HAPTIC SIZE DISCRIMINATION

An Investigation of Haptic Size Discrimination and Cue Combination

By KEON S. ALLEN, B.Sc.

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University

Doctor of Philosophy (2022)

Hamilton, Ontario (Department of Psychology, Neuroscience & Behaviour)

TITLE: An Investigation of Haptic Size Discrimination and Cue Combination

AUTHOR: Keon S. Allen (McMaster University)

SUPERVISOR: Dr. Daniel Goldreich

NUMBER OF PAGES: xv, 80

Lay Abstract

The sense of touch is understudied compared to the senses of sight and hearing. But simply reaching for a coin without looking involves complex calculations and decision-making. We studied how the brain may approach tasks like this. We were interested in how well the brain deals with multiple sources of information that do not always agree with each other. We investigated these questions in computer simulations and in experiments with undergraduate participants. Using carefully designed 3D-printed discs, we tested dozens of participants across 3 different experiments. Our results show that humans may use information in the best possible way and applications relevant to VR and robot-assisted surgery.

Abstract

Perception relies on the integration of numerous noisy inputs (cues). Cue combination has been relatively understudied in somatosensation, compared to vision and audition. Here, we investigated whether haptic cutaneous and hand configuration cues are combined optimally to discriminate between coin-sized discs of different sizes. When the hand is open such that the thumb and index fingers span the diameter of a disc to contact its perimeter, cutaneous cues occur from the indentation of the skin caused by the curvature of the disc (smaller discs cause greater indentation). Simultaneously, the hand configuration cue (relating to the perceived distance between fingers), provides an additional cue to size. These cues may vary in their reliability. In three experiments involving 34 participants, we measured these cues and considered three hypotheses for how humans may use them: humans rely solely on the least noisy cue (Winner-Take-All Model), humans combine cues based on a simple average (Average-Measurement Model), or humans combine cues via an optimal weighted average (Optimally-Weighted Model). Each experiment tested participants using a two-interval forced-choice (2IFC) paradigm with 3D printed disc stimuli. On each trial, under occluded vision, participants felt two discs sequentially and responded which felt larger. Participants were tested with each finger's cutaneous cue alone, the hand configuration cue alone, and all cues together. In two experiments, the presented discs were both circular. In a third experiment, unknown to participants, some of the presented discs were oval-like cue conflict stimuli. Participant performance was compared to predictions of the cue combination models. We conclude that humans

may combine haptic cutaneous and hand configuration cues optimally to judge the size of held objects.

Acknowledgements

I would like to express my deepest appreciation for my project advisor and mentor, Dr. Daniel Goldreich, without whom this undertaking could not have been possible.

I am extremely grateful for my committee members: Dr. Patrick Bennett and Dr. Aimee Nelson for their support and guidance.

Many thanks to the talented lab members who have inspired and helped me along my journey, with special thanks to Dr. Jonathan Tong, Dr. Lux Li, and Ky Gauder. I also had the pleasure of collaborating with and supervising many brilliant undergraduate thesis students: Bradley Karat, Maraam Haque, Khesraw Mohibi, and Arjun Patel.

A special thank you to my spouse, Keara Allen, for supporting me through all the ups and downs of my graduate career.

I am also extremely grateful to my mother and father, whose sacrifices allowed me to get this far. Thanks should also go to my sister and her family; my niece and nephew could always put a smile on my face.

Lastly, I would like to thank all my family and friends who have motivated and believed in me.

Contents

List of Figures

List of Tables

Abbreviations

2AFC *two-alternative forced-choice*

2IFC *two-interval forced-choice*

2PD *two-point discrimination*

2POD *two-point orientation discrimination*

AVG *Average-Measurement*

CDF *cumulative distribution function*

Comb. *Combined Cues*

Config. *Configuration Cue*

Cut.1 *Cutaneous Cue 1*

Cut.2 *Cutaneous Cue 2*

fMRI *functional magnetic resonance imaging*

LOI *letter of information*

OPT *Optimally-Weighted*

PC *Pacinian corpuscle*

PDF *probability distribution function*

PSE *Point of Subjective Equality*

RA *rapidly adapting*

ROC *receiver operating characteristic*

SA1 *slowly adapting type 1*

SA2 *slowly adapting type 2*

SDT *signal detection theory*

VPL *ventral posterolateral nucleus*

WTA *Winner-Take-All*

Declaration of Authorship

I, Keon S. Allen, declare that this thesis titled, "An Investigation of Haptic Size Discrimination and Cue Combination" and the work presented in it are my own. The work was completed with input from my supervisor and committee members.

Any exceptions are itemized here:

- Chapter 3: Undergraduate thesis students Bradley Karat and Maraam Haque assisted with conducting experiments. Photographs taken with Bradley Karat's help; his hand was also depicted in Figure 3.2.
- Chapter 4: Undergraduate thesis students Arjun Patel and Khesraw Mohibi assisted with conducting the experiment.

Chapter 1

Introduction

1.1 An Introduction to Somatic Perception

For humans, touch is the first of our senses to develop (Gallace & Spence 2010; Gottlieb 1971). Touch facilitates our interactions with the world and is critical for manipulating, interacting with, and experiencing our environment. Anything felt must be translated from interactions with the skin to neural impulses for the brain to decode.

1.1.1 Neuroanatomy and Neurophysiology

Four types of mechanoreceptors respond to different interactions upon glabrous (non-hairy) skin, sending signals to the brain through afferent nerve fibres. Merkel cell receptors respond to curvature, static indentations on the skin, and slow-moving stimuli. They are innervated by *slowly adapting type 1* (*SA1*) afferents. Ruffini corpuscles

are sensitive to the stretching of skin and are innervated by *slowly adapting type 2* (*SA2*) afferents. Meissner corpuscles are sensitive to low-frequency vibrations and are innervated by *rapidly adapting* (*RA*) afferents. *Pacinian corpuscle*s (*PC*s) are sensitive to high-frequency vibrations and are innervated by PC afferents (Bensmaia & Manfredi 2012; Yau et al. 2016). In regular interactions with the environment, all of these mechanorecptors contribute neural responses to some degree (Saal & Bensmaia 2014; Yau et al. 2016).

As the skin itself is a uniquely flexible and deformable organ which we move in order to feel during tactile exploration, the brain utilizes proprioceptive information, which relates to the configuration of limbs and digits in space. Muscle spindles, innervated by group Ia and II nerve fibres, respond to muscle stretch associated with joint movement. Golgi tendon organs, innervated by group Ib nerve fibres, respond to muscle tension (Fallon & Macefield 2007).

The touch and proprioceptive signals branch in the brainstem and thalamus along the way to the primary somatosensory cortex (Brodmann areas 3a, 3b, 1, and 2). Neurons in 3b primarily receive thalamic inputs from core regions of the *ventral posterolateral nucleus* (*VPL*) and respond best to cutaneous stimulation (Yau et al. 2016). Neurons in area 3a receive their inputs from the region surrounding the VPL and respond primarily to proprioceptive stimulation (Yau et al. 2016). However, this is not a strict rule. The thalamus is a complex structure in itself with many branching pathways at multiple levels (Guillery & Sherman 2002). While neurons further

along the pathway receive more convergent information, there is substantial horizontal interconnectivity as well (Buonomano & Merzenich 1998). As such, neurons in S1 can receive information from both cutaneous and proprioceptive sources (Yau et al. 2016). Ultimately, through this complex system of neuronal impulses and pathways, environmental stimuli relate to subjective perception.

1.1.2 Perception: Representation and Links

The brain is constantly, relentlessly bombarded by information. Perhaps luckily, humans are not always conscious of this. The feeling of clothes on the body, the weight of the limbs, and the effort of breathing are generally only apparent when attention is brought towards them. Nevertheless, the brain deals with the influx and endeavours to represent the world. The world we perceive, however, is not necessarily the world as it is. Various illusions across all sensory modalities, including touch, speak to that end (Lederman & Jones 2011). Cortical neurons can only represent the world based on their neural inputs. Palmer (1978) terms this representation the mental world. Thus, the 'real world' is a represented world and the mental world is a representing world which should reflect certain qualities of the represented world. When studying the brain, researchers often create representations of the mental world. Effectively, this becomes a representation of a representation of the represented world. In so doing, one should be clear about what aspects of the representation are being modeled, how the representations relate to each other, and what assumptions are being made.

When studying perception, researchers seek to examine the relationship between

stimuli, the resulting neural activation, and the perception itself. The idea that phenomenal states (perception) are lawfully related to physiological states (neural activation) is itself an assumption. This leads us to the concept of linking propositions. An important linking proposition, termed the Identity proposition states that statistically identical neural signals lead to statistically identical sensations (Teller 1984). The Contrapositive Identity proposition is essentially following the modus tollens argument of the Identity proposition. Thus, sensations that are not statistically identical imply physiological states that are not statistically identical (Teller 1984). These propositions are particularly relevant in perception research that involves detection or discrimination.

1.2 Frameworks for Investigating Detection and Discrimination

1.2.1 Signal Detection Theory

Modern *signal detection theory* (*SDT*) arose in the 1950s to analyze noisy data from radar readings. Nearly simultaneously, the theory emerged in the field of psychophysics as well, building on earlier contributions from researchers including Fechner and Gauss (Wixted 2020). Fechner, who was prominent in establishing the field of psychophysics, hypothesized that internal measurements of sensations existed and were affected by the same type of measurement error that Gauss described in physical measurements (Link 1994). According to *SDT*, when an experimenter delivers

a stimulus, the participant obtains sensory evidence—a measurement, which comes from the resulting state of their nervous system (Ehrenstein & Ehrenstein 1999; Gold & Shadlen 2007). Depending, for example, on whether a stimulus is present or absent, the magnitude of the measurement may differ. Due to noise in the nervous system, a repeated stimulus may evoke a different measurement each time. This measurement can be described as a random variable whose mean depends on whether a stimulus is present or absent (Gold & Shadlen 2007). Thus, the signal-plus-noise distribution can be described by a Gaussian distribution centred on the true signal value when the stimulus is present. Because of noise, when the stimulus is absent, there could still be evidence that the stimulus is present. Thus, the noise distribution can be described by a Gaussian distribution centred on 0 (no signal), when the stimulus is absent. The overlap of these distributions depend on the sensitivity of the participant and the difficulty of the task. To quantify the overlap, d' is a measure of how far apart the peaks are, relative to their spread (Ehrenstein & Ehrenstein 1999).

$$d' = \frac{\mu_{signal} - \mu_{noise}}{\sigma} \tag{1.1}$$

SDT assumes that a participant also has an implicit criterion: a measurement beyond which the participant reports the stimulus is present. Where this criterion is placed affects how often the participant will correctly detect the stimulus when it's present (hit), incorrectly detect the stimulus when it's absent (false alarm), correctly detect no stimulus when it's absent (correct rejection), and incorrectly detect no stimulus when it's present (miss). By plotting the hit rate together with the false

alarm rate, researchers can plot *receiver operating characteristic* (*ROC*) curves for the participant. This is useful to evaluate the sensitivity of the participant (d') because these curves are "isosensitivity functions". Different hit and false alarm rates can lead to the same curve and different curves correspond to different sensitivities. Alternatively, instead of measuring whether a stimulus is present or absent, and essentially comparing the noise distribution to the signal distribution, other tasks offer different advantages. A *two-alternative forced-choice* (*2AFC*) or *two-interval forced-choice* (*2IFC*) experiment allows participants to compare two measurements. There is no need to set an individual criterion, where the participant must subjectively have enough evidence to justify responding positively (stimulus present); the participant always responds positively in *2IFC* tasks (Ehrenstein & Ehrenstein 1999). *2IFC* testing has shown participants can differentiate between smaller stimulus intensities than would be captured by unforced testing (Ehrenstein & Ehrenstein 1999). Differences in sensation are also more easily detected than verbalizing absolute magnitudes (Luce & Krumhansl 1988; Wixted 2020).

1.2.2 Bayesian Inference

Bayes' Theorem

Bayes' theorem describes how to mathematically utilize conditional probabilities to reach some conclusion. It involves three key elements: priors, likelihoods, and posteriors. The principles of Bayes' theorem can be used to make inferences in the presence of uncertainty in a myriad of applications. The theorem was first written in its discrete

form by Pierre-Simon Laplace (Fienberg 2006; Laplace 1774).

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_{j=1}^{n} P(D|H_j)P(H_j)} \tag{1.2}$$

Equation 1.2 shows a modern form of the theorem. The Hs are the hypotheses. These are events or possibilities that have different probabilities associated with them. The priors, $P(H)$, are the probability of a given hypothesis before considering any data. When data are collected, you can then obtain likelihoods, $P(D|H)$, which correspond to the probability of obtaining those data if the hypothesis were true. Combining these together with Bayes' Theorem, you can ultimately obtain posterior probabilities, $P(H|D)$, for each hypothesis given the data. Combining prior knowledge and new data in this way yields optimal conclusions.

Bayesian Coin Examples

A couple of simple coin examples illustrate these principles. In the first example, there is a large jar filled with 700 fair coins (50% chance to land either heads or tails), and 300 heads-biased coins (specially weighted coins that land heads 80% of the time and tails, 20%). A randomly selected coin is chosen from the jar and flipped 20 times. If it landed on heads 15 times, which of the two coin-types was it?

$$H_1 = Fair, H_2 = Biased$$
$$P(H_1) = 0.7, P(H_2) = 0.3$$
$$P(D|H_1) = \binom{20}{15}0.5^{15} * 0.5^5, P(D|H_2) = \binom{20}{15}0.8^{15} * 0.2^5$$

$$P(H_1|D) = \frac{\binom{20}{15}(0.5^{15} * 0.5^5)(0.7)}{\binom{20}{15}(0.8^{15} * 0.2^5)(0.3) + \binom{20}{15}(0.5^{15} * 0.5^5)(0.7)} = 16.5\%$$
$$P(H_2|D) = \frac{\binom{20}{15}(0.8^{15} * 0.2^5)(0.3)}{\binom{20}{15}(0.8^{15} * 0.2^5)(0.3) + \binom{20}{15}(0.5^{15} * 0.5^5)(0.7)} = 83.5\%$$

Even though the biased coin was far less likely to be selected from the jar, there were enough data points (flips) to conclude that the selected coin was most likely to be biased. This is the optimal conclusion for maximizing correct responses.

What if we consider another extreme, however, with a uniform prior and limited data? There are now 500 coins in the jar which have a 90% chance of landing on heads and another 500 coins which have a 95% chance of landing heads. We're only going to flip a coin once, attempt to classify it, and return it to the jar. Every time a coin lands on heads, Bayesian inference would classify it as a 95-biased coin. Whenever a coin lands on tails, it will be classified as a 90-biased coin. As a result, many 90-biased coins will be misclassified as 95-biased coins. However, given the limited data available, this is still the optimal procedure in order to be correct most often. If we flip 1000 different coins from the jar, we would expect to encounter 500 90-biased coins. 450 of those would be expected to land on heads, and be misclassified as 95-biased coins. The remaining 50 would be correctly classified as 90-biased coins. We would also expect to encounter 500 95-biased coins. 475 of those would be expected to land on heads and be correctly classified as 95-biased coins. The remaining 25 would be misclassified as 90-biased coins. However, in the long run, we would expect an overall correct classification of $\frac{50+475}{1000} = 52.5\%$ which is better than chance. We still glean useful, optimal information in a simple, contrived, yet potentially counter-intuitive example.

As investigators, we're not only interested in applying Bayes' theorem to jars of coins. We use these methods to run simulations and analyze data. Rather than having a couple of hypotheses about coin-types, we can have thousands of hypotheses over different empirical parameters of interest. Further, many researchers have found the Bayesian framework to be useful in summarizing how the brain handles uncertainty and environmental statistics.

Priors and Perception

Humans are generally used to light shining more or less directly from above, whether it's artificial indoor light or noon sun. Adams et al. (2004) studied this visual light-from-above prior. Participants were shown bump-dimple visual stimuli with shadow details consistent with overhead light. Prior to training, participant responses were consistent with an overhead light prior. Participants were then trained with shifted light and given haptic feedback as to whether the stimuli were bumps or dimples. When shown the visual-only stimuli again, their responses were consistent with shifted light prior. This shift generalized to a novel task, which lends evidence against an alternative hypothesis that the change was a cognitive strategy. Adams et al. (2004) argue that the light-from-above prior can be modified, temporarily with training, and the brain updates its priors according to environmental data.

