# INVESTIGATING HAPTIC CATEGORIZATION USING BAYESIAN INFERENCE

# INVESTIGATING HAPTIC CATEGORIZATION USING BAYESIAN INFERENCE

By GRACE ARTHUR, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2024) Hamilton, Ontario (Psychology, Neuroscience, and Behaviour)

TITLE: Modelling a Haptic Categorization Task using Bayesian Inference

AUTHOR: Grace Arthur, B.Sc. (McMaster University)

SUPERVISOR: Dr. Daniel Goldreich, Ph.D.

NUMBER OF PAGES: xviii, 139

# Lay Abstract

We use our senses every day to accomplish numerous categorization tasks: categorizing footsteps as originating from an 'intruder' or a 'family member', a distant animal as a 'coyote' or a 'dog', a writing utensil as a 'pen' or a 'pencil', and so on. Despite performing countless categorization tasks each day, we often overlook their complexity. Our research investigates the processing behind these tasks, specifically those tasks completed using the sense of touch. We conclude that people combine the most reliable information from their environments to determine the identity of an unknown object or stimulus. Moving forward, we can apply this deepened understanding of tactile processing to advance research in special populations and robotic applications.

# Abstract

We rely heavily on our sense of touch to complete a myriad of tasks each day, yet past research focuses heavily on the visual and auditory systems, rarely concentrating on the tactile system. In the current study, we investigate human performance on a haptic categorization task and ask: what strategy do humans use to sense, interpret, and categorize objects using their sense of touch? During the experiment, participants complete 810 trials on which they receive a 3D printed object and categorize it as belonging to Category A or B. We sample the objects from a set of 25 objects, each of which differs in number of sides and dot spacing on one face. We define Categories A and B using overlapping Gaussian distributions, where Category A objects generally have fewer sides and smaller dot spacing, while Category B objects generally have more sides and larger dot spacing. Participants begin with no knowledge of the categories and learn them using feedback provided on each trial. We compared human performance to a Feature-Focused Bayesian

Observer that weights the sides and dots feature information based on their reliability. It combines information from one or both features to inform a final percept and categorize each object. Our results support the hypothesis that humans employ a feature-focused categorization strategy on this task, during which they learn the categories and consider one or both of an object's features based on their reliability. As participants complete more trials, they appear to maintain or switch to more optimal categorization strategies. Video analysis of hand movements during the experiment strongly supports these findings.

# Acknowledgements

I would like to express my gratitude to everyone who helped and supported me throughout this project. Primarily to my supervisor, Dr. Daniel Goldreich, whose knowledge, guidance, and mentorship developed both my understanding of and enthusiasm for the fields of sensory neuroscience and Bayesian reasoning. His unwavering support made this thesis a fantastic experience. I wish to further acknowledge my committee members, Dr. Patrick Bennett and Dr. Jonathan Cannon, whose guidance and suggestions helped to shape the direction of this project.

I would like to acknowledge Dr. Kyra Gauder, who began this project in the lab and first introduced me to the study, as well as my labmate Rose Jajarmi, who provided me with valuable advice and support. I further thank the undergraduate students who collaborated on this project and contributed significantly to our data collection efforts: Steven Yang, Lyljana Amely Nieto, Noah Ripstein, and Yao Maclean.

# Contents

# List of Figures

# List of Tables

# List of Acronyms and Variables

| | |
|---|---|
| $A$ | Category A |
| "$A$" | Category A Response |
| **ASD** | Autism Spectrum Disorder |
| $B$ | Category B |
| "$B$" | Category B Response |
| $d$ | Measured Dots Feature Value |
| $D$ | Actual Dots Feature Value |
| $d'$ | Disciminability |
| **DO** | Dots Only |
| **EP** | Exploratory Procedure |
| $f$ | Measured Feature Value |
| $\boldsymbol{f}$ | Set of Measured Feature Values |

| | |
|---|---|
| $F$ | Actual Feature Values |
| $g$ | Object Set Index |
| $H$ | Hypothesized Category Distribution |
| $i$ | Trial Number |
| $j$ | Computational Model Index |
| $k$ | Hypothesized Measured Feature Value Index |
| $M$ | Computational Model |
| **MSE** | Mean Squared Error |
| $\mu$ | Category Mean |
| $n$ | Category Hypothesis Index |
| $N$ | Number of As and Bs Presented |
| **PC** | Percent Correct |
| $\boldsymbol{r}$ | Set of Participant Responses on each Experiment Trial |
| $s$ | Measured Sides Feature Value |
| $S$ | Actual Sides Feature Value |
| $\sigma_{cat}$ | Category Sigma |
| $\sigma_{sens}$ | Sensory Sigma |

**SD** Sides & Dots

**SO** Sides Only

**SE** Standard Error

$\theta$ Category A Prior (or Prevalence)

$w$ Model Repetition Index

# Chapter 1

# Introduction

## 1.1 Sensory Perception

As humans, we use our senses to accomplish countless tasks each day. Consider everyday tasks, such as searching the pantry for a specific item, snoozing an early alarm before opening your eyes, and feeling a hot mug of coffee to determine whether it has cooled enough to take a sip. These tasks all employ one or more of a person's sensory systems. We often take our senses for granted, but without them these seemingly effortless tasks would not be possible.

## 1.2 Sensory Perception as Bayesian Inference

Humans gather and compile sensory information to make interpretations about the stimulus from which it originated. Because our sensory abilities have limitations, we base these interpretations on uncertain information acquired from our environments. Imagine, for example, that you go on an evening walk and encounter an animal. Your neighbour recently lost their dog, so while you initially assumed that this was an outdoor cat roaming the neighbourhood, you may have spotted the neighbour's missing dog. You must use the animal's appearance, as well as any sounds it produces, to determine its identity.

