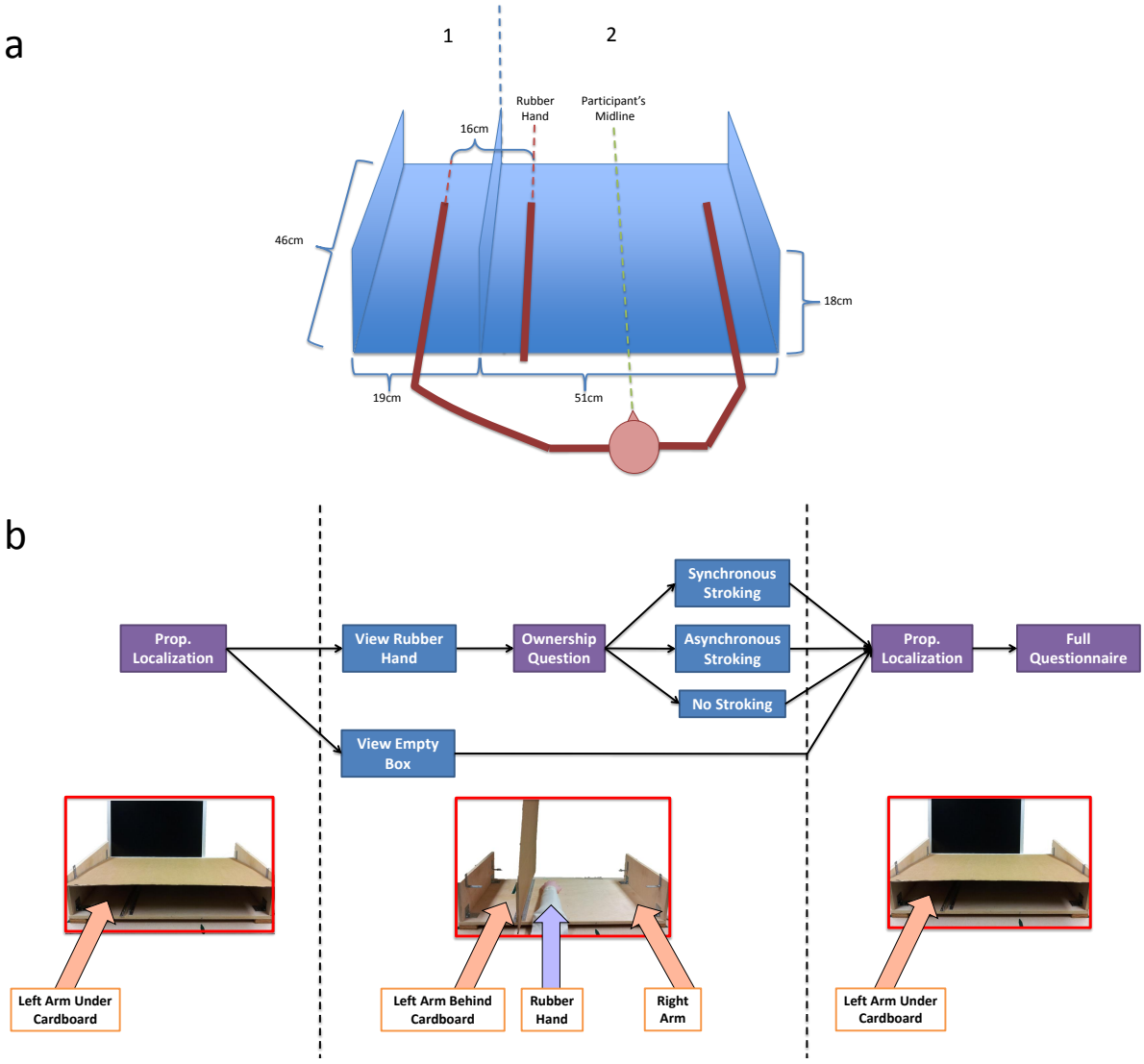


Figure 2.5: a) RHI Apparatus b) Experiment 1 procedural design

### Participants

Based on pilot data ($n = 9$), the expected effect size for proprioceptive drift was estimated to be 0.74, and therefore we aimed for a sample size of 22 subjects per condition to obtain a statistical power of 0.95. 90 psychology undergraduate students participated for course credit, and 6 were excluded for the following reasons: 3 for excessive hand movements, 1 for not understanding and complying with instructions, and 2 for outlier responses on proprioceptive localization. The exclusion criteria were determined prior to the start of data collection. Outliers were defined as those exceeding 3 standard deviations from the sample mean. After exclusions the dataset consisted of 84 participants (61 female, mean age = 20.83, 78 right-handers), with 21 in each group. All participants provided written informed consent and research was approved by the UCLA Institutional Review Board.

### Materials

A custom-built box was utilized for the induction of the rubber hand illusion. It measured $70 * 46 * 18\ cm^3$ and was split into two compartments as depicted in Figure 2.5a below. The box was designed to ensure that participants' body midline would be at the midpoint of compartment 2 in order to create symmetry between the rubber and contralateral hands. Two standard paintbrushes were used to administer tactile stimulation. Opaque black silicone goggles were used to block the view during the setup of the experiment for each participant. A large $3*3\ m^2$ black cloth was used to cover the interface between participants' arms and the box. A reinforced block of cardboard with dimensions $70 * 46\ cm^2$ was used to cover the box in one of two positions depending on the block, as described in the procedures below. A left

rubber hand was used (48.3$cm$ long from elbow to fingertip, RI Novelty, www.amazon.com). For additional details see SOM.

**Procedure**

In the pre-test phase, subjects were seated at a desk and instructed to wear the light-occluding goggles while the box was positioned in front of them in accordance with Figure 2.5. A cardboard sheet was used to cover the box and an opaque cloth was draped over the participant's shoulders and the proximal part of the box in order to eliminate visual position cues from the arms.

Subjects performed the exact same proprioceptive localization task in the pre-test and post-test (see Figure 2.5b). After the setup described above, the room was darkened to preclude subjects from using visual cues in the periphery to anchor their responses. Instructions were given and the task immediately commenced where subjects used a computer mouse with their right hand to move a cursor on the bottom edge of the screen to the position of their left index finger along the azimuth. Given the large variability of the proprioceptive estimate of hand location along azimuth observed in previous studies (Beers et al., 1999; van Beers et al., 2002) as well as our pilot data, we collected a large number of responses in order to get a reliable estimate by using the mean of all the responses. Therefore, the measurement was repeated 40 times and the task took 4 minutes to complete. Proprioceptive drift was calculated as the mean post-test localization minus the mean pre-test localization.

After the pre-test, subjects' eyes were covered by the goggles once more while the experimenter reconfigured the cardboard sheet in the vertical configuration to form a barrier between the two compartments in order to occlude observer's view of their left arm. Goggles

were then removed and participants verbally responded to a question probing their ownership of the rubber hand, namely the third question in the traditional rubber hand illusion questionnaire – "I feel like the rubber hand is my hand" (Botvinick and Cohen, 1998) – with response categories ranging from -3 (strongly disagree) to +3 (strongly agree). Depending on the experimental condition, synchronous or asynchronous visuotactile stimulation was then applied, or none at all. This stimulation was performed by the experimenter who applied brushstrokes to the real left hand and the rubber hand in a proximal to distal direction with each stroke lasting about one second at 1-second intervals. The strokes were performed on all fingers of the hand, moving from finger to finger pseudorandomly. In this phase, subjects were repeatedly instructed to refrain from all body movements (see SOM for more detail).

The post-test commenced immediately after the illusion phase of the experiment. Subjects' eyes were covered by the goggles once more while the box was reconfigured for the proprioceptive localization task. The vertical cardboard was repositioned to its horizontal configuration covering the hand. Then, the goggles were removed and subjects performed the same proprioceptive localization task that they performed in the pre-test. Participants were again instructed to refrain from moving. After the proprioceptive localization, subjects were given the full 9-item questionnaire, which they were instructed to respond to using the mouse (Botvinick and Cohen, 1998).

We chose to emphasize ratings on question three of the full 9-item questionnaire as this has been consistently found to be the question that most correlates with the other measures of the illusion as well as directly assess the subjective phenomenology of the experience (Longo et al., 2008).

## Results

We first tested proprioceptive drift (computed as mean post-test localization minus mean pre-test localization) against zero and found a statistically significant difference for the 'sync' group ($t_{20} = 4.66,\ p < 0.001$, Cohen's $d = 1.02$), the 'async' group ($t_{20} = 2.69,\ p = 0.014$, Cohen's $d = 0.59$), and the 'no-stroke' group ($t_{20} = 2.18,\ p = 0.041$, Cohen's $d = 0.48$), but not for the 'no-hand' group ($t_{20} = 1.15,\ p = 0.262$, Cohen's $d = 0.25$).

Next, we computed a one-way ANOVA on the proprioceptive drifts across the levels of the Group variable ('sync', 'async', 'no-stroke', and 'no-hand'). This analysis showed a statistically significant main effect of Group, $F(3, 80) = 3.53;\ p = 0.019$ (see Figure 2.7a). Three planned comparisons were performed in order to test the role of three factors in the induction of proprioceptive drift. The role of presence of the rubber hand, the role of stroking the rubber hand, and the role of synchronicity of stroking were examined by comparing the proprioceptive drift in group 'sync' with those of 'no-hand', 'no-stroke', and 'async', respectively. One-tailed independent groups t-tests showed a significantly larger proprioceptive drift in the 'sync' group compared to that of 'async' group ($t_{40} = 2.50,\ p = 0.009$, Cohen's $d = 0.77$), as well as that of the 'no-stroke' group ($t_{40} = 1.81,\ p = 0.039$, Cohen's $d = 0.56$), and that of 'no-hand' group ($t_{40} = 2.94,\ p = 0.003$, Cohen's $d = 0.91$). To confirm that the proprioceptive drift effect did not dissipate across the 40 trials of post-test localizations, we computed a dependent samples t-test on the means of the first fifteen and final fifteen trials of proprioceptive localization, which revealed no significant difference ($t_{83} = -1.38,\ p = 0.171$). Participants' baseline proprioceptive localizations are shown in Figure S1 of SOM.

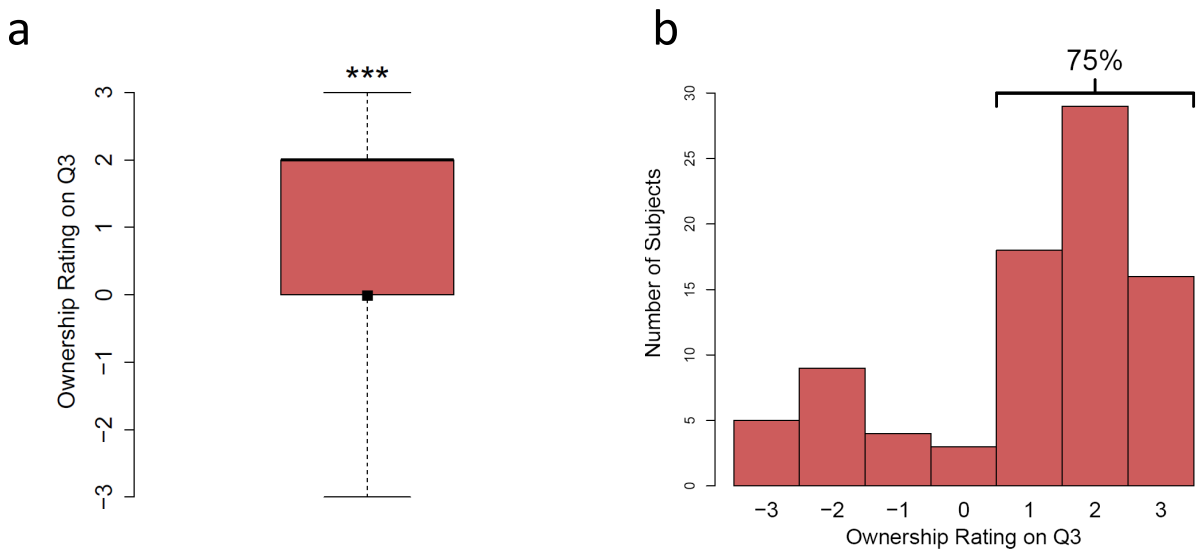


Figure 2.6: **Ownership Ratings Prior to Tactile Stimulation.** a) Median pre-test ratings from groups 'sync', 'async', and 'no-stroke' indicated by black square. Bars indicate interquartile range. b) Histogram of ownership ratings. The ratings are on a scale of -3 to 3, whereby -3 and +3 correspond to strong disagreement and strong agreement, respectively, with the statement "I feel as though the rubber hand is my hand." **** $p < 0.0001$

To address the question of whether illusion can occur in the absence of any tactile stimulation, we analyzed the pre-test (i.e., before any tactile stimulation was applied) measure of ownership from the groups that were presented with a rubber hand: 'sync', 'async', and 'no-stroke'. Since these groups did not differ at this point in the procedure, we collapsed the data across all three groups. This analysis revealed that 73% of participants rated the rubber hand as their own hand. As subjective ownership ratings are ordinal, a sign test was computed and revealed that the median of the pooled ownership ratings in the pre-test for the groups that were presented with a rubber hand (median = 2, indicating ownership) differed from zero, $p < 0.000$ (see Figure 2.6). Change in ownership scores were computed by subtracting the pre-test ratings from the post-test ratings, and were submitted to a Kruskal-Wallis one-way analysis of variance which revealed that there was a significant difference in median

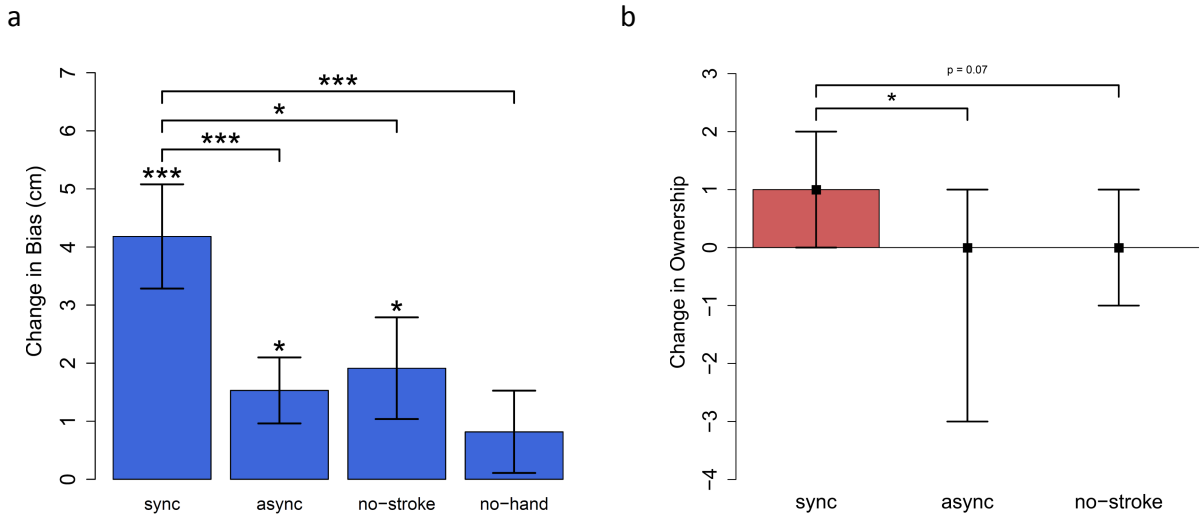


Figure 2.7: **Post-Test Results.** a) Proprioceptive Drift: The change in proprioceptive localization from pre-test to post-test. $n = 21$. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. b) Ownership: The median change in subjective ownership report from pre-test to post-test indicated by black squares. Bars display interquartile range. ** $p < 0.01$, *** $p < 0.001$

rating change between the three groups that saw the rubber hand, $\chi^2_2 = 8.7$, $p = 0.013$. Planned comparisons (sign tests) resulted in a significant median difference between 'sync' and 'async': $p = 0.049$, and a trend between 'sync' and 'no-stroke': $p = 0.077$ (see Figure 2.7b).

An interesting finding is that the 'async' group also exhibited a proprioceptive drift, albeit to a much smaller degree than that of the sync group. The 'async' group also showed only a trend for, and not a significant, decrease in ownership ratings. We believe the fact that asynchronous stroking did not entirely extinguish the perception of an illusion is due to the fact that the timing of the visual and tactile signals were 100% correlated. As shown by Parise and colleagues (Parise et al., 2012), this can induce the perception of a common cause. Alternatively, it may also be that this correlated stimulation caused a temporal recalibration between the two modalities and gradually brought the visual and tactile modalities in sync.

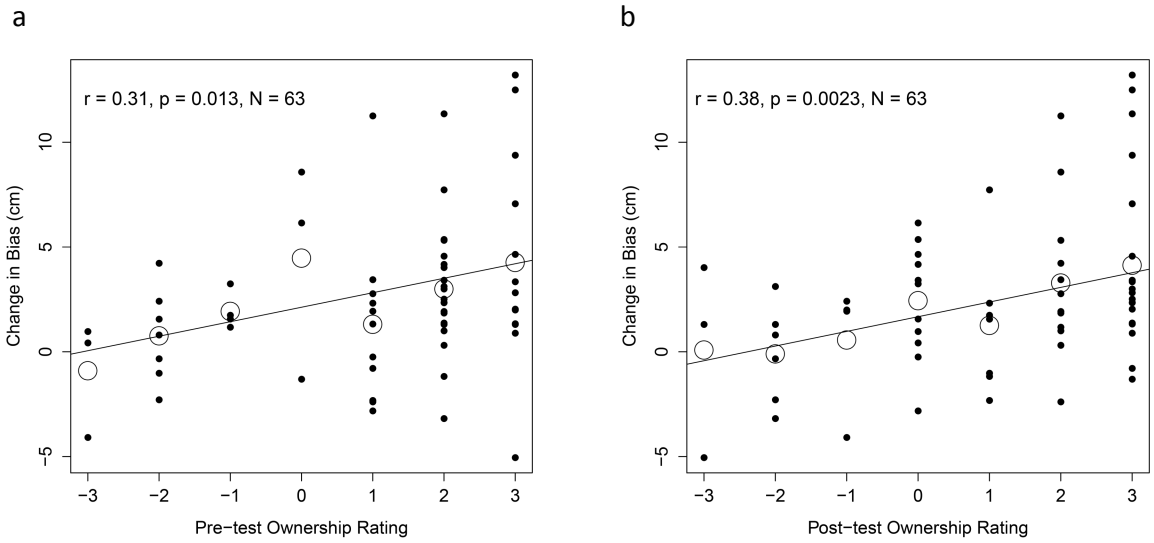


Figure 2.8: **Ownership and Proprioceptive Drift.** Scatterplot of ownership ratings plotted against proprioceptive drift in the (a) pre-test and (b) post-test from the three groups which were presented with a rubber hand. Large outlined circles represent means for those who gave the same ownership response.

Such fast recalibration of visual-tactile temporal synchrony has been previously reported (Fujisaki et al., 2004). We expect that a completely random relative timing of the rubber hand and real hand strokes would have more effectively suppressed the illusion and the consequent proprioceptive drift.

Finally, we examined correlations between the ownership ratings and the proprioceptive drifts separately for ratings from the pre-test and the post-test. The latter replicated previous reports of a significant correlation between drift and ownership (measured after application of tactile stimulation), $r = 0.38$; $p = 0.002$ (see Figure 2.8b). In addition, we found a significant correlation between the pre-test ownership ratings and drift, $r = 0.31$; $p = 0.013$ (see Figure 2.8a).

## 2.5 Experiment 2

The goal of this experiment was to examine whether RHI can occur in the absence of tactile stimulation (stroking) using skin conductance responses (SCR).

### Method

#### Design

We measured the participants' SCR in response to viewing of the rubber hand and to the threat to the rubber hand, and we collected questionnaire data as in Experiment 1. This experiment consisted of three between-group conditions, which differed only in the presentation of the rubber arm, as follows. The experimental condition was the same as the 'no-stroke' condition in Experiment 1, where the rubber hand was presented in an anatomically plausible position ('plausible-arm' condition). In order to examine the role of the illusion (ownership) in putative changes in SCR, we needed a control condition in which the illusion does not occur. As reported by previous studies, positioning the rubber hand in an anatomically implausible position would not induce the illusion (Tsakiris and Haggard, 2005). Therefore, in a control condition, we placed the rubber hand in front of subjects in a vertical orientation with the hand pointing downward, $104cm$ away from the subjects' shoulder and hanging from the bottom of a shelf mounted on the wall, outside of peripersonal space. We refer to this condition as 'hanging arm'. In our pilot study we noticed that for the illusion to be entirely eliminated the rubber hand needs to be outside the peripersonal space, hence the choice to position the arm at that distance. However, this difference in the distance of the rubber hand from the observer between the plausible-arm and hanging-arm conditions

meant that the simulated threat would also be at different distances from the observer in the two conditions, thus creating a confound for any potential difference in SCRs. To address this confound, we included an additional condition in which the threat was presented at the same location and distance from the observer as that of the plausible-arm condition, however no rubber hand was presented. We refer to this condition as the 'no arm' condition. If the SCR of the plausible-arm condition is higher than that measured in both the hanging-arm condition and the no-arm condition, it would indicate that the higher SCR is due to the percept of the illusion and cannot be attributed to the viewing of the rubber arm alone or scissors alone.

An increase in level of arousal (which can be induced by fear or surprise) is generally believed to result in an increase in SCR (Epstein and Roupenian, 1970; Staub et al., 1971). When observers perceive the rubber hand as their own hand, this is usually accompanied by a feeling of surprise and astonishment, thus raising arousal. Similarly, the observation of a threat to a (perceived) body part causes fear and increased arousal, and has been shown to increase SCR (Armel and Ramachandran, 2003). Therefore, we hypothesized that the majority of observers in the group that was presented with anatomically plausible rubber hand would report experiencing the illusion (as in Experiment 1), and these and only these participants would show an increased SCR to the viewing of the rubber hand and even a higher SCR response to the threat to the rubber hand.

**Participants**

We aimed for a sample size of 16-20 participants per condition following a previous study using the SCR measure of RHI (Armel and Ramachandran, 2003). 58 psychology

undergraduate students participated for course credit, and 7 were excluded due to technical malfunctions relating to running the code and electrode type used. After exclusions the dataset consisted of $N = 51$ (35 female, mean age = 20.7, 47 right-handers). Participants were pseudo-randomly assigned to three groups (see below for description), $n = 17$ per group. All participants provided written informed consent and research was approved by the UCLA Institutional Review Board.

### Materials

The same experimental setup and material were the same as those in Experiment 1. In addition, a custom built device composed of an electronic prototyping platform (Arduino SA, Italy) and 3M Red Dot Ag/AgCl electrodes was used for measuring skin conductance. This device was validated by running concurrent skin conductance measurements with an industry standard device (Biopac Systems, Inc.) and signals were correlated at $r = 0.58,\ p < 0.000$. Abrasive skin prep gel was used for electrode application and a pair of scissors was used to simulate a threat to the rubber hand.

### Procedure

The subjects were seated at a desk and skin prep gel was applied to the second joints of the palm side of index and middle finger of subjects' right hand. Two electrodes were placed on the prepped sites and connected to the skin conductance measuring device. Subjects were instructed to wear the light-occluding goggles, and to relax for 240 seconds. At this point, the subjects' skin conductance started being recorded at a sampling rate of $10Hz$ and was continued for the duration of the experiment. The experimenter then set up the

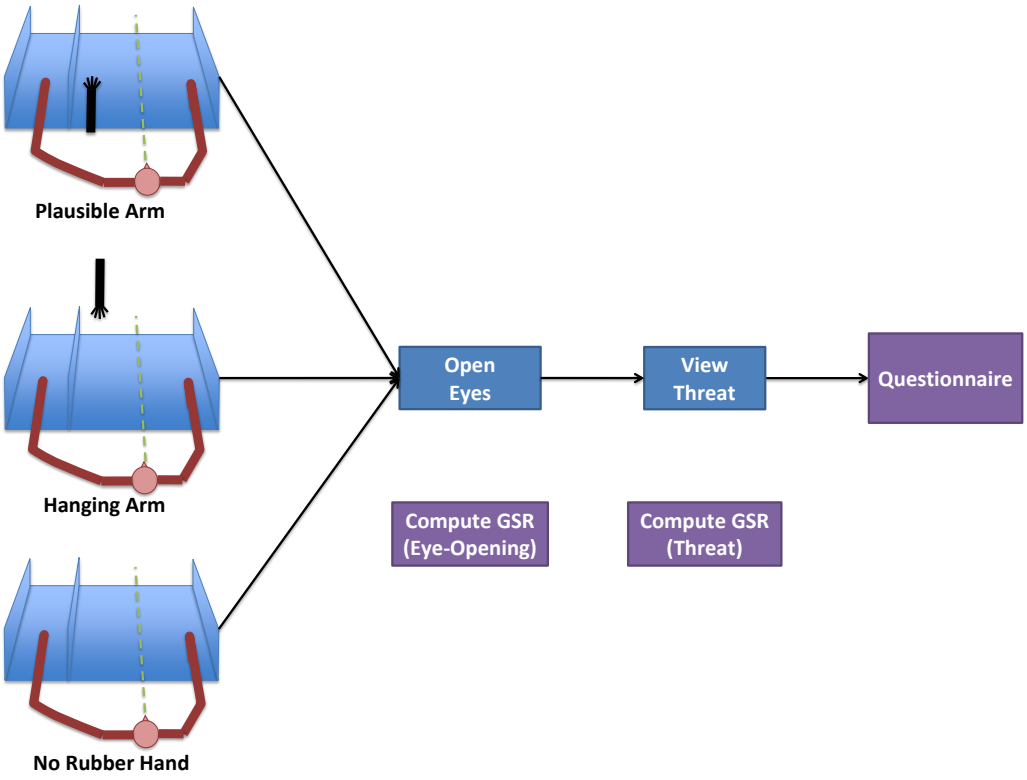


Figure 2.9: **Experiment 2 procedural design.**

experimental apparatus which was identical to the setup of Experiment 1. After this setup, the subjects were instructed to keep both their arms still and to relax for the remainder of the 240 seconds interval.