The rabbit illusion is a tactile illusion where a series of taps feel closer together as the time between them decreases (Geldard & Sherrick 1972). Goldreich and Tong (2013) explain the illusion as resulting from a low-speed prior on the skin. In daily

life, humans are used to slow-moving stimuli on the skin. Clothing may shift and slide along our skin, but generally without notable speed. Quick-moving stimuli, like those that produce the rabbit illusion, would be less likely. The brain, used to slow stimuli and not confident in the location of the stimulus (due to low acuity or weak taps), could infer that perhaps the stimuli weren't moving so quickly and instead the stimuli were closer together.

1.3 Empirically Exploring Perception

1.3.1 Investigating Perception in Cue Combination

Cue Combination and Cue Conflict

Generally, cue combination research has focused on multimodal cues between various senses and unimodal cues outside of haptics. Alais and Burr (2004) demonstrate optimal cue combination across auditory and visual stimuli. When localizing a spatial source by a beep and a flash, participants were better with both cues than with either alone. Even though vision was typically more reliable, audition provided useful information that was also integrated. Alais and Burr (2004) also experimentally manipulated the flash by degrading it and reducing its reliability. On some trials, they introduced cue conflicts where the beep and flash were displaced in opposite directions. Participants judged the location of the event to be closer to where the flash was than where the beep was. When the flash was degraded, however, participant localizations were pushed closer to where the beep was. Experimenters were able to

show that participants rely more on whichever of the two cues was more reliable. Hillis et al. (2004) demonstrate optimality for two visual cues to depth: texture and disparity. Many studies in these domains find that the reliability of cues is considered during the process of perception (Alais & Burr 2004; Buckley & Frisby 1993; Hillis et al. 2004; Jacobs 1999; Negen et al. 2018; van Beers et al. 2002).

Some studies have shown that a single cue can dominate perception (Colavita 1974; Girshick & Banks 2009; Hecht & Reiner 2009; Shams et al. 2000). Shams et al. (2000) found a scenario of auditory dominance. When a light, flashed once, is delivered with multiple auditory beeps, the light is perceived as having flashed multiple times. The *Colavita Effect* refers to a scenario of perceptual visual dominance. Participants were asked to press either a 'light key' or 'tone key' to indicate whether a light or tone was presented. On trials when both stimuli were presented simultaneously, participants tended to select the 'light key' even when controlling for different stimuli intensity and verbal instructions (Colavita 1974).

1.3.2 Investigating Perception in Haptics

Studies of touch can be organized into those that investigate affective touch (pertaining to emotion, eg. a caress), discriminative touch (for decision-making, eg. orientation discrimination), passive touch (without participant movement, eg. the Rabbit Illusion), active touch (with non-manipulative movement, eg. Braille reading), and haptics (active touch with grasping, manipulative movement, eg. stereognosis). We

imagine the literature surrounding somatosensation as spanning two orthogonal continua. One being a discriminative–affective continuum, moving from *what* you feel to *how* you feel. The other, relating to the degrees of freedom allowed in the study, a passive–active–haptic continuum. These continua can lead to many different types of sensory tasks (Table 1.1).

Table 1.1: Examples of Tactile Task Categorization

	Passive	**Active**	**Haptic**
Affective	Palm/forearm brushing (Löken et al. 2011)	Surface texture pleasantness (Chen et al. 2009)	Petting a dog
Discriminative	Tap localization	Braille reading	Stereognosis

Affective Touch

Touch can be used to explore, interact, and communicate. Affective touch relates to the emotion involved in touch. For example, many studies delve into how touch is perceived according to its speed. Löken et al. (2009) found that C-tactile fibres (unmyelinated, slow-conducting afferents innervating hairy skin) respond best to stimuli that move at intermediate speeds, not too fast nor too slow, within 1–10cm/s. This intermediate speed was also reported to be the most pleasant by participants. Björnsdotter and Olausson (2011) followed up on this study with a *functional magnetic resonance imaging* (*fMRI*) study. They found affective touch at speeds of 3cm/s to be more pleasant than at 30cm/s and found increased activation in the posterior insula, a cortical destination for C-tactile fibres. von Mohr et al. (2017), studying attachment and ostracism, show that slow, affective touch can reduce negative feelings of social

exclusion. Investigations of affective touch help us understand the emotional role of touch and the neurophysiology underlying it. In this thesis, however, we focus more on discriminative touch.

Passive Discriminative Touch

An example of passive discriminative touch is the *two-point orientation discrimination* (*2POD*) task which can be used to assess tactile spatial acuity and has medical applications (Tong et al. 2013). Traditionally, physicians who need to assess patient tactile acuity, which can be used as an indicator of serious nerve injuries, relied on *two-point discrimination* (*2PD*). The patient would be tapped with either one prong or two at progressively decreasing separations to find the threshold where they could no longer tell the difference. It's a useful task which could even be accomplished with a paper clip (Finnell et al. 2004). *2PD*, however, benefitted from a non-spatial cue that could inflate performance, making patients seem to have greater acuity than they should (Craig & Johnson 2000; Johnson & Phillips 1981; Tong et al. 2013). Tapping with one prong seems to result in stronger mechanoreceptor activation than with two, even at zero separation (Vega-Bermudez & Johnson 1999). By always using two prongs and switching orientation (vertical to horizontal) instead, this issue is avoided and the test should be more reliable. Investigations into passive discriminative touch provide insight into how the nervous system responds when forces are applied to it, and when humans must then make decisions.

Active Discriminative Touch

We define active touch as generally characterized by self-directed motion of the skin surface with low degrees of freedom. For example, Lederman and Taylor (1972) investigated the perception of roughness in active touch. Participants' arms rested on a padded swivel where they were allowed to move their fingertips horizontally, but not vertically, across grooved plates. A clever fulcrum design also allowed the experimenters to control the force participants used while scanning. Lederman and Taylor (1972) found that groove width was most important, but perceived roughness also increased with force. Since roughness perception can depend on the force applied, and the scanning speed, Yoshioka et al. (2011) examined how, given a rough surface, a constant perception of roughness arises from fluctuating forces and speeds. They found that the cutaneous input of the surface on the skin allowed for the perception of roughness, but proprioception from the motion of the hand itself was necessary for the perception of roughness *constancy*. Active touch literature explores areas of somatosensation that have some overlap with motion and proprioception while maintaining substantial control.

Haptic Discriminative Touch

Haptic touch is characterized by cutaneous stimulation, proprioception, and free motion. In the study of haptics, Lederman and Klatzky (1987) have shown that there are common techniques that people use to discern haptic cues for object identification. For example, when given an object to recognize, people first tend to grasp the object. This technique, termed "enclosure", allows for the extraction of structural

information including the overall 'global shape' of the object. Other techniques like "contour following", "pressure" and "static contact", allow more information to be extracted. These techniques are involved with manual stereognosis: the ability to use hands to identify objects without vision or audition, which is routinely assessed in neurological patients (Reidy et al. 2016). Reidy et al. (2016) examined sterognosis in older (over 65 years) and younger (under 30 years) adults with everyday items and euro coins. They were interested in how older adults, without apparent neurological impairment, fare against similarly healthy younger adults. There was no difference in older and younger adults in their ability to discriminate between object types like buttons, keys, coins, and rings. Older adults, however, were significantly worse at identifying individual euro coin denominations by touch alone. The study of haptics favours high degrees of freedom where the hand may freely move and explore stimuli.

The Questions at Hand

Our investigations sat firmly on the discriminative line, somewhere between active and haptic touch. Participants took part in 'controlled haptic' experiments where they were allowed to move toward, reach, and grasp stimuli in a controlled manner. We were interested in haptic cues relating to the sizes of different coin-sized discs. When feeling for a coin, without looking, to find a dime next to a quarter, the task may seem trivial. Finding a dollar next to a two-dollar coin may be more difficult, but still doable. When holding a coin between the index finger and thumb, there is a point of contact on each finger. These points of contact relate to the size of the coin through the curvature of the coin. The greater the curvature, the greater the indentation,

and the smaller the coin. These are cutaneous cues on the thumb, *Cutaneous Cue 1* (*Cut.1*), and index finger, *Cutaneous Cue 2* (*Cut.2*). Simultaneously, a proprioceptive cue relating to where your fingers are in space relates to the size of the coin as well. The further apart the fingers of the hand are configured around the coin, the larger the coin must be. This *Configuration Cue* (*Config.*) along with *Cut.1* and *Cut.2* may be combined in order to perceive the overall size of the coin. Do humans combine these cues optimally or in some other fashion? Upon which cues, if any, do humans most rely? We investigated three hypotheses for how these cues may be combined through a series of model simulations and three experiments.

Chapter 2

Simulations and Modelling

2.1 Three Hypothesized Models of Cue Combination

This thesis explores predictions of three hypothesized models of cue combination. If the brain has some internal understanding of its own sensory variability, as previous research in other modalities suggests, how might that information be used in a disc comparison task with three cues? Interactions with physical stimuli evoke afferent action potentials holding information, from various sources, that relate to physical properties of the stimuli (similar physiological states lead to similar phenomenal states). In the disc example, each point of contact is a source of information relating to object size, each of which gives a cutaneous cue to size (*Cutaneous Cue 1* (*Cut.1*) and *Cutaneous Cue 2* (*Cut.2*)). The configuration of the hand itself is a source of information which gives rise to a configuration cue to size (*Configuration Cue*

(*Config.*)). These cues are internal measurements relayed through afferents. These disc size measurements could be described by any relevant unit; we used mm radius. We denote the stimulus by s, the overall percept resulting from the cues by $\hat{s}$, the cue measurements on a given trial by x, and the standard deviation of the measurement or percept by σ. Perhaps the brain employs a strategy where the most reliable cue is fully relied upon in a form of sensory dominance. We refer to this possibility as the *Winner-Take-All* (*WTA*) model.

WTA: Humans rely solely on the cue with the highest reliability.

$$\hat{s} = x_{best} \tag{2.1}$$

$$\sigma_{\hat{s}} = \sigma_{smallest} \tag{2.2}$$

With such a model, the percept $\hat{s}$ is equivalent to the measurement of the most reliable of the three cues. Similarly, the sigma of the combined estimate is equivalent to the reliability of the most reliable cue.

Alternatively, perhaps the brain is unaware of its own reliability or does not use that information and instead favours a quick, simple average of the cues. We refer to this possibility as the *Average-Measurement* (*AVG*) model.

AVG: Humans combine cues based on an unweighted average of their cues.

$$\hat{s} = \frac{x_{Cut.1} + x_{Cut.2} + x_{Config.}}{3} \tag{2.3}$$

$$\sigma_{\hat{s}} = \frac{1}{3}\sqrt{\sigma^2_{Cut.1} + \sigma^2_{Cut.2} + \sigma^2_{Config.}} \tag{2.4}$$

With this model, the percept is the simple average of cues. The equation for the combined sigma follows from the percept formula (Appendix A1).

Perhaps, instead, the brain knows the reliability of each cue and optimally uses that information to weight the measurements from each of the individual cues. We refer to this possibility as the *Optimally-Weighted* (OPT) model.

OPT: Humans combine cues optimally, weighting each according to its reliability.

$$\hat{s} = \frac{\frac{x_{Cut.1}}{\sigma^2_{Cut.1}} + \frac{x_{Cut.2}}{\sigma^2_{Cut.2}} + \frac{x_{Config.}}{\sigma^2_{Config.}}}{\frac{1}{\sigma^2_{Cut.1}} + \frac{1}{\sigma^2_{Cut.2}} + \frac{1}{\sigma^2_{Config.}}} \tag{2.5}$$

$$\sigma_{\hat{s}} = \sqrt{\frac{1}{\frac{1}{\sigma^2_{Cut.1}} + \frac{1}{\sigma^2_{Cut.2}} + \frac{1}{\sigma^2_{Config.}}}} \tag{2.6}$$

Under the OPT model, the percept is the weighted average of the cues such that more reliable cues are relied upon more. The combined sigma formula follows from the percept formula (Appendix A2).

2.2 Model Predictions Depend on Relative Sigma Values

To explore the predictions of each model, we present a simple simulation of a *two-interval forced-choice* (*2IFC*) comparison task in which simulated participants are

asked to decide, for example, which of two discs has the larger diameter. Monte Carlo simulations reveal, depending on the relative sigmas of each cue, the models make similar or differing predictions. We used a range of stimulus values (in disc radius) from -5mm to +5mm, in increments of 0.1mm, relative to the reference. These were created using 1000 trials at each comparison level. At each of those comparison levels, the proportion of times the comparison measurement was greater than that of the reference is plotted in Figure 2.1. By sampling from Gaussian distributions centred on the stimulus value, with the sigma corresponding to the reliability of the cue, a cue is obtained. These cues are then combined according to the percept formulas above. On each trial, the comparison percept for each model is compared to the reference percept. Whichever had the larger value on that trial is taken as the response as to which stimulus was perceived to be greater. The proportion of times one stimulus is perceived greater than the other is plotted in the graph to create these psychometric functions for each model.

When one of the cue sigmas is exceptionally good, the *OPT* and *WTA* models make similar predictions (Fig. 2.1A). As the sigma (inverse of reliability) of a cue in the optimal measurement equation approaches zero, becoming more reliable, that cue is increasingly relied upon. In the limit that one cue's sigma approaches zero, the *OPT* model becomes similar to the *WTA* model. When the sigmas of each cue are similar, the *OPT* and *AVG* models give similar predictions (Fig. 2.1B). Notice in the measurement equation for the *OPT* model, what matters most is the relative differences in sigma between the cues. If each cue is weighted the same, this becomes

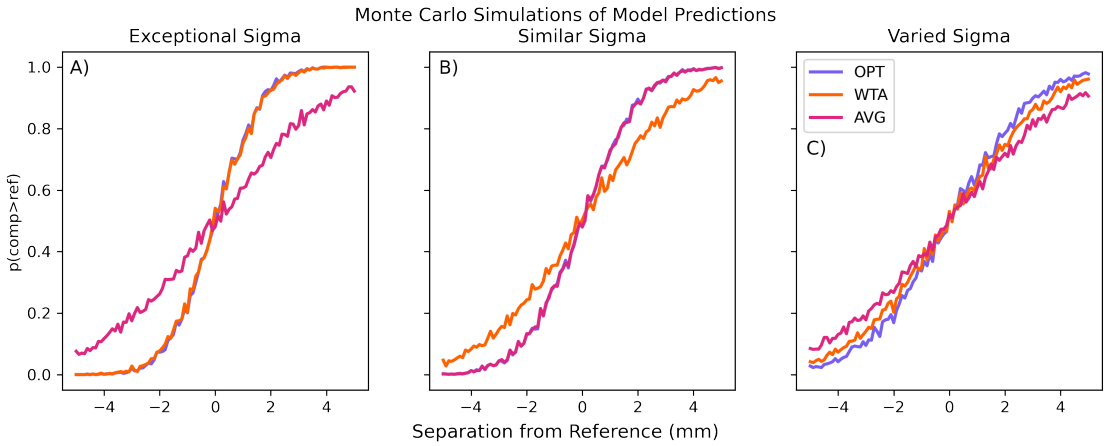

FIGURE 2.1: Monte Carlo simulations of model predictions under different relative sigma values. A) When the sigma of one cue is much smaller than the others, *OPT* and *WTA* models make similar predictions. B) When the sigmas of each cue are similar, *OPT* and *AVG* models make similar predictions. C) When the relative sigmas vary, each model makes different predictions.

equivalent to the *AVG* model. When the cue reliabilities vary, each model makes different predictions (Fig. 2.1C). Thus, it is possible to distinguish between all three hypotheses, depending on the relative sigmas of each cue. This simulation was done, however, by simulating thousands of trials at hundreds of comparison levels. Can we reasonably expect to be able to distinguish the *OPT* model from the WTA and *AVG* models using noisy human participants and a feasible number of trials per comparison level?

2.3 A Simulated Sigma Experiment Reveals Classification Difficulty

Here, we simulate a sigma experiment with thousands of Bayesian-optimal observer participants who will combine cues according to the *OPT* model. The observers will complete a simulated Method of Constant Stimuli (MCS) two-interval forced-choice (2-IFC) experiment. The experiment will be split into four conditions. In each condition, there will be 8 comparison levels with 20 trials per comparison level. The first three conditions will be the comparison task with each cue individually, so we can attempt to work backwards and estimate what the true sigmas of each observer are. The fourth condition will be a Combined Condition: the same comparison task using all three cues together. This will allow us to test to what extent we can then infer which model the observers are using.