Bayesian inference provides a rational mathematical framework to model this sensory process. In a situation with two or more hypotheses about the state of the world, Bayes' theorem quantifies the probability of each hypothesis, given the information at hand (posterior probability, or posterior) (Equation 1.1). In this example, we consider two hypotheses ('dog' and 'cat') and calculate a posterior for each hypothesis, given the animal's shape and sound production. Bayes' theorem quantifies each posterior using a (1) likelihood: the probability of perceiving this sensory information, given the hypothesis and a (2) prior probability (or prior): the probability of a hypothesis before collecting sensory information (Equation 1.1).

$$posterior\ probability \propto likelihood \times prior \tag{1.1}$$

We refer to the posterior and likelihood terms as conditional probabilities, which consider two events, A and B. In our case, A refers to the animal's identity (i.e. the hypothesis: dog or cat) and B refers to the sensory information (sound & shape). We denote the posterior as $P(A \mid B)$ or $P(dog \mid sound\ \&\ shape)$, read as the probability that the animal is a dog, given the sound & shape observed. We denote the likelihood as $P(B \mid A)$ or $P(sound\ \&\ shape \mid dog)$, read as the probability of observing the sound & shape, given the animal is a dog. We consider the same terms for the cat hypothesis. People commonly confuse these terms, falsely claiming that $P(A \mid B) = P(B \mid A)$. Consider the claim that $P(dog \mid fur) = P(fur \mid dog)$. Most (if not all) dogs have fur, making the probability that the animal has fur, given that it is a dog ($P(fur \mid dog)$), very large. However, many other animals (cats, wolves, lions, coyotes, etc.) also have fur, making $P(dog \mid fur)$ much smaller.

$$P(dog \mid sound\ \&\ shape) \propto P(sound\ \&\ shape \mid dog) \times P(dog) \quad (1.2)$$

$$P(cat \mid sound\ \&\ shape) \propto P(sound\ \&\ shape \mid cat) \times P(cat) \quad (1.3)$$

Before collecting any sound & shape information, you know that several people in the neighbourhood have outdoor cats. However, the area only has 1 lost dog. In our Bayesian model, we represent this information as priors, where the prior for cat is greater than that for dog ($P(cat) > P(dog)$).

As you approach this animal, you notice its small shape and arched back, which you commonly associate with cats. That is, the probability of observing a small shape and arched back when looking at a cat is greater than when looking at a dog ($P(sound \ \& \ shape \mid cat) > P(sound \ \& \ shape \mid dog)$). These likelihoods, paired with the greater prior for cat, produce a larger posterior probability for cat than dog ($P(cat \mid sound \ \& \ shape) > P(dog \mid sound \ \& \ shape)$) (equations 1.3 and 1.2). According to the Bayesian model that we have presented in this example, you should now identify the animal as a cat.

As you continue to approach the animal, you hear it bark. In a manner similar to a person's judgement, our Bayesian model incorporates this new information into its likelihood value. Because a dog is much more likely to bark than a cat, our probability of acquiring this sound & shape from a dog becomes much greater than from a cat ($P(dog \mid sound \ \& \ shape) > P(cat \mid sound \ \& \ shape)$), which in turn produces a greater posterior for dog than cat.

Sensory perception requires us to make interpretations using uncertain information. Consequently, such interpretations are sometimes inaccurate. In this vein, an effective mathematical model should simulate human behaviour, including both correct interpretations and incorrect interpretations of an environment. Bayesian models provide a promising mathematical-probabilistic framework for the integration and interpretation of multiple sources of uncertain sensory information (Rohe and Noppeney, 2015). We define an optimal strategy as one that recognizes its own uncertainty and takes steps to avoid or account for this uncertainty, therefore maximizing performance on a task. Conversely, we can define sub-optimal strategies as those that fail to account for uncertainty in an ideal manner, of which there are many. We can construct optimal Bayesian, sub-optimal Bayesian, and non-Bayesian models to compare to human performance on a task.

## 1.3 Tactile Perception

People often dismiss the sense of touch as being less accurate than other senses, in particular vision, and therefore deem touch to be less important. Researchers drew this flawed conclusion from studies that tested the haptic system on tasks geared toward the visual system (Lederman and Klatzky, 1987; Klatzky et al., 1985). For example, many studies evaluate participants' ability to feel and identify raised-line representations of common objects. A raised-line drawing depicts the outline of an object on a flat surface, with the lines recognizable by touch. People perform quite poorly on several variations of this task (Lebaz et al., 2012; Lederman et al., 1990). Humans can easily identify raised line drawings using vision, as the visual system can easily interpret shape with minimal uncertainty. However, touch relies heavily on additional cues, such as weight and texture, that two-dimensional object representations lack. This means that simple raised line stimuli eliminate highly informative cues. Further investigation suggests that during these tasks, participants convert the tactile cues to a visuospatial image, which the visual system then interprets. In other words, such tasks measure an individual's visuospatial imaging abilities, rather than their haptic abilities (Lebaz et al., 2012; Lederman et al., 1990).

Crucially, Klatzky et al. (1985) showed that humans can identify common three-dimensional objects with almost 100% accuracy in a matter of seconds using the sense of touch (Klatzky et al., 1985). Humans employ numerous strategies, each referred to as an exploratory procedure (EP), to extract information about a three-dimensional object's features (weight, texture, shape, etc.). Lederman and Klatzky (1987) defined eight EPs, each specialized for detecting and interpreting a specific object property (Table 1.1).