After the relaxation period, the goggles were removed and the participants attended to a location indicated by the experimenter. In the 'plausible arm' and 'hanging arm' condition, this was the rubber hand. In the 'no arm' condition, this location was the empty space where the rubber hand would have been placed. This moment was the first time-point at which a skin conductance response (SCR) was computed. We refer to this time-point as 'Eye Opening'. After 60 seconds of delay, the experimenter simulated a threat to the index finger of the rubber hand by pretending to aim to cut the finger using a pair of scissors, and maintained this simulated threat for 30 seconds. The beginning of this simulated threat is the onset of the second SCR time interval (we refer to this time-point as 'Threat'). In the 'no arm' condition, the experimenter applied the threat to the attended empty space. At the end of this exposure, subjects in the plausible-arm and hanging-arm groups were asked to rate their agreement with the statement 'I felt as if the rubber hand were my hand' on a scale ranging from -3 to 3 (see Figure 2.9).

**Analysis**

For each of the two time-points (removal of light-occluding goggles, application of threat), the skin conductance response (SCR) was calculated as the maximum skin conductance recorded within 1-5 seconds of that time-point, minus the minimum conductance during that same time window. To correct for non-normally distributed responses the following transformation was computed: log[SCR + 1] (Armel and Ramachandran, 2003; Christie and

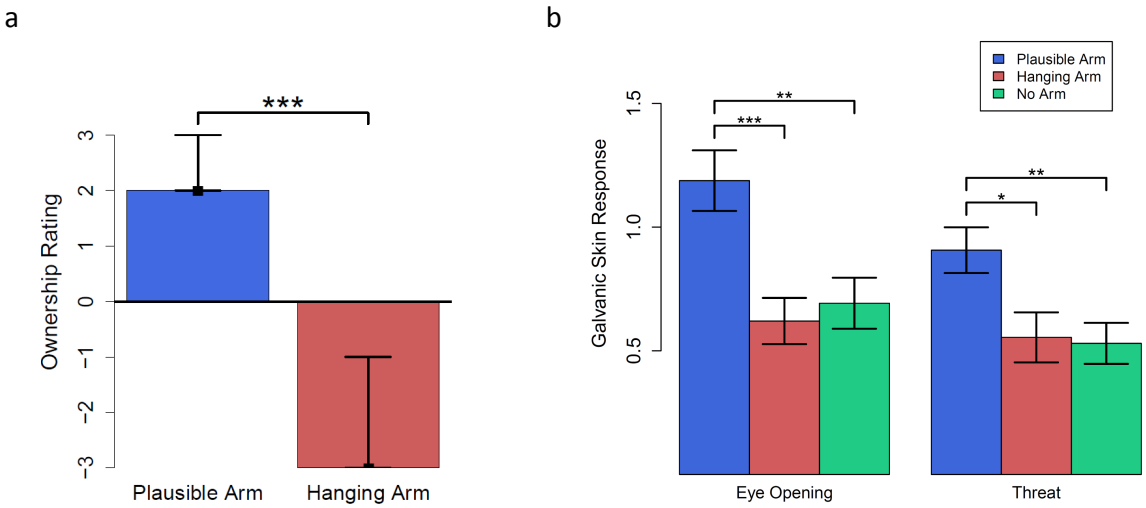


Figure 2.10: **Results.** a) Median ownership ratings after the end of the experiment indicated by black squares. Bars display interquartile range. b) Elicited SCR at two time points, "eye-opening" and "threat".

Venables, 1980).

## Results

As in Experiment 1, the majority (88% in this experiment) of the participants in the 'plausible-arm' condition reported ownership over the rubber hand (i.e., a positive rating on the ownership question). The median ownership rating of this group was 2.0, which a sign test revealed to be significantly different from zero ($p < 0.001$, see Figure 2.10a), suggesting a robust illusion for the participants in this group. The hanging-arm group, on the other hand, reported not experiencing an RHI as measured by the ownership ratings, the median of which was -3, which a sign test revealed to be significantly different from zero ($p < 0.001$, see Figure 2.10a). As expected, the plausible arm group showed significantly higher ownership ratings than the hanging arm group (paired sign test: $p < 0.000$).

Next, we examined the SCR responses at each of the two time points, eye-opening and

threat (see Figure 2.10b), across the three groups. One-way ANOVAs with the factor Condition (plausible-arm, hanging-arm, no-arm) at both time-points (eye-opening and threat) showed significant effect of condition ($F(2, 48) = 8.34,\ p < 0.001$; $F(2, 48) = 5.19,\ p = 0.009$, respectively). Planned comparisons between plausible-arm and the other two groups at both eye-opening and threat times showed a significantly higher SCR for the plausible-arm group compared to hanging-arm group (two-tailed independent groups t-tests, $t_{32} = 3.68,\ p < 0.001$, Cohen's $d = 1.26$, $t_{32} = 2.57,\ p = 0.007$, Cohen's $d = 0.88$, respectively) and no-arm group ($t_{32} = 3.1,\ p < 0.01$, Cohen's $d = 1.06$, $t_{32} = 3.04,\ p < 0.01$, Cohen's $d = 1.04$, respectively) (see Figure 2.10b). These results indicate that the increased SCR at eye-opening time cannot be explained by a general arousal from any visual stimulation (as in the no-arm group) or the surprise associated with seeing a rubber hand (as in the hanging-arm group). In fact, if the increased SCR was due to the observation of an odd stimulus, then the hanging-arm group should have exhibited the highest increase because that stimulus is arguably the most bizarre or unusual stimulus among the three conditions. Similarly, the increased SCR at the time of threat cannot be explained by the observation of movement of a sharp object per se (as the no-arm control), or the observation of action of a sharp object near a fake body arm (as in the hanging-arm condition). Therefore, the increased SCR appears to be associated with the ownership of the rubber arm.

If indeed the ownership of the rubber arm is the underlying factor for the observed increased SCR, then one would expect that a stronger sense of ownership would entail a stronger skin conductance response. We calculated the correlation between subjective ownership ratings and SCRs across participants in groups that were presented with a rubber-arm and provided subjective reports of ownership. There was a strong and statistically signifi-

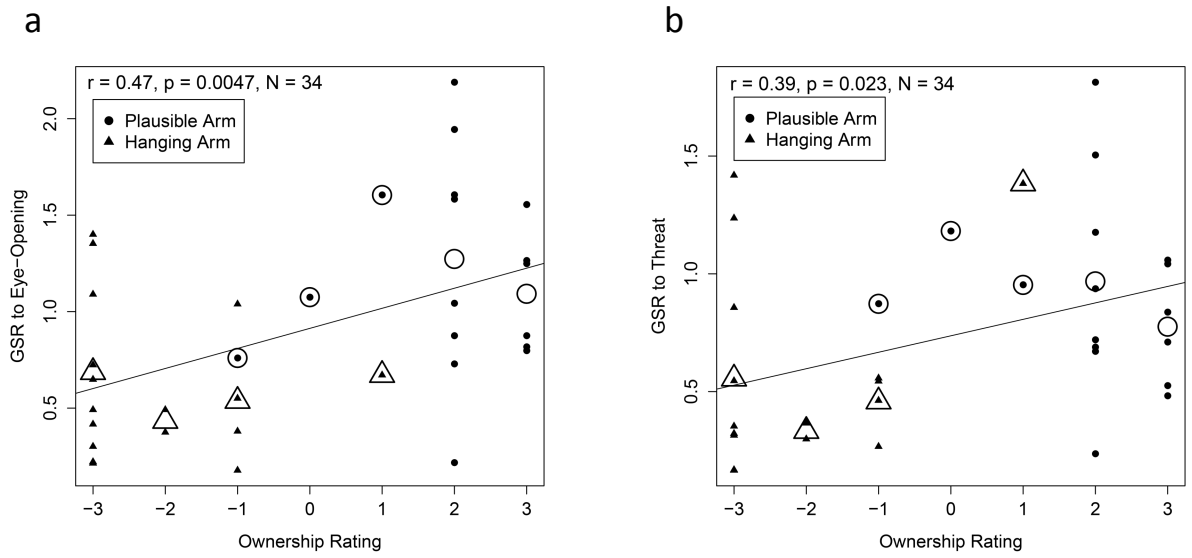


Figure 2.11: **Ownership and SCR.** Scatterplot of ownership ratings plotted against the logarithm of the SCR to Eye-Opening (a) and to Threat (b) from the two groups which were presented with a rubber hand. Large outlined shapes represent means for those who gave the same ownership response.

cant correlation, $r = 0.47,\ p = 0.005$, between the ownership ratings and the eye-opening SCRs, and also between the ownership ratings and the threat SCRs, $r = 0.39,\ p = 0.023$ (see Figure 2.11). Therefore, the objective and subjective measures of ownership consistently and strongly confirm the hypothesis that RHI can occur in the absence of tactile stimulation.

As can be seen in Figure 2.7b, the magnitude of the eye-opening SCR is comparable to that of threat SCR. We believe that the large SCR change at eye-opening time reflects the subjects' surprise at the dramatically changed appearance of what they perceive to be their hand, i.e., the rubber hand. The threat is presented after 60 seconds of delay. We speculate that this surprise and perhaps even the illusion fade with time and hence, the smaller SCR in response to the threat. It is also possible that the smaller SCR stemming from the threat may reflect a ceiling effect of the initial strong and sustained response to eye-opening. We interpret the strong skin conductance response to the first glimpse of the rubber hand as a

reflection of subjects' surprise at the heightened salience of the hand and the conflict that this produces with mental expectations about their hand appearance.

## 2.6 Discussion

While our intuition suggests that our sense of body ownership is in-born, fixed and immutable, recent research has shown otherwise. Simple and brief manipulations of our sensory experience can induce radical alteration of our body ownership and perception. We used one of these paradigms, namely RHI, to investigate the rules that govern body ownership. The Rubber Hand Illusion was discovered 16 years ago, and has been studied extensively since that time. However, the computational mechanisms of this illusion, which would provide insight into why this illusion occurs, have been largely unexplored and unaddressed to date.

In recent years, there has been much progress in our understanding of computational rules of multisensory perception. Specifically, it is now generally accepted that multisensory perception in natural environments involves two computational problems, the problem of causal inference – determining which signals are caused by the same source and which are caused by different sources – and the problem of integration – how to integrate the sensory signals originating from the same source. A Bayesian causal inference model (Kording et al., 2007; Shams and Beierholm, 2010; Shams et al., 2005), which addresses both of these problems in a normative and unified fashion, has been shown to account remarkably well for multisensory perception of the environment in spatial domain (Beierholm et al., 2009b; Kording et al., 2007; Wozny et al., 2010; Wozny and Shams, 2011a), and temporal domain (Shams et al., 2005; Wozny et al., 2008), and can account for two well-known illusions: Ventriloquist

illusion, and Sound-Induced Flash Illusion.

We adopted this model to examine whether it can account for the RHI. We included three modalities (proprioception, vision, touch) and both spatial and temporal information, the former provided by proprioception and vision and the latter provided by vision and touch. Our simulations accounted for the classic findings on RHI, namely that synchronous stroking produces the perception of a common cause for visual and tactile stimuli, and therefore the RHI is experienced, whereas asynchronous stroking produces the perception of independent causes, and no illusion is experienced. This provides the first computational account of the RHI. Furthermore, the model makes predictions about the spatial limit on the illusion, namely it predicts that the illusion will get weaker as the distance between the rubber hand and real hand increases, and starts to vanish as the distance approaches $30cm$ (see Figure 2.3) – a result which concords very closely with empirical findings reported in the literature (Lloyd, 2007).

To further explore the validity of this model, we investigated its untested predictions (but see (Rohde et al., 2011)). Specifically, the model predicted that the illusion can occur based purely on visual observation of the rubber hand, i.e., based purely on proprioceptive-visual integration. It is important to note that this prediction is in stark conflict with the common wisdom in RHI literature. It has been generally believed that visuotactile stimulation, in the form of synchronous stroking of the rubber hand and real hand, is required to induce the illusion. For instance, Holle and colleagues state that "it is rather uncontroversial that synchrony of touch with vision is a necessary condition" (Holle et al., 2011). Therefore, this prediction would provide a strong test of the model. We tested this prediction in two experiments that used different measures of ownership. The results of both experiments

indicated that a majority of participants experienced a vivid illusion despite not receiving any tactile stimulation. These results strongly confirm the prediction of the model. The model also predicted that synchronous stroking should strengthen the perception of the illusion. Indeed both proprioceptive drift data (between-group, Figure 2.7a) and ownership data (within-subject, Figure 2.7b) strongly confirm this prediction.

Proprioceptive drift, which is generally associated with the perception of the RHI (and has often been used as a measure of RHI), is a form of spatial recalibration of proprioception by the visual modality. Proprioceptive adaptation has been the subject of several studies (Hay and Pick Jr., 1966; Held and Hein, 1958; van Beers et al., 2002), although none investigated the relationship between inference of a common cause and the degree of recalibration. However, the relationship between the perception of a common cause and recalibration has been investigated in an audiovisual spatial task (Wozny and Shams, 2011a). This study showed that the magnitude of the visually-induced auditory spatial recalibration was significantly larger when a common cause is inferred (Wozny and Shams, 2011a). Consistent with these previous findings, here we found that ownership ratings both prior to and after stroking were significantly correlated with proprioceptive drift, indicating that the stronger the sense of a common cause for the proprioceptive and visual signals the stronger the recalibration of proprioception is by vision. The fact that pre-test ownership ratings correlated with the subsequently observed proprioceptive drift supports our conclusion that the illusion occurs in the absence of tactile stimulation, and that these pre-test ownership ratings index the same illusion that previous studies have assessed after stroking, despite solely arising from visuo-proprioceptive integration in this case.

The baseline proprioceptive bias that we have observed (see Figure S1) is consistent

with previous research reporting an accumulating drift in proprioceptive localization in the direction of the body midline (Paillard and Brouchon, 1968; Wann and Ibrahim, 1992). However, this bias has not been found by other studies that used different settings (Desmurget et al., 2000) and the exact factors/conditions underlying the bias remain unclear and require further study.

There are several procedural differences between our experiments and previous studies of RHI. We believe that some of these procedural aspects greatly enhance the illusion and may be the reason why we obtain the illusion in the absence of stroking whereas previous studies have not. First, the passive placement of the arm in the box under conditions of visual occlusion dampens the proprioceptive signal and reduces accuracy of localization (Paillard and Brouchon, 1968). This may serve to facilitate the integration of this noisy signal with the very reliable visual signal. Second, in our experiments the rubber hand's position was symmetrical with respect to the contralateral real hand (see Figure 2.5a), which may have the effect of increasing the anatomical and postural plausibility of the rubber hand. Third, in our experiments, the subjects did not see the rubber hand prior to the beginning of the trial (at which time the rubber hand was already in place and the real hand was already hidden) and did not see the experimenter hiding their real hand. Their eyes were covered throughout the time of experimental setup. This prevents the formation of a perceptual decision regarding the real and rubber hands prior to the subjects' exposure to them in the experimental positions. Our between-groups design additionally did not allow the observer to form such a perceptual decision in a different condition. Finally, we took great care to ensure that subjects were not able to see the proximal discontinuity between the rubber hand and their own body and to remove cues indicating that their hand was hidden behind

the cardboard divider, by covering this entire region (from shoulder downward to the rubber hand) with a thick black cloth. We believe these factors collectively resulted in a strong boost to visual-proprioceptive integration that gave rise to the illusion of ownership prior to the application of any tactile stimulation. In addition, we wonder whether the design of past studies of rubber hand illusion, in which the questionnaire is invariably administered only after stroking is applied, may have precluded the detection of the visuo-proprioceptive illusion in those participants who did experience it.

Notably, a recent study of RHI reported a proprioceptive drift in a condition that did not include tactile stroking (Rohde et al., 2011). However, subjects anecdotally reported not experiencing ownership of the rubber hand. Furthermore, the magnitude of proprioceptive drift in this condition was not smaller than that in the synchronous stroking condition. While the finding of the proprioceptive drift in the no-touch condition[1] is consistent with our findings, the absence of increase in the drift in the synchronous stroking condition, and the apparent lack of illusion in the no-touch condition are at odds with our results. We suspect that some of the same factors discussed above may play a role in these differences. For example, in Rohde et al.'s study, the rubber hand's position was aligned with subjects' midlines, rather than with the shoulder and this may strain the postural plausibility. Proprioceptive drift was computed based on only three measurements in each of the pre-test and post-test proprioceptive localization. Finally, and perhaps most importantly, the within-subject experimental design and the fact that the no-touch condition was preceded by synchronous and asynchronous stroking conditions may have caused carry-over effects in the no-touch

[1]The subjective ownership ratings were not obtained in that condition, and the debriefing data were not reported in the paper.

condition. In contrast, in the present study, proprioceptive drift was computed from the mean of 40 measurements (in pre-test and post-test each), all participants in rubber hand conditions had no prior exposure to the rubber hand (due to the between-groups design), and were asked about their experience of ownership immediately after being presented with the rubber hand.

The RHI has been studied extensively and several studies have shed light on the factors that can modulate the strength of the illusion. For example, it has been shown that asynchronous stroking or a large distance between the rubber hand and the real hand can weaken or disrupt the illusion (Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005). While the RHI has been viewed as a manifestation of visual-tactile integration, in the absence of a computational framework there has been no explanation for why the aforementioned factors matter and whether there are other factors that can influence the illusion. The current study fills this void, and provides a coherent understanding of the various facets of the illusion.

The Bayesian causal inference model shows that several factors contribute to the perception of a common cause and hence, the rubber hand illusion. These include the overlap between the proprioceptive and visual spatial estimates, which depends on both the spatial proximity of the rubber hand and the real hand as well as the degree of proprioceptive noise (and visual noise – although in most individuals fairly negligible), the congruency between tactile and visual sensations, and the a priori tendency to integrate crossmodal stimuli. This model predicts that the illusion is stronger the nearer the fake and real hand are to each other, the noisier the proprioception modality is, the more congruent the temporal pattern of stroking is across visual and tactile modalities, and the stronger the tendency to integrate signals. If one of these factors is weak, however, it will not necessarily break the illusion, as

the other factors can compensate and collectively provide sufficient evidence for a common cause. It is the strength of the overall evidence for a common cause that determines the probability of inferring a common cause and the illusion, and not any individual factor by itself. The finding that the pretest proprioceptive responses were biased by approximately $3.15cm$ towards the midline (see Figure S1) – and thereby towards the to-be-seen rubber hand – may provide a clue as to why the majority of our participants experienced the illusion before any tactile stroking was applied. If their proprioceptive estimate of their hand location is both imprecise and inaccurately skewed in this manner, the visual signal of the rubber hand would be more likely to be integrated with it. The model not only provides a quantitative description of the conditions that give rise to the illusion, but also explains that the RHI occurs as a result of optimal statistical inference about the causal structure and spatiotemporal properties of the sensations (with an explicit specification of the cost function that is being optimized).

While the Bayesian model presented here was intended only to model RHI in its standard form, the framework is nevertheless general and extendable to incorporate additional variables and to account for the RHI's variants. In the movement-induced RHI (Kalckert and Ehrsson, 2012; Walsh et al., 2011) (wherein the synchronous movement of the rubber hand and the real hand induces the illusion), the spatial conflict between the proprioceptive and visual estimates is compensated for by the temporal congruence of the kinesthetic and visual estimates. The kinesthetic signals would substitute for the tactile signals in the current model. In the self-touch RHI (Ehrsson et al., 2005) (wherein the active hand touches the rubber hand synchronously with a touch of the passive hand), there is a spatial conflict between two proprioceptive estimates, that of the passive hand, and that of the actively

touching hand. This spatial conflict is compensated for by the synchrony of the two tactile signals, the one felt by the passive hand and the one felt by the actively touching hand. In the invisible hand illusion (Guterstam et al., 2013), there is no rubber hand and the stroking is applied to empty space, inducing the illusion of ownership of an invisible hand. The main difference between this illusion and the conventional RHI is in the visual object recognition computations that result in perception of a hand in the conventional RHI and no object in this variant. As the current model does not include these computations, but rather assumes these object processing steps have already been completed and provided the perception of a posturally congruent hand, the model in its current form is not equipped to capture this difference. Having said that, if we nonetheless assume that the kinematic details of the stroking of the invisible hand convey sufficient information to the hand recognition module in the brain which would in turn infer the existence of a transparent hand, then the output of this object recognition module would indeed provide the visual signal that our model uses as input, though in a degraded form.

It has also been shown that if the rubber hand is positioned in an anatomically implausible way (Tsakiris and Haggard, 2005), the illusion does not occur. The model in its current form makes the simplifying assumption that the rubber hand has an anatomically plausible and congruent posture. The model currently does not incorporate hand posture as a variable and therefore, is not equipped to incorporate the congruency in posture as a factor contributing to the inference of a common cause (and hence, the illusion). Should the model be extended to incorporate posture as an additional random variable, the incongruence between the posture of the real and fake hand would decrease the probability of a common cause and can break the illusion. Finally, this model is not intended to capture the full temporal dynamics of the

emergence of the illusory percept, reported by several studies to be 5-10 seconds after the administration of stroking (Ehrsson et al., 2004; Lloyd, 2007; Ehrsson et al., 2005; Guterstam et al., 2013). However, it can be extended to do so. As the evidence for the synchrony of stroking increases, so does the evidence for common cause, strengthening the illusion. In cases where the inference of common cause had not yet exceeded $p = 0.5$ (i.e., where there is no experience of the RHI), sustained stroking in synchrony would accumulate the evidence and could eventually tip the balance towards inference of a common cause.

In conclusion, a normative Bayesian model that makes an inference about the causal structure of sensory stimuli, namely visual, proprioceptive and tactile signals, based on the similarity of the stimuli and prior knowledge can account for the rubber hand illusion. Moreover, several predictions of this model were confirmed empirically providing strong support for the notion that a Bayesian causal inference process is involved in the perception of body and experience of body ownership. More specifically, these results suggest that when the spatio-temporal information conveyed by the senses are sufficiently congruent, a common cause for the sensations is inferred by the nervous system leading to the experience of unified source and body ownership. If an incongruity is artificially introduced between two of the senses (e.g., between visual and proprioceptive spatial information) as in the studies of rubber hand illusion or out-of-body experience, then additional information providing support for a common cause, such as congruent tactile temporal information, may be needed to provide sufficient "evidence" for a common cause and the perception of body ownership, and hence the illusion.

The studies of body ownership such as rubber hand illusion and out-of-body experience (Armel and Ramachandran, 2003; Botvinick and Cohen, 1998; Ehrsson, 2007; Lenggenhager

et al., 2007) have already revealed that humans' body representation and sense of body ownership is remarkably malleable. What the current findings show is that this process can be modeled as a sophisticated and statistically optimal rule of inference (Bayesian causal inference) which also appears to govern other perceptual processes. Therefore, it appears that our perception and consciousness of self is no different in principle than our perception of the outside world: it follows the same rules, and it can be altered in the same fashion.

## Acknowledgments

We would like to thank Lawrence Rosenblum and Martin Monti for their valuable feedback and insightful comments on the manuscript.

# Supporting Information

**Dataset S2. Experiment 1 Data.** Comma Separated Values (.csv) File containing collected dataset from experiment 1.

**Dataset S3. Experiment 2 Data.** Comma Separated Values File (.csv) containing collected dataset from experiment 2.

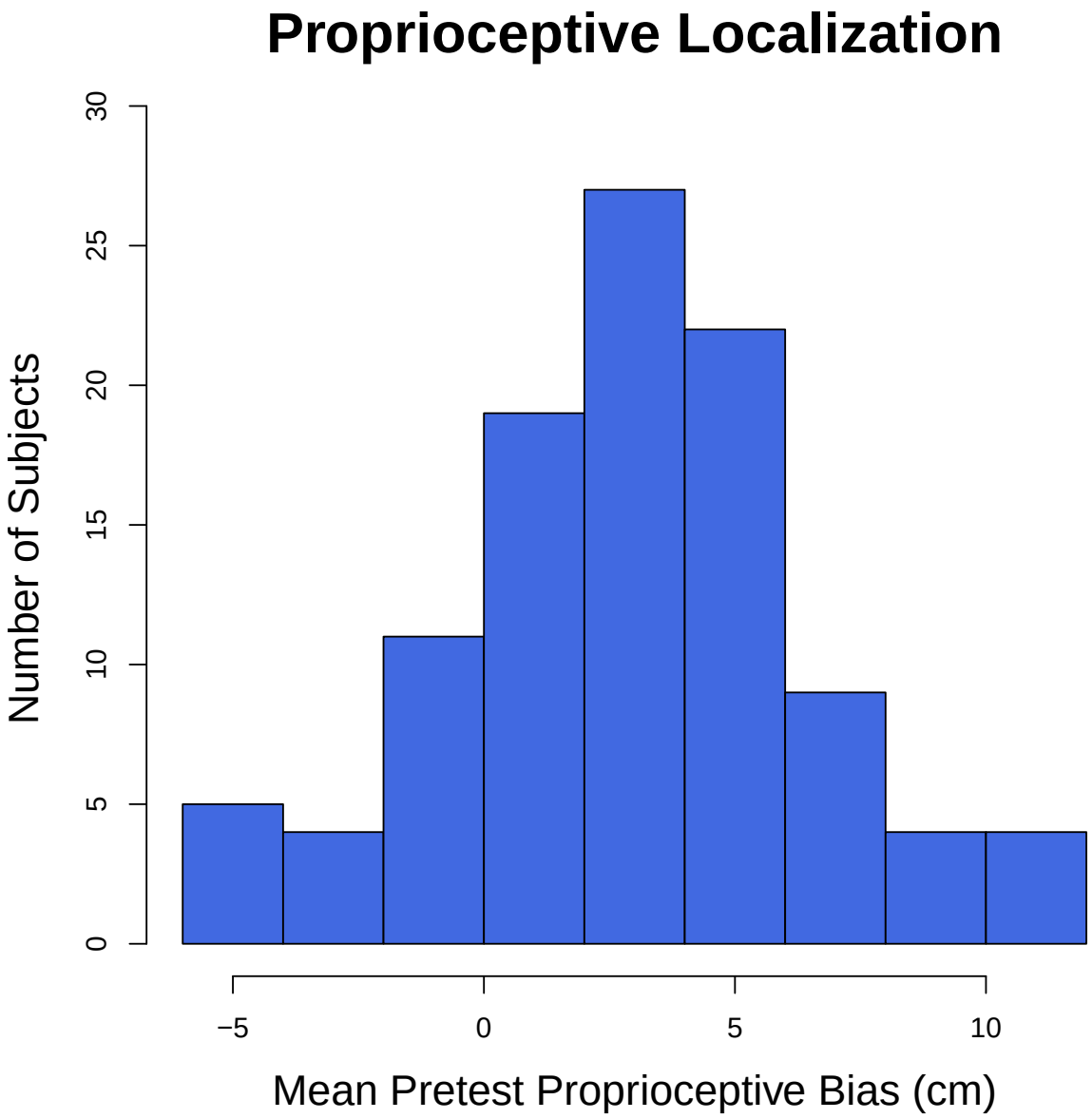


Figure 2.12: **Figure S1. Pre-Test Proprioceptive Bias.** Proprioceptive localization responses of all participants during pre-test revealed a statistically significant bias towards the midline ($t_{83} = 7.88$, $p < 0.0001$). The average bias across participants was $3.15cm$, and the average standard deviation of subjects' 40 localization responses on this task was $1.3cm$. **** $p < 0.0001$.