To begin, we set a Gaussian distribution for the mean sigma of each cue for the overall population of simulated Bayesian-optimal observer participants (Fig. 2.2A). The sigmas for each cue of any individual observer are sampled from these population-level Gaussians. In turn, each observer, defined by their respective sigmas for each cue, completes the full simulated experiment. On a given trial, each observer compares a reference stimulus with a comparison stimulus. In the Combined Condition, for example, the measurement distribution for each cue is a Gaussian centred on the true radius of the piece, with a standard deviation equal to the sigma a given observer has for that cue (Fig. 2.2B). The three cues to reference size are combined according

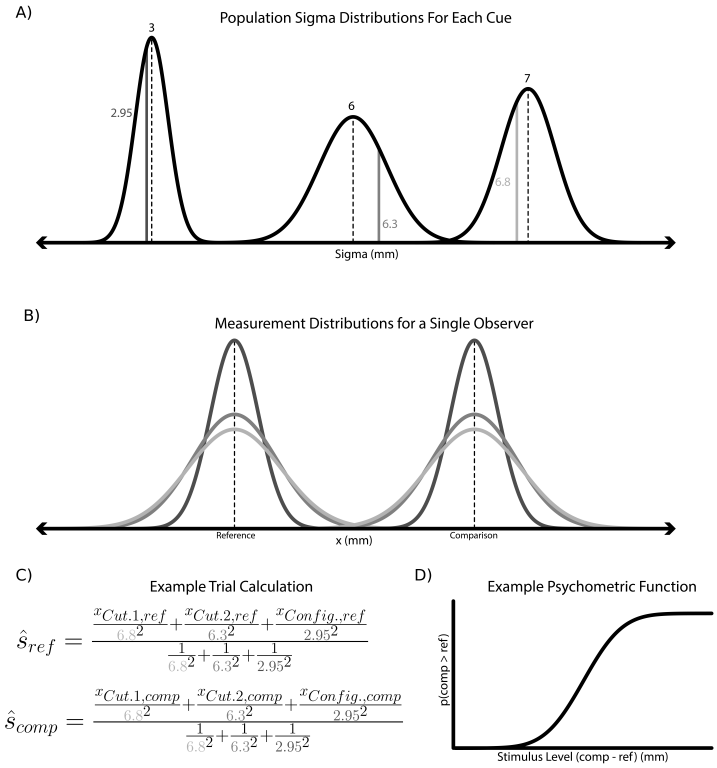

Figure 2.2: Not to scale, for illustrative purposes. A) Example of population-level distributions of possible sigmas for observers. From left to right, *Config.*, *Cut.2*, *Cut.1*. The sigmas for each cue of an observer are sampled from these Gaussians. Depicted is one such sample to define the sigmas of a single observer. The sigmas of all 1000 observers are determined in this way. B) Depiction of measurement distributions for the example observer in a Combined Condition where the mean of the measurement for each cue is centered on the true stimulus value and the sigmas are equivalent to the sigmas of each cue for that observer. Light grey: *Cut.1*, medium grey: *Cut.2*, dark grey: *Config.* C) Example calculation of a trial to see what the observer believes was the greater stimulus. D) The proportion of times $\hat{s}_{comp} > \hat{s}_{ref}$ results in a psychometric function.

to the optimal percept formula as are the three cues for the comparison (Fig. 2.2C). The proportion of times the comparison is perceived greater than the reference leads to a psychometric function (Fig. 2.2D). Data for each observer are analyzed as any real-world participant would be. We apply Bayesian inference to the data in each individual and combined cue condition which results in a Posterior *probability distribution function* (*PDF*) over sigma (Appendix B1). The mode of that *PDF* also defines the best-fit cumulative normal distribution for each psychometric function. By integrating over the sigma *PDF* for each of the individual-cue conditions, we calculate the sigma *cumulative distribution function*s (*CDF*s). By uniformly sampling from the ordinate of the sigma *CDF*s and interpolating to the abscissa, we can sample sigma values from the sigma *PDF*s. This sampling procedure is preferable to using the mode of the sigma *PDF* as a single estimate because the samples collectively reflect the confidence we have in our estimate. A single sample from each of the three sigma *PDF*s, from the three individual-cue conditions in the simulation, yields a triplet. Sampling thousands of these, we combine each triplet according to the three hypothesized model equations. This results in three different combined sigma values—one for each model. Thus, each model defines a different combined-cue psychometric function. We calculate the probability that the data collected in the Combined Condition came from each of the three prospective combined psychometric functions. Across the thousands of sampled triplets, we use Bayesian model comparison to calculate a posterior probability for each model for each observer.

The results across all observers are found in the boxplots (Fig. 2.3). The entire

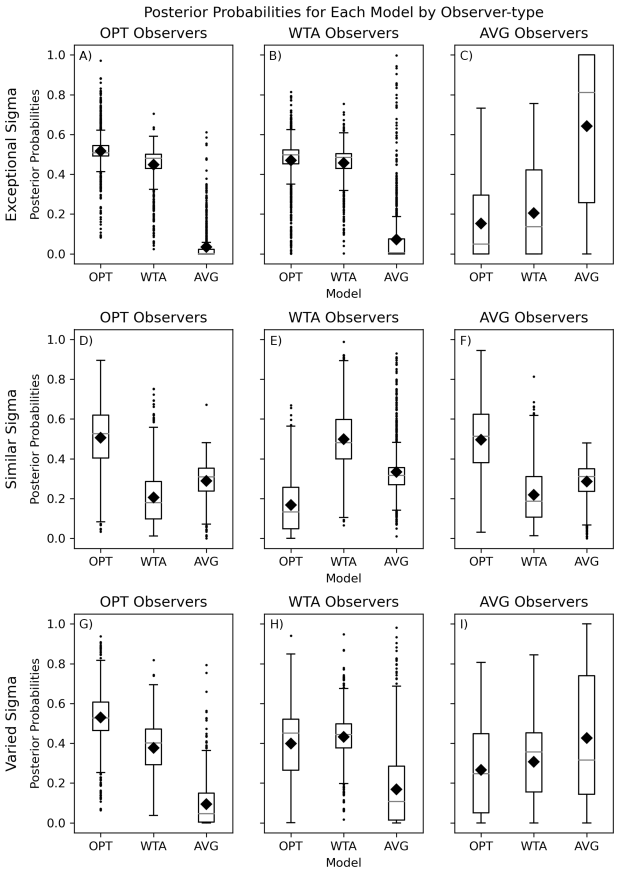

FIGURE 2.3: Boxplot of Posterior Probabilities in the simulated sigma experiment for each Model by Observer-type and Population Parameter. Grey line = median; Black diamond = mean. A-C) Top row. Posterior probabilities for each model by observer-type when participants tend to have one cue with a relatively exceptional sigma. D-F) Middle row. Posterior probabilities for each model by observer-type when participants tend to have similar sigmas. G-I) Bottom row. Posterior probabilities for each model by observer-type when participants tend to have varied sigmas.

simulation was completed with observers combining cues according to each of the models under three different population-level conditions: one sigma was exceptional ($M_{Cut.1} = M_{Cut.2} = 8mm, S_{Cut.1} = S_{Cut.2} = 2mm; M_{Config.} = 2mm, S_{Config.} = 1mm$),

the relative sigmas were similar ($M_{Cut.1} = M_{Cut.2} = M_{Config.} = 4mm; S_{Cut.1} = S_{Cut.2} = S_{Config.} = 0.1mm$), or relative sigmas varied ($M_{Cut.1} = 7mm, S_{Cut.1} = 2mm; M_{Cut.2} = 6mm, S_{Cut.2} = 3mm; M_{Config.} = 3mm, S_{Config.} = 1mm$). Since we know for certain which model each observer is using, these simulations should give a reasonable expectation for what proportion of participants could be correctly or incorrectly classified. The distance between the boxes in the boxplots helps provide a sense of how confidently observers are categorized. From the boxplots in the first column of Fig. 2.3, we see that the classifications are not strongly confident, but are correct the majority of the time when identifying *OPT* observers. Table 2.1 shows an overall classification accuracy across observers of 59.13% with *Exceptional Sigma*, 53.87% with *Similar Sigma*, and 53.30% with *Varied Sigma*. 67-79% of *OPT* observers are correctly identified as such regardless of population scenario. *WTA* and *AVG* observers, however, each have one scenario where more observers are likely to be classified as optimal than as *WTA* or *AVG* respectively.

Table 2.1: Observer Classification in Sigma Simulation

Classification	OPT Observer	WTA Observer	AVG Observer	Population Parameters
OPT Classification	**677**	516	98	Exceptional Sigma
WTA Classification	315	**445**	250	
AVG Classification	8	39	**652**	
OPT Classification	**793**	125	763	Similar Sigma
WTA Classification	139	**758**	172	
AVG Classification	68	117	**65**	
OPT Classification	**740**	438	254	Varied Sigma
WTA Classification	241	**471**	358	
AVG Classification	19	91	**388**	

2.3.1 Non-optimal Observer Classification is Sensitive to Underlying Population Parameters

When evaluating *WTA* observers, we find that, as the first simulation suggested, it's difficult to differentiate between the *OPT* and *WTA* models when one cue is much better than the others. We can see that although the scale often tips towards the *OPT* model, there is a substantial amount of uncertainty shown through the boxplots, as the mean and median posteriors for *OPT* and *WTA* are close to 50% (Fig 2.3B).

As revealed in the first simulation, when the sigmas are similar, the *AVG* observers are hard to distinguish from the *OPT* observers. In the Similar Sigma scenario only, the group of observers classified as *AVG* had the least members.

These simulations reveal that such a task can provide useful information even when distilled to a manageable task with limited trials per comparison. The classifications across observer-type and scenario-type are informative and usually correct, with two caveats. When one cue sigma is much better than the others, *WTA* observers are harder to classify as such. Also, when all the cues have similar sigmas, *AVG* observers are harder to classify as such.

We can move forward acknowledging and appreciating that this inverse problem, figuring out which model most likely produced the data collected, is a difficult one to solve. Considering the trial to trial variability, it should be no surprise that perfectly distinguishing between all three possibilities is not a trivial undertaking. While there are a few misclassifications, the inference is still informative, optimal, and correct most

of the time, across all three models. Consider the coin example from Section 1.2.2. Further, such a task would reveal which of the three population parameter scenarios are likely. This can provide further information to consider the trustworthiness of classifications. We do, however, want to see more confident, correct classifications within each model. By running the sigma experiment with real participants, we can learn some information about the underlying population parameters and which pitfalls to avoid. Additionally, there are other experimental variations, like a cue conflict experiment, which can aid our investigation.

2.4 A Simulated Cue Conflict Experiment Better Differentiates Between Models

By modifying and extending the previous simulation, we simulate a full cue conflict experiment with thousands of Optimal, WTA, and Average-model observers. Using the same range and values for population parameters as above, we extend the simulation into a cue conflict paradigm. To create the conflict, we will set up our cues such that each cue has a different 'true value' and the respective measurement distributions are centered on that value. The measurement distributions for *Cut.1* and *Cut.2* will be centered on the same value some fixed distance from the reference. Simultaneously, *Config.* will be centered on a value separated by the same absolute distance but in the opposite direction. As such, the -3 conflict condition will be defined as the condition where *Config.* corresponds to a value -3mm from the normal (circular, non-conflict) reference and *Cut.1* and *Cut.2* correspond to a +3mm value from reference. For the

-2 conflict condition, *Config.* is -2mm from the reference and *Cut.1* and *Cut.2* are +2mm from the reference. For the +2 conflict condition, *Config.* is +2mm from the reference and *Cut.1* and *Cut.2* are -2mm from the reference. For the +3 conflict condition, *Config.* is +3mm from the reference and *Cut.1* and *Cut.2* are -3mm from the reference. These 'virtual' conflict discs used for the simulation were defined in the same way for the real discs used in a later experiment (Fig. 2.4).

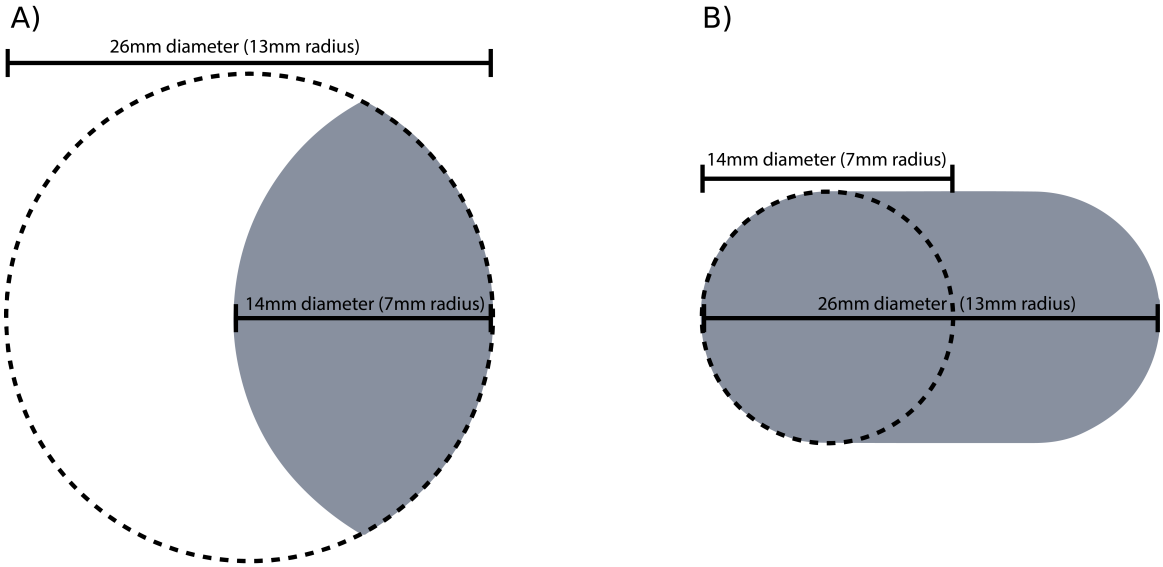

Figure 2.4: A) Illustration of a **-3** Conflict Disc. The diameter of the disc is equal to a **-3mm** radius disc (stimulus level, Δ = mm radius separation from reference). The curvature is that of a +3mm disc. B) Illustration of a **+3** Conflict Disc. The diameter of the disc is equal to a **+3mm** disc. The curvature is that of a -3mm disc.

For the cue conflict conditions, a different conflict disc will replace the reference in each condition and be compared to normal, circular comparison discs. Creating conflict discs with cue values that are equidistant in opposite directions from the reference is convenient because simply seeing which side of 0 (the reference) the *Point*

of Subjective Equality (*PSE*) is in each conflict condition tells us whether *Cut.1* and *Cut.2* together or *Config.* had more influence on the observer.

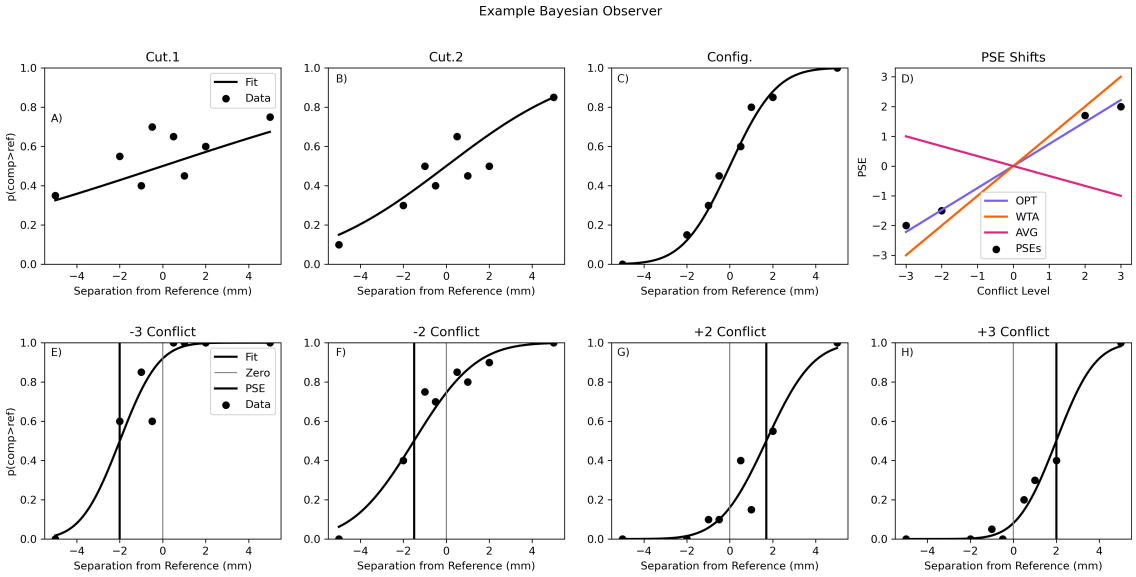

FIGURE 2.5: Example of cue conflict with Bayesian optimal observer under varied sigma. A-C) Sigma portion of the experiment, used to estimate the sigma of each cue in order to make the predictions for the cue conflict experiment. E-H) Cue conflict experiments under the 4 different cue conflicts. *PSE*s indicated by black vertical line, zero indicated by grey line. D) Summary of the *PSE*s from each of the conflict conditions along with the predictions of each model, based on the estimated sigmas.