**Table 1.1: Exploratory Procedures** Brief description and purpose of the 8 EPs defined by Lederman and Klatzky (1987).

| **Exploratory Procedure (EP)** | **Description of Hand Movement** | **Primary Property Under Investigation** |
|---|---|---|
| Lateral Motion | Hand swipes across an object's surface | Texture |
| Pressure | Hand applies force to an object's surface | Hardness |
| Static Contact | Hand maintains contact with an object with no lateral motion | Temperature |
| Unsupported Holding | Hand lifts and independently supports object | Weight |
| Enclosure | Hand encloses object to make as much contact as possible | Volume, Global Shape |
| Contour Following | Hand traces around the perimeter of an object | Volume, Exact Shape |
| Part Motion Test | Hand applies force to part of an object to produce movement | Part Motion |
| Function Test | Hand applies force to part of an object to produce movement associated with a specific object function | Specific Function |

Studies using two-dimensional raised-line representations restrict participants to EPs that can interpret shape. When feeling a three-dimensional object, participants are provided with more properties to investigate (eg. texture, hardness, temperature, weight, and function) and can employ any of the EPs defined by Lederman and Klatzky (1987). Some studies suggest that participants preferentially use EPs best suited to extract the desired information (Lederman and Klatzky, 1987, 1993; Withagen et al., 2013). For example, a participant tasked with differentiating objects of various textures primarily uses lateral motion to explore the object (because it captures the most information), even though static contact pressure could also provide information about object texture (Lederman and Klatzky, 1990; Schwarzer et al., 1999).

## 1.4 Sensory Integration

Real-world stimuli typically consist of two or more physical properties that humans can interpret. Different stimuli have different properties available for interpretation. For example, we can consider face-to-face speech perception as an audiovisual cue integration task. It requires a person to consider an auditory cue (speaker's voice) and a visual cue (speaker's lip movement) to interpret the stimulus (speech content).

Researchers commonly use sensory illusions to investigate cue integration (Alais and Burr, 2004; Bejjanki et al., 2011). If our sensory systems make certain assumptions to process an environment, sensory illusions demonstrate cases in which our sensory systems generate incorrect perceptions of the environment. This provides us with opportunity to investigate the assumptions and processing patterns of our sensory systems that would generate these illusions. The McGurk effect describes a classic audiovisual illusion in which participants interpret an auditory cue differently, depending on the associated visual cue (Mcgurk and Macdonald, 1976). When participants hear the auditory cue, 'da', while watching a speaker's lips mouth the same sound, then they interpret the sound as 'da'. However, when participants hear the same auditory cue while watching a speaker's lips mouth the sound 'ba', then they disproportionately interpret the sound as 'ba'. When the auditory and visual cues become increasingly ambiguous, some participants perceive cues between 'ba' and 'da', such as 'va' or 'ga'. This speech perception task requires participants to integrate an auditory and a visual cue for the sounds 'ba' and 'da', a process studied by Massaro and Cohen (1983). Using a number of auditory stimuli, ranging from 'ba' to 'da' with several intermediate sounds, and either a 'ba' or 'da' visual indicator, they showed participants pairs of audiovisual stimuli. Their results suggest that participants consider both cues to form a speech percept – when the visual stimulus appears as 'ba', then the person is likely to perceive the sound as 'ba' unless the sound very distinctly resembles 'da'. Bejjanki et al. (2011) propose a Bayesian model, which simulated human performance differentiating these sounds and suggests that humans combine these cues in a Bayesian optimal manner.

Alais and Burr (2004) investigate multisensory integration in an alternative audiovisual task where participants combine sensory cues to determine their origin in space. This task simulates the ventriloquism effect, traditionally described as the illusion produced when a puppeteer produces speech while their puppet's mouth moves. In this experiment, participants located a sound source using an auditory and a visual cue that varied in their origin. They effectively modelled this task using an optimal Bayesian observer with the following properties:

- Weight sensory cues according to their reliability. The model weights an ambiguous cue lighter than a cue that reliably represents the hypothesized stimulus. Humans tend to weight visual cues more heavily, so according to their model, people consider the puppet's mouth movement more heavily than the sound source. If, however, the lip movement of the puppet becomes blurred, people begin to interpret the puppeteer as the speaker because the auditory cue becomes more reliable than the visual cue.

- Perform better when integrating both sensory cues than when relying on a single cue.

- Sensed stimuli can differ from the actual stimulus, as the nervous system has sensory noise that leads to imperfect measurements. For example, the actual stimulus may originate in one location, while you may sense that stimulus as originating from a position slightly displaced from the actual origin.

Additional research investigates optimal Bayesian models of sensory integration in additional audiovisual tasks and alternative multisensory combinations, such as visuotactile, sensorimotor, and audiotactile tasks

(Arnold et al., 2019; Ernst and Banks, 2002; Gepshtein and Banks, 2003; Knill and Saunders, 2003; Körding and Wolpert, 2004; Petrini et al., 2014). Their findings suggest that across a variety of sensory integration tasks, Bayesian models effectively simulate human performance.

## 1.5 Perceptual Categorization

Sensory integration involves several levels of interpretation. After perceiving a stimulus, we often strive to further categorize that stimulus. Consider again the animal from Section 1.1. We can combine auditory and visual cues produced by the animal, but maintain an overarching goal to categorize it as a dog or cat. Each category (dog and cat) includes several possible sizes and breeds of animal, making this a surprisingly complex task.

In order to accomplish these types of categorization tasks, we must learn and store categorical representations in the nervous system. In other words, we need to learn and remember categories. Past studies used functional neuroimaging to investigate the neural structures and pathways involved in both unimodal and multisensory category learning (Li et al., 2020; Lim et al., 2019). While these studies reveal the brain's activation patterns during categorization tasks, we still don't fully understand the processing accomplished by brain regions during these tasks. Rosch and Mervis (1975) proposed prototype theory, an early theory of categorization, which argues that the resemblance of an object to a single prototype determines category membership. More recently, researchers have developed a number of different mathematical models to study categorization. These models propose frameworks for category learning that incorporate the calculation of decision boundaries based on the features of previously presented stimuli (Anderson, 1991; Ashby and Gott, 1988).