# Chapter 3

# Visual-Somatotopic Interactions in Spatial Perception

## 3.1 Abstract

Ventriloquism is a well-studied multisensory illusion of audiovisual spatial perception in which the perceived location of an auditory stimulus is shifted in the direction of a synchronous but spatially discrepant visual stimulus (Howard and Templeton, 1966). This effect is due to vision's superior acuity in the spatial dimension, but has also been shown to be influenced by the perception of unity of the two signals (Wallace et al., 2004). We sought to investigate whether a similar phenomenon may occur between vision and somatosensation along the surface of the body, as vision is known to possess superior spatial acuity to somatosensation. We report the first demonstration of the visuotactile ventriloquist illusion: subjects were instructed to localize visual stimuli (small white disks) or tactile stimuli (brief localized vibrations) that were presented concurrently or individually along the surface of

the forearm, where bimodal presentations included spatially congruent and incongruent stimuli. Subjects showed strong visual-tactile interactions. The tactile localization was strongly biased in the direction of the visual stimulus, and the magnitude of this bias decreased as the spatial disparity between the two stimuli increased. The Bayesian causal inference model which has previously been shown to account for auditory-visual spatial localization and ventriloquism effect also accounted well for the present data. Therefore, crossmodal interactions involving spatial representation along the surface of the body follow the same rules as crossmodal interactions involving representations of external space (auditory-visual).

## 3.2 Introduction

Ventriloquism is the effect that occurs when a puppet's moving mouth causes an audience to misperceive the sounds they hear as having originated in the puppet, rather than in their true source, the puppeteer's mouth. Although this illusion may appear to represent an error on behalf of the perceptual system, it has been shown (Kording et al., 2007; Wozny and Shams, 2011a; Wozny et al., 2010) that it is an epiphenomenon of a Bayes optimal spatial perception.

Recently, we have demonstrated that the same Bayesian inference model (Bayesian causal inference) that has accounted for observers' auditory-visual spatial (as well as temporal) perception, can also explain the Rubber-hand illusion which involves perception of body ownership and body part location (Samad et al., 2015). However, this previous study probed the perception of the hand in allocentric spatial coordinates. In other words, while the spatial inference task involved location of a body part, it was still in an allocentric frame of reference.

Here, we aimed to investigate the rules governing spatial perception in a different reference frame: that of the surface of the body, or somatotopic coordinates.

Several previous studies have investigated visuotactile interactions both neurally and behaviorally. A series of single-cell recording experiments by Michael Graziano and others (Graziano and Botvinick, 2002; Graziano et al., 2000; Graziano and Gross, 1998) has shown that across a network of brain regions, including the ventral premotor cortex, the putamen and the intraparietal sulcus region, there exist populations of bimodal neurons that respond to both visual and tactile stimuli; the response of these neurons is facilitated when the stimuli are spatially and temporally congruent (Avillac et al., 2007). Additionally, it has been reported that the visual receptive fields of these neurons dynamically remap so as to stay spatially coincident with the tactile receptive fields (Graziano et al., 1994). These intriguing results suggest that the computations involved ought to include a step to infer whether or not a tactile and a visual stimulus came from the same location along the surface of the body. A parallel can be drawn between these findings and those of Meredith and Stein reporting similar audio-visual facilitatory activity in the superior colliculus, and the postulation of these circuits as the neural mechanism underlying auditory-visual ventriloquist illusion (Meredith and Stein, 1986).

Behavioral studies have also reported a variety of visuotactile interactions. For example, viewing one's hand reduces tactile detection reaction times (Tipper et al., 1998) and improves tactile two-point discrimination despite its task-irrelevance (Kennett et al., 2001).

In light of these findings, it is not unlikely that visual and tactile modalities also interact in perception of space along the surface of the body. Therefore, we asked whether something akin to the ventriloquist illusion operates in somatotopic coordinates. In particular, our

research question was: do visual and somatotopic representations interact in spatial perception? We hypothesized that the less precise tactile representations will be shifted towards the more precise visual representations, and that as in auditory-visual interactions, this interaction will primarily occur for intermediate disparities (when a common cause is perceived), and not when the disparity is too large to allow the perception of a common cause (Kording et al., 2007; Wozny and Shams, 2011a; Wozny et al., 2010; Samad et al., 2015). This pattern of results would be consistent with Bayesian causal inference that has previously been shown to account for audiovisual ventriloquism very well (Kording et al., 2007; Wozny and Shams, 2011a; Wozny et al., 2010). We predicted that Bayesian Causal Inference would provide a unifying account of multisensory integration, whether it involves representations encoded in an external reference frame or somatotopic coordinates. Alternatively, it is possible that perception of one's own body is special: it may involve hard-wired representations, not be as prone to crossmodal interactions, or such interactions may be governed by different rules. In that case, we would not observe interactions between tactile and visual modalities akin to the ventriloquist illusion, and the Bayesian Causal Inference would not be able to account for the data.

## 3.3 Method

### Participants

Twenty-one UCLA psychology undergraduate students (16 females; average age 20.1, ages ranging 19-22) participated in the experiment and received course credit for their par-

ticipation. All participants were screened to ensure normal or corrected to normal vision, and consented using procedures that were approved by the Institutional Review Board at UCLA.

## Materials

Our experimental setup consisted of three components: a tactor array for tactile stimuli, a projector with mirror mount for visual stimuli, and a computer running Matlab with PsychToolBox for control of the stimuli and recording responses (see Figure 3.1A).

The projector was mounted on a shelf directly above the subject, with a mirror angled at 45 degrees so as to reflect the display onto the subject's forearm below.

The tactor array consisted of 5 tactors (Pico Vibe $9mm$ Vibration Motor - 25mm Type; Model Number: 307-103) embedded into a soft foam material, spaced apart by 41.1mm, a distance that subtended 12° of visual angle – thus spanning the positions -24°, -12°, 0°, 12°, 24° – and driven by a custom built controller circuit that used an Arduino (Arduino SA, Italy) to interface with Matlab. The tactor array was then affixed to a piece of acrylic measuring 30.5 x 10 $cm^2$, and which was itself attached by a hinge joint to a vertical mounted piece of wood. Each subject placed their arm under the tactor array, with their wrist upon a thick foam layer in order to flatten the upper surface of the arm as much as possible. Then, the foam array was lowered onto the forearm and pressed down using a $750ml$ filled bottle weighing approximately $2kg$. This was done so as to ensure that half of the surface of the forearm was covered by the foam array, and thus would feel the tactor stimulation, and the other half would be exposed to the projector's display, and thus would allow presentation

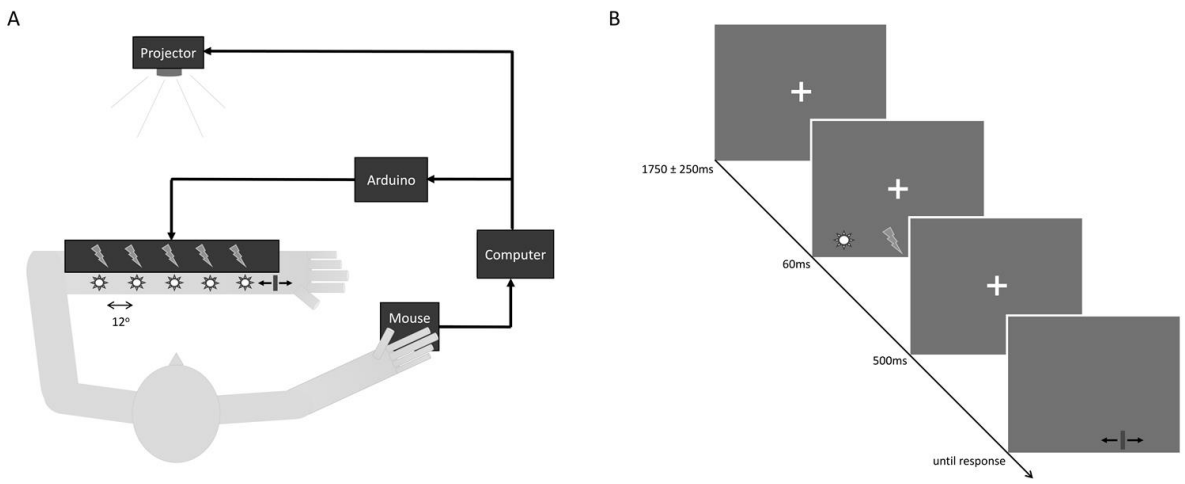


Figure 3.1: **A. Experimental Setup.** Diagram depicting the experimental setup used to present the stimuli, including a tactor array that is driven by a microcontroller (Arduino) for tactile stimulus delivery, and a projector for visual stimulus delivery, and a computer for control of both. **B. Trial Design.** This shows the sequence for a given trial: A fixation cross is presented on the screen for a variable interstimulus interval, after which stimuli are presented for $60ms$. After a $500ms$ interval, the fixation cross disappears and a cursor appears for subjects to make their responses with.

of visual stimuli. This enabled the bisensory stimuli to be as proximate to each other as possible while allowing for both stimuli to be presented directly on the surface of the arm.

Finally, Gaussian noise audio was played at 70 $dB$ simultaneously with stimuli on a pair of headphones worn by the subject in order to eliminate the audibility of the vibration of the tactors. The volume of this noise mask was determined in our pilot experiments such that the location of the tactile stimuli could not be determined based on the tactor noise. Subjects' had their head position fixed by means of a chin-rest that was placed 195mm away from the tactor array. Subjects were allowed to adjust the height of the chair and/or the chin-rest to achieve a comfortable posture.

## Procedure

13 repetitions of all 35 stimulus conditions (5 x 5 = 25 possible bisensory presentations and 5 + 5 = 10 possible unisensory presentations) were presented in a pseudorandom order.

Stimuli were presented for a duration of 60ms and intertrial intervals were sampled from the normal distribution: $\mathcal{N}(\mu = 1.5s, \sigma = 0.25s)$. Subjects were given a 1-2 minute break every 150 trials.

The subjects' task was to localize the visual and tactile stimuli on each trial using a mouse cursor that was restricted to movement in the azimuth and bounded to the region of stimulus appearance spanning 93.4 degrees. The order of these localizations was counterbalanced across subjects. The color of the mouse cursor corresponded to the modality to be localized, blue to indicate a visual localization and red to indicate a tactile localization. The mouse cursor appeared 500 ms after stimulus offset and they were instructed to move the mouse rapidly to the position they perceived the stimulus to have been presented at and click with the left mouse button to indicate their response (see Figure 3.1B).

A white fixation cross was presented at a position 30 degrees above the middle tactor position. To ensure fixation, the observers were asked to also perform a fixation task throughout the experiment. On 10% of trials, the fixation cross changed color from white to red simultaneously with stimulus appearance and for 500 $ms$ longer. Participants were instructed to press the middle mouse button whenever they detected the change of color of the plus sign. The duration of the presentation of the colored cross was adjusted during piloting such that it was only detectable if the observer directly looked it and not if they gazed away from the fixation point. The fixation cross disappeared during responding and

subjects were only asked to fixate during stimulus presentation and between trials.

### Modeling

We used the Bayesian Causal Inference model as described in (Wozny et al., 2010) to fit the data from each individual observer. The perceptual decision-making strategy (model averaging, probability matching, model selection) was also fitted to the data of each observer.

## 3.4 Results

The data from the fixation task showed that observers detected the change of color 96.2% of the time on average and therefore fixated on the fixation point throughout the experiment.

We first analyzed unisensory trials by computing the standard deviation of responses for each modality at each of the five spatial positions. For tactile unisensory trials, the average standard deviation across subjects and positions, with its 95% confidence interval was $6.18° \pm 0.47$. In contrast, for visual unisensory trials, it was $2.83° \pm 0.46$. Thus, participants' visual responses were more than twice as precise as their tactile responses.

To investigate crossmodal interactions, from incongruent bisensory trials, we computed the bias in each modality as follows:

$$\Delta V = \frac{\hat{V} - V}{T - V}, \qquad \Delta T = \frac{\hat{T} - T}{V - T}$$

where $\Delta V$ and $\Delta T$ denote the bias in visual and tactile responses, respectively and where $\hat{V}$ and $\hat{T}$ denote the visual and the tactile responses, respectively, and where $V$ and $T$ denote the true visual and tactile stimulus locations, respectively.

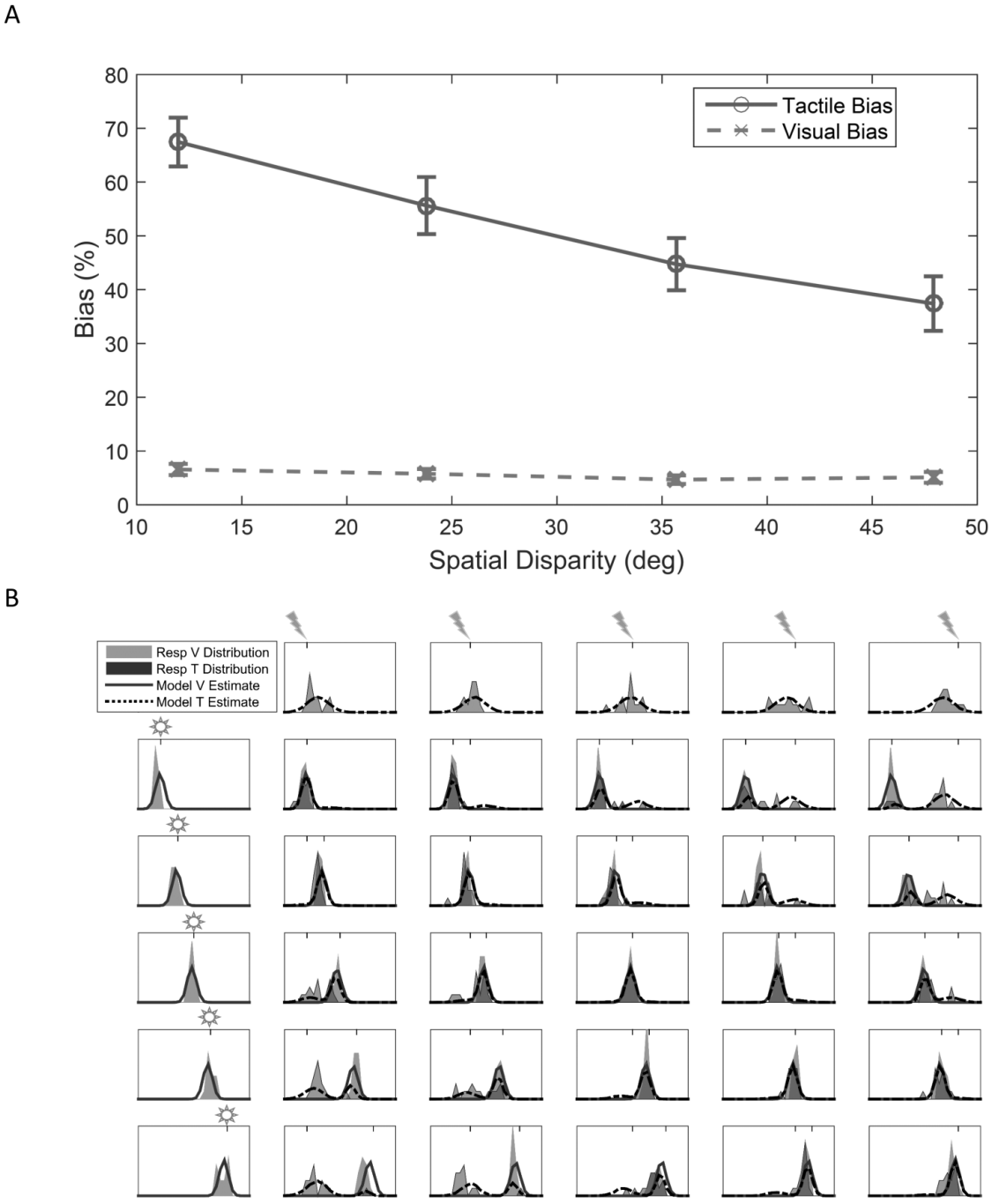


Figure 3.2: **Visual-Tactile Interactions. A.** Average tactile bias and visual bias as a function of spatial disparity between the stimuli across all participants. Error bars represent S.E.M. **B.** Response distributions and Model Fits for a Representative Observer. Each panel represents data and model fits for one of the stimulus conditions. The first row and the first column correspond to unisensory tactile and unisensory visual conditions, respectively. The remaining panels correspond to bisensory conditions. In each panel, the horizontal axis represents the spatial position, and the vertical axis denotes response probability. Shaded areas represent distribution of responses for each modality. Thick solid and dashed lines represent model fits for each modality. Tick marks at the top of every inset represent true stimulus locations. Resp, response.

Average bias across all disparities and all participants was 51.3% for touch and 5.5% for vision. Figure 3.2A shows the average bias for each modality as a function of spatial disparity between the two modalities. As can be seen the visual bias is small for all disparities, which is not surprising given the much lower reliability in the visual modality. The tactile bias is large for small disparities and decreases as a function of disparity (regression slope, $\beta_{disparity} = -0.89$, $t_{418} = -5.83$, $p < 10^{-8}$). This is consistent with previous findings from audiovisual ventriloquist studies (Kording et al., 2007).

Figure 3.2B shows the response distributions of a representative participant for all stimulus conditions and the model fits. As can be seen in the unisensory tactile conditions (first row), the responses are shifted towards the center relative to the veridical position of the stimulus.

A linear regression analysis using data from all subjects confirmed a consistent bias in the tactile perception towards the center (regression slope, $\beta_{tact} = -0.3$, $t_{1363} = -22.3$, $p < 10^{-94}$), indicating that positions in the periphery were shifted towards the center by as much as 30%.

The Bayesian Causal Inference accounted for the data very well. This can be seen in Fig. 3.2B for a representative subject. The thick solid lines indicate that the distributions of the model predictions are in very close concordance with this subject's actual response distribution (depicted by the shaded areas). The average generalized goodness of fit coefficient across all observers was $R^2 = 0.812$ with 95% CI [0.798, 0.825] (Nagelkerke, 1991).

## 3.5 Discussion

We report evidence for a visual-tactile ventriloquist illusion, in which visual and tactile stimuli spatially interact along the surface of the body. In summary, when subjects were presented with visual and tactile stimuli along their forearm, their localization judgments of the tactile stimuli were significantly biased towards the visual stimulus location. Moreover, this integration was consistent with the rules of optimal Bayesian statistical inference, as the noisier tactile signal was biased towards the more reliable visual signal, and crossmodal interactions occurred according to an inference about the causal structure of the stimuli.

As both sensory modalities are represented in a spatially topographic map, the presence of these interactions makes sense and is consistent with similar interactions that have been observed between visual and auditory modalities (Wallace et al., 2004; Kording et al., 2007; Wozny and Shams, 2011a). This is especially true considering the evidence from single cell recordings and behavioral studies (Graziano et al., 2000; Graziano and Botvinick, 2002; Graziano and Gross, 1998; Avillac et al., 2007; Ladavas, 2002) indicating interactions between the two modalities at the single-unit level of representation. In addition, recent studies suggest that the interaction between the visual and somatosensory modalities may happen at a very early processing level (Mahoney et al., 2015; Ley et al., 2015; Sieben et al., 2013).

To our knowledge, this is the first demonstration of a spatial biasing of a tactile representation by a visual representation along the surface of the skin. The results of unisensory tactile trials in the present study also showed a spatial compression in perception of stimuli. This finding is consistent with several previous findings (Green, 1982), some dating back to at least 1834, with Weber's observation that distances along the surface of the skin are

underestimated on regions that have poorer acuity – a phenomenon he referred to as spatial compression (Ross and Murray, 1978).

At any given moment, the nervous system is typically busy processing multiple sensory stimuli, and the perceptual system has to determine a) which of these signals are caused by the same object/event–a problem referred to as "the causal inference problem", and b) for those sensory signals that are inferred to have the same cause, how to integrate them in order to achieve the best estimate of objects/events in the environment–a problem referred to as "the integration problem". The Bayesian Causal Inference model is a normative model derived from fundamental statistical principles that solves both of these problems in a coherent and unified fashion. The inference about the causal structure (common cause vs. independent causes) is based on the consistency between the signals as well as the a priori knowledge about the world (i.e., the prior tendency or bias for perceiving a common cause). The inference involved in the integration process similarly depends on the sensory signals (their reliabilities), and the prior knowledge about their occurrence. Another noteworthy feature of this model is its parsimony. It explains both causal inference and integration processes, and accounts quantitatively for individual observers' data remarkably well with minimal model complexity. The model has only four free parameters and accounts for 780 data points.

The Causal Inference model has previously been quantitatively compared with other models of multisensory perception and has been shown to be substantially superior at accounting for behavioral data (Kording et al., 2007). Traditional models assume integration – sometimes also called forced fusion – and are therefore, incapable of accounting for partial integration and segregation phenomena (Beierholm et al., 2009a; Shams and Beierholm,

2011).

Finally, the fact that this model has been shown to account for a wide variety of multisensory phenomena, ranging in tasks and sensory modality combinations (Samad et al., 2015; Beierholm et al., 2009b; Wozny et al., 2008) strongly suggests that it is a computation that evolution has adopted in solving a variety of perceptual problems, thus, in widespread use across brain regions. It provides a normative and unifying account for how nervous systems combine information from the various modalities and coordinate systems, both in the external world and on the surface of the body.

In conclusion, the data reported here show that ventriloquist illusion is not limited to interaction of modalities in external space, and extends also to somatotopic representations of the space. And more importantly, the results provide yet another demonstration of Bayesian causal inference in multisensory perception, where the causal structure and the estimate of the stimuli are inferred optimally from available sensory inputs and prior knowledge. This demonstrates that the integration of visual and tactile information that is encoded in somatotopic coordinates is governed by the same rules of statistical inference that many other perceptual processes have been shown to obey (Samad et al., 2015; Beierholm et al., 2009b; Wozny et al., 2008).

# Chapter 4

# Recalibrating the Body: Visuotactile Ventriloquism Aftereffect

## 4.1 Abstract

Visuotactile ventriloquism is a recently reported effect showing that somatotopic tactile representations (namely, representation of location along the surface of one's arm) can be biased by simultaneous presentation of a visual stimulus in a spatial localization task along the surface of the skin. Here we investigated whether the exposure to discrepancy between tactile and visual stimuli on the skin can induce lasting changes in the somatotopic representations of space. We conducted an experiment investigating this question by asking participants to perform a localization task that included unisensory and bisensory trials, before and after exposure to spatially discrepant visuotactile stimuli. Subjects localized brief flashes of light and brief vibrations that were presented along the surface of their forearms, and were presented either individually (unisensory conditions) or were presented simultaneously at the

same location or different locations. We then compared the localization of tactile stimuli in unisensory tactile conditions before and after the exposure to discrepant bisensory stimuli. After exposure, subjects exhibited a shift in their tactile localizations in the direction of the visual stimulus that was presented during the exposure phase. These results demonstrate that the somatotopic spatial representations are capable of rapidly recalibrating after a very brief exposure to visually discrepant stimuli.

## 4.2 Introduction

The nervous system is at all times playing a guessing game with the aim of identifying which sensations should be integrated and which ought to be segregated. For example, consider what happens when a sound and sight are concurrently processed by the brain. If they originate from different sources, but the brain erroneously infers that they have a common cause, this can lead to an illusion known as the ventriloquist illusion, wherein the perceived location of the sound is captured by the location of the visual stimulus (Alais and Burr, 2004). A similar phenomenon has been shown to occur between auditory and tactile representations (Caclin et al., 2002). However, when some of the sensations involve somatosensation, these guesses may also lead to aberrant bodily percepts, such as the rubber hand illusion (Botvinick and Cohen, 1998; Armel and Ramachandran, 2003; Samad et al., 2015). Moreover, we have previously shown that visual and tactile stimuli interact in somatotopic coordinates such that estimations of tactile stimulus location are biased towards concurrently presented visual stimuli (Samad and Shams, 2016).

In the audiovisual ventriloquist illusion, prior work has uncovered evidence that there is

an accompanying aftereffect that develops as a result of exposure to audiovisual stimulus pairs (Lewald, 2002; Recanzone, 1998). In brief, this effect is interpreted as a recalibration of the mapping between auditory and visual spatial representations. In a study that examined this, subjects were given an exposure phase where they were presented with audiovisual stimulus pairs that always had a constant disparity between them for no longer than 10 minutes (Wozny and Shams, 2011b). Results from this study showed that subjects' localizations after this exposure were significantly biased in the direction of the visual stimulus that was paired with the auditory stimulus during this exposure phase (Wozny and Shams, 2011b).

While this kind of spatial recalibration has been shown for auditory and visual spaces, it is not yet clear whether the somatotopic space is similarly malleable. We designed an experiment to test the hypothesis that visuotactile ventriloquism induces an aftereffect such that prolonged exposure to spatially incongruent visuotactile stimuli results in a measurable recalibration of tactile representations. Given that our previous study (Samad and Shams, 2016) identified a vigorous interaction between visual and tactile stimuli in the somatotopic space, we hypothesized that an aftereffect would also be observable such that tactile representations would be biased, dependent on the disparity between the visuotactile stimuli that were presented during the exposure phase.

## 4.3 Method

### Participants

Thirty-seven individuals (23 female) with a mean age of 21.3 gave written consent to participate for course credit. All participants had normal or corrected-to-normal vision. The experimental methods were approved by the UCLA IRB. One participant was excluded from the experiment for non-compliance with instructions. The remaining participants ($N = 36$) were randomly assigned to two groups, VT recalibration ($N = 18$) and TV recalibration ($N = 18$).