Previously, with the sigma experiment, each model makes a prediction for how the sigmas of the individual cues relate to the sigma of the combined cues. With a cue conflict, each model will also make a prediction for what the *PSE* should be in each conflict condition. We now have two parameters to aid in the model comparison: the

sigma and the mu (*PSE*). Figure 2.5 shows an example of the cue conflict with a Bayesian optimal observer.

2.4.1 The Cue Conflict Simulation is More Accurate and Resilient to Population Parameters

With limited noisy data, the sigma simulation struggled to confidently and accurately classify certain observers. Completing the cue conflict simulation with each of the three population parameter scenarios, and the same process of creating observers according to the different models, reveals greater classification confidence and accuracy than in the sigma experiment. Table 2.2 shows how many observers were classified as each model. It reveals an overall classification accuracy across observers of 90.07% with *Exceptional Sigma*, 66.73% with *Similar Sigma*, and 93.50% with *Varied Sigma*. The simulation reveals that in all but one scenario, the cue conflict experiment allows us to correctly identify the majority of observers regardless of the population sigma parameters. The simulation only fails to correctly identify *OPT* observers in the Similar Sigma scenario. This matches our expectation, from earlier simulations, that the *OPT* and *AVG* models are difficult to tell apart when sigmas are similar. Nevertheless, not only is accuracy higher, but the inference is also more confident, as evidenced by the boxplots in Figure 2.6. Mean and median posteriors for the correct classification are generally high, except for the *OPT* and *AVG* observers in the Similar Sigma scenario. This is important to note since confidence seems to appropriately correlate with correctness.

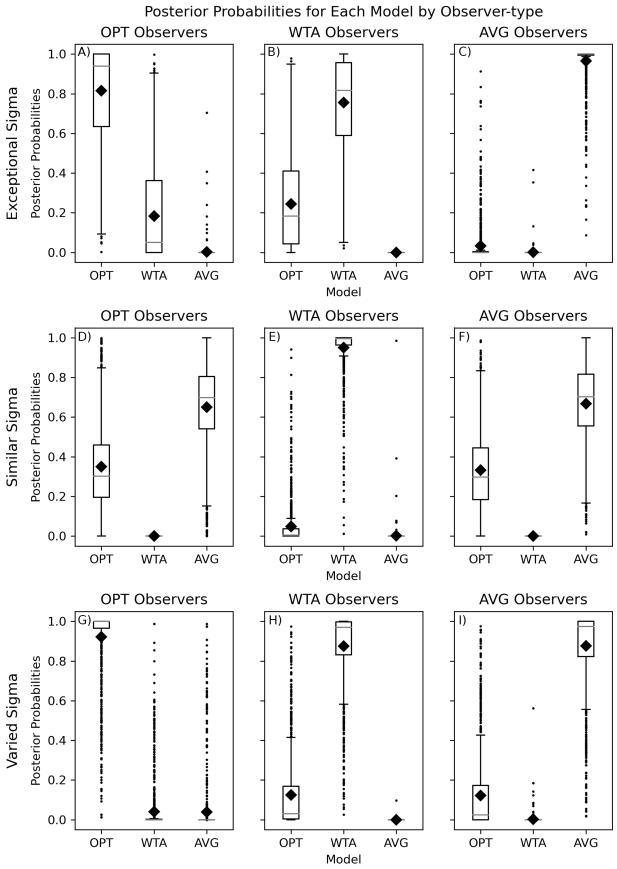

Figure 2.6: Boxplot of Posterior Probabilities in the simulated cue conflict experiment for each Model by Observer-type and Population Parameter. Grey line = median; Black diamond = mean. A-C) Top row. Posterior probabilities for each model by observer-type when participants tend to have one cue with a relatively exceptional sigma. D-F) Middle row. Posterior probabilities for each model by observer-type when participants tend to have similar sigmas. G-I) Bottom row. Posterior probabilities for each model by observer-type when participants tend to have varied sigmas.

These simulations reinforce the intuition we gained earlier by exploring the different predictions each model makes. It goes a step further, simulating a reasonable experiment and showing the differences population-level parameters can have on the

Table 2.2: Observer Classification in Cue Conflict Simulation

Classification	OPT Observer	WTA Observer	AVG Observer	Population Parameters
OPT Classification	**870**	157	10	Exceptional Sigma
WTA Classification	129	**843**	1	
AVG Classification	1	0	**989**	
OPT Classification	**217**	18	196	Similar Sigma
WTA Classification	0	**981**	0	
AVG Classification	783	1	**804**	
OPT Classification	**940**	69	65	Varied Sigma
WTA Classification	25	**931**	1	
AVG Classification	35	0	**934**	

results. Additionally, it shows how perfectly optimal observers can, on occasion, still appear non-optimal even when they really are optimal.

2.4.2 The Cue Conflict Paradigm Reveals a Unique Phenomenon

An interesting observation of note, is that about 10% of the observers in the cue conflict experiment exhibit a positive slope *PSE* shift while all model predictions predict a negative slope. Figure 2.7 shows an example of one such observer. Upon further investigation, this occurs when *Config.* is better than *Cut.1* or *Cut.2*, but noise causes the estimated sigma for *Config.* to be worse than either *Cut.1* or *Cut.2*. The predictions are based on the sigma estimated in the sigma portion of the experiment, but the conflicts are completed using the actual sigmas that define the observer. We can estimate the sigma of participants from noisy data, and our estimates may occasionally be inaccurate, but of course participants are not using our estimates in the task—they're using the sigma that they actually have.

Using the insight and intuition gained through these simulation exercises, we can better explore and interpret data from participants with real humans.

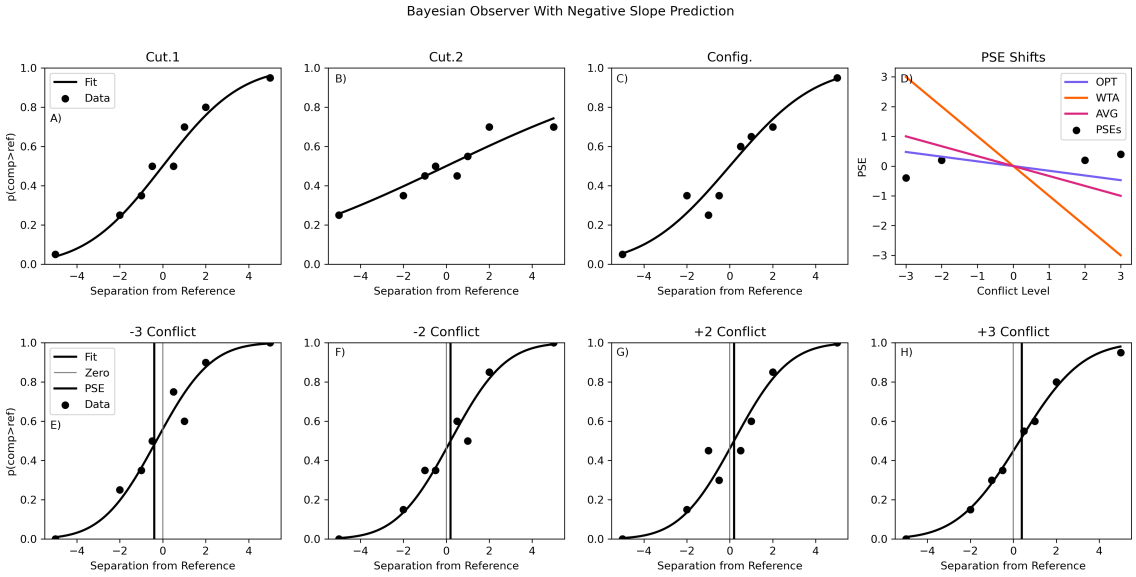

FIGURE 2.7: An example observer whose predictions are negative (flipped) with respect to the slope of its *PSE*s. A-C) Sigma portion of the experiment, used to estimate the sigma of each cue in order to make the predictions for the cue conflict experiment. E-H) Cue conflict experiments under the 4 different cue conflicts. *PSE*s indicated by black vertical line, zero indicated by grey line. D) Summary of the *PSE*s from each of the conflict conditions along with the predictions of each model, based on the estimated sigmas.

Chapter 3

Haptic Size Discrimination Sensory Sigma Experiments

3.1 Materials and Methods

3.1.1 Participants

Participants were McMaster University undergraduate students. Persons with any one or more of the following conditions did not participate, as these conditions are known to adversely affect tactile acuity or the ability to perform tactile tasks: diabetes, nervous system disorder or injury (tremor, epilepsy, multiple sclerosis, stroke, etc.), learning disability, dyslexia, attention deficit disorder, cognitive impairment, carpal tunnel syndrome, arthritis of the hands, and hyperhidrosis. These entrance criteria were clearly listed in the online study announcement on the SONA system and in the *letter of information* (*LOI*). Twelve participants took part in Experiment 1 (3

women, 9 men; mean age 19.3 years). Eleven participants took part in Experiment 2 (1 woman, 10 men; mean age 18.8 years). Participants gave signed informed consent to participate and received monetary compensation or course credit for their time. This study was approved by the McMaster Research Ethics Board.

3.1.2 Materials

All stimuli were designed in OpenSCAD, a 3D script-based modeller. Stimuli were 3D printed with precision using the Ultimaker 2 Go 3D printer. In Experiments 1 and 2, the stimuli were flat coin-shaped discs. The thickness of the discs were consistent within each experiment while radii varied. In each experiment, discs were fastened to a wheel that rotates to present the correct discs for a given trial. This wheel was controlled by an ISM-7411 NEMA 23 National Instruments Integrated Stepper motor that interfaced with the experiment program via an ethernet cable and the National Instruments LabVIEW SoftMotion module. Plastic thimbles, 0.5mm in thickness, were designed and 3D printed, using the above resources, in a range of sizes such that each participant found two thimbles that fit comfortably over their index finger and thumb. An adjustable laboratory scissor jack acted as a rest and guide for the hand and was affixed to a Heavy Duty 20-inch linear bearing slide rail (Firgelli Automations Canada). The rail guided participants' hands to the appropriate disc as their arms extended. An FE7B series photoelectric retroreflective infrared light sensor (Honeywell) sent a voltage change to a USB 6210 data acquisition device (National Instruments) when its beam was broken or unbroken. Thus, the computer was aware of hand positioning relative to the wheel. A GO PRO Session 5 camera

automatically took pictures of participant hand position on some trials. The camera was controlled through the experiment program via WiFi. Participant responses were recorded with a wireless Bluetooth clicker that interfaced with the experiment program to save their responses. The experiment programs and Bayesian analyses were coded in National Instruments LabVIEW 2018 on PC. Frequentist stats were completed using Python and relevant packages and modules including pandas, numpy, statsmodels, and matplotlib.

3.1.3 Procedure

The discs, wheel, and their mechanisms were hidden from participant view behind the white box shown in Figure 3.1B. Each experiment tested participants using a *two-interval forced-choice* (*2IFC*) paradigm. On a given trial, participants extended their right hand to feel one disc in the first interval and a different disc in the second interval. Their task was always to report which disc felt larger, using the Bluetooth clicker. Participants in Experiments 1 and 2 were tested in four different conditions measuring the reliabilities of the cues alone and altogether. The four conditions were as follows: the *Cutaneous Cue 1* (*Cut.1*) condition, measuring the cutaneous cue on the thumb; the *Cutaneous Cue 2* (*Cut.2*) condition, measuring the cutaneous cue on the index finger; the *Configuration Cue* (*Config.*) condition, measuring the hand configuration cue; and the *Combined Cues* condition, measuring the sigma when participants have access to all cues (Figure 3.2). In *Cut.1* and *Cut.2*, participants used only the thumb or index finger, respectively, of their right hand to feel the edge perpendicular to the face of the disc. The discs were left-aligned for the thumb and right-aligned for the

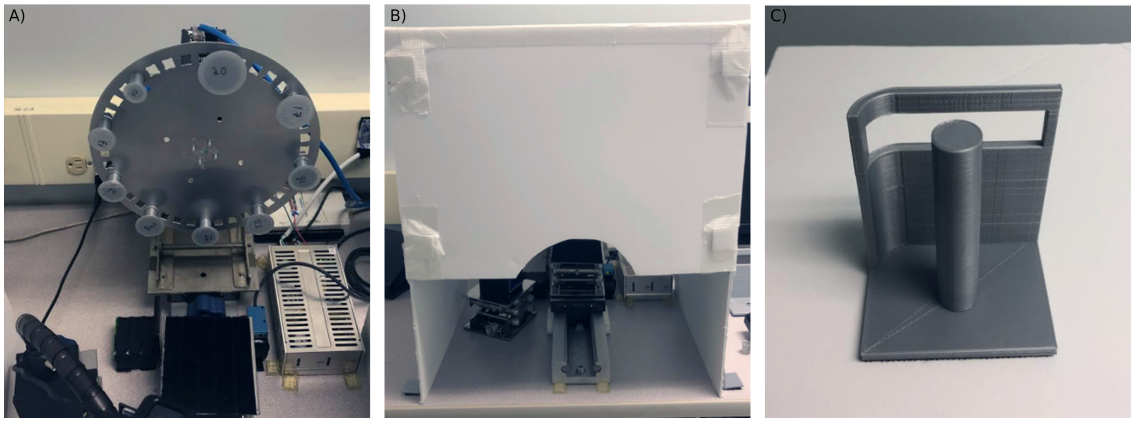

Figure 3.1: A) View of the rotating wheel with disc stimuli attached. B) View of the covered apparatus with all stimuli and electronics hidden from participants. C) View of the hand rest.

index such that the time and distance to reach the disc did not vary systematically with disc size. In the *Configuration Cue* condition, participants wore thimbles on their index finger and thumb before grasping the disc edge. Thus, they felt the distance between their fingers without feeling the curvature of the object. In the *Combined Cues* condition, participants did the same task with thimbles removed, having access to both cutaneous and hand configuration cues to size. On each trial, the order of presentation of the reference and comparison disc (ie. which came first) was randomized. Participants reported via Bluetooth clicker which disc felt larger. Participants completed up to thirty practice trials before starting each condition.

Experiment 1

Experiment 1 took place over two sessions a week apart. The conditions were partially counterbalanced such that, *Config.*, *Comb.*, and *Cut.1* or *Cut.2* conditions had equal

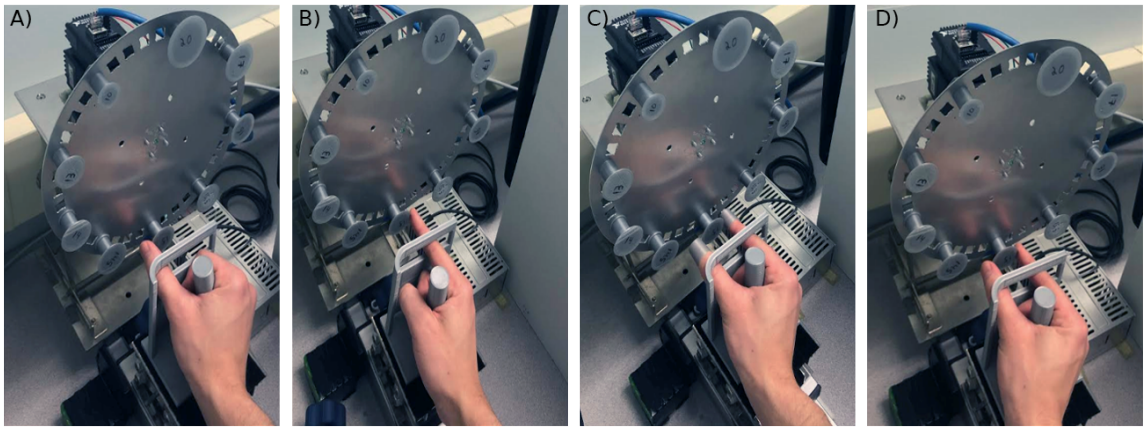

Figure 3.2: A) Demonstration of Cut.1 (cutaneous cue on the thumb finger). B) Demonstration of Cut.2 (cutaneous cue on the index). C) Demonstration of Config. (configuration cue without cutaneous stimulation). D) Demonstration of Combined Condition (all cues present).

chances of occurring at the beginning, middle, or end of the experiment. Although *Cut.1* and *Cut.2* conditions always occurred in succession, their order was also randomized. Participants completed two blocks of each condition across two days for a total of four blocks of 40 trials each. Discs in Experiment 1 had the following radii in millimetres: 10, 13, 14, 14.5, 15, 15.5, 16, 17, and 20. The 15mm disc was the reference to be compared with any of the other discs during each trial. In each block, the stimuli were delivered by Method of Constant Stimuli, with each comparison disc being delivered five times in pseudo-random order. Accordingly, each participant was tested for a total of 160 trials, 20 times with each of the 8 comparison discs, in each condition.