Previously, I outlined the audiovisual cue integration involved in the McGurk effect, where a visual cue impacts sound perception. This effect also includes a categorization aspect. In addition to integrating two uncertain sensory measurements, a participant must map the stimulus onto established categories with their own variance. People learn the two categories, 'ba' and 'da', through experience. Bejjanki et al. (2011) model this categorization process using an optimal Bayesian observer that weights the sensory cues based on both sensory and category variation. Cues with less variability are more reliable, and therefore weighted more heavily in the observer's categorization decision. The researchers could, for example, present participants with visual stimuli from people who had the same accent as the participant, but auditory stimuli from a person with a different accent. Presumably, the person spends more time around people with the same accent and therefore can more reliably interpret cues from individuals with the same accent. In this case, the auditory cure would become less reliable, and they would weight the visual cue more heavily in their categorization decision. Qualitatively, this model provides a promising framework for sensory categorization. Because the 'ba' and 'da' categories occur naturally, we cannot easily quantify the variance within each category, making it difficult to quantitatively compare this model to humans. Bankieris et al. (2017) propose a similar Bayesian model for an audiovisual categorization task with two novel categories. Their Bayesian categorical model, which considers sensory and categorical variance when weighting cues, best simulates human performance on the task.

## 1.6 Prior Probability of Categories

In the real world, each individual's experience with categories varies. If we again consider the McGurk effect, prior experiences involving the sounds 'ba' and 'da' can further influence a participant's perceived sound. For example, a person named David may hear 'da' more than a person with a different name because, after years frequently hearing and responding to a name starting with the same sound, his prior experience suggests that 'da' occurs more commonly than 'ba'. People acquire this information before even hearing the sound stimulus.

Bayesian models incorporate this life experience variable as a prior. Because every individual has unique life experience, prior probabilities prove difficult to quantify. Many researchers design studies with uniform priors for all hypotheses to avoid the need to quantify this prior.

Some existing studies implement optimal Bayesian models of audiovisual, sensorimotor, and spatiotemporal tasks with variable priors. Their models consider the prior for each possible response based on completed trials (Beierholm et al., 2009; Berniker et al., 2010; Petzschner et al., 2012; Tassinari et al., 2006; Guo et al., 2004; Gredin et al., 2021; Miyazaki et al., 2005). When sensory information becomes less reliable, priors begin to influence our perception more heavily (Hansen et al., 2012). For example, a participant may rely upon their prior for an event more heavily if the current visual stimulus becomes increasingly blurry. Results from these studies suggest that Bayesian priors simulate both the incorporation of existing knowledge in decision making, as well as the process of learning these priors during early testing trials (Beierholm et al., 2009; Berniker et al., 2010; Petzschner et al., 2012;

Tassinari et al., 2006; Guo et al., 2004; Gredin et al., 2021; Miyazaki et al., 2005). When the prior distribution was Gaussian over all hypotheses, Berniker et al. (2010) found that participants can learn the mean of a prior distribution quickly, but take more trials to learn the variance of the prior distribution. Nagai et al. (2012) extended this research to model a tactile temporal order judgement task with two different sources of prior knowledge that can inform participants' decisions. Their results suggests that an optimal Bayesian model, which considers two different visual sources of prior information, effectively models human behaviour.

In categorization studies, researchers commonly design novel categories with nonsense names to study a categorization task in which all participants have no prior experience with the categories (Bankieris et al., 2017). This allows them to control the participant's expectation regarding the probability of receiving a stimulus from each category. Gifford et al. (2014) extended these findings to categorization tasks. They designed an auditory task in which participants group auditory stimuli into one of two overlapping uniform categories. The probability of hearing a stimulus from each category (i.e. the category priors) varied throughout the experiment. Contrary to expectations, their Bayesian probability matching model (initially considered sub-optimal) best simulated human performance, outperforming a Bayesian MAP estimate model. While this Bayesian MAP estimate model categorized stimuli according to the category with the greater posterior probability, the Bayesian probability matching model sampled from the posterior distributions to determine a stimulus' category. They hypothesize that either the combination of category and prior uncertainty, or potentially incorrect stationary prior assumptions made by their model, cause the Bayesian MAP estimate model to poorly simulate human performance.

## 1.7 Current Study

Recent research from Gauder (2024) investigated performance on a haptic categorization task where human participants used their sense of touch to interpret tactile cues from 3D printed objects and sorted them into one of two novel categories, A and B. They defined these categories as overlapping 2D Gaussian distributions, where objects further from the mean belonged to that category less commonly. This research suggested that a Bayesian Observer that experiences sensory measurement noise effectively modelled human performance on their entirely haptic categorization task.

In the present study, we made modifications to this same haptic categorization task. We tested human participants on two different variations of this task, where we manipulated either variation within each category or the prior probability of the categories. Using several variations of a novel Bayesian Observer that learned the task categories trial to trial, each of which uniquely integrated tactile cues, we studied the process by which participants sense, integrate, and categorize these objects.

# Chapter 2

# Computational Models

## 2.1 Introduction

The research cited in Chapter 1 demonstrates that Bayesian models can effectively simulate human performance on several sensory categorization tasks (Bejjanki et al., 2011; Gifford et al., 2014). Despite this large body of work that supports sensory perception as Bayesian inference, researchers have yet to propose a Bayesian model of a solely tactile categorization task to simulate category learning and cue combination required for task performance.

In this chapter, we propose a Feature-Focused Bayesian Observer that models performance on our solely tactile categorization task. We made several variations of this computational observer to reflect potential categorization strategies used to complete the task. We further describe a Bayesian model comparison process, which aimed to determine whether our Feature-Focused Bayesian Observer effectively simulated participant performance on our task. We classified large groups of simulated participants, each of which used a known categorization strategy, to validate the accuracy of our model comparison.