### Stimuli and Apparatus

We used a the same setup that was described in Samad and Shams (2016). It comprised three components: a tactor array that will be described further below, a ceiling mounted projector that was redirected downward onto subjects' forearms via a 45° angled mirror, and the experimental computer that was running Matlab with PsychToolBox for stimulus presentation.

The tactor array was composed of a soft foam material measuring 30.5 by 10 $cm^2$, in which five tactors (Pico Vibe 9 $mm$ vibration motors – 25 $mm$ type; model number 307-103; Precision Microdrives, London, UK) were embedded, spaced apart 41.1 $mm$, a distance that subtended 12° of visual angle. Thus the five locations were -24°, -12°, 0°, 12°, 24° with respect to fixation. The tactors were driven by a custom-build controller circuit that used an Arduino (Arduino, Salerno, Italy) to interface with Matlab. The foam block was itself mounted on a piece of acrylic of the same dimensions that was fixed to the tabletop with

the use of a hinge joint that allowed the block to be pressed onto subjects' forearms. This was aided by a 750 $mL$ opaque bottle filled to a weight of $\sim 2$ $kg$, that was used as a ensure a complete contact of the tactors with the forearm. The visual stimulus was a white disk of light subtending 1.5 degrees, presented by the projector at a location that was 30 degrees below fixation, at one of five points coinciding with the positions of the tactors. Care was taken to ensure that each participant placed their forearm into the setup such that half of the forearm lengthwise was under the foam block, and would thus feel the vibrotactile stimuli, and the other half would be exposed to the projector's screen and would thus have the visual stimuli displayed directly upon it. This enabled the bisensory stimuli to be as proximate to one another while allowing for both to be presented directly on the surface of the arm (see Figure 1A).

Additionally, Gaussian white noise at $\sim 70$ $dB$ was used to mask the audibility of the tactors by being played on headphones worn by the subjects simultaneously with stimulus presentation. The volume was determined in pilot experiments such that the location of the tactile stimuli could not be determined on the basis of the tactor noise alone. Participants had their head position fixed by means of a chin-rest that was placed 195 mm away from the tactor array. Participants were allowed to adjust the height of the chair and/or the chinrest to achieve a comfortable posture.

## Procedure

The experiment consisted of three blocks: pre-test, exposure, and post-test. The total duration of the experiment was 2 hours, and all three blocks ran consecutively in the

same session. During pre- and post-test blocks, subjects localized visual and tactile stimuli delivered to their arm in both unisensory and bisensory conditions that were randomly interleaved. The post-test block contained some top-up exposure trials interleaved. During the exposure block, subjects were exposed to visual-tactile stimulus pairs that were always spatially incongruent and with a constant disparity between them ($\pm$ 12 degrees) in the hopes of inducing an aftereffect. In order to familiarize subjects with the task they were to perform, we included 15 trials of practice before the pre-test and exposure blocks.

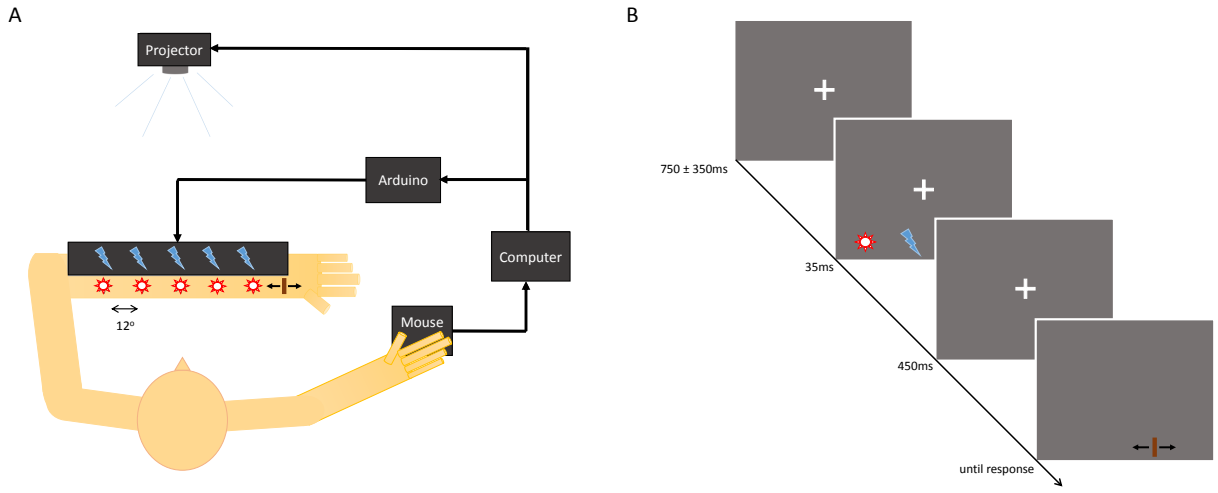


Figure 4.1: **A. Pre- and Post-test Experimental Setup.** Diagram depicting the experimental setup used to present the stimuli, including a tactor array that is driven by a microcontroller (Arduino) for tactile stimulus delivery, and a projector for visual stimulus delivery, and a computer for control of both. **B. Trial Design.** This shows the sequence for a given trial: A fixation cross is presented on the screen for a variable interstimulus interval, after which stimuli are presented for 35 $ms$. After a 450 $ms$ interval, the fixation cross disappears and a cursor appears for subjects to make their responses with.

The pre-test block consisted of 420 trials and took about 35 minutes to complete. Every possible pairing of visual-tactile positions was repeated 12 times in a pseudorandomized order (5 x 5 = 25 bisensory trial types), and we also interleaved 12 repetitions of each of the unisensory stimulus positions (5 + 5 = 10 unisensory trial types). Subjects were to report

the location of the stimuli presented, using a mouse cursor that spanned the same space where the visual stimuli were presented. The color of the cursor indicated to the subject which stimulus to respond to, blue for visual stimuli and red for tactile stimuli. The order of appearance of these two cursors was counterbalanced across participants.

Each trial started with the presentation of a fixation cross 30 degrees above the middle stimulus position, and was followed by the stimulus that was on the screen for 35 $ms$. The fixation cross was taken off the screen and the cursor appeared at a random horizontal location spanning the stimulus space to eliminate any biasing effects of a consistent starting location 450 $ms$ after the stimulus offsets. Subjects moved the cursor using a Bluetooth wireless mouse and were instructed to "move the cursor as quickly and accurately as possible to the position where you saw/felt the stimulus and click the left mouse button".

The exposure block consisted of 40 trials and took approximately 10 minutes to complete. On every trial, a train of 20 successive spatially incongruent stimulus pairs were presented to subjects at the same location. A uniform distribution was used to select a random pair between the 5th and the 15th pairs that would be presented with the visual stimulus 100% brighter than on other pairs. Subjects were instructed to report having seen the brighter visual stimulus by clicking the left mouse button. Upon a successful detection, the stimulus pair changed position whilst maintaining the same spatial disparity between the visual and the tactile stimuli, and the same train of stimuli was presented, with a newly selected random pair to be presented with the brighter visual stimulus, until the next successful detection. Failures to detect the brighter visual stimuli caused the train of stimulus pairs to repeat until a brighter stimulus was detected.

Finally, subjects performed 420 trials of spatial localization with 90 top-up trials in-

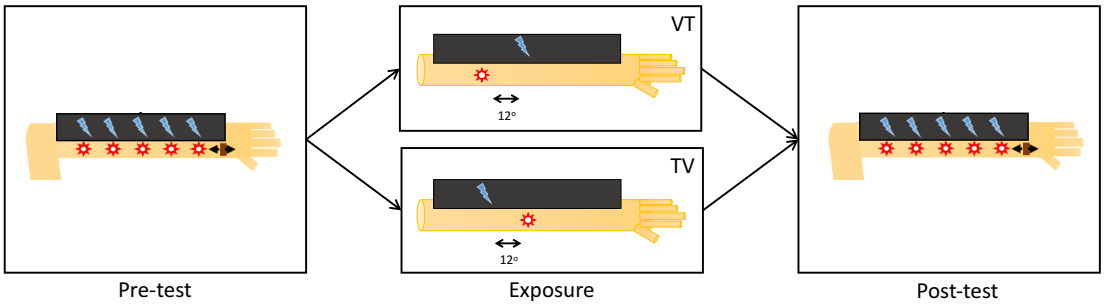


Figure 4.2: During pre- and post-test blocks, subjects performed a visuotactile localization task. During the exposure block, subjects had to passively attend to visuotactile pairs that were spatially discrepant such that the visual stimulus was either 12 degrees displaced towards the elbow (VT-group) or towards the wrist (TV-group) with respect to the tactile stimulus.

terleaved during the post-exposure block. The localization trials were identical to those performed in the pre-exposure block and the top-up trials were identical to those performed in the exposure block. The top-up trials were performed after every 40 localization trials had been completed. The post-test block took a total of 45 minutes to complete.

On 10% of all trials, the fixation cross changed color to red simultaneously with stimulus presentation and subjects were asked to report when this change occurred by clicking the right mouse button, which advanced them to the next trial. This was done to ensure that subjects were fixating during stimulus presentation.

## 4.4 Results

We first analyzed the effects of exposure on unimodal tactile localizations, by computing the difference between the tactile localization on tactile-only trials in the pre-exposure and

the post-exposure blocks. These differences were coded so that they are positive when the tactile localization is shifted towards the elbow, which is where the visual stimulus would have been presented for the VT group. Thus, we expect a positive shift for the VT group and a negative shift for the TV group. These results are shown in Figure 3.

We first analyzed the difference between tactile localizations on tactile-only trials from pre-test to post-test using an ANOVA with repeated measures with between-subjects factor "group" (2 levels: VT-group and TV-group) and within-subjects factor "position" (5 levels: -24, -12, 0, 12, 24). This analysis revealed a significant effect of "group", $F(1, 34) = 7.41, p = .01, BF = 4.24$, and a trend for "position", $F(4, 136) = 2.36, p = .057, BF = .64$. A Bayes Factor between 0.33 and 3 indicates that the evidence is inconclusive as to whether the null hypothesis or the alternative hypothesis is better supported. Therefore, in the absence of a strong effect of stimulus position, we computed a one-tailed independent samples t-test on the change in unimodal tactile localization from pre-test to post-test averaged across positions, $t(34) = 2.72, p = .005,$ Cohen's $d = 0.91$ (see Figure 3). This analysis therefore shows that there was a statistically significant effect on the shift in the unisensory tactile localizations from pre-test to post-test, such that subjects that were exposed to disparities in opposite directions exhibited shifts in correspondingly opposite directions.

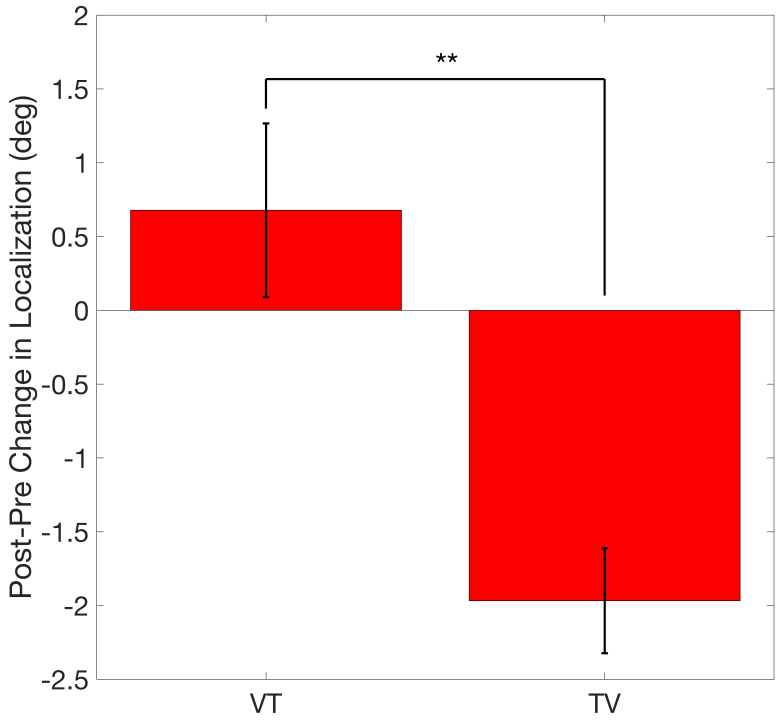


Figure 4.3: **Unisensory Tactile Recalibration Results:** Average change in localization of unisensory tactile stimuli from pre-test to post-test, collapsed across stimulus position, for both groups of subjects. A positive shift indicates a shift towards the elbow, which is in the direction where the visual stimulus was presented with respect to the tactile stimulus during the exposure phase for the VT group. ** indicates statistical significance at $p < 0.01$.

## 4.5 Discussion

Results demonstrate that subjects who were briefly exposed to synchronous but spatially incongruent visual-tactile pairs of stimuli exhibited a subsequent bias in their localizations of tactile stimuli presented in isolation, which corresponded to the direction of the spatial incongruence that they were exposed to. This phenomenon can be described as a visual-tactile ventriloqusit aftereffect and closely parallels the audiovisual ventriloquism aftereffect (Alais and Burr, 2004; Wozny and Shams, 2011b; Lewald, 2002; Recanzone, 1998). This demonstrates the generality of the rules of integration and plasticity throughout the nervous

system, and that the rapid recalibration of sensory maps to each other is not restricted to exteroceptive modalities but is also an actively utilized process in the mapping between the somatotopic and visual representational spaces.

It is notable when comparing the magnitude of this effect with what has been reported in the audiovisual modalities that the latter is generally a larger effect (Wozny and Shams, 2011b). We believe that several methodological differences may be behind the weaker effect observed here. First, the frequency of breaks that we gave our subjects differed from that of Wozny and Shams (2011b) due to the need to allow our participants to rest their arms during the two-hour long experimental sessions. The increased frequency of these breaks may have diluted the effect slightly by providing subjects with episodes without any visuotactile spatial incongruence.

In conclusion, we have demonstrated that visual-somatotopic spatial recalibration occurs after very brief exposure to spatially discrepant stimuli and therefore that the somatotopic space is more malleable than was previously thought.

# Chapter 5

# The Bayesian Body Hypothesis

## 5.1 Summary of Experimental Results

In the preceding chapters, I have described three experimental investigations into body-related representations and posited computational frameworks to account for them. The general guiding principle throughout all of this has always been to arrive at a clearer understanding of how the brain solves the problem of identifying which object is its own body, and which objects are not.

In chapter 2, I described an experiment that revealed several important facts about the rubber hand illusion. First, I showed that it is possible to induce the illusion without the requirement of tactile stroking. This was a surprising finding that went against the conventional wisdom in the field and shed new light on the nevertheless well-studied phenomenon. More importantly, however, was the fact that this novel effect was predicted by a form of the Bayesian causal inference model (Kording et al., 2007; Beierholm et al., 2009b; Wozny et al., 2010) that was extended to operate over both space and time in order to simulate the effect

of temporally synchronized touches. The space upon which this model was constructed is that which arises from the integration of vision with proprioception, a space most adequately described by the notion of peripersonal space. The temporal dimension is characterized by the timing of the felt and seen stroking of both real and rubber hands.

In chapter 3, I described an experiment that was designed to investigate the space that lines the surface of the body, and as such can be considered an orthogonal space to that which I concerned myself with in chapter 2. The experiment was thus concerned with the way in which somatotopically encoded visual and tactile representations interact, and specifically whether the same principles of statistical inference could be successfully applied. Indeed, we found that a version of the model that was designed to operate over this somatotopic space was capable of adequately accounting for our collected data.

Finally, in chapter 4 I described an experiment that further investigated the malleability of the somatotopic space using an adaptation paradigm aimed at the recalibration of the visual-tactile mapping. This was achieved by exposing subjects to spatially discrepant but temporally synchronous visual-tactile stimulus pairs for about 10 minutes and assessing the resultant change in their tactile localization bias. Therefore, this experiment relied on the same somatotopic space model that was developed for the experiment reported in chapter 3.

Thus, across the three preceding chapters, I have made use of three variants of the Bayesian causal inference model, which operate within the two spatial fields of relevance to body representations, namely the peripersonal space and the somatotopic space. This current chapter concerns itself with the way in which to reconcile these two spaces, which under some considerations can be seen to be orthogonal to one another, and to construct a generalized causal inference model that performs inferences across both spaces.

## 5.2 Body Representations: Fixed or Dynamic?

The results described above can be synthesized into the following general statement regarding the perceptual system's representation of the body: it is not immutable and fixed but is rather a dynamic estimate produced by the constant sampling of the bodily senses and their calibration with the external senses. This idea challenges our natural intuitions, which posit that the perception of our own bodies ought to be hardwired and absolute, immune to the trickeries that the rest of the perceptual system is so demonstrably susceptible to. Illustrative of this intuitive difficulty are accounts such as that espoused by Glenn Carruthers (2008), which posits a distinction between online and offline body representations, with the former being comprised of the sorts of things that are updated with sensory information about how the body is moving, and the latter being a relatively fixed and hardwired representation that is immune to the effects of sensory stimulation. Glenn is very careful to distinguish his account from the older *body image/body schema* dichotomy (Head and Holmes, 1911), claiming that the offline representations are unlike the body schema in that they do not contain any information about the current position of the body, but rather underlie the conscious sense of embodiment. In support for this idea is the fact that patients who have had limbs amputated continue to feel the phantom presence of the missing limb, a condition known as phantom limb syndrome, as though an offline representation has failed to be adequately updated (Ramachandran et al., 1995).

Thus, it seems that researchers have for a while now been cognizant of the plasticity of some body representations. But indeed, even in patients with phantom limb syndrome, merely using a mirror to reflect the image of the intact arm so that it appears to occupy

the position that the amputated arm would have is enough to produce a stark improvement in the diagnosis (Ramachandran et al., 1995). So, despite the influence of an offline semi-permanent representation of the body, it seems that even they are not immune to the effects of learning. However, it is important to note that there are some constraints on this rapid re-wiring. While one of the early papers on the rubber hand illusion seemed to show evidence for the ability to embody any object regardless of their identification as body parts (Armel and Ramachandran, 2003), accumulating evidence challenges this initial finding and suggests instead a requirement that the object to be embodied be a semantically congruent body part (Tsakiris and Haggard, 2005; Tsakiris, 2010; Haans et al., 2008).

The rubber hand illusion is perhaps the strongest and most salient phenomenological effect demonstrating the malleability of the body representation. It provides an illustrative contrast with the phenomenology of the phantom limb pain syndrome, both of which are nevertheless often discussed under the same light. The contrast can be highlighted by recognition of the fact that although they are both demonstrations of anomalous body representation, the error in the case of the RHI is substitutive, whereas in the case of phantom limb syndrome it is residual. In other words, the arm that is owned in the RHI is substituted for the veridical arm, whereas the limb that is owned in the case of the phantom limb syndrome is the shadow of the previously attached limb. In both cases, we can see the influence of a top-down constraint on the process of body representation, in that the identity of the body part is used to constrain the incoming bottom-up sensations – in the RHI to enable the substitution, and in the phantom limb to override the sensory evidence for the lacking limb. Thus, it is clear that there is an interplay between the two processes and that this may be at the heart of the vigorous theorizing by researchers regarding how many body

representations there are and what is their nature. So, on the one hand it is clear that the body representation is modifiable and can be tricked by sensory illusion paradigms, but at the same time there is a strong influence of prior knowledge that serves to bound the space of possible body representations.

That perception of the external world is an inherently probabilistic process is a fairly uncontroversial claim, and recent advances in computer vision and vision science have in fact relied on formulating and implementing highly sophisticated statistical inference methods to compute what is most likely from amongst the full space of possible scenes generating the current sensations. When the conversation turns to the perception of our own bodies, it has not yet been unequivocally demonstrated that this too relies on an inference process. The work described in this dissertation has shown that the Bayesian model can account for these phenomena well, and this can be taken as yet another sign that there is nothing privileged about these bodily representations. Rather, they too are the nervous system's best guess as to the latent structure that produces incoming sensations. Insomuch as a sensation is a sensation, it will have a characteristic amount of noise determining its reliability, and the nervous system is highly adept at utilizing this in the process of making its estimates. Whether the sensation arrives from mechanoreceptors in the skin, or whether it arrives from photoreceptors in the retina, in both cases it carries information that can be used to reconstruct its emitting object in the reality beyond the nervous system's epistemic horizon. What we are in pursuit of here is a way to formulate the inference process so that it combines the top-down constraints we have been discussing with the bottom-up reliability-based cue combination under the same overarching statistical framework.

A caveat, which will be discussed in further detail in Section 5.3 below, is worth men-

tioning: a crucial difference between the signals arriving from mechanoreceptors and those arriving from photoreceptors regards the location of origin of the event that triggered them. In the case of a photoreceptor, the originating body may be located anywhere, at any distance away from the nervous system in the external world. A mechanoreceptor, on the other hand, provides information regarding events that occur on the body itself. This point will come to represent a crucial component in the model of body ownership being proposed herein as it will be the differentiating element that enables the system to invoke specialized systems that match body templates to specific objects.

## 5.3 The Bayesian Body Hypothesis

At this junction, I would like to argue for a synthesis from the results and conclusions scattered throughout the various experiments documented within this dissertation. In what follows, I will attempt to construct a synthetic account of body ownership based on the logic of Bayesian causal inference and will consider two alternative approaches for doing so. In the interest of clarity, I will hereafter refer to this effort as the Bayesian body hypothesis.

Let us begin by re-examining the model of the rubber hand illusion that was provided in Chapter 2. In that model, we made several simplifying assumptions, which we will now spell out more carefully in an extended form of that model. In particular, we had assumed that a previous calculation would have already recognized the foreign object as a body part, namely an arm, that matches the internal template for the body part. This assumption was necessary so that the azimuthal disparity between the visual and proprioceptive signals could be directly assessed. If the visual object were not categorized as a body part to

be embodied, there would be no sense in the calculation of disparity between it and the proprioceptive spatial location. We also assumed that the touch applied to the rubber hand was congruent with the touch applied to the real hand in somatotopic coordinates. This too was a necessary simplifying assumption as the model was restricted to only utilizing the temporal components of the touches, and could only do so if it was established that their spatial components were congruent. Therefore, to allow for an adequate formalization of the model of the rubber hand illusion, we chose to disregard these potential sources of incongruence because in the experiments that were conducted to test the model, care was taken to experimentally preclude such incongruencies.

Figure 5.1 shows the extended graphical model that incorporates these additional computations that were previously disregarded. Thus, we have supplemented the original graphical model from Figure 2.1 with three additional signals. In the original model, we had a signal for $V_{azim}$, encoding for spatial location along azimuth, and have added to that here a signal for $V_{som}$, encoding for spatial location along somatotopic coordinates, as well as temporal information regarding the timing of the touches. We had previously modeled the tactile signal, $T_{som}$, as conveying only temporal information regarding the timing of the touches, but have here expanded it to also encode the spatial location along somatotopic coordinates. Note that $V_{azim}$ and $P_{azim}$ can also be expanded to contain temporal information if the hands are moving, as would be possible in a virtual hand illusion setup. Finally, we also have a signal regarding the hand similarity metric that was discussed above. Thus, this framework models ownership as arising when an inference of common cause is made for all these five signals, four of which are spatiotemporal vectors and one of which is categorical.

However, if we begin to delve a little deeper into the inner workings of the model, as

would be necessary for a full formalization of the process shown in the graphical model in Figure 5.1, we will observe some potential difficulties, which my attempt to address will lead to the second possible formulation of the Bayesian body hypothesis, as will be discussed at length below. First, we will observe that the four spatiotemporal signals are actually represented across two spaces, the peripersonal and the somatotopic. This would seem to imply that somatotopic coordinates can be extracted independently from the peripersonal space. In particular, this arrangement would require that the visual system be able to observe a hand-like object be touched, and then to convert that into its corresponding representation along somatotopic space, so that the comparison with the actual felt touch can proceed in determining the congruence of these signals, or lack thereof. Is there any evidence that the visual system computes these mappings for every hand it sees getting touched anywhere in its field of view? Or would we find it more parsimonious to postulate this special retinotopic-to-somatotopic transformation only for regions of visual space that are plausible candidates for where *my* hand may be located? This is certainly not an uncontroversial question, although my scales tip slightly more towards to the latter notion, the full implications of which will be developed in the second possible formulation of the Bayesian body hypothesis below.

In fact, there is substantial evidence for the existence of a population of neurons (see Section 1.5) capable of rapidly and dynamically remapping their visual receptive fields as the arm moves around in space. In modern parlance, these neurons are the foundation of the system called peripersonal space, whose existence implies that the parts of the visual field that correspond to the somatotopic coordinates depend on this remapping that is always occurring, and whose role is to center itself upon the location of the body part (Graziano et al., 2000). This seems to suggest that the two processes of integration here (the visuo-

proprioceptive and visuo-tactile) are actually two stages rather than one, and that the visuo-tactile depends upon the results of the visuo-proprioceptive. Therefore, as an alternative to the single process causal inference model described above, a reformulation guided by this dual-process approach will be advanced in what follows.

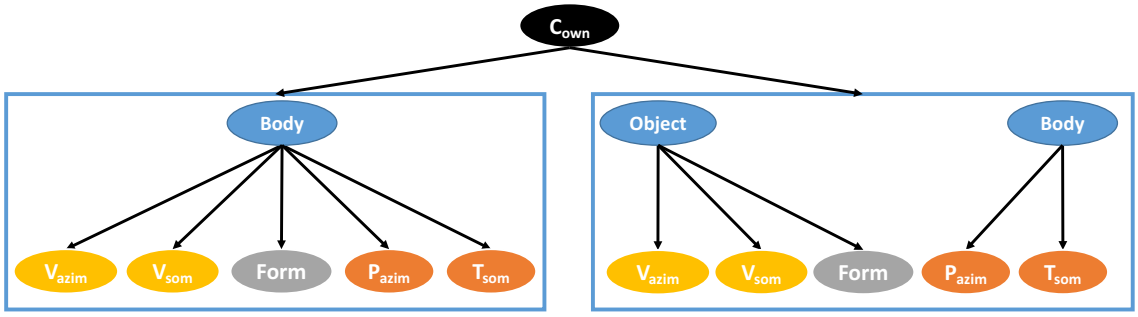


Figure 5.1: **Extended Rubber Hand Illusion Model** Extended version of the graphical model for the Bayesian causal inference model of the rubber hand illusion that was utilized in Chapter 2. Yellow ovals indicate visual signals providing information regarding spatial location of the arm in azimuth, $V_{azim}$, and spatial location of the stroking in somatotopic coordinates, $V_{som}$. Orange circles indicate somatosensory signals providing information regarding the spatial location of the arm in azimuth, $P_{azim}$, and spatial location of the stroking in somatotopic coordinates, $T_{som}$. The gray circle indicates visual form information of the object to be embodied, and specifically provides information regarding its similarity with the template for the body part of relevance. The blue circles denote the possible sources of the signals, namely a body part or a foreign object. And finally, the black circle denotes a binary variable relating to the number of causes that generated the signals, which if inferred to be equal to 1, indicates a feeling of ownership over the visual object.