Experiment 2

Experiment 2 took place in one 2.5-3 hour session. The conditions were partially counterbalanced in the same way as in Experiment 1. Each condition consisted of two blocks of 40 trials. Discs in Experiment 2 had the following radii in millimetres: 5, 8, 9, 9.5, 10, 10.5, 11, 12, and 15. The 10mm disc was the reference to be compared with any of the other discs during each trial. The stimuli were delivered using a Bayesian Adaptive Procedure which allowed us to present the disc on each trial from which the most information about the participant's psychometric function could be learned (Goldreich et al. 2009; Kontsevich & Tyler 1999). The Bayesian adaptive procedure can be applied to any psychometric function parameterization; we defined our psychometric functions as cumulative normal distributions of the stimulus level, Δ (Eq. 3.1).

$$P(comp > ref|\Delta) = \int_0^{\inf} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\Delta_m - \Delta)^2}{2\sigma^2}\right) d\Delta_m = \Psi(\Delta) \tag{3.1}$$

3.2 Results

3.2.1 Experiment 1

We applied Bayesian parameter estimation to the data in order to estimate each participant's sigma for each condition. Each participant's sigmas were combined according to each model's formulas to generate the model prediction. Figure 3.3 summarizes these data, displaying the mean estimates for each condition along with the

mean expected sigma, given those estimates, for each model. Each model aims to predict what performance would be expected in the Combined Condition, given Cutaneous and Configuration Conditions. One-way repeated measures ANOVAs reveal significant effects of condition ($p = 0.006$, $F(3, 33) = 4.961$) and model prediction ($p = 0.001$, $F(2, 22) = 9.038$). Paired-sample t-tests demonstrate that the *Average-Measurement* (*AVG*) model prediction is statistically significantly different than both the *Optimally-Weighted* (*OPT*) ($p = 0.010$, $t(11) = 3.094$) and *Winner-Take-All* (*WTA*) model predictions ($p = 0.013$, $t(11) = 2.922$). There was, however, no significant difference between the predictions of the *WTA* and *OPT* models ($p = 0.050$, $t(11) = 2.197$). There was a significant difference between the Combined Condition and the *AVG* model prediction ($p = 0.010$, $t(11) = 3.099$) but not between Combined and *OPT* ($p = 0.730$, $t(11) = 0.354$) or Combined and *WTA* ($p = 0.206$, $t(11) = 1.343$).

Table 3.1: Exp. 1 Model Posterior Probabilities by Participant

Participant	OPT	WTA	AVG	Classification
1	0.472	0.527	0.001	WTA
2	0.458	0.540	0.002	WTA
3	0.507	0.493	0.000	OPT
4	0.535	0.463	0.002	OPT
5	0.548	0.447	0.004	OPT
6	0.052	0.182	0.766	AVG
7	0.588	0.045	0.367	OPT
8	0.481	0.519	0.000	WTA
9	0.548	0.451	0.000	OPT
10	0.560	0.440	0.000	OPT
11	0.665	0.118	0.217	OPT
12	0.528	0.472	0.000	OPT

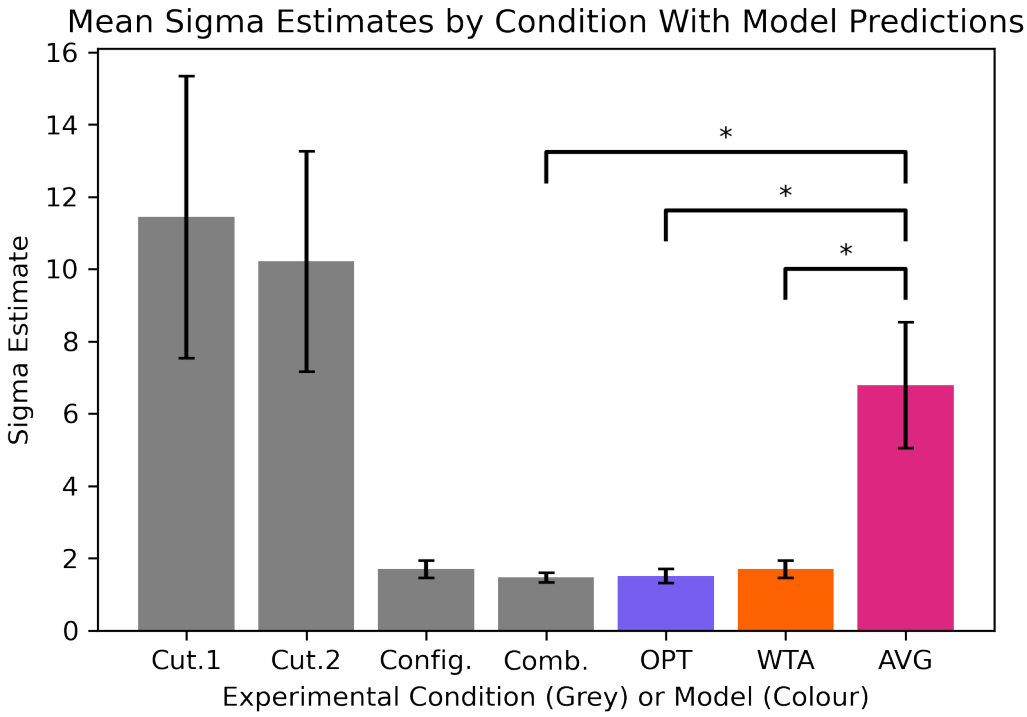

Figure 3.3: Mean sigma estimates from each participant in Experiment 1 (Grey). Mean model predictions for performance in Combined Condition, based on the Cutaneous and Configuration conditions (Colour). Error bars = +/- SE. Asterisk (*) = Significant t-test.

Applying Bayesian model comparison, we calculated posterior probabilities for each participant for each model. The individual posterior probabilities for each participant in Experiment 1 are found in Table 3.1. Median posterior probabilities corroborate the trends from the above frequentist statistics. The median posterior probabilities are as follows: $OPT = 0.532$, $WTA = 0.457$, $AVG = 1.39 \cdot 10^{-3}$. Overall, the Bayesian model comparison categorized 8 participants as OPT, but the posterior probabilities were often quite close for OPT and WTA.

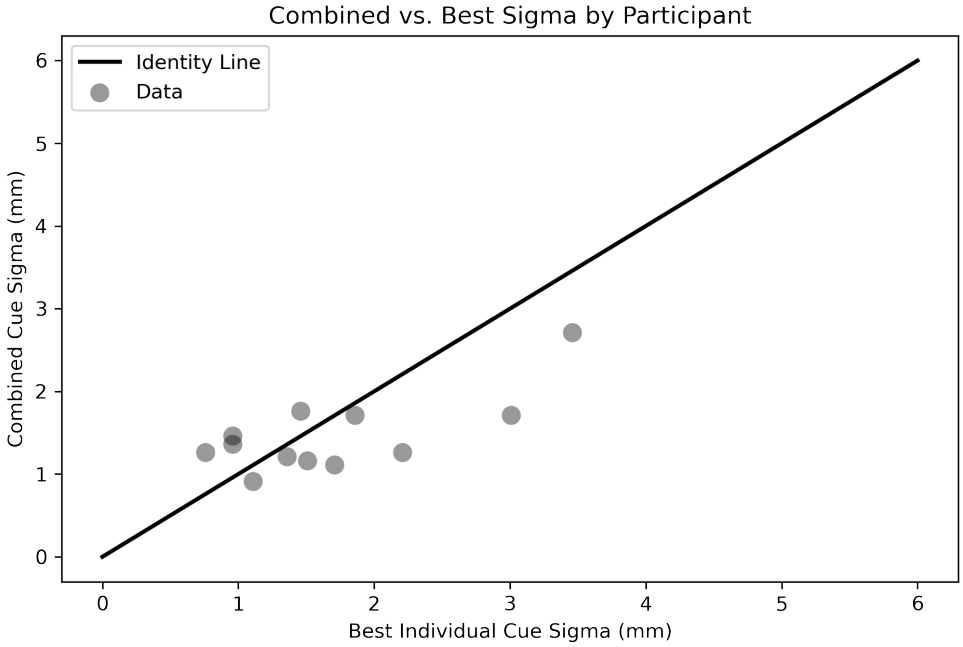

Figure 3.4: Scatterplot of participant sigma estimates for the Combined Condition plotted against their respective sigma estimates for their best (lowest) individual sigma in Experiment 1.

The frequentist and Bayesian statistics agree that the *AVG* model does not convincingly account for participant data. We find, however, that it is harder to distinguish between the *OPT* and *WTA* models. While the *OPT* model is slightly more favoured according the Bayesian inference, we would like the answer to be more definitive. These results are somewhat unsurprising, however, considering the model predictions from 2.1. We appear to be dealing with a scenario where we have an exceptional sigma. The *Config.* condition sigma is much more reliable than *Cut.1* and *Cut.2* as shown in Figure 3.3. Still, 8 of the 12 participants did better in the Combined Condition than they did in their best individual cue condition (Figure 3.4). A paired-samples

t-test on the data from Figure 3.4 would be identical to that of the Combined and *WTA* statistic reported above ($p = 0.206$, $t(11) = 1.343$). Although not statistically significant, the scatter plot suggests a possible trend. As with the Bayesian analysis, this observation suggests that the majority of participants were better described as *OPT* than *WTA*.

3.2.2 Experiment 2

Experiment 1 provided evidence against the *AVG* model but was unable to convincingly distinguish between the *OPT* and *WTA* models. We reasoned that this difficulty may have resulted from the fact that *Config.* had a much smaller sigma than *Cut.1* and *Cut.2* in most participants. In Experiment 2, with the goal of enhancing *Cut.1* and *Cut.2*, we made the discs smaller so that they would indent the fingers more sharply. We hoped to reduce the sigma of the cutaneous cues so that the cutaneous and configuration cues would have more similar sigma values, facilitating the classification of participants as either *OPT* or *WTA*.

Figure 3.5 shows a reduction in sigma in the Cutaneous Conditions. The reliabilities of the cues in the Cutaneous and Configuration Conditions are relatively closer together and we begin to observe a greater difference between the predictions of the *OPT* and *WTA* models.

One-way repeated measures ANOVAs reveal non-significant effects of condition ($p = 0.104$, $F(3, 30) = 2.236$) and model prediction ($p = 0.187$, $F(2, 20) = 1.825$). Paired-sample t-tests demonstrate that the *AVG* model prediction is not statistically

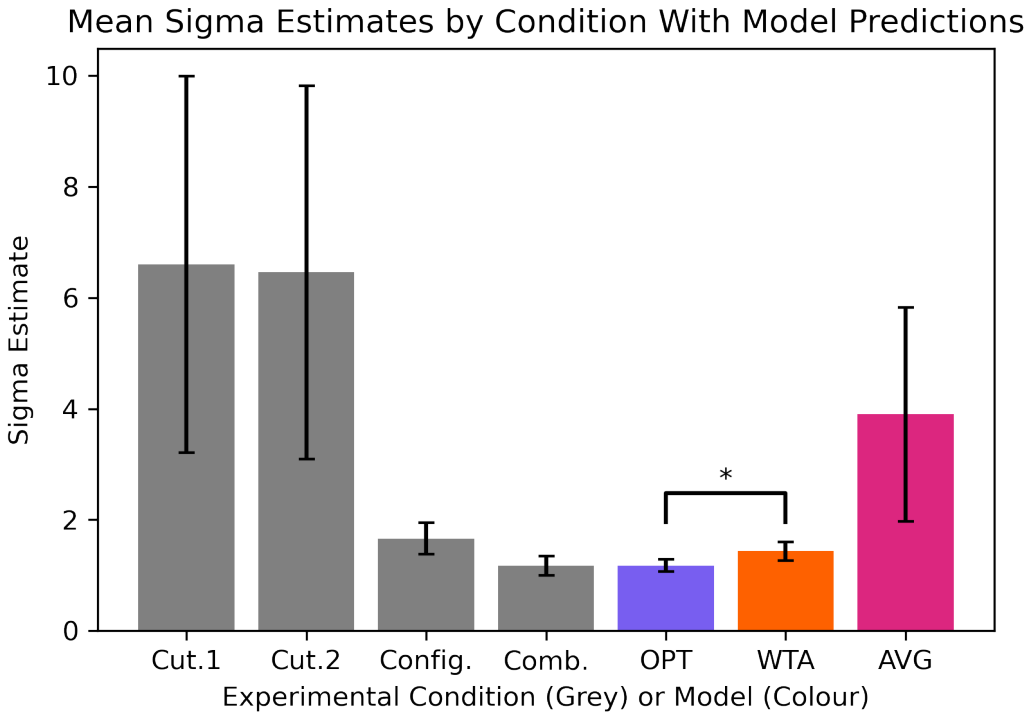

Figure 3.5: Mean sigma estimates from each participant in Experiment 2 (Grey). Mean model predictions for performance in Combined Condition, based on the Cutaneous and Configuration conditions (Colour). Error bars = +/- SE. Asterisk (*) = Significant t-test.

significantly different than the *OPT* (p = 0.185, t(10) = 1.425) or *WTA* predictions (p = 0.233, t(10) = 1.269). There was, however, a significant difference between the predictions of the *WTA* and *OPT* models (p = 0.004, t(10) = 3.666). There was no significant difference between the Combined Condition and the *AVG* model prediction (p = 0.193, t(10) = 1.395), between Combined and *OPT* (p = 0.997, t(10) = 0.004), or between Combined and *WTA* (p = 0.180, t(10) = 1.441).

Calculated posterior probabilities provide additional nuance and context to the

Table 3.2: Exp. 2 Model Posterior Probabilities by Participant

Participant	OPT	WTA	AVG	Classification
1	0.500	0.482	0.019	OPT
2	0.188	0.385	0.427	AVG
3	0.010	0.659	0.331	WTA
4	0.199	0.370	0.430	AVG
5	0.838	0.131	0.031	OPT
6	0.455	0.437	0.109	OPT
7	0.773	0.044	0.183	OPT
8	0.538	0.462	0.000	OPT
9	0.720	0.269	0.011	OPT
10	0.502	0.498	0.000	OPT
11	0.524	0.362	0.114	OPT

above findings. The individual posterior probabilities for each participant in Experiment 2 are found in Table 3.2. The median posterior probabilities are as follows: $OPT = 0.502$, $WTA = 0.393$, $AVG = 0.097$). Overall, the Bayesian model comparison categorized 8 of the 11 participants as OPT. While the posterior probabilities for OPT and WTA were often close, the posterior probabilities from Experiment 2 were generally more strongly in favour of OPT over WTA than were the posterior probabilities from Experiment 1.