We begin here with a brief outline of the procedure followed in our human experiments. Then, we describe our computational models and experiment simulations. We report the results of the actual human experiments in Chapters 3 and 4.

## 2.2 Methods

### 2.2.1 Haptic Categorization Task

#### Object Set

In this haptic categorization task, we used 3D printed polygons ('objects') designed and printed by Gauder (2024) that differed in number of sides and dot spacing on one textured face (Figure 2.1). Number of sides ranged from 6 to 10, while dot spacing ranged from 4mm to 8mm in increments of 1mm. Our object set included 25 objects, one with each combination of feature measurements.

**(a)** Sample Object **(b)** Full Object Set

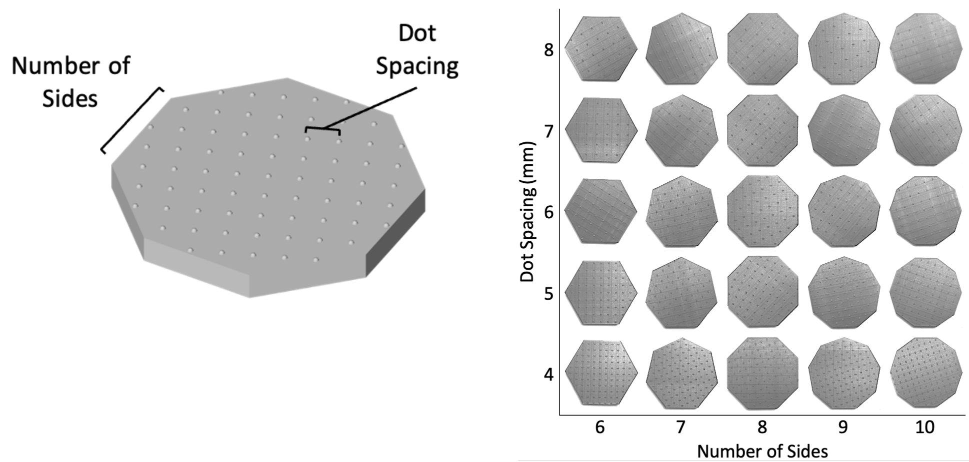



**Figure 2.1: Object Set** (a) Sample object with 8 sides and 6 mm dot spacing. (b) Full set of 25 objects. Each object had a unique combination of sides and dots measurements.

#### Categories

During this experiment, participants categorized several objects as belonging to either Category A or Category B, where the task's premise and procedure closely resemble that described by Gauder (2024). We defined each category using a unique 2D Gaussian distribution, centred at a mean number of sides and dot spacing, with no correlation between the 2 features. Category A was centred at 5mm dot spacing and 7 sides, while category B was centred at 7mm dot spacing and 9 sides (Figure 2.2). Objects with feature measurements closer to the mean were more likely to belong to that category. However, variation within each category caused the categories to overlap, meaning that any object could belong to either category. An object with 7 sides, for example, had a high probability of belonging to Category A and a lower probability of belonging to Category B. We could increase the standard deviation of these category distributions to make them broader with more overlap or decrease the standard deviation to make them narrower and decrease overlap.

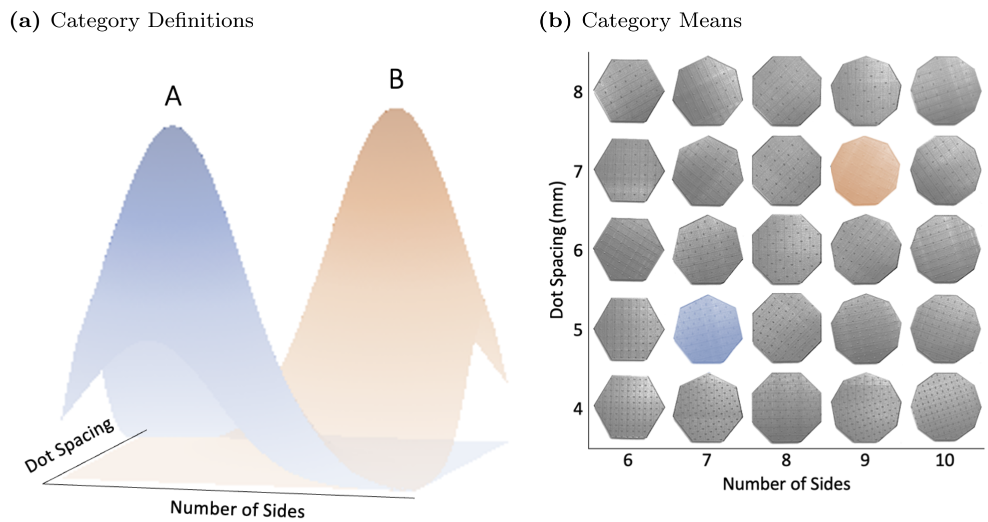



**Figure 2.2: Category A and B Definitions** (a) 2D Gaussian distributions that defined categories A and B. (b) Category A (blue) was centred at 5mm dot spacing and 7 sides, while category B (orange) was centred at 7mm dot spacing and 9 sides. These distributions illustrate category distributions with a sample standard deviation of 0.75 in both dimensions. In our task, standard deviation could differ to make categories broader or narrower.