Before we delve into that detailed proposal, however, and in an attempt to draw inspiration from a highly tangential source, let us remind ourselves of the quote that we began this dissertation with, and which was first stated by the most noble of those who have pondered the mysteries of the mind:

*"The foot feels the foot when it feels the ground."* — Buddha

An innocently sounding aphorism that at first glance affords little insight and conveys something seemingly trivial, reveals itself to contain tremendous profundity upon closer examination. Much of the spirit of the argument that I will make in what follows will be inspired by that trite saying of the Buddha's. But, before I move on, let us consider a more recent statement that conveys much of the same spirit:

> *"When our hand touches an object in our environment, we know that the object and our hand are at the same location"* — Smeets et al. (2006)

The guiding principle that underlies the present modeling endeavor posits that body ownership is the result of a probabilistic process that operates on the interplay between exteroceptive and interoceptive sensations, namely when visual and/or auditory representations are to be integrated with somatosensory and proprioceptive representations. Causal inference, thus, operates on the sensations in order to attribute them to their proper causes both in the body and in the environment. The visual object that the nervous system decides to integrate with the somatosensory signals can only therefore be one of two objects: either it is a part of my own body, or it is the external object that touched me. Thus, the lesson we can learn from the literature we have surveyed and from the insightful quotes above is that rather than attempting to force two processes into one and the same causal inference model, as the rubber hand illusion model from Chapter 2 does, we can expand the scope of the framework by analyzing these two inferences separately. In other words, it may well prove advantageous to treat the situation as involving two processes of causal inference, seeing as the event of a touch involves two components coming into contact.

In order to more thoroughly describe this hypothesis, let us take recourse to an example:

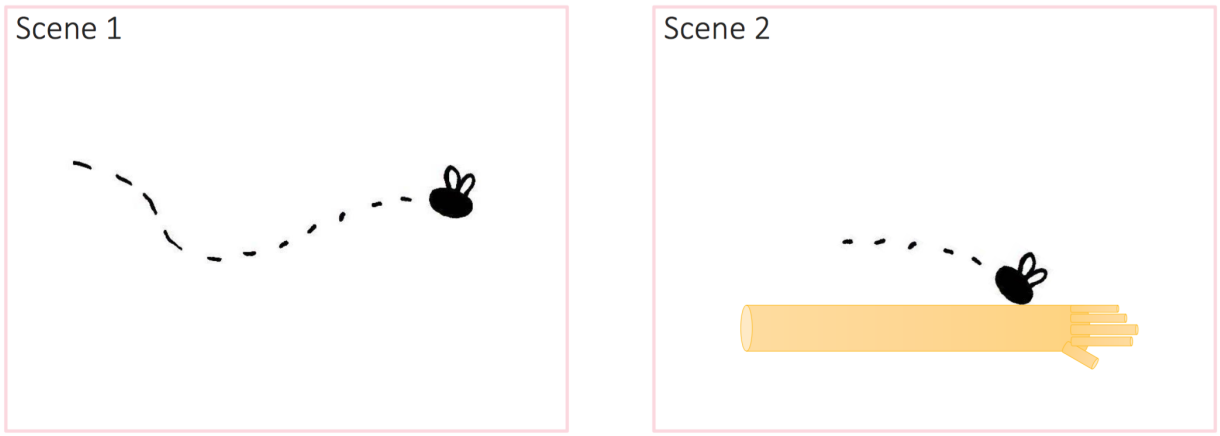


Figure 5.2: **Two Illustrative Vignettes:** Scene 1) a fly passes over the visual field emitting visual and auditory signals. Scene 2) the fly lands on the surface of the body causing a somatosensory signal to be transduced in addition to the audiovisual ones.

a fruit fly passes across my visual field and lands on my forearm, triggering visual, auditory, and tactile sensations (see Figure 5.2). Let us consider these two episodes in turn: first, it flies across my visual field giving off photons and sound waves that my nervous system detects and begins to process. At this point, audiovisual integration processes are presumably computing the source most likely to have generated these exteroceptive sensations and thereafter attributing them to it. As such, they provide information relating to events occurring outside the body. Therefore, it is always possible to move the source of a given signal progressively farther away while commensurately increasing its signal intensity and the nervous system would be presented with presumably identical sensations (putting aside for simplicity ancillary differences such manipulations are sure to introduce). Of course, the redundancy in information arriving from two modalities will serve to constrain this space of possible sources. Specifically, there are other structural constraints that this dual presentation to the nervous system helps to provide, such as spatial and temporal register, as well as higher order correspondences such as the semantic dimensions of the signals.

Let us contrast this purely audiovisual episode with the subsequent one that includes the tactile sensation of the fly having landed on my forearm. In this case, the mere inclusion of a somatosensory signal provides much stronger constraints on the space of possible sources for the audiovisual signal, assuming of course that it was inferred to have had a common cause with the tactile one. This would have to imply that the audiovisual object has arrived at an interface with the body for it to have triggered the tactile signal. Thus, these integrated percepts occupy a privileged status among the perceptual world, in that they give information regarding this interface between our bodies and the environment – our skin. So, the nervous system's task is to parse the event of interest into two components: 1) the external event that triggered the somatosensory signal, and 2) the body part that was the site of contact with this external event. Therefore, with the exception of self-touches, **every tactile signal is an indication of physical contact between a foreign body and one's own.**

## 5.4 Modeling Considerations

Let us now return our attention to the formal architecture that would underlie a model of the kind of expanded scope that is at present being argued for. A crucial question arises immediately regarding the ordering of the two operations under consideration: does the brain infer the location of the arm first, and then proceed to infer the location of the fly on the surface of the arm, or vice versa? If we are to reconcile the two inferences that operate across the peripersonal and the somatotopic spaces, then the answer to this question regarding the order of inferences would seem to be a highly important one to address. I can think of three possible schemes: 1) the two proceed simultaneously and mutually influence each other, 2)

the referral – or transformation – of proprioception into external space coordinates calibrates the retinotopic and somatotopic spaces to one another, or 3) the spatial coregistration of a visuo-tactile stimulus helps to refer proprioception into external space coordinates.

The most intuitive route to adopt in approaching this problem is that which takes its cue directly from the literature on peripersonal space that was reviewed in Section 1.5, and therefore posits that the visuo-proprioceptive integration step is computed first, before any referral of a visual touch to the resultant hand-centered coordinates may proceed. That said, however, it remains possible – indeed, likely – that the subsequent stages may still exert the kind of mutually constraining effects discussed under option 1 above. In other words, if the visuo-tactile inference does not yield an integrated percept, this may in turn provide evidence against the remapping of proprioceptive space to the location of the visual stimuli. The key question to consider, however, is whether the visuo-tactile computation has any chance of yielding an integrated percept in the absence of the peripersonal space, or rather if the latter is a critical prerequisite that enables the retinotopic-to-somatotopic mapping, as discussed above. Nevertheless, the separation of the model into this dual-process framework enables the consideration of the two inferences spaces independently of one another, and makes some predictions that allow for the arbitration between it and the single-process model described above, which will be discussed in greater detail below.

Therefore, in order to more fully describe this dual process form, let us first consider each of the inference spaces separately. Looking at the diagram shown in Figure 5.3, we can begin to parse the overarching framework into its two constituent inferences: the inference of the causal structure generating the external object, shown in the figure as the portion of the diagram above the blue dotted line – and the inference of the causal structure generating

the body part that has been touched, shown as the portion below the dotted line. Let us begin with the former.

The yellow and orange ovals represent the visual and tactile sensations, respectively. The blue ovals producing arrows directed towards the sensations depict the causes of those sensations, which in this case, are in the external world. Thus, these two "leaves" occupying the upper portion of the diagram in Figure 5.3 correspond to the scenarios where one or two events external to the body gave rise to the visual and tactile sensations. And the black oval labeled $C_{\text{ext}}$ is the binary variable determining which is the case, and whose latent state the model attempts to infer. This inference is guided by the congruence of the visual and tactile representations with each other, evaluated on the basis of proximity in somatotopic space and time. As mentioned above, this requires a coordinate system transformation between the retinotopic visual field and the somatotopic field for this comparison to be made. And needless to say, the requirement to transform between spaces as just mentioned appears to necessitate that visuo-proprioceptive integration have commenced already, in order to extract this mapping function from its results. Let us now turn to that step.

Looking back at Figure 5.3, again we have sensations represented by orange and yellow ovals and sources by blue ovals. Notably, however, the sensations are now the visual and proprioceptive sensations coming from the body part itself. And, the sources are now located within the body's borders, as they are drawn in the lower portion of the diagram, which is on the body's side of the skin. Thus, the inference in this case is over the hidden state of another binary variable, the black oval labeled $C_{\text{body}}$, which determines whether the visual and proprioceptive sensations share a common cause or separate causes. If the former is inferred to be most likely, then the visual object must be the image that that body

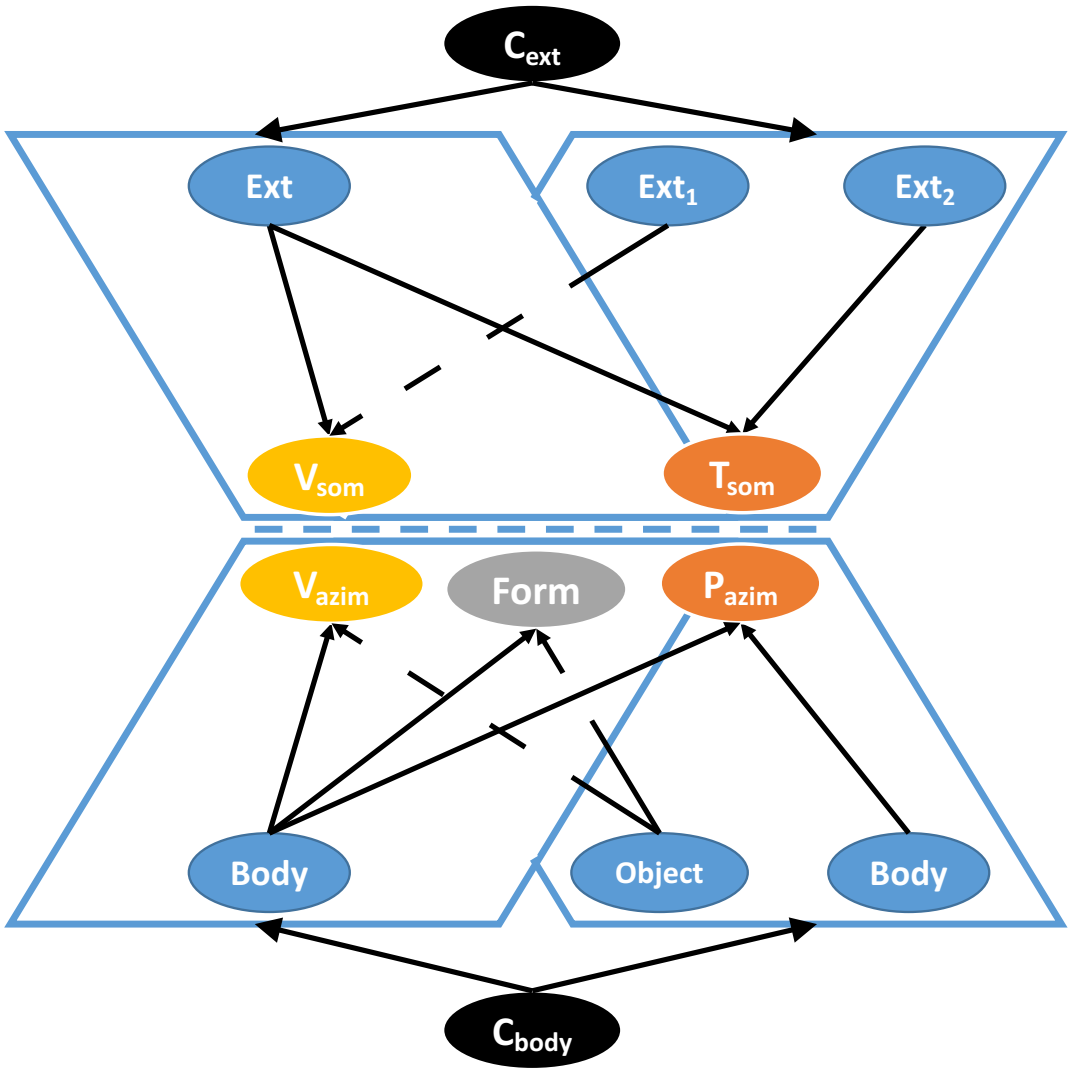


Figure 5.3: **The Bayesian Body Hypothesis:** Proposed graphical model for the generalization of Bayesian causal inference to inference regarding body ownership. The dashed line indicates the interface across which a tactile stimulus ($T$) signals contact, i.e. the skin. The pair of visual stimuli that sit on either side of this dashed line, $V_{som}$ and $V_{azim}$ indicate the object and the body part that have been brought into contact, and encoded in somatotopic and peripersonal spaces, respectively. $C_{ext}$ is a binary variable representing the causal structure for the visual and tactile representations from the external object and $C_{\text{body}}$ is a binary variable representing the causal structure for the visual and proprioceptive representations from the body part.

part therefore casts upon the retina. Importantly, this step relies not just on spatial and temporal proximity, but an additional requirement, namely that the visual representation resemble the body part, a judgment conveyed by the variable labeled *Form*. Nevertheless, we should hasten to add that the spatial comparison that is involved here also involves a coordinate transform, this time operating in between the retinotopic and the proprioceptive space. It is illuminating that we do not have a comparable -topy word for the topology of the proprioceptive space; it highlights the paucity of data regarding its primary sensory characteristics. This makes our computational approach all the more challenging, since we do not know how to properly formulate the encoding of sensations in this space, let alone determine the mapping function between it and the visual space.

If the process of visuo-proprioceptive integration step was under-constrained, or difficult to solve for any reason, could the visuo-tactile integration step inform it? Specifically, if the visual and tactile representations of signals from the external world were found to have been more likely to share a common cause, this would be an additional source of information aiding the anchoring of proprioception within that same exteroceptive space. Knowing that a touch signifies contact between the body and a foreign object enables this "knowledge-transfer", if you will, between domains. How are we to characterize the mechanism by which this information is utilized? We can choose to implement it as a additional signal to be combined with the visual and proprioceptive signals coming from the body itself. Perhaps the most direct way to do this would be to use the information gained by the other step to update the spatial prior in the current step.

Finally, to turn to the notion of hand similarity and its effect on the visuo-proprioceptive integration process, we can speculate on possible ways to instantiate probabilistic inference

machinery to achieve the desired effect. This sort of calculation appears to be best captured by the sorts of computer vision systems that have become very popular in recent years. We can plausibly imagine that what is required is a classifier that is trained to recognize the body part of relevance – say, a hand – and can also output a confidence metric on its categorization. This enables the use of a continuous space upon which to perform the inference of casual structure yielding the proprioceptive and visual signals. That is, they need not only be congruent in their spatiotemporal profiles, but also need to be congruent on this dimension of *handness*, quantified as the confidence in the "hand" category of a classifier trained on a large enough database of hand models.

The question may arise as to the added value of this model framework as compared to the model that was used to account for the rubber hand illusion in Chapter 2. We can now offer a more direct answer, as follows: the causal inference that identifies the visually presented hand as one's own relies on two integrative steps, and there is evidence to suggest that they can operate independently of one another, and thus that it does not suffice to assume their mutual determination. In particular, the visuo-proprioceptive integration yields the peripersonal space – a prioritized zone encoded in body-part centered coordinates, and visuo-tactile integration happens on the basis of this space that is defined with respect to the body part's location. In the section that follows, we will discuss recent studies that seem to show evidence for the ability of these two steps to be dissociated, and therefore the utility of this modeling framework in accounting for this.

In this way, this model provides a quantitative innovation to the neurocognitive model espoused by Tsakiris (2010) (see Figure 1.2). In that work, the proposed model goes through several sequential stages in order to determine first the incorporeability of the ob-

ject (the aforementioned *handness* similarity component of our visuo-proprioceptive integration process), then the postural congruency (the spatiotemporal component of the visuo-proprioceptive integration process), and finally the referral of the felt touch to the new hand-centered coordinates (the visuo-tactile integration process). This framework, thus, offers a normative probability theoretic formulation for that influential neurocognitive model, and moreover enables a much stronger falsifiability criterion in its ability to make quantitative predictions that can be tested experimentally, a topic we will turn to next.

## 5.5 Testable Predictions

A model is only as useful as its testable predictions, and in fact becomes utterly unscientific if it does not have that all-important characteristic of being falsifiable.

One prediction that seems to come out of this framework concerns the independence of the dual inferences. More specifically, a situation may be possible where an object's visual and tactile signals are integrated into a coherent object on the surface of the body, but yet where the complementary visual object is not integrated with proprioception. This could perhaps be achieved in a virtual setting where a non-body object is touched by another such that the subject feels the touch at the location of the non-body object, yet does not feel ownership nor does proprioception recalibrate. We see hints of this in the three-arm illusion (Guterstam et al., 2011), where subjects are able to see both real and rubber hands while the illusion is being induced (see Figure 5.3). Results showed that subjects did indeed feel ownership for all three arms, brought about by the synchronous visuo-tactile stroking, but also revealed a curious dissociation in the pattern of responding to the questionnaire

items. They reported less ownership than what is typically seen in the conventional rubber hand illusion paradigm, but greater levels of referral of touch to the rubber hand. Using this framework, we would explain the phenomenon as a case of inference of common cause for the external object – the spatiotemporal profile of the paintbrush – while simultaneously computing a much weaker inference of common cause for the body-object, since the real hand is in plain sight and provides a much better match for integration with proprioception.

**Figure 1. Illusion set-up.**

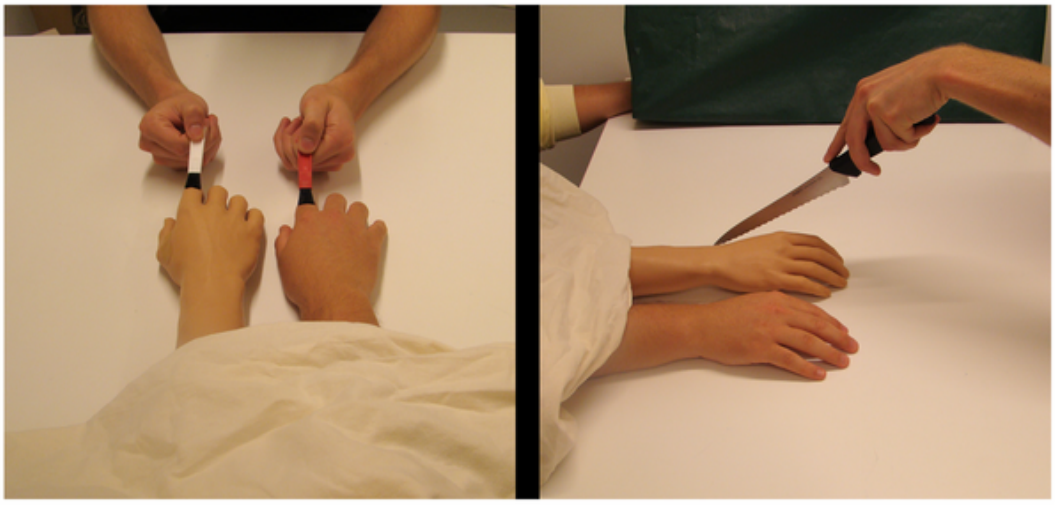



Figure 5.4: **Supernumerary Hand Illusion:** Reproduced from (Guterstam et al., 2011) depicting the experimental setup that was used to induce the supernumerary hand illusion. Note that both the real and the rubber hands are visible to the subject while the stroking is being applied and throughout the experiment.

Conversely, a situation should also be possible where visuo-proprioceptive integration

happens, thus shifting the perceived location of the hand, and yet visuo-tactile stimuli at the hand's new location are not deemed to have had a common cause. The asynchronous stroking condition, and the difficulty associated with fully eliminating the illusion by it, may be an indication of something like this taking place. In fact, the accumulation of recent evidence pointing to the failure of proprioceptive drift and ownership reports to be causally related (Rohde et al., 2011; Abdulkarim and Ehrsson, 2016) seems to provide evidence for the independence of the visuo-proprioceptive and the visuo-tactile integration steps. To speculate, we may claim that when subjects' proprioception drifts, they are performing a readout from the visuo-proprioceptive process, but the full-fledged ownership experience requires integration across both steps to occur.

Finally, we have the recent magnetic touch illusion experiment where the experimental setup was much the same as the conventional rubber hand illusion, with the modification that the seen touch did not actually make physical contact with the rubber hand, but rather floated a variable distance away from it (Guterstam et al., 2016). Subjects felt an invisible force connecting the distant paintbrush with the rubber hand, as though a magnet were transmitting the touch across the space, and this sensation correlated strongly with their reports of ownership. Under the current framework, we would say that the visuo-proprioceptive congruence – and hand similarity – provided a strong inference of common cause for the body-part, which consequently remaps the peripersonal space to be centered upon the rubber hand, and that visuo-tactile integration proceeds subsequently in this newly calibrated space. This is indeed consistent with the account the authors provide for their results, in that they also refer to the well known phenomenon of anchoring of tactile receptive fields to the visual location of the arm, and that this can account for the visuo-tactile

integration that occurs in peripersonal space (Graziano et al., 1994, 2000; Graziano and Gross, 1998).

More directly, however, we can attempt to provide a specific experimental proposal whose explicit purpose would be to determine whether the single or dual process approach is better supported by the evidence. In particular, we can employ an experimental design that independently manipulates the visuo-proprioceptive congruence and the visuo-tactile congruence, with a dependent measure that records a subject's localization of both their arms as well as the location of the stroking on the surface of the arm. In this way, we may observe the dissociation between these measures that might indicate integration in one space without the concomitant integration in the other, thus providing evidence in favor of the dual process approach illustrated in Figure 5.3. In contrast, if we observe that the integration/segregation behavior of the subjects is perfectly correlated across the somatotopic and proprioceptive localizations, then this would lend its support for the single process model, as shown in Figure 5.1.

## 5.6 Conclusion

So to recap, the Bayesian body hypothesis states that the sense of body ownership arises from the processes of statistical inference that help to make the best guess as to what objects and events gave rise to the myriad sensations the nervous system receives, only insomuch as the sensations include the somatosensory modality and a non-somatosensory object that matches the template for the relevant body part. In particular, the inference process can be described by an analysis of the two integrative steps that are its component parts, and

which serve to mutually constrain one another. Namely, these two steps are as follows: 1) retinotopic-to-somatotopic transformation that allows for the visuo-tactile integration in the localization of objects touching the body, 2) retinotopic-to-peripersonal space transformation that allows for localization of the body as it is touched by objects. It is clear that there is much interplay between these two processes, in that each transformation serves to constrain the other, since knowing where in the visual field the body part is aids in then knowing where in the visual field various regions of the skin are, and vice versa. But it is also becoming more and more clear that they can be manipulated independently of one another, as seen recently by the demonstrations of illusions of visuo-proprioceptive integration that don't also manipulate ownership or vice versa. Therefore, the model presented in this chapter is a first step towards the generalization of the spatiotemporal Bayesian causal inference model of the rubber hand illusion that was presented in Chapter 2, such that it can more completely describe its component inferences that had previously only been partially described. Future experiments testing its predictions and utilizing this framework can therefore start to provide greater clarity on the computational structure underlying the emergence of the subjective sense of ownership for one's body parts.

# Chapter 6

# Summary and Conclusions

## 6.1 Summary of Main Findings

This dissertation has documented three experimental investigations into the mechanisms of multisensory integration that relate to the perception of our own bodies. In the first of these (Chapter 2), I have demonstrated that the rubber hand illusion falls under the umbrella of multisensory phenomena that can be accounted for by the Bayesian causal inference model, and moreover uncovered the first evidence of the elicitation of the illusion without the use of brushstrokes, an effect that was predicted by the model. In the second investigation (Chapter 3), I demonstrated that the mapping between the visual field and the somatotopic space that maps the surface of the skin is just as susceptible to the illusory phenomenology of spatial ventriloquism as the audiovisual mapping, and moreover that it is also governed by the Bayesian causal inference model. Finally, Chapter 4 documents evidence that mere exposure to discrepant visual-tactile stimuli can induce an aftereffect in this visuo-tactile mapping.

In Chapter 5, I propose a novel theoretical framework for body ownership, the Bayesian body hypothesis, that incorporates elements from the computational modeling approach I used in earlier chapters, as well as some of the cognitive constraints that others had previously proposed (Makin et al., 2008; Tsakiris, 2010), thus providing the first complete theoretical framework for the investigation of body ownership.

## 6.2 Limitations and Considerations

Let us now turn our attention briefly to the merits and limitations of the present body of work. As with any research project, there are many grains of salt with which the conclusions ought to be taken. Seeing as we have now arrived at the concluding chapter of the dissertation, it would seem to be a most appropriate moment to spend a little time in this consideration.

#### Merits and Flaws in Chapter 2

In Chapter 2, I proposed the first computational account of the rubber hand illusion. Since then, that work has already started to generate much interest in the field as researchers investigating body ownership seek to rest their accounts on firmer mathematical foundations. The model that was first proposed in Chapter 2 has the potential of being extended to account for a great variety of facets of body ownership (see Chapter 2 Discussion for more on this). Therefore, the work described therein provides the much needed first step to understanding the mechanisms of perception and ownership over the whole body and paves the way for that line of work. Another merit was the discovery of the touchless illusion and validation

of the model that this aided with.