For most participants, the configuration cue was most informative in their task. Figure 3.6 compares performance on the participants' *best* individual condition to performance in the *Combined* condition. Participants tended to perform better when having access to all cues than any one cue alone. Points that fall below the identity line in Figure 3.6 are participants whose Combined Condition sigmas were lower than that of their best individual cue. This more closely follows a prediction of the

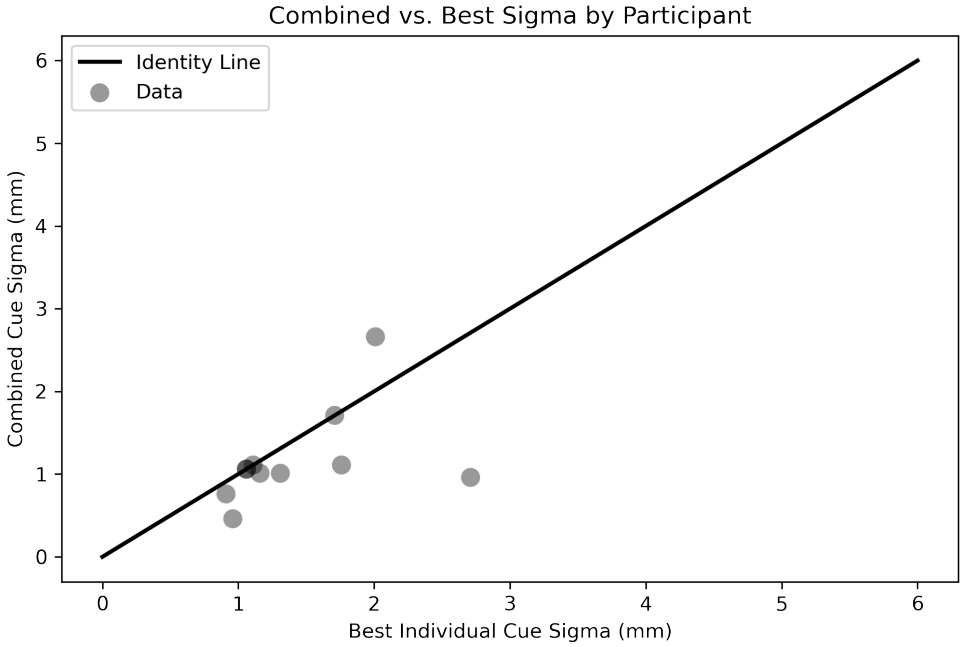

Figure 3.6: Scatterplot of participant sigma estimates for the Combined Condition plotted against their respective sigma estimates for their best (lowest) individual sigma in Experiment 2.

OPT Model than the *WTA* Model since information is apparently gained even from relatively poorer cues. A paired-samples t-test on the data from Figure 3.6 would be identical to that of the Combined and *WTA* statistic reported above ($p = 0.180$, $t(10) = 1.441$).

3.3 Discussion

In Experiment 1, a novel investigation, the disparity between the cue sigmas was greater than anticipated. This experiment gave us better insight into the perceptive

capabilities of humans, provided evidence in support of the *OPT* and *WTA* models over the *AVG* model, and prepared us to design Experiment 2.

Experiment 2 reduced overall disc size to more sharply indent the skin. The manipulation appeared to have an effect with the mean sigmas for *Cut.1* and *Cut.2* being reduced by 4.84mm and 3.75mm respectively. In contrast, the mean sigma of *Config.* decreased by only 0.038mm. As we intended, the manipulation of disc size appears to preferentially affect the Cutaneous Cues over the Configuration Cue. This is consistent with earlier research using spherical stimuli. The responses of *slowly adapting type 1* (*SA1*) afferents increase with curvature (Goodwin & Wheat 2004).

In both experiments 1 and 2, the frequentist and Bayesian statistics generally agreed, but the latter seemed to be more powerful at discerning between model predictions. The frequentist stats were computed using our best single estimate for each participant's cue sigma to investigate the repeated measures data on the participants as a group. The Bayesian statistics, however, utilize all of the raw data for each participant to do full Bayesian inference on each participant as an individual.

These experiments provide evidence regarding the extent to which humans can utilize these haptic cues. The results are weakly in favour of *OPT* over *WTA*. In the following chapter, we set forth to better differentiate between these models and learn the extent to which these cues may be weighted.

Chapter 4

Haptic Size Discrimination Cue Conflict Experiment

Experiment 3 was designed as a cue conflict experiment that uniquely allowed us to test different, but related, predictions of the models. For regular, non-conflict discs, the *Point of Subjective Equality* (*PSE*), where a comparison disc feels subjectively equal to the reference, is at a 0mm difference from the reference. That is, the reference feels equal in size to a disc of its own size. This *PSE* can change, however, as conflict stimuli are introduced.

4.1 Materials and Methods

4.1.1 Participants

Participants were McMaster University undergraduate students. Persons with any one or more of the following conditions did not participate, as these conditions are known to adversely affect tactile acuity or the ability to perform tactile tasks: diabetes, nervous system disorder or injury (tremor, epilepsy, multiple sclerosis, stroke, etc.), learning disability, dyslexia, attention deficit disorder, cognitive impairment, carpal tunnel syndrome, arthritis of the hands, and hyperhidrosis. These entrance criteria were clearly listed in the online study announcement on the SONA system and in the *letter of information* (*LOI*). Eleven participants took part in Experiment 3 (2 women, 9 men; mean age 19.1 years). Participants gave signed informed consent to participate and received monetary compensation or course credit for their time. This study was approved by the McMaster Research Ethics Board.

4.1.2 Materials

All stimuli were designed in OpenSCAD, a 3D script-based modeller. Stimuli were 3D printed with precision using the Ultimaker 2 Go 3D printer. The materials used were the same as in Experiment 2 with the exceptions noted here. The comparison and reference discs were the same as in Experiment 2. A set of cue conflict stimuli were also introduced, however. The cue conflict stimuli were designed such that each indicates a curvature and radius in conflict with one another. Figure 4.1 illustrates the design of **-3** and **+3** conflict stimuli. The magnitude of the number indicates the

difference in radius (mm) from the reference, and the sign indicates in which direction the lateral distance across the face of the disc changed. Thus, a **-3** conflict disc (Fig. 4.1A) has the lateral diameter of a disc with a radius that is -3mm from the reference of 10mm: a 7mm radius. Simultaneously, it has the lateral *curvature* of a disc +3mm from the reference of 10mm: a 13mm radius. A **+3** conflict disc (Fig. 4.1B) has the lateral diameter of a disc with a radius that is +3mm from the reference of 10mm: a 13mm radius. Simultaneously, it has the lateral *curvature* of a disc -3mm from the reference of 10mm: a 7mm radius. The same logic applies to the **-2** and **+2** conflict discs.

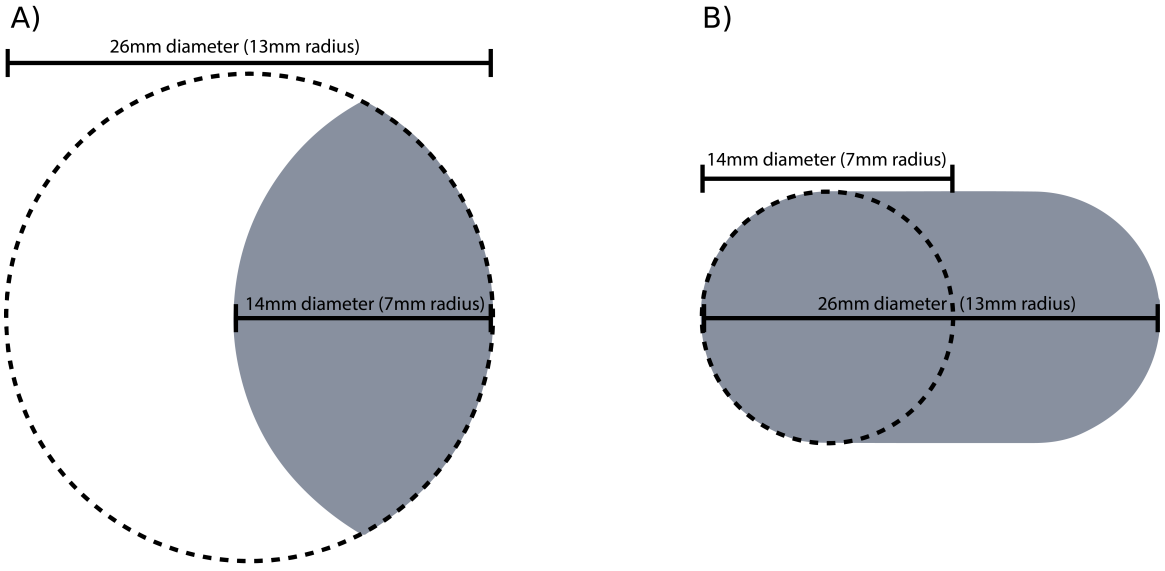

Figure 4.1: Reproduced from Fig. 2.4. A) Illustration of a **-3** Conflict Disc. The diameter of the disc is equal to a **-3mm** radius disc (stimulus level, Δ = mm radius separation from reference). The curvature is that of a +3mm disc. B) Illustration of a **+3** Conflict Disc. The diameter of the disc is equal to a **+3mm** disc. The curvature is that of a -3mm disc.

4.1.3 Procedure

The methodology used was generally the same as in Experiment 2 with the exceptions noted here. Instead of 2 blocks of 40 trials for each of 4 conditions, participants completed one block of 40 trials for each of 7 different conditions. The first 3 conditions were the same as the earlier experiments, to estimate participant sigmas for the individual cues (*Cutaneous Cue 1* (*Cut.1*), *Cutaneous Cue 2* (*Cut.2*), and *Configuration Cue* (*Config.*)). The remaining 4 conditions are all Cue Conflict Conditions. From the participant's point of view, these were no different from the Combined Conditions of the earlier experiments. In each of these conflict conditions, however, the 10mm radius reference disc was swapped out for 4 different cue conflict stimuli. The conditions were partially counterbalanced with restrictions: individual cue conditions occurred at the beginning of the experiment, but *Cut.1* and *Cut.2* conditions occurred successively in random order; conflict conditions occurred at the end of the experiment in random order, but ± 2 conditions occurred together and ± 3 conditions occurred together. The task was otherwise carried out in the same manner as before.

4.2 Results

Figure 4.2 demonstrates an overview of participant performance in each experimental condition. Observing the mean sigma estimates across participants for each condition, we found similarities to earlier experiments. The mean sigma of the cues decreases from *Cut.1* to *Cut.2* to *Config.* The four Cue Conflict Conditions are analagous to the earlier Combined Conditions of previous experiments because participants have

access to all cues. As an aside, sigmas in the Conflict Conditions generally appear to be lower than the the sigmas of any individual cue, with the exception of the particularly noisy **+3** Condition. Nevertheless, Experiment 3 was not intended as a replication of Experiments 1 and 2. The predictions and planned analyses are different than in the earlier experiments.

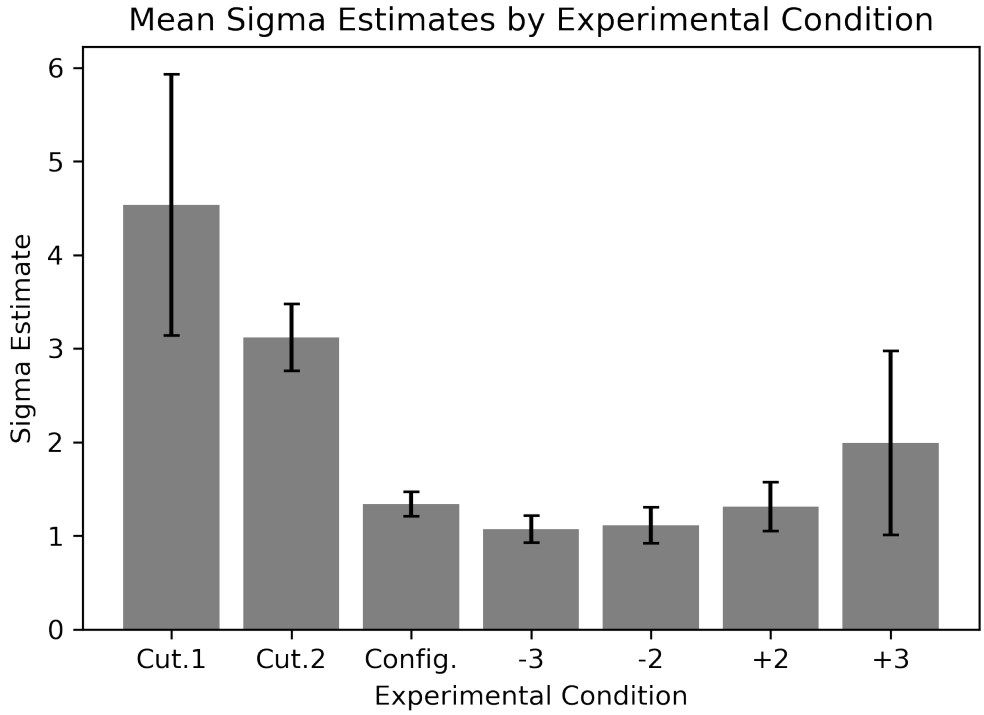

Figure 4.2: Sigma estimates for each individual cue condition and each of the four conflict conditions.

As the reference in each of the Conflict Conditions is one of four conflict discs, each model predicts that the *PSE* should shift linearly in the direction of the cues upon which the participant most relies. For example, if a participant is relying solely

on *Config.*, a **+3** conflict disc would feel subjectively equal to a 13mm radius disc. Since, in a *two-interval forced-choice* (*2IFC*) task, the response is to which disc felt larger, the participant would be guessing when comparing the **+3** conflict disc to a 13mm comparison disc. On average, the proportion of times one disc is reported to be greater than the other would be 0.5: a rightward shift of 3mm in the *PSE*.

The mean *PSE* shift across participants, as a function of conflict condition, is shown in Figure 4.3. Each model predicts a different linear *PSE* shift. The *Average-Measurement* (*AVG*) model predicts a negative slope because of the influence of both *Cut.1* and *Cut.2*, which produce measurements counter to *Config.* The *Winner-Take-All* (*WTA*) model predicts a positive slope equal to 1, since as noted above, the magnitude of the *PSE* shift would be equal to that of the Conflict Condition. On average, the shifting *PSE*s most closely match the predictions of the *Optimally-Weighted* (*OPT*) model. The prediction of the slope for the optimal model is based on Equation 4.1 where s is the Conflict Condition and $\hat{s}$ is the PSE (Appendix A3). The positive slope indicates that *Config.* was more heavily relied upon than *Cut.1* or *Cut.2*. Additionally, the slope is not equal to 1, which implies that *Cut.1* and *Cut.2* did indeed influence perception and cues were combined.

$$\hat{s} = s \left(\frac{\frac{1}{\sigma^2_{Config.}} - \frac{1}{\sigma^2_{Cut.2}} - \frac{1}{\sigma^2_{Cut.1}}}{\frac{1}{\sigma^2_{Cut.1}} + \frac{1}{\sigma^2_{Cut.2}} + \frac{1}{\sigma^2_{Config.}}} \right) \tag{4.1}$$

Mounting evidence from Experiments 1, 2, and 3 has allowed us to determine that the *AVG* model does not adequately predict or describe participant data. Subsequent

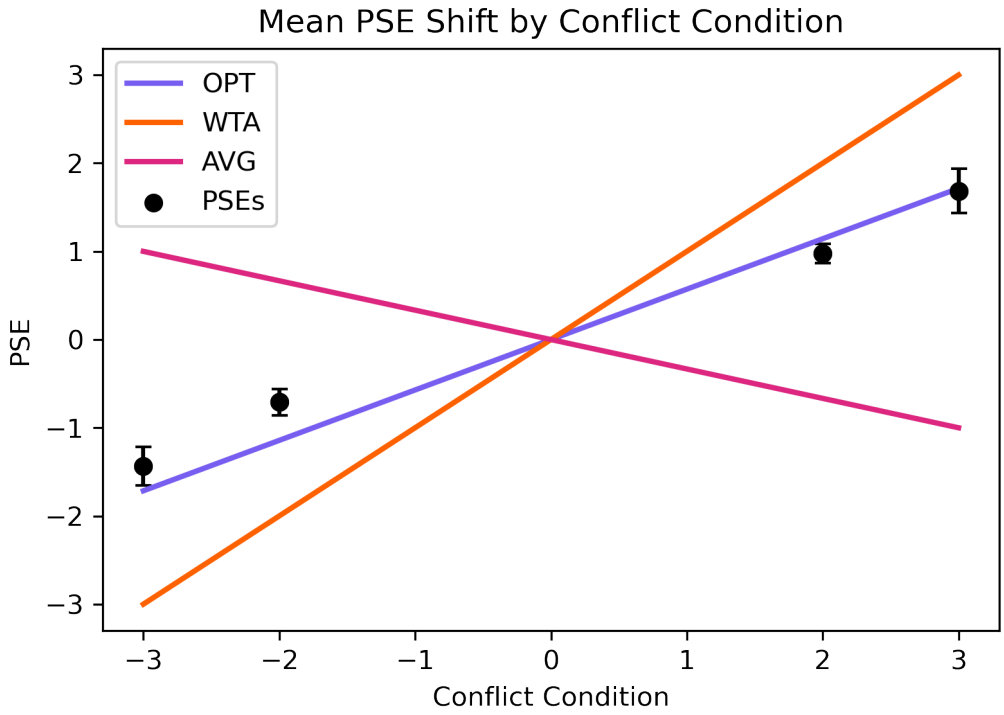

Figure 4.3: The mean PSE shift across participants by Conflict Condition (Error bars = ± SE)

analyses are focused on distinguishing between *OPT* and *WTA*. Figure 4.4 shows the *PSE* shifts for each individual participant. Occasionally, the models predicted a negative slope, but that is not unexpected given our investigations in 2.4.2. There, we found that if sigma estimates themselves are noisy enough, *Cut.1* or *Cut.2* can occasionally appear to have a lower sigma than *Config.* even when that does not reflect reality. If this does occur, then the model predictions which are based on those sigma estimates can be negative. The *PSE* shifts, however, are still likely to be positive if *Config.* was truly more relied upon. Notably, the actual recorded PSEs

of each participant have positive slopes even if the model prediction was negative. This matched our results from Figure 4.2 and earlier experiments, where *Config.* has the lowest sigma of any individual cue, on average. Thus, participants appear to combine these cues, relying on *Config.* more than the others, although *Cut.1* and *Cut.2* generally have influence on perception as well. Table 4.1 displays the posterior probabilities for each participant across the three models, rounded to two decimal places. The majority of participants were strongly *OPT* over *WTA* or *AVG*. Participant 11, however, appeared to be strongly *WTA*.