**Experimental Procedure**

We ran this experiment using LabVIEW v18.0. During the task, each participant completed 810 trials of our haptic categorization task. On each trial, the participant reached their hands through a hole in the bottom of a large, opaque screen. This screen blocked their field of vision and ensured that they used only their sense of touch to complete the task. A 50/50 draw, similar to a coin flip, determined whether the experimenter presented the participant with an object from Category A or B. After selecting a category, the computer randomly sampled feature values (i.e. number of sides and dot spacing) from the 2D Gaussian distribution defining that category. Values close to the mean were more likely to be sampled, while values far from the mean were less likely to be sampled. We presented the corresponding object to the participant to feel for up to 5 seconds, after which a beep sounded to indicate to the participant that their time for haptic exploration had elapsed. The participant categorized the object as A or B and received auditory feedback in the form of a ding or a buzz to indicate a correct or incorrect response, respectively. Because all participants began the task with no knowledge of the categories, they needed to guess the category on the first trial. As they completed trials and received feedback, participants could learn the categories and perform better on the task.

### 2.2.2 Overview of Feature-Focused Bayesian Computational Observer

In this section, we outline the computational basis for a Feature-Focused Bayesian Observer that aimed to simulate human performance on our task. This Observer considered both the proportion of A's and B's presented on past trials and sensory information about the current object's features (sides and dots) to determine its category identity. In Section 2.2.3, we outline the mathematical foundation for this Observer.

To consider several possible categorization strategies that a participant could employ on this task, we generated three versions of this Observer: a Sides & Dots (SD) Observer, a Sides Only (SO) Observer, and a Dots Only (DO) Observer. Our SD Observer considered information about both haptic features in its categorization decision, while our SO and DO Observers considered only one of the two haptic features in categorization decisions. Because this observer interpreted sensory measurements of the object's actual sides and dots feature values, we needed to quantify the sensory noise involved in interpreting the features. A previous Graduate Student conducted a study with 16 participants to quantify the sensory noise associated with each feature (Table 2.1) (Gauder, 2024). On average, participants experienced more sensory noise when feeling the sides feature than the dots feature. Further, participants experienced more sensory noise when attending to both the sides and dots features than when attending to either sides or dots. We referred to the sensory noise associated with attending to a single feature as light sensory noise, while we labelled the sensory noise associated with attending to multiple features simultaneously as heavy sensory noise. This association between sensory noise and number of features under investigation may result from an increased cognitive load associated with attending to multiple features, which manifests as increased sensory noise in participants.

We modelled an SD, an SO, and a DO Observer with heavy sensory noise. To consider the possibility that a participant could focus their attention on a single feature to minimize sensory noise when employing an SO or a DO categorization strategy, we also modelled an SO and a DO Observer with light sensory noise.

**Table 2.1: Sensory Noise Measurements** Average magnitude of sensory noise experienced when feeling the sides and/or dots features of objects, determined by a previous study of 16 subjects using the same object set (Gauder, 2024). We considered sensory noise when a participant attends to both features and when they attend to a single feature.

| **Feature** | **Number of Features Under Investigation** | |
|---|---|---|
| | Multiple (Attending to Sides and Dots) | Single (Attending to Sides or Dots) |
| Number of Sides | 1.65 | 1.26 |
| Dot Spacing (mm) | 1.18 | 0.78 |

### 2.2.3 Feature-Focused Bayesian Observer

Our Feature-Focused Bayesian Observer aimed to simulate human performance on a haptic categorization task. Like humans, sensory noise altered the observer's perception, causing it to sense 'noisy' feature measurements ($f$) of the object's number of sides ($s$) and dot spacing ($d$). To generate these 'noisy' feature measurements, we randomly sampled a value from a Gaussian distribution centred at the object's actual feature value, $F$ (actual number of sides, $S$, and actual dot spacing, $D$), with standard deviation equal to the amount of sensory noise associated with the sides ($\sigma_{sensS}$) and dots features ($\sigma_{sensD}$) (see Table 2.1 for values). On each trial, our Feature-Focused Bayesian Observer received 'noisy' sides and dots measurements as input.

Because Categories A and B overlapped, both category and sensory noise limit the ability of an observer to discriminate between A's and B's. If we assumed that the category distributions were continuous, extended infinitely, and had equal sigma values in both dimensions (i.e. our 0.75/0.75 and 1.25/1.25 conditions), then we could quantify this discriminability as $d'$ (Equation 2.1). $d'$ depended on the amount of category ($\sigma^2_{catS}$ and $\sigma^2_{catD}$) and sensory ($\sigma^2_{sensS}$ and $\sigma^2_{sensD}$) variance, as well as the the distance between the means of Categories A ($\mu_{S,A}$ and $\mu_{D,A}$) and B ($\mu_{S,B}$ and $\mu_{D,B}$).

$$d' = \sqrt{\frac{(\mu_{S,A} - \mu_{S,B})^2}{\sigma^2_{sensS} + \sigma^2_{catS}} + \frac{(\mu_{D,A} - \mu_{D,B})^2}{\sigma^2_{sensD} + \sigma^2_{catD}}} \tag{2.1}$$

Larger $d'$ values occurred with less variance and/or more distance between category means, resulting in the increased ability of an observer to discriminate between A's and B's. Conversely, low $d'$ values indicated a decreased ability of an observer to discriminate between A's and B's. Note that we use $d'$ to conceptualize discriminability in our experimental conditions, but the modelling outlined in this chapter did not incorporate $d'$ in its calculations.

Once our Observer 'felt' the object, it had two goals: (1) learn the categories and (2) categorize the object. The Observer completed these two steps on each trial of the experiment.

#### Learning the Categories Using Bayesian Parameter Estimation

**Hypotheses** The Feature-Focused Bayesian Observer began this experiment with no knowledge of the two categories. It assumed that the categories were Gaussian and generated $n = 1225$ hypotheses for each category ($H_{n,A}$, $H_{n,B}$). Each hypothesis defined four parameters describing a unique 2D Gaussian distribution: mean number of sides ($\mu_S$ ranging from 6 to 10 in increments of 1 side), mean dot spacing ($\mu_D$ ranging from 4 to 8 in increments of 1mm), standard deviation in the sides dimension ($\sigma_{catS}$ from 0.5 to 2 in increments of 0.25), and standard deviation in the dots dimension ($\sigma_{catD}$ from 0.5 to 2 in increments of 0.25) (Figure 2.3).