With regard to limitations, the model was not quantitative and thus serves purely as a proof of principle at this early stage. Of course, the next step is collect sufficient data to be able to conduct model fitting using the spatiotemporal Bayesian causal inference model, a feat that was not yet possible at the time that project was conducted. An additional limitation of that work was that we did not adequately explain the large SCR signal at eye-opening. In particular, why that time point should generate an even larger SCR than the threat time point continues to be mysterious and deserves to be investigated further. In that work, we interpreted the eye-opening signal as a measure of the illusion, especially because the difference between it and our control conditions at that time point was very large. However, the threat had not yet been presented, and therefore the interpretation requires more nuance in terms of a justification for the large conductance signal. At the time, we reinterpreted the signal as a measure of surprise, provoking a physiological arousal that manifested as an SCR, and left it at that. Thus, this is clearly a potential avenue for further elucidation of the phenomenon.

**Merits and Flaws in Chapter 3**

In Chapter 3, I described a project demonstrating a form of ventriloquism that occurs on the surface of the body. As such, it represents the discovery of a novel phenomenon of multisensory perception. In addition, we also showed that the Bayesian was a good fit to the data and thus represents the generalizability of that model to a new paradigm, extending its applicability to multisensory phenomena. In that regard, we also got one step closer to the goal of providing a quantitative Bayesian causal inference model for a body related

phenomenon, although this was a one-dimensional model operating in space only. The goal of quantifying the spatiotemporal model remains to be done.

One noteworthy limitation of that work was the lack of a satisfactory explanation for the off-center tactile prior. We have repeatedly noticed in this paradigm, the intriguing pattern of localization errors that compresses unisensory tactile localizations, and moreover, whose center of gravity is not the center of the space, but is rather a point somewhere between the elbow and the center of the forearm. We attempted to account for this by recourse to mechanisms akin to cortical magnification arising out of the differential distribution of mechanoreceptors in the skin, but further investigations are greatly needed.

**Merits and Flaws in Chapter 4**

Finally, in Chapter 4, I described an extension of the work on visuotactile ventriloquism that showed the existence of a visuotactile ventriloquism aftereffect. This is a highly interesting finding showing that very brief exposure to discrepancy between the visual and tactile stimuli can cause a measurable recalibration between the two spaces, just as a similar phenomenon has been observed for the other forms of ventriloquism.

## 6.3 Suggestions for Future Work

Taken as a whole, this dissertation represents only the very beginnings of the effort towards understanding the computational principles that underlie body representation. In that regard, the contribution of this dissertation has to be interpreted as paving the way for future studies seeking to continue this line of work. Therefore, we would be very well advised

to learn from the lessons contained within these pages, and to inform the future generations of scientists interested in advancing the knowledge of this subject matter.

## The Origins of Unisensory Biases

First, it is suggested that future research tackle the outstanding question regarding the origin of the biases that have been repeatedly observed in localization paradigms akin to the visuotactile ventriloquism paradigm that has been described in Chapters 3 and 4 of this dissertation. The origin of these biases remains an open question especially regarding the idiosyncratic patterns that they often display. In the tactile modality, I have observed biasing of unisensory localization estimates that seem to be centered at a point that is midway between the middle of the forearm and the elbow. We have speculated on the origin of this off-center attractor point as perhaps arising from the distribution of mechanoreceptors across the skin – a line of argument reminiscent of Weber's spatial compression observation (Ross and Murray, 1978). However, a rigorous study of this phenomenon is highly recommended if the question is to be satisfactorily answered.

## A Quantitative Spatiotemporal Bayesian Model

Another suggestion relates to the results reported in Chapter 2 of this dissertation, wherein the rubber hand illusion was shown to be compatible qualitatively with a spatiotemporal form of the Bayesian causal inference model. Recall that a major limitation of that work was our inability to conduct a more quantitative analysis of the suitability of the model, due to the fact that only spatial estimates are collected from our subjects, and

moreover that there is only one such estimate per subject. In order for model fitting to be feasible, we would require first vastly more data points, but in addition, we would require temporal as well as spatial estimates. Whether this would remain a workable requirement within the framework of the rubber hand illusion or not remains to be seen, but it would provide the field a great service if such a line of research was attempted.

## Generalizations of Bayesian Causal Inference

Another line of research that I wish I had had more time to devote to relates to extending the generality of the causal inference model. What is most problematic about the model is its specificity to the case where there are only two sensations presented to the observer, whose inference thus only needs to consider whether they would have had 1 or 2 causes in the environment. The reality of this process, however, is evidently far less constrained, receiving as we do many more than just two sensations at any moment in time, and often perceiving many objects and events in the world that are collectively clustered from our sensations. Nevertheless, the explicit mathematics of the model make it so that any consideration of more than 2 causes renders the computations intractable due to a combinatorial explosion that is necessitated by the model's explicit representation of the entire hypothesis space upon which the inference proceeds. Therefore, some alternative mechanism ought to exist to quickly and efficiently explore this space and perform the inferences as the brain innately appears to do. I had briefly ventured into a variety of non-parametric forms of the causal inference model, which led me to unsupervised clustering algorithms as well as the Expectation-Maximization algorithm and Variational Bayes. But I have seemingly run out of time and will not be able

to continue this pursuit. However, I believe that this will be a breakthrough in the near future that will enable the field to begin to apply the process of causal inference to more ecological stimuli.

In pursuit of this goal, I had also begun conducting a mathematical analysis of the computations of Bayesian causal inference, and noticed an appealing similarity of it with the computations involved in signal detection theory (SDT). The notion of $d'$, which is intended to convey a measure of discriminability between two signals, offers a potential avenue to heuristically shortcutting the explicit representation of the hypothesis space. I feel very strongly that profoundly more powerful models will be devised with the aid of heuristics related to this, all the more so given the mathematical affinity between the two that I have briefly observed, but have not pursued. On a related note, the information theoretic notion of Kullback-Leibler divergence also offers a potential heuristic shortcut to the formal theoretical causal inference process. A rigorous study of the ways in which these and related heuristics can be substituted for the full model would be a very important contribution to this field.

## Coordinate System Transformations

I have also been interested in investigating more deeply the spatial prior that we have always used in our Bayesian models. In particular, I have always wished to explore whether a coherent model can be contrasted with it that includes modality-specific priors. More generally, the question that continues to vex me relates to understanding the characteristic biases we observe in these localization paradigms, as discussed above, and importantly, whether these biases can be accounted for with a model that incorporates modality-specific

encoding followed by coordinate-transformation to read out the posterior into the task-relevant reference frame. This is important because in our paradigms there has always been an inherent asymmetry such that stimuli are presented from multiple sensory modalities, but the response relies only on the visual modality. In this, we have always assumed that the posterior is amodal, and thus can be read out into either of the source modalities. However, it may be that signals are always modality-specific, and that the integrative process relies on arbitrating between their influences, such that a response in one modality requires an active transformation from the coordinate system of the other modality to its own. Such a model would challenge the Bayesian causal inference model described throughout this dissertation because it would require modality-specific priors and would introduce a new node in the graphical model with the express purpose of capturing this transformation (See Figure 6.1).

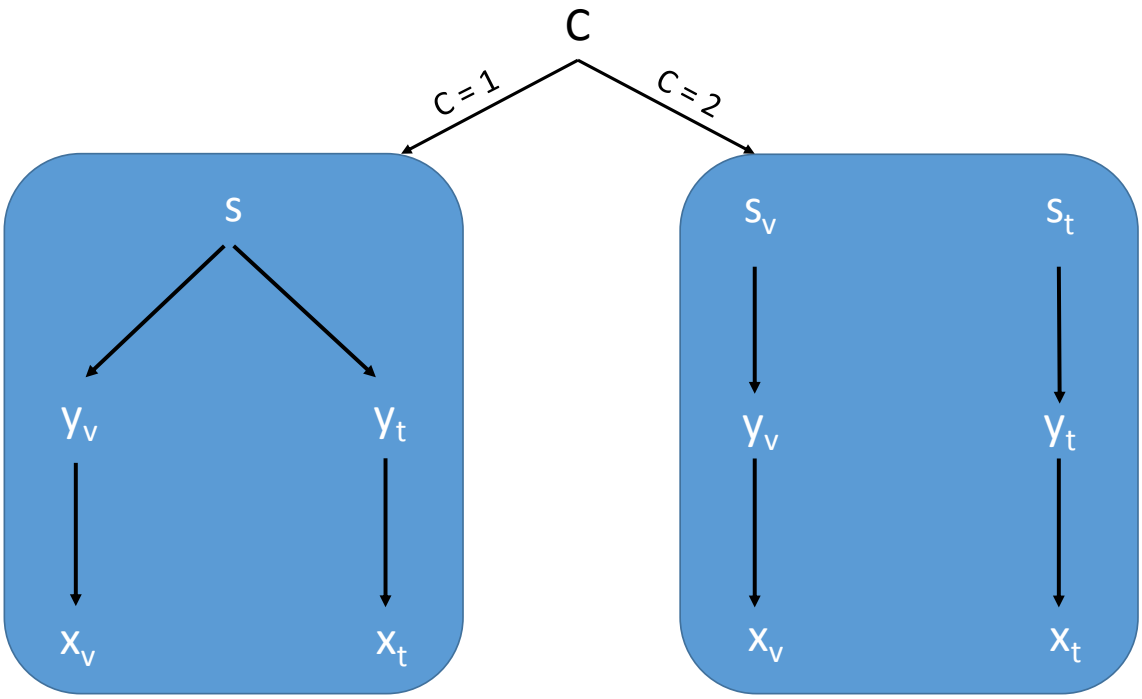


Figure 6.1: **Bayesian Causal Inference with Coordinate System Transformation.** This directed acyclic graph depicts the generative model for multisensory integration with coordinate system transformations. In short, the variable $C$ denotes the number of causes in the environment. If $C = 1$, then there is one true source, $s$, that generates the sensory signals $x_v$ and $x_t$. Note that since $x$ is represented in the task-relevant reference frame, we require an intervening coordinate transformation, $x_v = f(y_v)$ and $x_t = f(y_t)$, in order to account for systematic differences in the way different modalities encode information. Thus, we can refer to the step from $s$ to $y$ as **encoding**, and the step from $y$ to $x$ as **coordinate transformation**. In the case where $C = 2$, the environment contains two sources, $s_v$ and $s_t$, which are therefore independently encoded in their modality-specific reference frames as $y_v$ and $y_t$, and then subsequently transformed into the task-relevant reference frame as $x_v$ and $x_t$.

## 6.4 Significance

The body of work presented in this dissertation has broad relevance and applicability to a variety of fields of inquiry. As mentioned in the introduction, there are numerous aberrations from normality in regard to the psychological and perceptual representation of our own bodies. The more that the basic principles that underlie body ownership are understood, the higher are the chances that these syndromes may be alleviated. As a few examples, there are the psychiatric conditions known as body dysmorphic disorder and anorexia and its associated eating disorders, both of which involve abnormalities in the way that body-related information is processed by the perceptual system. At the very least, a model of the healthy functioning of the system can provide for a principled and quantitative way to both diagnose, as well as treat, these and related disorders. In addition, there are the many sufferers of limb amputation that experience the excruciating agony of phantom limb pain, for whom interventions that manipulate the body ownership have already proven to be efficacious (Ramachandran et al., 1995). In the future, investigations like those documented in this dissertation may well build on this and also help such patients incorporate prosthetics more fully into their body ownership.

Beyond its ability to contribute to the understanding and treatment of disorders, knowledge of the way in which the nervous system processes body related information and develops a sense of self-consciousness grounded therein has applications in the fields of robotics and artificial intelligence. As technology progresses, there may well come a time when a program similar to the one the human nervous system uses will have to be implemented in an autonomous agent that uses this as the stepping stone towards a full fledged consciousness

oriented around this nascently emergent self.

## 6.5 Conclusion

In conclusion, this dissertation was a multi-pronged attempt to study various aspects of body ownership and body representation from the perspective of multisensory integration and the computational models that have been proposed as accounts of such. To summarize, the rubber hand illusion and a novel multisensory illusion we have called the visuotactile ventriloquist effect have been observed and modeled using two forms of the Bayesian causal inference model which differ only in terms of the space that they model, namely peripersonal space in the former and somatotopic space in the latter. Resultantly, a comprehensive model has been proposed in a few different varieties, which aims to incorporate the inference over both spaces into one overarching causal inference model of body ownership. It is my sincere hope and wish that this contribution to the field be given its consideration under the uncompromising light of the scientific method, and that it becomes a stepping stone in service of those who strive dispassionately and unfalteringly towards the asymptotic goal of complete knowledge of the natural world.

# Appendix A

# Bayesian Causal Inference Toolbox (BCIT) for MATLAB

## A.1 Abstract

BCIT is a software extension built for the MATLAB platform, intended to facilitate running simulations and model fitting using the Bayesian causal inference model, a statistical framework for determining whether to integrate or segregate information arriving to the nervous system from different sensory channels. This is of relevance for many multisensory phenomena that result from the tricking of the sensory-perceptual system such as ventriloquism, the flash-beep illusion, the rubber hand illusion, and the Mckgurk effect. These illusions can all be accounted for under the normative framework of the Bayesian causal inference model. The aims of this program are 1) to provide the user with a graphical user interface by which the inner workings of the model can be made intuitive and easy to understand, and 2) to provide the user with the necessary machinery to be able to run an optimization procedure

to obtain optimal model fits to user-supplied datasets. Therefore, the primary intention with releasing this toolbox is to enhance the acquisition of the intuition behind the computational framework, which we hope to achieve by the implementation of user interface elements to control various parameters in the model, and to instantaneously observe the effect they have on the output from the model. Noting that the model is typically used to account for experimentally collected data, and would thus be intended to be fit to such data, we are also hoping to eliminate the barriers that prevent researchers from implementing the model and conducting fitting. Here, we present the toolbox and provide a description of the models that are implemented in it, and we also document a validation procedure demonstrating the convergence of the fitting procedure to the approximately correct parameters. Therefore, this toolbox provides a powerful platform for the rapid implementation of the Bayesian causal inference model.

## A.2 Introduction

The Bayesian causal inference model is a well-established computational model of perception that performs a statistical inference to determine whether signals across different sense modalities originated from the same cause, and thus ought to be integrated, or otherwise, and thus ought to be segregated. It was established nearly a decade ago and has been widely used since then to account for a wide range of multisensory perception phenomena (Kording et al., 2007; Beierholm et al., 2009b,a; Wozny et al., 2010; Samad et al., 2015; Samad and Shams, 2016; Rohe and Noppeney, 2015; Kilteni et al., 2015).

Specifically, the causal inference model is a statistical inference model that performs an

arbitration between integration and segregation, based on the spatiotemporal congruence of the signals – or indeed congruence along any suitably defined space – as well as a prior tendency to integrate/segregate. As such, this model represents a significant advance in the field of computational modeling from the method commonly known as Maximum Likelihood Estimation (MLE), which assumes that the signals of interest ought to always be combined, and thus, that they were generated by a common cause (Ernst and Banks, 2002). In contrast, the causal inference model makes no such assumptions but rather infers whether the situation of having been generated by a common cause or separate causes is more likely and then estimates the stimulus attributes accordingly.

This model has been shown to account for a wide range of multisensory phenomena across many domains including numerosity judgments in the flash-beep illusion (Wozny et al., 2008), spatial localization judgments in an audiovisual task (Wozny et al., 2010) and a visuotactile task (Samad and Shams, 2016), the size-weight illusion (Peters, 2014) and the rubber hand illusion (Samad et al., 2015), and has even been shown to account for the Mcgurk-Mcdonald Effect (Magnotti et al., 2013). The dissemination and distribution of this toolbox will, therefore, provide a very important service to the field of perception, and by extension computational neuroscience. In recent years, we have seen an explosion of interest in this computational framework by research groups from all around the world. Providing them with an interface as user-friendly as ours will dramatically reduce the friction with which they will be able to make use of its powerful computations.

Across these different domains, three general forms of the model have been in use, and are therefore provided to the user with the current software release. Namely, we are referring to the models of localization and numerosity across a variety of modalities. These have been

modeling using the Bayesian causal inference model that is characterized over a continuous or a discrete space. A third variant was introduced by Samad et al. (2015) in order to account for the rubber hand illusion and thus operates over a two dimensional (spatiotemporal) continuous space. Therefore, in what follows, we will concern ourselves with these three forms.

## Mathematical Formulation

The model utilizes the following form of Bayes Rule:

$$p(C|x_1, x_2) = \frac{p(x_1, x_2|C)p(C)}{p(x_1, x_2)} \tag{A.1}$$

where $x_1$ and $x_2$ are two signals received by the nervous system, and $C$ is a binary variable denoting the number of causes in the environment, 1 or 2.

Therefore, the posterior probability of the signals having a single cause in the environment is computed as:

$$p(C = 1|x_1, x_2) = \frac{p(x_1, x_2|C = 1)p(C = 1)}{p(x_1, x_2|C = 1)p(C = 1) + p(x_1, x_2|C = 2)(1 - p(C = 1))} \tag{A.2}$$

where the likelihood probability is:

$$p(x_1, x_2|C = 1) = \iint p(x_1, x_2|X)p(X)dX \tag{A.3}$$

and $p(C = 1)$ is the prior probability of a common cause. $X$ denotes the attributes of the stimuli in the dimension of relevance, and which gives rise to the neural representations $\{x_1, x_2\}$. It is modeled as a continuous random variable and has the following prior: $\mathcal{N}(\mu_X, \sigma_X)$, where $\mathcal{N}(\mu, \sigma)$ stands for a normal distribution with mean $\mu$ and standard deviation $\sigma$. Equation A.2 shows that two factors contribute to the inference of a common

cause: the likelihood (the first term in the numerator) and the prior (the second term in the numerator). A high likelihood (Equation A.3) occurs if the sensory signals are similar. The prior probability of a common cause, $p(C = 1)$, on the other hand, is independent of the present sensations, and depends on the observer's prior experience.

Note that the rest of the full mathematical formulation is left out from this document as it has appeared previously in print, and we direct the inquisitive reader thereto for further elucidation. In the tables which follow, any ambiguous formulation will be paired with an Equation number in Wozny and Shams (2011a) that describes the formulation in full detail.

## A.3 Program Description

BCIT is structured into two main types of operations: simulation and model fitting. The former permits the user to simulate from a selection of the three most commonly used variants of the model. These will be described in much greater detail below. The model fitting aspect permits users to use the `fminsearchbnd.m` optimization method on their own data sets, or alternatively, on a sample data set created from the `create_data.mat` file that is included. Thus, the user interface is structured into a main panel that allows the user to choose from among different model types, and whether simulation or fitting is desired. From there, the user navigates through some additional panels to achieve the desired computation, which will be described in the sections that follow.

## The graphical user interface

The main menu that the user sees upon first opening up the program is illustrated in Figure A.1. Here the user is presented with a choice from amongst the three most commonly used variants of the model: one dimensional continuous space, one dimensional discrete space, and two dimensional continuous space.

The user is also provided with the option to run simulations using these all three of the variants and can launch separate windows to view these simulations by having selected the desired model and pressing the "Simulate" button. In addition, the user is provided with the ability to conduct a model fitting procedure for a dataset of choice, using either of the one dimensional models, by selected one of them and pressing the "Fit Model" button. Note that a model fitting routine for the two dimensional model was not provided as this has not yet been performed due to the difficulty with acquiring a suitable dataset.

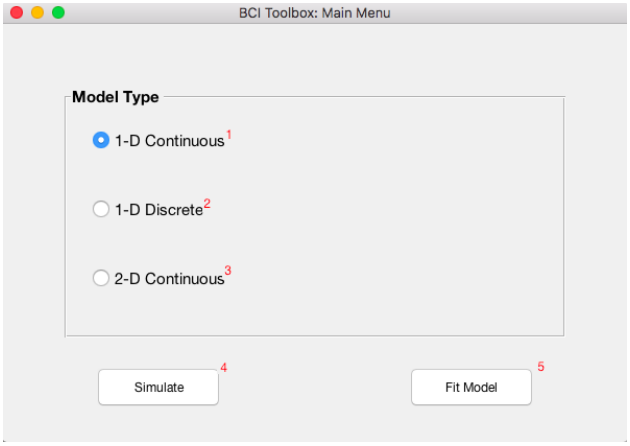


Figure A.1: Main Menu of the GUI

## The simulation panels

The primary aim of this project is to provide the user with an interface by which the inner workings of the Bayesian causal inference model can be made intuitive and easy to understand. Therefore, the primary intention with releasing this toolbox is to enhance the acquisition of the intuition behind the computational framework, and is thus primarily to be used as an educational tool. To that end, we hope to be able to provide the user with interface elements to control various parameters in the model, and instantaneously be able to observe the effect they have on the output from the model.

The simulation panels are split into three parts.

| | | |
|---|---|---|
| **Model Elements** | Response Distribution | The model output: a distribution of the estimates of the positions of the stimuli based on the likelihood and prior |
| | Stimulus Encoding | The probability density functions representing the encoding of the stimuli, modeled as Gaussian distributions. See equations 1 and 2 in Wozny and Shams (2011a) |
| | Spatial Prior | The probability density function representing the expected stimulus location, modeled as a Gaussian distribution. See equation 3 in Wozny and Shams (2011a) |
| **Model Estimates** | Mode | Most probable estimated response |
| | Mean | Mean estimated response |
| **Strategies** | Selection | Model selection is when the observer selects the most likely causal structure and estimates the stimulus location wholly on the basis of the selected model. See equation 16 in Wozny and Shams (2011a) |
| | Averaging | Model averaging is when the observer weights the estimates of the stimulus locations by the inferred probabilities of their causal structure. Considered the most optimal strategy. See equation 15 in Wozny and Shams (2011a) |
| | Matching | Probability matching is a strategy that choses the estimates from either causal structure based on their inferred probabilities. Although this method is suboptimal, it appears to be the most frequently used in cognitive tasks. See equation 17 in Wozny and Shams (2011a) |

Table A.1: **Simulation Panels:** Overview of Settings

| | | |
|---|---|---|
| **Stimulus Position** | Stimulus 1 | The true position of the stimulus (modality 1) |
| | Stimulus 2 | The true position of the stimulus (modality 2) |
| **Parameters** | P(C=1) | The prior probability that both signals can be attributed to one cause |
| | SD(1) | The standard deviation of the Gaussian distribution of the likelihood for modality 1 |
| | SD(1) | The standard deviation of the Gaussian distribution of the likelihood for modality 2 |
| | SD(Prior) | The standard deviation of the Gaussian distribution of the prior (the anticipated location of the stimuli ) |
| | Mean(Prior) | The mean of the Gaussian distribution of the prior |
| | Additional Parameters | Additional parameters specific to the models will be discussed in the detailed model descriptions below |
| **Bottom Panel Buttons** | Screenshot | Saves a copy of the screenshot (figure, all parameters) with user-set filename to the current directory |
| | Reset | Resets all parameters and figure to default settings |
| | Return | Returns to the main menu |

Table A.2: **Simulation Panels:** Overview of UI Elements

### One Dimensional Continuous

This model is one dimensional and continuous. This represents the most basic form of the model and is most akin to the form that was introduced in the seminal paper in 2007 (Kording et al., 2007). Over the years, we have produced several variants of it that were tailored for particular tasks and domains. But it is best we begin our discussion of the core computations with reference to this initial form.

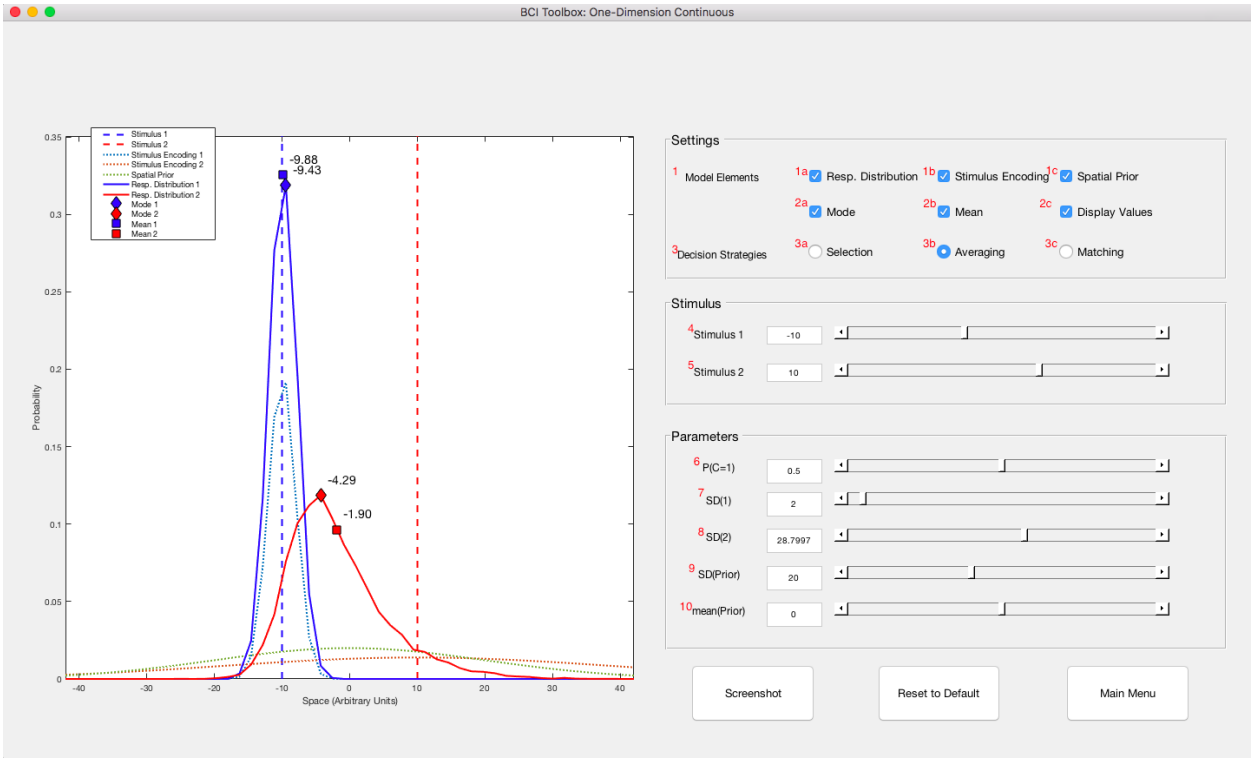


Figure A.2: Simulation Panel 1: One Dimensional Continuous

For all parameters: Both the boxes and sliders can be used to manipulate values. Sliders can be manipulated either by using the arrow keys attached to the left and right or by pressing the spaces to the left or right of the value indicator, and the increments of the sliders are given below.

| | | |
|---|---|---|
| **Model Elements (1)** | Response Distribution (1a) | Indicated by the solid blue and red lines |
| | Stimulus Encoding (1b) | Indicated by the dotted blue and red lines |
| | Spatial Prior (1c) | Indicated by the dotted green line |
| **Model Estimates (2)** | Mode (2a) | Value indicated by red and blue diamonds |
| | Mean (2b) | Value indicated by red and blue squares |
| | Display Values (2c) | Shows the value of the model estimate of probability on the figure |
| **Strategies (3)** | Selection (2a) | See explanations in "Description" section, under "Simulation Panels" |
| | Averaging (3b) | |
| | Matching (3c) | |
| **Stimulus Position** | Stimulus 1 (4) | Stimulus position ranges from -40 to 40. Sliders increment by 1 |
| | Stimulus 2 (5) | |
| **Elements Parameter** | P(C=1) (6) | Probability values range from 0 to 1. Slider increments by 0.01 |
| | SD(1) (7) | Standard deviation of X1 signal ranges from 0.1 to 50. Slider increments by 0.1 |
| | SD(1) (8) | Standard deviation of X2 signal ranges from 0.1 to 50. Slider increments by 0.1 |
| | SD(Prior) (9) | Standard deviation of prior signal ranges from 1 to 50. Slider increments by 0.1 |
| | Mean(Prior) (10) | Average of prior signal ranges from -40 to 40. Slider increments by 1 |

Table A.3: **One-Dim Continuous Simulation Panel:** Description of UI Elements

### One Dimensional Discrete

This model is also one dimensional and discrete.