Table 4.1: Exp. 3 Model Posterior Probabilities by Participant

Participant	OPT	WTA	AVG	Classification
1	1.000	0.000	0.000	OPT
2	1.000	0.000	0.000	OPT
3	1.000	0.000	0.000	OPT
4	0.999	0.001	0.000	OPT
5	1.000	0.000	0.000	OPT
6	0.994	0.006	0.000	OPT
7	0.999	0.000	0.001	OPT
8	0.973	0.027	0.000	OPT
9	0.871	0.001	0.127	OPT
10	1.000	0.000	0.000	OPT
11	0.000	1.000	0.000	WTA

4.3 Discussion

Consistent with the results in Chapter 2, the cue conflict paradigm appears to be a more powerful method for differentiating between different cue combination hypotheses. The *PSE* shifts appear to be robust to the noise that makes sigma estimation

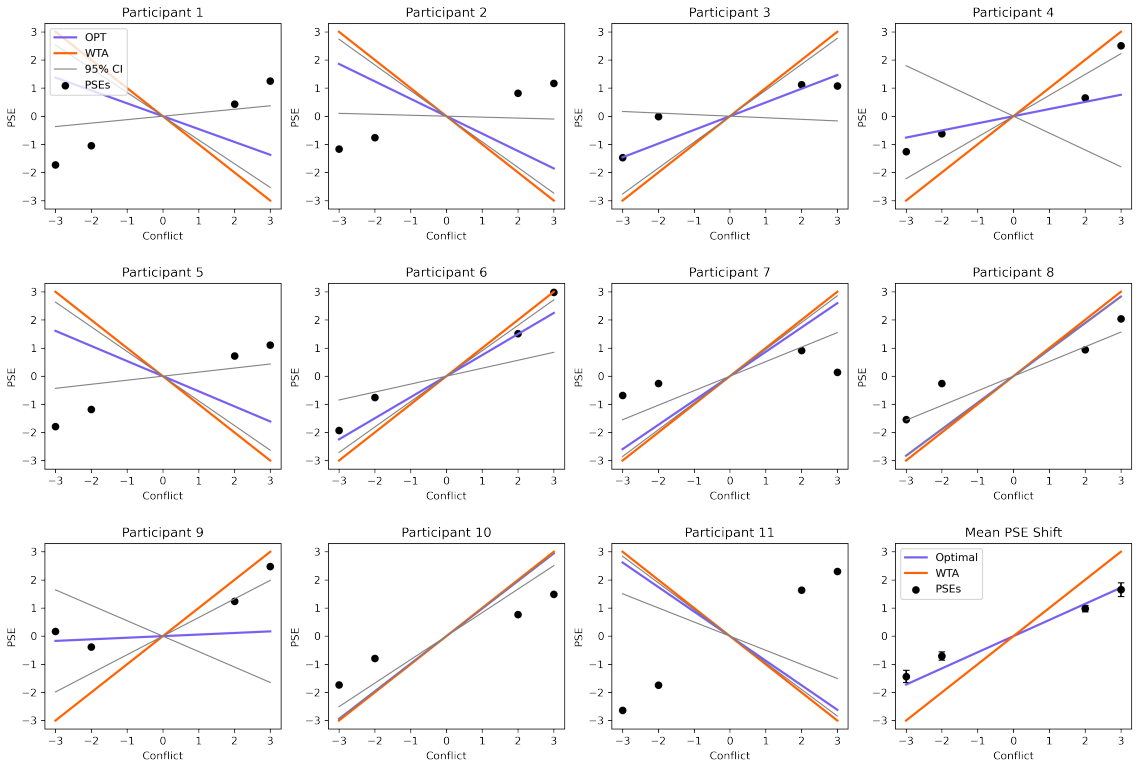

FIGURE 4.4: The PSE shifts of each participant plotted with the WTA prediction, the OPT prediction, and 95% Bayesian Credible Intervals around the OPT prediction. The AVG prediction was omitted from these subplots to avoid clutter. The mean PSE shifts (Fig. 4.3) replotted for quick reference.

difficult. Of the models considered, the *OPT* model has strong support in every participant, with the exception of Participant 11.

Reflecting on Participant 11's *PSE* shifts in Figure 4.4 shows that the *PSE* shifted linearly with a slope almost equal to 1. Participant 11 appears to be the only participant who did not combine cues. There are a few possible reasons for this. Participant 11 is either *WTA* or did not perceive the conflict stimuli as discs, but some object

with two sources of information, rather than one.

Additionally, we've observed a trend where the sigmas in the Conflict Conditions appear to increase successively. As the Conflict Condition number increases, so does the lateral curvature and the diameter of the stimulus. We've observed that increased curvature leads to a reduction in sigma, as the cutaneous cue becomes more reliable. This led to the question: does sigma increase for *Config.* as diameter increases? While this is theoretically plausible, a repeated measures ANOVA over Conflict Condition yielded a non-significant result ($p = 0.532$, $F(3, 30) = 0.749$).

Chapter 5

Discussion

5.1 Summary of Main Findings

In Chapter 2, we found that each of the considered models makes predictions that are similar or different to the predictions of other models depending on their relative sigma values. We considered three different scenarios for relative sigmas: *Exceptional*, *Similar*, and *Varied Sigmas*. We found that the model classification inference struggled to correctly identify observers within models, although its classification was better than chance within each observer population. The uncertainty of the classifications, however, was captured in the posterior probabilities for each model as well. When simulating a cue conflict experiment, we found that the classifications were generally correct within models and within population. The improved classifications were coupled with posterior probabilities that implied greater certainty in the classifications. Altogether, this chapter revealed limitations due to noisy data and how that can affect

model predictions. It also demonstrated that our analyses and statistics are able to accurately differentiate between the models despite noise and population parameters.

In chapter 3, Experiment 1 showed the relative sigmas of each cue. The data suggested we were dealing with an *Exceptional Sigma* scenario where *Configuration Cue* (*Config.*) was much better than *Cutaneous Cue 1* (*Cut.1*) and *Cutaneous Cue 2* (*Cut.2*). Our simulations revealed that, in such a scenario, the *Winner-Take-All* (*WTA*) and *Optimally-Weighted* (*OPT*) models make similar predictions. Although we found strong evidence against the *Average-Measurement* (*AVG*) model, we were unable to convincingly differentiate between the *WTA* and *OPT* models. In order to tell them apart, we aimed to lower the sigmas of the cutaneous cues. Our intention was that this would cause *OPT*, but not *WTA*, to predict a lower sigma, permitting us to differentiate between these two models. Experiment 2 demonstrated that reducing overall disc size did preferentially reduce the sigmas of the cutaneous cues. Our inferences also became more in favour of *OPT* over *WTA*.

In Chapter 4, we concluded our investigation with a cue conflict experiment. By introducing conflict stimuli such that the cutaneous cues were in conflict with the configuration cue, we were able to glean information from *Point of Subjective Equality* (*PSE*) shifts. The *PSE*s showed that participants generally rely more on *Config.* than *Cut.1* or *Cut.2* but information from all cues are integrated in a way consistent with *OPT*. Our inferences in this experiment were notably confident in their classifications. These results mirror the results of our model simulations and lend credence to the results. Participant 11, however, was more strongly *WTA* than *OPT* or *AVG*.

One of the assumptions we made in our experiments is that participants will perceive the cues as all originating from the same source. We believed this to be a reasonable assumption, given that the cues do arise from the same physical source: a 3D-printed disc. Other cue combination studies, like Alais and Burr (2004), pair two different stimuli together, which really do come from independent sources, and found that humans can combine those cues too. Bayesian causal inference, however, is a theory that involves "competitive priors" between single and multiple cue sources and can account for many studies on multisensory perception (Shams & Beierholm 2022). In the conflict experiment, it's possible that the conflict stimuli could break the expectations of some participants and cause them to perceive the cues as arising from two independent sources. Participant 11's data could have arisen from the participant's choice to pay attention to only *Config.*, either because they misunderstood the task, or because they felt the cues as arising from two independent sources and chose to use *Config.*

5.2 Paths for Future Investigation

The conflict experiment was extremely informative and generated new questions. Are some fraction of participants *WTA* or resistant to the experimental manipulation? How large can a cue conflict be made before participants begin to notice it? Does *Config.* sigma increase with disc diameter? Future experiments can include these questions in their investigations.

In the experiments we conducted, we manipulated disc size to lower the cutaneous

sigmas and bring their performance closer to *Config.* performance. A future study could aim to lessen the disparity between cutaneous and configuration cues by preferentially worsening *Config.* and making its sigma higher. Potentially, this could be accomplished by numbing or vibrating select areas of the hand, but this manipulation would not be trivial. Ideally, if the manipulations resulted in each of the three cues having *Similar Sigmas*, then the difference between the predictions of the *WTA* and *OPT* models would be at their greatest.

Future experiments could also investigate heteroscedasticity in the measurement. As a first approximation, our experiments assume that participants have a single value for sigma for each cue. It's possible, however, that the increasing sigma, as conflict increases, reflects a real trend. An investigation into the matter could consider whether participants have a 'fractional sigma', analagous to a Weber fraction (Norwich 1987). The 'fractional sigma' could be multiplied by the stimulus level to give the sigma (mm) for that cue and stimulus.

Our experiments involve controlled haptics. The hand moves towards and grasps the target, but participants could not freely explore the stimuli. Future experiments could reduce the level of control in favour of greater external validity. Participants could be allowed time to grasp, scan with all of their fingers, and feel the weight of different objects as they see fit. This would allow the participant to have access to not only more of the cues we have currently, but additional cues including weight and texture. Would *Config.* still be most informative? Would participants be consciously aware of the cue upon which they most relied?

In our experiments, from one interval to another within a trial, there is a short delay between stimulus presentations. Thus, participants must engage working memory to hold the perceived size of the first disc in order to compare it to the second. Could participant performance improve if we delivered the reference and comparison in each hand simultaneously? If so, how much improvement could be seen?

Our results support the *OPT* model over the *WTA* and *AVG* models. Similar to research discussed in Section 1.3.1, the data suggest that participants have some internal monitor of cue reliability and use that information when integrating sensory signals. Because it is currently unknown how the brain achieves this, we point to this direction as ripe for future research as well.

5.3 Applications for Current and Upcoming Fields

As haptic feedback for virtual reality seems to be on the cusp of great improvement, learning more about how haptic cues are combined is especially timely. Some of the most advanced haptic VR gloves, like ManusVR and HaptX, are either lacking in cutaneous stimulation or limited by the number tactors and the range of motion of the hand itself. Our research suggests ways to aid both virtual object discrimination and immersion. While the configuration cue is generally the better cue, a substantial amount of information and influence comes from the cutaneous cues as well. By integrating even a few different curvatures or stimuli to the finger in conjunction with the configuration cues delivered by the gloves, novel shapes could presumably be perceived and disambiguating information could be gleaned to grant the user a more immersive

experience. We have shown that cutaneous and configuration cues can be integrated even when in conflict with one another. Adding a few different cutaneous cues, and pairing them with the configuration cues, even when in conflict, can produce unique sensation that users are likely to integrate. By creatively employing cue conflicts, it may be possible to stretch the use of the limited tactile actuators some VR gloves use. This opens up new possibilities, experiences, and would likely improve performance on VR tasks that require discrimination.

At the conjunctions of robotics and medicine, telesurgery and robot-assisted surgery, haptic feedback is an important aspect. Progress in the area of robot-assisted, minimally-invasive cardiac surgery has been hampered by the lack of haptic feedback (Bethea et al. 2004). Without feedback, surgeons using the da Vinci surgical system, a no-feedback system, struggled to apply appropriate force when performing surgical tasks, like suturing, leading to damaged tissues (Bethea et al. 2004). The da Vinci system is the most prevalent today (El Rassi & El Rassi 2020). Until recently, robot-assisted surgical devices have often lacked any haptic feedback at all, forcing surgeons to over-rely on visual cues to the detriment of their craft (El Rassi & El Rassi 2020). When the option is available, surgeons find haptic feedback valuable, preferring it over no feedback and over audio feedback (Koehn & Kuchenbecker 2015). In the study by Koehn and Kuchenbecker (2015), haptic feedback was delivered in the form of vibrations during contact of the tools with tissue and other tools. Over the past decade, researchers have been investigating ways to deliever cutaneous and kinaesthetic cues

to achieve 'transparency' or 'telepresence', a precise recreation of the remote environment where the surgeon feels as though they are in the remote environment (El Rassi & El Rassi 2020). Force feedback, cutaneous feedback, and vibrotactile feedback all improve performance alone compared to no feedback, but a system that combines them all should lead to more improvement (El Rassi & El Rassi 2020). El Rassi and El Rassi (2020) raise the issue of the difficulty of representing cutaneous feedback and trying to replicate the performance of multiple mechanoreceptors. We argue that the solution does not have to be perfectly veridical. Even sparse cutaneous cues in conjuction with other haptic cues can influence perception, improve discrimination, and be integrated together with the other cues into an overall percept.