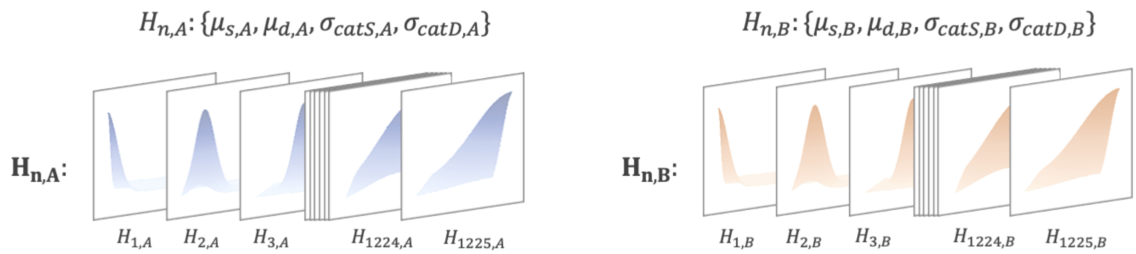
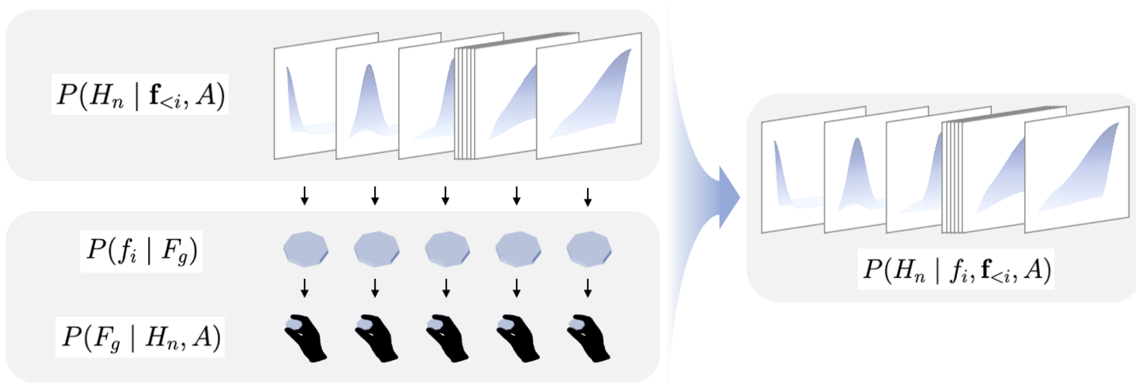



**Figure 2.3: Category Hypotheses** Five examples of hypothesized 2D Gaussian distributions that could describe Category A and Category B. Each distribution was centred at a different mean sides and mean dots measurement: (A) $\mu_S = 6$, $\mu_D = 4$ mm, $\sigma_{catS,D} = 0.75$ (B) $\mu_S = 7$, $\mu_D = 5$ mm, $\sigma_{catS,D} = 0.75$ (C) $\mu_S = 8$, $\mu_D = 6$ mm, $\sigma_{catS,D} = 0.75$ (D) $\mu_S = 9$, $\mu_D = 7$ mm, $\sigma_{catS,D} = 2.0$ (E) $\mu_S = 10$, $\mu_D = 8$ mm, $\sigma_{catS,D} = 2.0$. This Feature-Focused Bayesian Observer considered 1225 hypothesized categories, each a set of 4 values: $\mu_S$,$\mu_D$,$\sigma_{catS}$,$\sigma_{catD}$.

The Feature-Focused Bayesian Observer quantified a likelihood and, in turn, a posterior for each hypothesis, as outlined in Figure 2.4.

**(a)** Overview of Category A Learning




**(b)** Overview of Category B Learning

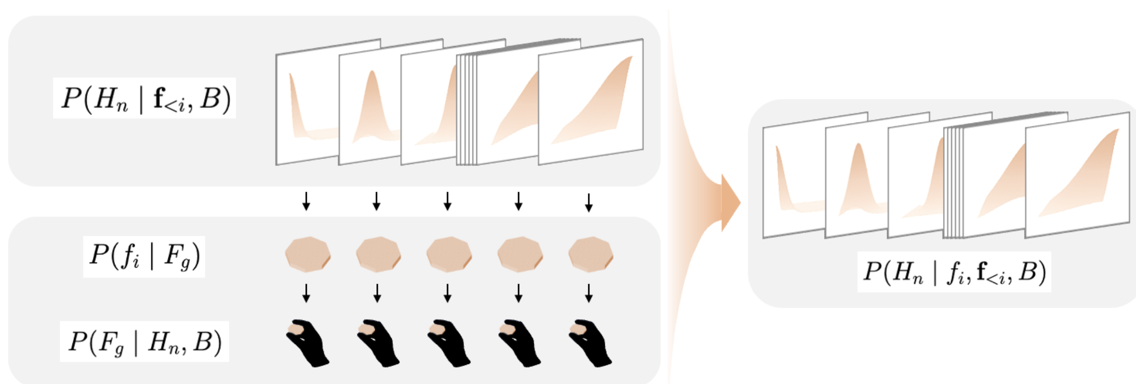



**Figure 2.4: Overview of Category Learning** To learn the Category A and B distributions, this Feature-Focused Bayesian Observer used Bayesian parameter estimation to determine the mean and standard deviation of each category distribution. (a) and (b) outline this learning for Categories A and B, respectively. The Observer first generated hypotheses for each distribution as sets of 4 variables, $\{\mu_S, \mu_D, \sigma_{catS}, \sigma_{catD}\}$, each defining a potential distribution for one of the categories ($P(H_n \mid f_i, \mathbf{f}_{<i}, A)$ and $P(H_n \mid f_i, \mathbf{f}_{<i}, B)$). For each hypothesis, the Observer quantified a likelihood as the probability of sampling the object presented from the current trial from the hypothesized distribution and the probability of 'feeling' the measured feature value from that object ($P(f_i \mid H_n, A)$ and $P(f_i \mid H_n, B)$). Using each prior (posterior probability for that hypothesis from the previous trial) and likelihood, the observer determined the posterior probability that the hypothesis correctly describes the actual category distribution ($P(H_n \mid f_i, \mathbf{f}_{<i}, A)$ and $P(H_n \mid f_i, \mathbf{f}_{<i}, B)$).