For all parameters: Both the boxes and sliders can be used to manipulate values. Sliders can be manipulated either by using the arrow keys attached to the left and right or by pressing the spaces to the left or right of the value indicator. The increments of the sliders

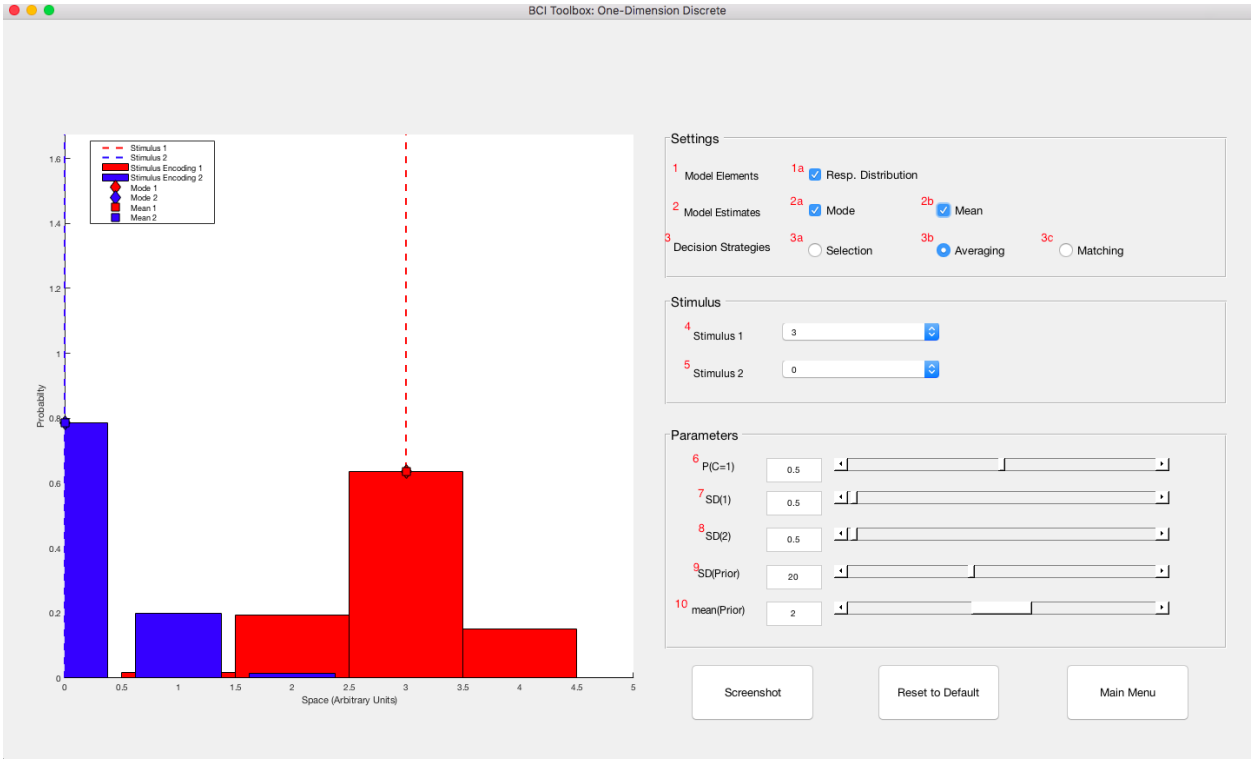


Figure A.3: Simulation Panel 2: One Dimensional Discrete

are given below.

| | | |
|---|---|---|
| **Model Elements (1)** | Response Distribution (1a) | Model estimate indicated by red and blue bars |
| **Model Estimates (2)** | Mode (2a) | Value indicated by red and blue diamonds |
| | Mean (2b) | Value indicated by red and blue squares |
| **Strategies (3)** | Selection (2a) | See explanations in "Description" section, under "Simulation Panels" |
| | Averaging (3b) | |
| | Matching (3c) | |
| **Stimulus Position** | Stimulus 1 (4) | Discrete stimuli can be specified as: 0, 1, 2, 3 or 4 |
| | Stimulus 2 (5) | |
| **Elements Parameter** | P(C=1) (6) | Probability values range from 0 to 1. Slider increments by 0.01 |
| | SD(1) (7) | Standard deviation of X1 likelihood ranges from 0.1 to 50. Slider increments by 0.1 |
| | SD(1) (8) | Standard deviation of X2 likelihood ranges from 0.1 to 50. Slider increments by 0.1 |
| | SD(Prior) (9) | Standard deviation of prior ranges from 1 to 50. Slider increments by 0.1 |
| | Mean(Prior) (10) | Mean of prior ranges from -40 to 40. Slider increments by 1 |

Table A.4: **One-Dim Discrete Simulation Panel:** Description of UI Elements

### Two Dimensional Continuous

This model is used for spatiotemporal causal inference.

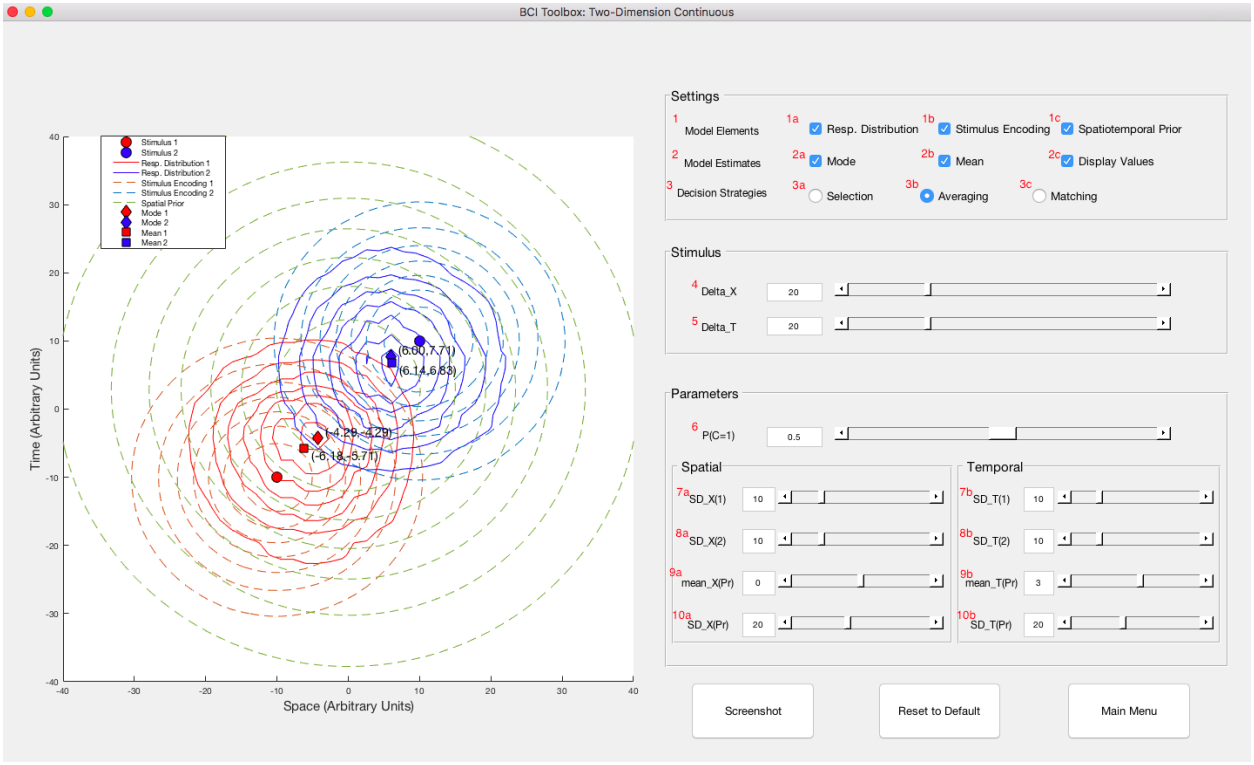


Figure A.4: Simulation Panel 3: Two Dimensional Continuous

| | | |
|---|---|---|
| **Model Elements (1)** | Response Distribution (1a) | Indicated by the solid blue and red lines |
| | Stimulus Encoding (1b) | Indicated by the dotted blue and red lines |
| | Spatiotemporal Prior (1c) | Indicated by the dotted green line |
| **Model Estimates (2)** | Mode (2a) | Value indicated by red and blue diamonds |
| | Mean (2b) | Value indicated by red and blue squares |
| | Display Values (2c) | Shows the model estimates on the figure |
| **Strategies (3)** | Selection (2a) | See explanations in "Description" section, under "Simulation Panels" |
| | Averaging (3b) | |
| | Matching (3c) | |
| **Stimulus Position** | Delta_X (4) | This value changes the difference between the stimuli along the x-axis |
| | Delta_T (5) | This value changes the difference between the two stimuli along the y-axis |
| **Probability** | P(C=1) (6) | Probability values range from 0 to 1 |
| **Spatial** | SD_X(1) (7a) | Standard deviation of X1 likelihood ranges from 0.1 to 50 |
| | SD_X(2) (8a) | Standard deviation of X1 likelihood ranges from 0.1 to 50 |
| | Mean_X(Prior) (9a) | Spatial mean of the prior, ranges from -40 to 40 |
| | SD_X(Prior) (10a) | Spatial standard deviation of the prior, ranges from 0.1 to 50 |
| **Temporal** | SD_T(1) (7b) | Temporal standard deviation of modality 2 likelihood, ranges from 0.1 to 50 |
| | SD_T(2) (8b) | Temporal standard deviation of modality 2 likelihood, ranges from 0.1 to 50 |
| | Mean_T(Prior) (9b) | Temporal mean of the prior, ranges from -40 to 40 |
| | SD_T(Prior) (10b) | Temporal standard deviation of the prior, ranges from 0.1 to 50 |

Table A.5: **Two-Dim Continuous Simulation Panel:** Description of UI Elements

## The fitting panel

Noting that the model is typically used to account for experimentally collected data, and would thus be intended to be fit to such data, we secondarily intend to provide our users with the necessary machinery to be able to achieve this stated purpose of the model, given that the data they have collected conforms to our nominal conventions, which are to be specified in the documentation. While there are a great many variants of the model that have been implemented over the years for fitting purposes, we will provide a restricted set of the eight most commonly used parameters and give the user control over which of them to include in the fitting procedure, as well as the three most commonly used decision strategies by which the model estimates are read out.

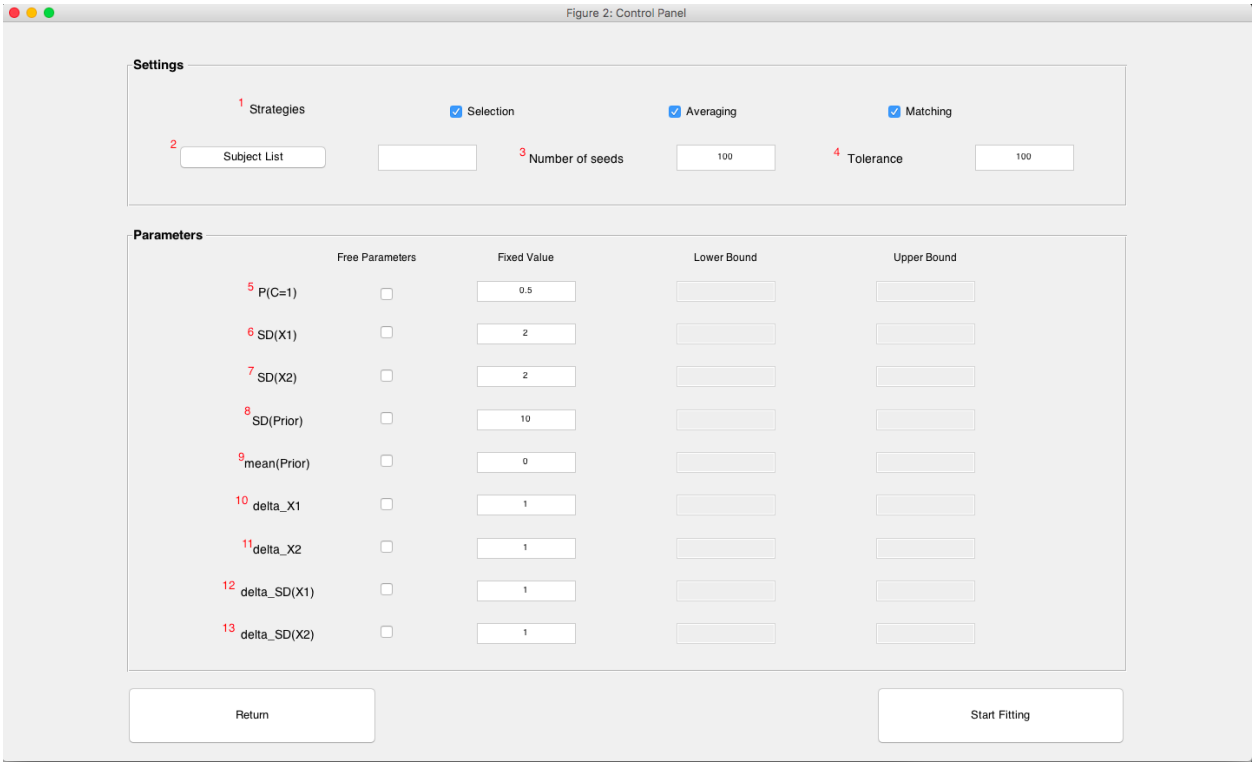


Figure A.5: Fitting Panel

| | | |
|---|---|---|
| **Strategies** (1) | Selection | For strategy descriptions, see explanations in "Description" section, under "Simulation Panels" |
| | Averaging | |
| | Matching | |
| **User Inputs** | Subject List (2) | Users can upload their data using a specific layout in a .mat format (see A Note on How to Format Data) |
| | Number of Seeds (3) | Sets the number of seeds used to analyze user data, increasing this number will require more processing time. Number of seeds will be consistent and independent for all strategies the user runs |
| | Tolerance (4) | The lower bound on changes in error that the optimizer uses as a criterion for convergence |
| **Parameters** The user can designate all parameters as either "Free Parameters" or "Fixed Values". Free parameters will be allowed to vary between a user-specified lower bound and upper bound. Using more free parameters will increase the model fit to the data, however it will increase the processing time required. | P(C=1) (5) | The prior probability that both signals can be attributed to one cause |
| | SD(X1) (6) | The standard deviation of the Gaussian distribution of the sensory encoding for modality 1 |
| | SD(X2) (7) | The standard deviation of the Gaussian distribution of the sensory encoding for modality 2 |
| | SD(Prior) (8) | The standard deviation of the Gaussian distribution of the prior |
| | Mean(Prior) (9) | The mean of the Gaussian distribution of the prior |
| | Delta_X1 (10) | A multiplicative factor that scales the mean of the Gaussian distribution for the sensory encoding for modality 1 |
| | Delta_X2 (11) | A multiplicative factor that scales mean of the Gaussian distribution for the sensory encoding for modality 2 |
| | Delta_SD(X1) (12) | A multiplicative factor that scales the standard deviation of the Gaussian distribution for the sensory encoding for modality 1 as a function of the space |
| | Delta_SD(X2) (13) | A multiplicative factor that scales the standard deviation of the Gaussian distribution for the sensory encoding for modality 2 as a function of the space |

Table A.6: **Fitting Panel:** Description of UI Elements

### Description of Model Fitting Procedure

Our implementation of model fitting relies on the use of the included function `fminsearchbnd.m`, which is based on the built-in Matlab function `fminsearch.m`. Briefly, this function implements the Nelder & Mead Simplex algorithm for derivative-free optimization, that is, for use with objective functions that are difficult or impossible to calculate gradients for. In the case of this toolbox, the objective function is based on the calculation of the negative log likelihood of the data to be fit given the predicted response distribution that is estimated from the model parameters. Given the complexity of the model structure, it is not readily apparent how a gradient of this error quantity with respect to the parameters can be computed, thus making `fminsearchbnd.m` a good choice of optimization algorithm. The difference between `fminsearchbnd.m` and `fminsearch.m` is that the former adds a way to constrain the space within which the algorithm searches for parameters, thus requiring the user to specify these bounds individually for each parameters to be optimized. This is an important and desired property for our purposes because it prevents the optimizer from diverging too wildly in its estimates and helps minimize the variance of the optimized parameters.

As with `fminsearch.m`, `fminsearchbnd.m` requires the user to input initial values from which the optimizer begins its descent down the error hill. These are chosen by random sampling from the uniform distribution encompassing the space defined by the bounds on each parameter. The greater the number of such initial random **Seeds**, as they are called, the less likely that the optimizer will be stuck in a local minimum, and thus the greater will be the modeler's faith that it will converge onto a global optimum. Another factor to

consider here is also the criterion on error changes that the optimizer uses to terminate the procedure and consider its parameters converged. This is often referred to as the **Tolerance** and a careful setting of its value can greatly aid in the efficient optimization of a set of parameters. If it is set too low, the optimizer will waste computational resources chasing negligible reductions in error that do not add significant improvements, but in contrast, if the value is too high, the optimizer will terminate very rapidly without generating a satisfactory fit.

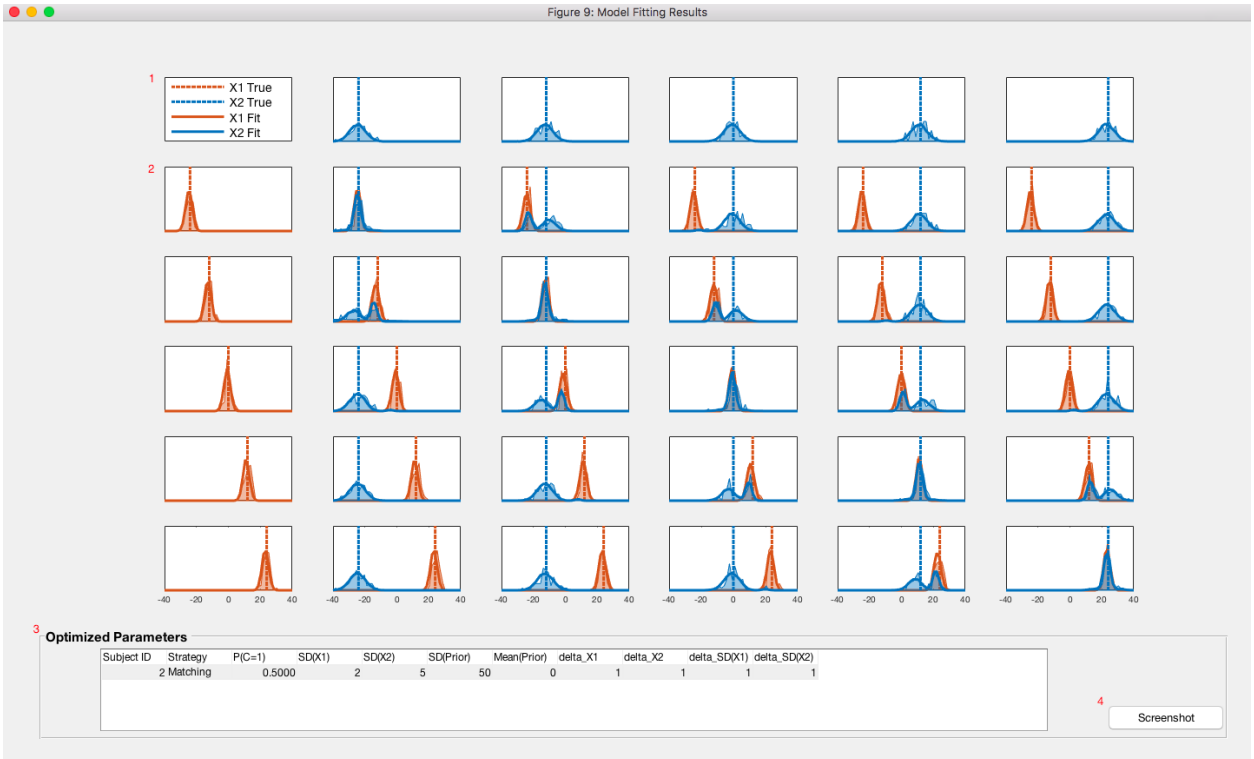


Figure A.6: Model Fits and Optimized Parameters

## A Note on How to Format Data

If researchers wish to use their own data for conducting model fitting using this toolbox, a few very important considerations regarding the format of the data will arise. In order for

| | | |
|---|---|---|
| **Legend** (1) | X1 True | True position of the first stimulus represented by the red dotted line |
| | X2 True | True position of the second stimulus represented by the blue dotted line |
| | X1 Fit | Model estimate for first stimulus location probability represented by solid red line |
| | X2 Fit | Model estimate for second stimulus location probability represented by solid red line |
| **Plots** (2) | Axes | Axes mirror that of the simulation panels for the 1-dimension continuous model, the X corresponding to spatial location ranging from -40 to 40 degrees and the Y corresponding to the normalized probability of perceived location of the stimuli |
| | Layout | The plots are displayed such that moving horizontally rightward corresponds to shifting the true position of the second stimuli rightward. Conversely moving downwards vertically corresponds to shifting the true position of the first stimuli rightward |
| **Optimized Parameters** (3) | Subject ID | The subject number for which the parameters were optimized for |
| | Strategy | The optimal strategy used for parameters (calculated in model fitting based on the selected strategies in the control panel) |
| | Parameters | The optimized parameter values |
| **Buttons** (4) | Screenshot | The screenshot button will save both the plots as well as the optimization parameters as a 300 dpi .png file (named by user) to the current directory. The button will disappear as the image is being saved and then reappear after the process is complete |

Table A.7: **Model Fitting Results:** Description of Figure Output

the built-in model fitting routines to properly function, the data format conventions that we have outlined below must be adhered to. The included file in the toolbox `create_data.mat`,

can be examined for a demonstration of how this format is implemented. In brief, the data must be stored as a MATLAB data structure, the required details of which will be explained in more detail in what follows.

Firstly, the data structure variable must be assigned the name "data", and it is critical that the data structure contain the following required fields, which the model fitting routine expects and utilizes for setting up and performing the fitting procedure. (1) A field with the name "N" indicating the number of Monte Carlo samples to take in computing the model estimates – setting this equal to 10,000 will suffice in most cases. (2) A field with the name "space" that contains the discretized continuum upon which the data and model fits will be represented – it is often useful to construct this so as to yield 1 degree of visual angle per discrete unit. (3) A field with the name "stim_locs" that contains the stimulus positions within the space from where stimuli can be presented – note that the first element must be a NaN so as to indicate the possibility of unisensory conditions. (4) A field with the name "conds" whose columns specify all the possible unique stimulus combinations – the two rows represent the stimulus positions for the two modalities, respectively, with a NaN indicating the lack of the presentation of a stimulus from that modality, thus, a unisensory condition. (5) A field with the name "stim" indicating the order of trials, it is essentially a pseudorandomly generated permutation of the columns of the "conds" field – thus it has as many columns as there were trials in the experiment and two rows for the two modalities that stimuli could be presented from. (6) A field with the name "resp", which has the same dimensions as the "stim" field, but which stores the responses on each of those trials that are indicated therein. (7) A field with the name "cond_resps" and containing an array that stores all the responses sorted into columns according to the conditions indicated by the

"conds" field. Thus "cond_resps" has as many columns as "conds", as many rows as there were repetitions for each particular stimulus configuration, and has dimension 2 along the third index separating the data based on whether the responses were from modality 1 or modality 2. (8) A field with the name "subject" that specifies a numeric identifier for the subject that will be used to identify the fitting results.

## A.4 Fitting Validation

### Method

To demonstrate the standard operation of the fitting procedure that is built-in to the toolbox, as well as to validate its proper functioning, we first created some example data using the `create_data.mat` file that is included. This file utilizes the one dimensional continuous Bayesian causal inference model, supplied with some parameters that the user is able to input. For this simulation, we used the following parameters: $\{p(C = 1) = 0.5,\ \sigma_{X1} = 2,\ \sigma_{X2} = 5,$ prior $\sim \mathcal{N}(0, 15)\}$, and used a strategy of probability matching. The `create_data.mat` file simulated 2030 trials of localization that included both unisensory and bisensory trials. There were five candidate positions where stimuli from both modalities could be presented at, which were separated by 12°, thus providing for 35 possible stimulus configurations ($(5 \times 5 = 25$ bisensory pairs) and $(5 + 5 = 10$ unisensory stimuli) $= 35$ total). The script runs through a pseudorandomized and balanced order of stimulus configurations and generates simulated response distributions using the supplied parameters. This distribution is then sampled in accordance to its probability distribution in order to generate the dataset.