Appendix A

Supplemental Derivations

A1 Chapter 2 Supplement: *AVG*

From equation 2.3, the measurements are averaged (summed and divided by three):

$$\hat{s} = \frac{x_{Cut.1} + x_{Cut.2} + x_{Config.}}{3} \tag{A.1}$$

Measurement distributions are modeled as Gaussian. When Gaussian distributions are added, their variances add.

$$\hat{s} = x_{Cut.1} + x_{Cut.2} + x_{Config.} \Rightarrow \sigma_{\hat{s}}^2 = \sigma_{Cut.1}^2 + \sigma_{Cut.2}^2 + \sigma_{Config.}^2 \tag{A.2}$$

The standard deviation (sigma) is the square root of the variance:

$$\sigma_{\hat{s}} = \sqrt{\sigma_{Cut.1}^2 + \sigma_{Cut.2}^2 + \sigma_{Config.}^2} \tag{A.3}$$

Lastly, when multiplying a Gaussian sample by a constant, the standard deviation becomes multiplied by the same constant. So when multiplying $\frac{1}{3}$ to $\hat{s}$, we must do the same to $\sigma_{\hat{s}}$:

$$\sigma_{\hat{s}} = \frac{1}{3}\sqrt{\sigma^2_{Cut.1} + \sigma^2_{Cut.2} + \sigma^2_{Config.}} \tag{A.4}$$

A2 Chapter 2 Supplement: *OPT*

From equation 2.5:

$$\hat{s} = \frac{\frac{x_{Cut.1}}{\sigma^2_{Cut.1}} + \frac{x_{Cut.2}}{\sigma^2_{Cut.2}} + \frac{x_{Config.}}{\sigma^2_{Config.}}}{\frac{1}{\sigma^2_{Cut.1}} + \frac{1}{\sigma^2_{Cut.2}} + \frac{1}{\sigma^2_{Config.}}} \tag{A.5}$$

This equation can be conceptualized as xs multiplied by some weights, ws:

$$\hat{s} = w_{Cut.1}x_{Cut.1} + w_{Cut.2}x_{Cut.2} + w_{Config.}x_{Config.} \tag{A.6}$$

The x-values can be factored out to obtain the weights. For example:

$$w_{Cut.1} = \frac{\frac{1}{\sigma^2_{Cut.1}}}{\frac{1}{\sigma^2_{Cut.1}} + \frac{1}{\sigma^2_{Cut.2}} + \frac{1}{\sigma^2_{Config.}}} \tag{A.7}$$

Over many trials, the xs approximate a Gaussian sample. When multiplying a Gaussian sample by a constant, the result is a Gaussian whose standard deviation had been multiplied by the same constant. Thus:

$$w_{Cut.1}\sigma_{Cut.1} = \sigma_{Cut.1}\frac{\frac{1}{\sigma^2_{Cut.1}}}{\frac{1}{\sigma^2_{Cut.1}} + \frac{1}{\sigma^2_{Cut.2}} + \frac{1}{\sigma^2_{Config.}}} = \frac{\frac{1}{\sigma_{Cut.1}}}{\frac{1}{\sigma^2_{Cut.1}} + \frac{1}{\sigma^2_{Cut.2}} + \frac{1}{\sigma^2_{Config.}}} \tag{A.8}$$

When adding Gaussian samples together, their *variances* add. Thus:

$$\sigma^2_{\hat{s}}=\left(\frac{\frac{1}{\sigma_{Cut.1}}}{\frac{1}{\sigma^2_{Cut.1}}+\frac{1}{\sigma^2_{Cut.2}}+\frac{1}{\sigma^2_{Config.}}}\right)^2+\left(\frac{\frac{1}{\sigma_{Cut.2}}}{\frac{1}{\sigma^2_{Cut.1}}+\frac{1}{\sigma^2_{Cut.2}}+\frac{1}{\sigma^2_{Config.}}}\right)^2+\left(\frac{\frac{1}{\sigma_{Config.}}}{\frac{1}{\sigma^2_{Cut.1}}+\frac{1}{\sigma^2_{Cut.2}}+\frac{1}{\sigma^2_{Config.}}}\right)^2 \quad \text{(A.9)}$$

Simplifying:

$$\sigma^2_{\hat{s}}=\frac{\frac{1}{\sigma^2_{Cut.1}}+\frac{1}{\sigma^2_{Cut.2}}+\frac{1}{\sigma^2_{Config.}}}{\left(\frac{1}{\sigma^2_{Cut.1}}+\frac{1}{\sigma^2_{Cut.2}}+\frac{1}{\sigma^2_{Config.}}\right)^2}=\frac{1}{\frac{1}{\sigma^2_{Cut.1}}+\frac{1}{\sigma^2_{Cut.2}}+\frac{1}{\sigma^2_{Config.}}} \quad \text{(A.10)}$$

The square root of the variance gives us the standard deviation and Equation 2.6:

$$\sigma_{\hat{s}}=\sqrt{\frac{1}{\frac{1}{\sigma^2_{Cut.1}}+\frac{1}{\sigma^2_{Cut.2}}+\frac{1}{\sigma^2_{Config.}}}} \quad \text{(A.11)}$$

A3 Chapter 4 Supplement

From equation 2.5:

$$\hat{s}=\frac{\frac{x_{Cut.1}}{\sigma^2_{Cut.1}}+\frac{x_{Cut.2}}{\sigma^2_{Cut.2}}+\frac{x_{Config.}}{\sigma^2_{Config.}}}{\frac{1}{\sigma^2_{Cut.1}}+\frac{1}{\sigma^2_{Cut.2}}+\frac{1}{\sigma^2_{Config.}}} \quad \text{(A.12)}$$

Let s denote the stimulus level of the conflict stimuli (can have values of -3, -2, 2, or 3). In the conflict conditions, the *Configuration Cue* (*Config.*) is displaced opposite to *Cutaneous Cue 1* (*Cut.1*) and *Cutaneous Cue 2* (*Cut.2*) but with the same magnitude, the percept formula gives the following:

$$\hat{s}=\frac{\frac{-s}{\sigma^2_{Cut.1}}+\frac{-s}{\sigma^2_{Cut.2}}+\frac{+s}{\sigma^2_{Config.}}}{\frac{1}{\sigma^2_{Cut.1}}+\frac{1}{\sigma^2_{Cut.2}}+\frac{1}{\sigma^2_{Config.}}} \quad \text{(A.13)}$$

Rearranged:

$$\hat{s} = \frac{\frac{s}{\sigma^2_{Config.}} + \frac{-s}{\sigma^2_{Cut.1}} + \frac{-s}{\sigma^2_{Cut.2}}}{\frac{1}{\sigma^2_{Cut.1}} + \frac{1}{\sigma^2_{Cut.2}} + \frac{1}{\sigma^2_{Config.}}} \tag{A.14}$$

Factored out, the form gives Equation 4.1:

$$\hat{s} = s\left(\frac{\frac{1}{\sigma^2_{Config.}} - \frac{1}{\sigma^2_{Cut.2}} - \frac{1}{\sigma^2_{Cut.1}}}{\frac{1}{\sigma^2_{Cut.1}} + \frac{1}{\sigma^2_{Cut.2}} + \frac{1}{\sigma^2_{Config.}}}\right) \tag{A.15}$$

Appendix B

Bayesian Analyses

B1 Estimating Sigma

In the Sigma experiments, the reference was always at 0. The mean of the psychometric function does not change, so the psychometric function was defined by the sigma alone. We began by generating a list of hypotheses for sensory sigma across an extensive range of values.

$$\sigma_1 = 0.01mm, \sigma_2 = 0.06mm, \ldots, \sigma_{801} = 40.01mm \tag{B.1}$$

We applied a uniform prior across the sigma hypotheses.

$$P(\sigma_1) = P(\sigma_2) = \ldots = P(\sigma_{801}) = \frac{1}{801} \tag{B.2}$$

On each trial, participants reported which of two discs felt larger. These data are recorded as the proportion of times the comparison was judged greater than the reference. Each sigma hypothesis produces a probability that the comparison will be perceived greater than the reference in a given trial through Equation 3.1:

Assuming the Gaussian random sample

$$P(comp > ref|\Delta) = \int_0^{\text{inf}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\Delta_m - \Delta)^2}{2\sigma^2}\right) d\Delta_m = \Psi(\Delta) \tag{B.3}$$

As depicted in Figure 2.2B, Δ_m comes from a Gaussian measurement distribution centered on the true stimulus value, defined as the difference in mm between the comparison and the reference (size of the comparison disc minus the size of the reference). The Δ is the true stimulus level. The Δ-value of the reference is 0 (it is the size of the reference minus the size of the reference). By integrating from 0 to inf, we calculate the probability that the participant will obtain a comparison measurement greater than the reference measurement.

When participants feel a stimulus in each interval on a given trial, the measurement they receive is a Gaussian sample centered on the true level (Δ). Reporting which of two stimuli felt larger is analagous to taking the difference score of the measurements of each interval in a trial. The variances add. Thus, when estimating the psychometric function, we multiply our sensory sigma hypotheses by $\sqrt{2}$ to find the psychometric sigma.

We calculate the likelihoods for each sigma hypothesis by calculating how probable their data are given the curve defined by σ_i, number of times perceived greater k, and total number of comparisons n.

$$P(D|\sigma_i) = \prod_{\Delta} \Psi(\Delta)^{k_\Delta}(1 - \Psi(\Delta))^{n_\Delta - k_\Delta} \tag{B.4}$$

Subsequently, using Bayes' formula (1.2), with the likelihoods and priors above, allows us to generate posterior *probability distribution function*s (*PDF*s) over sigma. We find which hypothesis number had the greatest posterior, and that sensory sigma is our best estimate for the participant's sigma.

B2 Bayesian Model Comparison Formulae

The marginal likelihood formula below results from the sigma sampling method described in section 2.

$$P(D|M_i) = \sum_{\sigma_{\hat{s}}} P(D|\sigma_{\hat{s}}, M_i) \cdot P(\sigma_{\hat{s}}|M_i) \tag{B.5}$$

For the Cue Conflict, the likelihood formula includes $\mu_{\hat{s}}$.

$$P(D|M_i) = \sum_{\sigma_{\hat{s}}, \mu_{\hat{s}}} P(D|\sigma_{\hat{s}}, \mu_{\hat{s}}, M_i) \cdot P(\sigma_{\hat{s}}, \mu_{\hat{s}}|M_i) \tag{B.6}$$

Using the marginal likelihoods in Bayes' formula results in the posterior probabilities.

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{\sum_{j=1}^{3} P(D|M_j)P(M_j)} \tag{B.7}$$

Bibliography

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience*, *7*(10), 1057–1058.

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

Bensmaia, S., & Manfredi, L. (2012). The sense of touch. *Encyclopedia of Human Behavior: Second Edition, Edited by V.S. Ramachandran*, 379–386.

Bethea, B. T., Okamura, A. M., Kitagawa, M., Fitton, T. P., Cattaneo, S. M., Gott, V. L., Baumgartner, W. A., & Yuh, D. D. (2004). Application of haptic feedback to robotic surgery. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, *14*(3), 191–195.

Björnsdotter, M., & Olausson, H. (2011). Vicarious responses to social touch in posterior insular cortex are tuned to pleasant caressing speeds. *Journal of Neuroscience*, *31*(26), 9554–9562.

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, *33*(7), 919–933.

Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience*, *21*(1), 149–186.

Chen, X., Barnes, C. J., Childs, T. H. C., Henson, B., & Shao, F. (2009). Materials' tactile testing and characterisation for consumer products' affective packaging design. *Materials & Design*, *30*(10), 4299–4310.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409–412.

Craig, J. C., & Johnson, K. O. (2000). The two-point threshold: Not a measure of tactile spatial resolution. *Current Directions in Psychological Science*, *9*(1), 29–32.

Ehrenstein, W. H., & Ehrenstein, A. (1999). Psychophysical methods. *Modern Techniques in Neuroscience Research, Edited by U. Windhorst & H. Johansson*, 1211–1241.

El Rassi, I., & El Rassi, J.-M. (2020). A review of haptic feedback in tele-operated robotic surgery. *Journal of Medical Engineering & Technology*, *44*(5), 247–254.

Fallon, J. B., & Macefield, V. G. (2007). Vibration sensitivity of human muscle spindles and golgi tendon organs. *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, *36*(1), 21–29.

Fienberg, S. E. (2006). When did bayesian inference become "bayesian"? *Bayesian Analysis*, *1*(1), 1–40.

Finnell, J. T., Knopp, R., Johnson, P., Holland, P. C., & Schubert, W. (2004). A calibrated paper clip is a reliable measure of two-point discrimination. *Academic Emergency Medicine*, *11*(6), 710–714.

Gallace, A., & Spence, C. (2010). The science of interpersonal touch: An overview. *Neuroscience & Biobehavioral Reviews*, *34*(2), 246–259.

Geldard, F. A., & Sherrick, C. E. (1972). The cutaneous "rabbit": A perceptual illusion. *Science*, *178*(4057), 178–179.

Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: Weighted averaging and robustness as optimal percepts. *Journal of Vision*, *9*(9), 8–8.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Goldreich, D., & Tong, J. (2013). Prediction, postdiction, and perceptual length contraction: A bayesian low-speed prior captures the cutaneous rabbit and related illusions. *Consciousness Research*, *4*, 221.

Goldreich, D., Wong, M., Peters, R. M., & Kanics, I. M. (2009). A tactile automated passive-finger stimulator (TAPS). *JoVE (Journal of Visualized Experiments)*, (28), e1374.

Goodwin, A. W., & Wheat, H. E. (2004). Sensory signals in neural populations underlying tactile perception and manipulation. *Annual Review of Neuroscience*, *27*, 53.

Gottlieb, G. (1971). Ontogenesis of sensory function in birds and mammals. *The Biopsychology of Development*, 67–128.

Guillery, R. W., & Sherman, S. M. (2002). The thalamus as a monitor of motor outputs. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1428), 1809–1821.

Hecht, D., & Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research*, *193*(2), 307–314.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12), 1–1.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(21), 3621–3629.

Johnson, K. O., & Phillips, J. R. (1981). Tactile spatial resolution. i. two-point discrimination, gap detection, grating resolution, and letter recognition. *Journal of Neurophysiology*, *46*(6), 1177–1192.

Koehn, J. K., & Kuchenbecker, K. J. (2015). Surgeons and non-surgeons prefer haptic feedback of instrument vibrations during robotic surgery. *Surgical Endoscopy*, *29*(10), 2970–2983.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737.

Laplace, P.-S. (1774). Mémoire sur les suites récurro-récurrentes et sur leurs usages dans la théorie des hasards. *Mémoires de l'Académie Royale des Sciences Paris*, *6*, 353–371.

Lederman, S. J., & Jones, L. A. (2011). Tactile and haptic illusions. *IEEE Transactions on Haptics*, *4*(4), 273–294.

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology*, *19*(3), 342–368.

Lederman, S. J., & Taylor, M. M. (1972). Fingertip force, surface geometry, and the perception of roughness by active touch. *Perception & Psychophysics*, *12*(5), 401–408.

Link, S. W. (1994). Rediscovering the past: Gustav fechner and signal detection theory. *Psychological Science*, *5*(6), 335–340.

Löken, L. S., Evert, M., & Wessberg, J. (2011). Pleasantness of touch in human glabrous and hairy skin: Order effects on affective ratings. *Brain Research*, *1417*, 9–15.

Löken, L. S., Wessberg, J., McGlone, F., & Olausson, H. (2009). Coding of pleasant touch by unmyelinated afferents in humans. *Nature Neuroscience*, *12*(5), 547–548.

Luce, R. D., & Krumhansl, C. L. (1988). Measurement, scaling, and psychophysics. *Stevens' Handbook of Experimental Psychology*, *1*, 3–74.

Negen, J., Wen, L., Thaler, L., & Nardini, M. (2018). Bayes-like integration of a new sensory skill with vision. *Scientific Reports*, *8*(1), 1–12.

Norwich, K. H. (1987). On the theory of weber fractions. *Perception & Psychophysics*, *42*(3), 286–298.

Palmer, S. (1978). Fundamental aspects of cognitive representation. *Cognition & Categorization. Edited by E. Rosch & B.B. Lloyd*.

Reidy, M., Kinane, J., Bradley, D., Harbison, J., & McDonagh, R. (2016). Cold, hard cash: Clinical assessment of stereognosis using common objects and coins in older subjects. *European Geriatric Medicine*, *7*(2), 180–182.

Saal, H. P., & Bensmaia, S. J. (2014). Touch is a team effort: Interplay of submodalities in cutaneous sensibility. *Trends in Neurosciences*, *37*(12), 689–697.

Shams, L., & Beierholm, U. (2022). Bayesian causal inference: A unifying neuroscience theory. *Neuroscience & Biobehavioral Reviews*, 104619.

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

Teller, D. Y. (1984). Linking propositions. *Vision research*, *24*(10), 1233–1246.

Tong, J., Mao, O., & Goldreich, D. (2013). Two-point orientation discrimination versus the traditional two-point test for tactile spatial acuity assessment. *Frontiers in Human Neuroscience*, *7*, 579.

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

Vega-Bermudez, F., & Johnson, K. O. (1999). Surround suppression in the responses of primate SA1 and RA mechanoreceptive afferents mapped with a probe array. *Journal of Neurophysiology*, *81*(6), 2711–2719.

von Mohr, M., Kirsch, L. P., & Fotopoulou, A. (2017). The soothing function of touch: Affective touch reduces feelings of social exclusion. *Scientific reports*, *7*(1), 1–9.

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(2), 201.

Yau, J. M., Kim, S. S., Thakur, P. H., & Bensmaia, S. J. (2016). Feeling form: The neural basis of haptic shape perception. *Journal of Neurophysiology*, *115*(2), 631–642.

Yoshioka, T., Craig, J. C., Beck, G. C., & Hsiao, S. S. (2011). Perceptual constancy of texture roughness in the tactile system. *Journal of Neuroscience*, *31*(48), 17603–17611.