Next, we ran the model fitting routine in order to observe whether the model would converge on the parameters that we used in generating the simulations. We conducted this using 100 random initial values that we used as seeds to the fitting procedure. These initial values were selected by random sampling using a uniform distribution between the lower and upper bounds. For this test, the bounds used were $\{p(C = 1) \in [0, 1]$, $\sigma_{X1} \in [1, 10]$, $\sigma_{X2} \in [1, 10]$, prior $\sim \mathcal{N}(\mu \in [-40, 40], \sigma \in [10, 100])\}$. Here, we should hasten to add that we ran this test with all three decision strategies selected. Thus, the optimization routine will use each randomly generated initial seed three times as it attempts the model fitting for all three strategies in turn, and selects the best fitting set of parameters and decision strategy at the end.

In addition, we set the tolerance on error changes that would be used as a criterion for convergence to 100. Note that since this error is computed as negative log likelihood of the data under the simulated response distribution, the absolute value of the tolerance depends strongly on the number of points that the model attempts to fit and would therefore require modification for the particular dataset at hand. The value of 100 herein was chosen so as to optimally balance speed of fitting with accuracy of fitted parameter values.

## Results

The model fitting procedure took 10 minutes to complete on a Macbook (Retina, 15-inch, Mid 2015, 2.8 GHz Intel Core i7, 16 GB 1600 MHz DDR3). In the table below, we report the optimized parameters from the 8 runs of model fitting that we conducted. As can be clearly seen, despite a little variation across runs, the parameters appear to converge satisfactorily

onto the true values.

| | $\mathbf{p(C=1)}$ | $\sigma_{\mathbf{X1}}$ | $\sigma_{\mathbf{X2}}$ | $\mu_{\mathbf{prior}}$ | $\sigma_{\mathbf{prior}}$ | **Strategy** | **Error** |
|---|---|---|---|---|---|---|---|
| **True Values** | 0.5 | 2 | 5 | 0 | 15 | Probability Matching | |
| **Fitting Run 1** | 0.67 | 2.00 | 4.51 | 0.00 | 15.73 | Probability Matching | 7156.1 |
| **Fitting Run 2** | 0.45 | 2.03 | 5.21 | 0.00 | 14.05 | Probability Matching | 7136.9 |
| **Fitting Run 3** | 0.40 | 1.91 | 5.17 | 0.00 | 16.04 | Probability Matching | 7142.2 |
| **Fitting Run 4** | 0.38 | 1.96 | 4.70 | 0.00 | 13.68 | Probability Matching | 7154.7 |
| **Fitting Run 5** | 0.54 | 2.09 | 5.71 | 0.00 | 14.74 | Probability Matching | 7156.3 |
| **Fitting Run 6** | 0.45 | 1.91 | 4.84 | 0.00 | 14.14 | Probability Matching | 7144.0 |
| **Fitting Run 7** | 0.53 | 2.11 | 5.26 | 0.00 | 18.02 | Probability Matching | 7143.2 |
| **Fitting Run 8** | 0.52 | 2.10 | 5.04 | 0.00 | 15.01 | Probability Matching | 7130.3 |

Table A.8: **Results:** Optimized Parameter Fits

## A.5 Outlook

Here we present a new computational tool designed for the purpose of aiding researchers to better understand and implement the Bayesian causal inference model as an explanatory framework where they might have data suitable for this purpose. Many of the paradigms of multisensory research are highly amenable to being modeled by this framework and we, therefore, expect this tool to have widespread utility across the field. Aside from its educational function as an intuition-building software package, this tool also provides researchers the ability to generate fits of the model to their own data, gaining insights into the behavior of subjects through the optimized parameters. In the past decade of work on this model, it still remains a complex framework to grasp, and there is often a steep barrier to its utilization by research groups interested in doing so. We, therefore, expect this tool to be of use to researchers in a variety of disciplines such as experimental psychology, computational neu-

roscience and cognitive science, who may wish to implement it for the study of multisensory phenomena across a wide range of paradigms.

# Bibliography

Abdulkarim, Z. and Ehrsson, H. H. (2016). No causal link between changes in hand position sense and feeling of limb ownership in the rubber hand illusion. *Attention, Perception & Psychophysics*, 78(2):707–720.

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology: CB*, 14(3):257–262.

Alsius, A., Navarra, J., Campbell, R., and Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current biology: CB*, 15(9):839–843.

Armel, K. C. and Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proceedings of the Royal Society B: Biological Sciences*, 270(1523):1499–1506.

Asai, T., Mao, Z., Sugimori, E., and Tanno, Y. (2011). Rubber hand illusion, empathy, and schizotypal experiences in terms of self-other representations. *Consciousness and Cognition*, 20(4):1744–1750.

Aspell, J. E., Heydrich, L., Marillier, G., Lavanchy, T., Herbelin, B., and Blanke, O. (2013). Turning body and self inside out: visualized heartbeats alter bodily self-consciousness and

tactile perception. *Psychological Science*, 24(12):2445–2453.

Avillac, M., Hamed, S. B., and Duhamel, J.-R. (2007). Multisensory Integration in the Ventral Intraparietal Area of the Macaque Monkey. *The Journal of Neuroscience*, 27(8):1922–1932.

Baily, J. S. (1972). Arm-body adaptation with passive arm movements. *Perception & Psychophysics*, 12(1):39–44.

Bedford, F. L. (1989). Constraints on learning new mappings between perceptual dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):232–248.

Beers, R. J. v., Sittig, A. C., and Gon, J. J. D. v. d. (1999). Integration of Proprioceptive and Visual Position-Information: An Experimentally Supported Model. *Journal of Neurophysiology*, 81(3):1355–1364.

Beierholm, U., Kording, K. P., Shams, L., and Ma, W. J. (2009a). Comparing Bayesian models of multisensory cue combination without mandatory integration. In *Advances in neural information processing systems*, volume 20, pages 81–88. MIT Press, Cambridge, MA.

Beierholm, U. R., Quartz, S. R., and Shams, L. (2009b). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, 9(5):23–23.

Bekrater-Bodmann, R., Foell, J., Diers, M., and Flor, H. (2012). The perceptual and neuronal stability of the rubber hand illusion across contexts and over time. *Brain Research*, 1452:130–139.

Botvinick, M. and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, 391(6669):756.

Brown, L. E., Rosenbaum, D. A., and Sainburg, R. L. (2003). Movement speed effects on limb position drift. *Experimental Brain Research*, 153(2):266–274.

Bruns, P., Spence, C., and Roder, B. (2011). Tactile recalibration of auditory spatial representations. *Experimental Brain Research*, 209(3):333–344.

Butler, J. S., Smith, S. T., Campos, J. L., and Bulthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision*, 10(11):23–23.

Caclin, A., Soto-Faraco, S., Kingstone, A., and Spence, C. (2002). Tactile "capture" of audition. *Perception & Psychophysics*, 64(4):616–630.

Carruthers, G. (2008). Types of body representation and the sense of embodiment. *Consciousness and Cognition*, 17(4):1302–1316.

Christie, M. and Venables, P. (1980). Electrodermal Activity. In *Techniques in Psychophysiology*, pages 2–67. John Wiley, New York.

Costantini, M. and Haggard, P. (2007). The rubber hand illusion: sensitivity and reference frame for body ownership. *Consciousness and Cognition*, 16(2):229–240.

Craig, A. D. B. (2010). The sentient self. *Brain Structure & Function*, 214(5-6):563–577.

Dadarlat, M. C., O'Doherty, J. E., and Sabes, P. N. (2015). A learning-based approach to artificial sensory feedback leads to optimal integration. *Nature Neuroscience*, 18(1):138–144.

Davies, A. M. A., White, R. C., Thew, G., Aimola, N. M. V., and Davies, M. (2010). Visual Capture of Action, Experience of Ownership, and the Illusion of Self-Touch: A New Rubber Hand Paradigm. *Perception*, 39(6):830–838.

de Vignemont, F. (2010). Body schema and body image–pros and cons. *Neuropsychologia*, 48(3):669–680.

Desmurget, M., Vindras, P., Grea, H., Viviani, P., and Grafton, S. T. (2000). Proprioception does not quickly drift during visual occlusion. *Experimental Brain Research*, 134(3):363–377.

Dinh, H. Q., Walker, N., Hodges, L. F., Song, C., and Kobayashi, A. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *, IEEE Virtual Reality, 1999. Proceedings*, pages 222–228.

Dummer, T., Picot-Annand, A., Neal, T., and Moore, C. (2009). Movement and the Rubber Hand Illusion. *Perception*, 38(2):271–280.

Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science (New York, N.Y.)*, 317(5841):1048.

Ehrsson, H. H., Holmes, N. P., and Passingham, R. E. (2005). Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(45):10564–10573.

Ehrsson, H. H., Spence, C., and Passingham, R. E. (2004). That's my hand! Activity

in premotor cortex reflects feeling of ownership of a limb. *Science (New York, N.Y.)*, 305(5685):875–877.

Epstein, S. and Roupenian, A. (1970). Heart rate and skin conductance during experimentally induced anxiety: The effect of uncertainty about receiving a noxious stimulus. *Journal of Personality and Social Psychology*, 16(1):20–28.

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

Feldman, H. and Friston, K. J. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, 4.

Ferri, F., Chiarelli, A. M., Merla, A., Gallese, V., and Costantini, M. (2013). The body beyond the body: expectation of a sensory event is enough to induce ownership over a fake hand. *Proceedings. Biological Sciences / The Royal Society*, 280(1765):20131140.

Folegatti, A., Farne, A., Salemme, R., and de Vignemont, F. (2012). The Rubber Hand Illusion: two's a company, but three's a crowd. *Consciousness and Cognition*, 21(2):799–812.

Frissen, I., Vroomen, J., de Gelder, B., and Bertelson, P. (2003). The aftereffects of ventriloquism: Are they sound-frequency specific? *Acta Psychologica*, 113(3):315–327.

Fujisaki, W., Shimojo, S., Kashino, M., and Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, 7(7):773–778.

Geldard, F. and Sherrick, C. (1972). The cutaneous "rabbit": a perceptual illusion. *Science*,

178(57):178–179.

Gentile, G., Guterstam, A., Brozzoli, C., and Ehrsson, H. H. (2013). Disintegration of Multisensory Signals from the Real Hand Reduces Default Limb Self-Attribution: An fMRI Study. *The Journal of Neuroscience*, 33(33):13350–13366.

Gepshtein, S., Burge, J., Ernst, M. O., and Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of vision*, 5(11):1013–1023.

Graziano, M., Yap, G. S., and Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science*, 266(5187):1054–1057.

Graziano, M. S. and Botvinick, M. M. (2002). How the brain represents the body: insights from neurophysiology and psychology. *Common mechanisms in perception and action: Attention and performance*, XIX:136–157.

Graziano, M. S. A. and Cooke, D. F. (2006). Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia*, 44(6):845–859.

Graziano, M. S. A., Cooke, D. F., and Taylor, C. S. R. (2000). Coding the Location of the Arm by Sight. *Science*, 290(5497):1782–1786.

Graziano, M. S. A. and Gross, C. G. (1998). Visual responses with and without fixation: neurons in premotor cortex encode spatial locations independently of eye position. *Experimental Brain Research*, 118(3):373–380.

Green, B. G. (1982). The perception of distance and location for dual tactile pressures. *Perception & Psychophysics*, 31(4):315–323.

Guterstam, A., Gentile, G., and Ehrsson, H. H. (2013). The invisible hand illusion: multisensory integration leads to the embodiment of a discrete volume of empty space. *Journal of Cognitive Neuroscience*, 25(7):1078–1099.

Guterstam, A., Petkova, V. I., and Ehrsson, H. H. (2011). The Illusion of Owning a Third Arm. *PLOS ONE*, 6(2):e17208.

Guterstam, A., Zeberg, H., ÃŰzÃğiftci, V. M., and Ehrsson, H. H. (2016). The magnetic touch illusion: A perceptual correlate of visuo-tactile integration in peripersonal space. *Cognition*, 155:44–56.

Haans, A., Ijsselsteijn, W. A., and de Kort, Y. A. W. (2008). The effect of similarities in skin texture and hand shape on perceived ownership of a fake limb. *Body Image*, 5(4):389–394.

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., and Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, 15(1):20–29.

Hay, J. C. and Pick Jr., H. L. (1966). Visual and proprioceptive adaptation to optical displacement of the visual stimulus. *Journal of Experimental Psychology*, 71(1):150–158.

Head, H. and Holmes, G. (1911). Sensory Disturbances from Cerebral Lesions. *Brain*, 34(2-3):102–254.

Held, R. and Hein, A. V. (1958). Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. *Perceptual and Motor Skills*, 8(3):87–90.

Helmholtz, H. v. (1867). *Handbuch der physiologischen Optik*. Leopold Voss, Leipzig.

Herrera, G., Jordan, R., and Vera, L. (2006). Agency and Presence: A Common Dependence on Subjectivity? *Presence*, 15(5):539–552.

Hirsh, I. J. and Sherrick, C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, 62:423–432.

Holle, H., McLatchie, N., Maurer, S., and Ward, J. (2011). Proprioceptive drift without illusions of ownership for rotated hands in the "rubber hand illusion" paradigm. *Cognitive Neuroscience*, 2(3-4):171–178.

Holmes, N. P., Crozier, G., and Spence, C. (2004). When mirrors lie: "visual capture" of arm position impairs reaching performance. *Cognitive, Affective & Behavioral Neuroscience*, 4(2):193–200.

Holmes, N. P. and Spence, C. (2005). Visual bias of unseen hand position with a mirror: spatial and temporal factors. *Experimental Brain Research*, 166(3-4):489–497.

Howard, I. and Templeton, W. (1966). *Human spatial orientation*. John Wiley & Sons, Oxford, England.

Jones, S. A. H., Cressman, E. K., and Henriques, D. Y. P. (2010). Proprioceptive localization of the left and right hands. *Experimental Brain Research*, 204(3):373–383.

Kalckert, A. and Ehrsson, H. H. (2012). Moving a Rubber Hand that Feels Like Your Own: A Dissociation of Ownership and Agency. *Frontiers in Human Neuroscience*, 6.

Kammers, M. P. M., de Vignemont, F., Verhagen, L., and Dijkerman, H. C. (2009). The rubber hand illusion in action. *Neuropsychologia*, 47(1):204–211.

Kennett, S., Taylor-Clarke, M., and Haggard, P. (2001). Noninformative vision improves the spatial resolution of touch in humans. *Current Biology*, 11(15):1188–1191.

Kilteni, K., Maselli, A., Kording, K. P., and Slater, M. (2015). Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience*, 9.

Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, 2(9).

Ladavas, E. (2002). Functional and dynamic properties of visual peripersonal space. *Trends in Cognitive Sciences*, 6(1):17–22.

Ladavas, E., Pellegrino, G. d., Farne, A., and Zeloni, G. (1998). Neuropsychological Evidence of an Integrated Visuotactile Representation of Peripersonal Space in Humans. *Journal of Cognitive Neuroscience*, 10(5):581–589.

Landy, M. S., Maloney, L. T., Johnston, E. B., and Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, 35(3):389–412.

Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science (New York, N.Y.)*, 317(5841):1096–1099.

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

Ley, P., Steinberg, U., Hanganu-Opatz, I. L., and Roder, B. (2015). Event-related potential

evidence for a dynamic (re-)weighting of somatotopic and external coordinates of touch during visual-tactile interactions. *European Journal of Neuroscience*, 41(11):1466–1474.

Lloyd, D. M. (2007). Spatial limits on referred touch to an alien limb may reflect boundaries of visuo-tactile peripersonal space surrounding the hand. *Brain and Cognition*, 64(1):104–109.

Longo, M. R., Schuur, F., Kammers, M. P. M., Tsakiris, M., and Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition*, 107(3):978–998.

Macaluso, E. and Maravita, A. (2010). The representation of space near the body through touch and vision. *Neuropsychologia*, 48(3):782–795.

Magnotti, J. F., Ma, W. J., and Beauchamp, M. S. (2013). Causal inference of asynchronous audiovisual speech. *Frontiers in Psychology*, 4:798.

Mahoney, J. R., Molholm, S., Butler, J. S., Sehatpour, P., Gomez-Ramirez, M., Ritter, W., and Foxe, J. J. (2015). Keeping in touch with the visual system: spatial alignment and multisensory integration of visual-somatosensory inputs. *Frontiers in Psychology*, 6.

Makin, T. R., Holmes, N. P., and Ehrsson, H. H. (2008). On the other hand: dummy hands and peripersonal space. *Behavioural Brain Research*, 191(1):1–10.

Makin, T. R., Holmes, N. P., and Zohary, E. (2007). Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(4):731–740.

Mamassian, P. and Landy, M. S. (1998). Observer biases in the 3d interpretation of line

drawings. *Vision Research*, 38(18):2817–2832.

Mamassian, P. and Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Research*, 41(20):2653–2668.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W. H. Freeman.

Mcgurk, H. and Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588):746–748.

Meredith, M. A. and Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, 56(3):640–662.

Moseley, G. L., Gallace, A., and Iannetti, G. D. (2012). Spatially defined modulation of skin temperature and hand ownership of both hands in patients with unilateral complex regional pain syndrome. *Brain: A Journal of Neurology*, 135(Pt 12):3676–3686.

Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., and Spence, C. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proceedings of the National Academy of Sciences of the United States of America*, 105(35):13169–13173.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692.

Navarra, J., Soto-Faraco, S., and Spence, C. (2007). Adaptation to audiotactile asynchrony.

*Neuroscience Letters*, 413(1):72–76.

Noel, J.-P., Pfeiffer, C., Blanke, O., and Serino, A. (2015). Peripersonal space as the space of the bodily self. *Cognition*, 144:49–57.

Ocklenburg, S., Peterburs, J., Ruther, N., and Gunturkun, O. (2012). The rubber hand illusion modulates pseudoneglect. *Neuroscience Letters*, 523(2):158–161.

O'Doherty, J. E., Lebedev, M. A., Ifft, P. J., Zhuang, K. Z., Shokur, S., Bleuler, H., and Nicolelis, M. A. L. (2011). Active tactile exploration using a brain-machine-brain interface. *Nature*, 479(7372):228–231.

Paillard, J. and Brouchon, M. (1968). Active and passive movements in the calibration of position sense. *The neuropsychology of spatially oriented behavior*, 11:37–55.

Parise, C. V., Spence, C., and Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Current biology: CB*, 22(1):46–49.

Peters, M. A. K. (2014). *Hierarchical Bayesian Causal Inference and Natural Statistics Explain Heaviness Perception*. University of California, Los Angeles.

Petkova, V. I. and Ehrsson, H. H. (2008). If I Were You: Perceptual Illusion of Body Swapping. *PLoS ONE*, 3(12).

Press, C., Heyes, C., Haggard, P., and Eimer, M. (2008). Visuotactile learning and body representation: An ERP study with rubber hands and rubber objects. *Journal of cognitive neuroscience*, 20(2):312–323.

Ramachandran, V. S., Rogers-Ramachandran, D., and Cobb, S. (1995). Touching the phantom limb. *Nature*, 377(6549):489–490.

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(3):869–875.

Renzi, C., Bruns, P., Heise, K.-F., Zimerman, M., Feldheim, J.-F., Hummel, F. C., and Roder, B. (2013). Spatial Remapping in the Audio-tactile Ventriloquism Effect: A TMS Investigation on the Role of the Ventral Intraparietal Area. *Journal of Cognitive Neuroscience*, 25(5):790–801.

Reuschel, J., Drewing, K., Henriques, D. Y. P., Rosler, F., and Fiehler, K. (2010). Optimal integration of visual and proprioceptive movement information for the perception of trajectory geometry. *Experimental Brain Research*, 201(4):853–862.

Revonsuo, A. (1999). Binding and the Phenomenal Unity of Consciousness. *Consciousness and Cognition*, 8(2):173–185.

Rincon-Gonzalez, L., Buneo, C. A., and Helms Tillery, S. I. (2011). The Proprioceptive Map of the Arm Is Systematic and Stable, but Idiosyncratic. *PLoS ONE*, 6(11).

Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V. (1997). The Space Around Us. *Science*, 277(5323):190–191.

Rohde, M., Di Luca, M., and Ernst, M. O. (2011). The Rubber Hand Illusion: Feeling of Ownership and Proprioceptive Drift Do Not Go Hand in Hand. *PLoS ONE*, 6(6).

Rohe, T. and Noppeney, U. (2015). Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLoS Biol*, 13(2):e1002073.

Rosenthal, O., Shimojo, S., and Shams, L. (2009). Sound-Induced Flash Illusion is Resistant to Feedback Training. *Brain Topography*, 21(3-4):185–192.

Ross, H. and Murray, D. (1978). *EH Weber: The sense of touch*. Academic Pr.

Samad, M., Chung, A. J., and Shams, L. (2015). Perception of Body Ownership Is Driven by Bayesian Sensory Inference. *PLoS ONE*, 10(2):e0117178.

Samad, M. and Shams, L. (2016). Visual-Somatotopic Interactions in Spatial Perception. *Neuroreport*, (27):180–185.

Sanchez-Vives, M. V., Spanlang, B., Frisoli, A., Bergamasco, M., and Slater, M. (2010). Virtual Hand Illusion Induced by Visuomotor Correlations. *PLoS ONE*, 5(4).

Serwe, S., Drewing, K., and Trommershauser, J. (2009). Combination of noisy directional visual and proprioceptive information. *Journal of Vision*, 9(5):28–28.

Seth, A. K., Suzuki, K., and Critchley, H. D. (2012). An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology*, 2.

Shams, L. and Beierholm, U. (2011). From Integration to Segregation: When and How the Human Nervous System Combines Crossmodal Sensory Signals. In *Sensory Cue Integration*, Computational Neuroscience. Oxford University Press.

Shams, L. and Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, 14(9):425–432.

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

Shams, L., Kamitani, Y., and Shimojo, S. (2002). Visual illusion induced by sound. *Brain Research. Cognitive Brain Research*, 14(1):147–152.

Shams, L., Ma, W. J., and Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, 16(17):1923–1927.

Shimada, S., Fukuda, K., and Hiraki, K. (2009). Rubber Hand Illusion under Delayed Visual Feedback. *PLoS ONE*, 4(7).

Sieben, K., Roder, B., and Hanganu-Opatz, I. L. (2013). Oscillatory Entrainment of Primary Somatosensory Cortex Encodes Visual Control of Tactile Processing. *The Journal of Neuroscience*, 33(13):5736–5749.

Smeets, J. B. J., Dobbelsteen, J. J. v. d., Grave, D. D. J. d., Beers, R. J. v., and Brenner, E. (2006). Sensory integration does not lead to sensory calibration. *Proceedings of the National Academy of Sciences*, 103(49):18781–18786.

Snijders, H. J., Holmes, N. P., and Spence, C. (2007). Direction-dependent integration of vision and proprioception in reaching under the influence of the mirror illusion. *Neuropsychologia*, 45(3):496–505.

Spence, C., Pavani, F., Maravita, A., and Holmes, N. (2004). Multisensory contributions to the 3-D representation of visuotactile peripersonal space in humans: evidence from the crossmodal congruency task. *Journal of Physiology-Paris*, 98(1-3):171–189.

Staub, E., Tursky, B., and Schwartz, G. E. (1971). Self-control and predictability: Their effects on reactions to aversive stimulation. *Journal of Personality and Social Psychology*, 18(2):157–162.

Stein, B. E. and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience*, 9(4):255–266.

Suzuki, K., Garfinkel, S. N., Critchley, H. D., and Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia*, 51(13):2909–2917.

Tajima, D., Mizuno, T., Kume, Y., and Yoshida, T. (2015). The mirror illusion: does proprioceptive drift go hand in hand with sense of agency? *Frontiers in Psychology*, 6.

Taylor-Clarke, M., Kennett, S., and Haggard, P. (2002). Vision Modulates Somatosensory Cortical Processing. *Current Biology*, 12(3):233–236.

Tipper, S. P., Lloyd, D., Shorland, B., Dancer, C., Howard, L. A., and McGlone, F. (1998). Vision influences tactile perception without proprioceptive orienting. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*, 9(8):1741–1744.

Tsakiris, M. (2010). My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia*, 48(3):703–712.

Tsakiris, M. and Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology. Human Perception and Performance*, 31(1):80–91.

Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P., and Fink, G. R. (2007). Neural signatures of body ownership: a sensory network for bodily self-consciousness. *Cerebral Cortex (New York, N.Y.: 1991)*, 17(10):2235–2244.

Tsakiris, M., Jimenez, A. T., and Costantini, M. (2011). Just a heartbeat away from one's body: interoceptive sensitivity predicts malleability of body-representations. *Proceedings of the Royal Society B: Biological Sciences*, 278(1717):2470–2476.

Tsakiris, M., Prabhu, G., and Haggard, P. (2006). Having a body versus moving your body: How agency structures body-ownership. *Consciousness and Cognition*, 15(2):423–432.

van Beers, R. J., Sittig, A. C., and Denier van der Gon, J. J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, 122(4):367–377.

van Beers, R. J., Wolpert, D. M., and Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Current biology: CB*, 12(10):834–837.

van der Hoort, B., Guterstam, A., and Ehrsson, H. H. (2011). Being Barbie: The Size of One's Own Body Determines the Perceived Size of the World. *PLoS ONE*, 6(5).

Violentyev, A., Shimojo, S., and Shams, L. (2005). Touch-induced visual illusion. *Neuroreport*, 16(10):1107–1110.

Vroomen, J., Keetels, M., de Gelder, B., and Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Research. Cognitive Brain Research*, 22(1):32–35.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., and

Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2):252–258.

Walsh, L. D., Moseley, G. L., Taylor, J. L., and Gandevia, S. C. (2011). Proprioceptive signals contribute to the sense of body ownership. *The Journal of Physiology*, 589(Pt 12):3009–3021.

Wann, J. P. and Ibrahim, S. F. (1992). Does limb proprioception drift? *Experimental Brain Research*, 91(1):162–166.

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6):598–604.

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

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636.

Wozny, D. R., Beierholm, U. R., and Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision*, 8(3):24–24.

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

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

Wozny, D. R. and Shams, L. (2011b). Recalibration of auditory space following milliseconds of crossmodal discrepancy. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(12):4607–4612.