**Likelihoods** The Observer calculated the likelihood of each hypothesis, that is, the probability that the feature measurements came from an object sampled from the hypothesized category and presented on the current trial, $i$ (Equation 2.2). Because our Feature-Focused Bayesian

Observer was aware of its own sensory noise, the likelihood of a hypothesis depended on both the probability of 'feeling' the measured object features from a particular object ($P(f_i \mid F_g)$) and the probability of sampling that object from the hypothesized category distribution ($P(F_g \mid H_n, A)$). The observer was more likely to 'feel' feature measurements close to the actual object's feature values, but could theoretically sense a given feature measurement from any of the 25 objects. Like humans completing this task, the observer sensed only noisy feature measurements and therefore remained unaware of the actual feature values of the actual object presented. Equation 2.2 quantifies a likelihood for each hypothesized distribution, that is, the probability of 'feeling' the measured features, given that they resulted from the hypothesized distribution.

$$P(f_i \mid H_n) = \sum_{g=1}^{25} [P(f_i \mid F_g) \times P(F_g \mid H_n)] \tag{2.2}$$

Note that here we show the calculations for Category A. We performed equivalent calculations for Category B.

Our SD Observer considered both object features and quantified $P(f_i \mid F_g)$ as the product of two independent Gaussian distributions (Equation 2.3). Similarly, this SD Observer quantified $P(F_g \mid H_n, A)$

as the product of two independent Gaussian distributions, one for each feature (Equation 2.4). The mean and standard deviation values considered in each Gaussian of Equation 2.4 equaled those defined by the current hypothesis. Our SO and DO Observers considered only one object feature and therefore quantified $P(f_i \mid F_g)$ and $P(F_g \mid H_n, A)$ using a single Gaussian distribution, corresponding to the single feature considered.

$$P(f_i \mid F_g) = \frac{1}{\sigma_{sensS}\sqrt{2\pi}} e^{\frac{-(s_i - S_g)^2}{2\sigma^2_{sensS}}} \times \frac{1}{\sigma_{sensD}\sqrt{2\pi}} e^{\frac{-(d_i - D_g)^2}{2\sigma^2_{sensD}}} \tag{2.3}$$

$$P(F_g \mid H_n, A) = \frac{1}{\sigma_{catS,A}\sqrt{2\pi}} e^{\frac{-(Sg - \mu_{S,A})^2}{2\sigma^2_{catS,A}}} \times \frac{1}{\sigma_{catD,A}\sqrt{2\pi}} e^{\frac{-(Dg - \mu_{D,A})^2}{2\sigma^2_{catD,A}}} \tag{2.4}$$

Our Feature-Focused Bayesian Observer was more likely to sense feature values close to the mean of a hypothesized category distribution. The probability of sensing a feature measurement from an object with very dissimilar feature value was greater in a model with more sensory noise than one with less sensory noise.

**Posterior Probabilities** We calculated the posterior probability for each hypothesized distribution for a given category using Equation 2.5, which defines Bayes' theorem for an $n$ hypothesis problem considering data from both the current trial ($f_i$) and data from all previous trials. Data from previous trials included both sensory measurements ($\mathbf{f}_{<i}$) and the number of As and Bs presented ($N_{<i}$).

$$P(H_n \mid f_i, \mathbf{f}_{<i}, A) = \frac{P(f_i \mid H_n, A) \times P(H_n \mid \mathbf{f}_{<i}, A)}{\sum_{H_n} \left[P(f_i \mid H_n, A) \times P(H_n \mid \mathbf{f}_{<i}, A)\right]} \tag{2.5}$$

On trial 1, all hypotheses had an equal probability of correctly describing the category (i.e. uniform priors, $P(H_{1,A}) = P(H_{2,A}) = ... = P(H_{1225,A})$). This simulated a participant's lack of knowledge regarding the category definitions. Accordingly, the posterior probability of each hypothesis equaled its likelihood divided by the sum of the likelihoods for all hypothesized categories (Equation 2.5).

**Updating Probabilities of the Category Distributions** Following trial 1, the Feature-Focused Bayesian Observer possessed information that informed its understanding of the Category A and B distributions. For example, we may sample an object from category A with 6 sides and 4mm dot spacing to present on a given trial. Its features are closer to the mean of Category A than Category B, providing the Observer with evidence that the mean of Category A is close to 6 sides and 4mm dot spacing.

As the observer completed trials, hypotheses were continuously re-weighted to reflect the observer's knowledge of Categories A and B. We quantified this acquired knowledge as prior probabilities for each hypothesis. At the end of each trial, the model learned through feedback whether the object belonged to Category A or B. If the object belonged to Category A, the Feature-Focused Bayesian Observer used the posteriors for each hypothesized distribution ($P(H_n \mid f_i, \mathbf{f}_{<i}, A)$) as priors on the following trial. If the object belonged to Category B, the Observer used the posteriors for each hypothesized distribution ($P(H_n \mid f_i, \mathbf{f}_{<i}, B)$) as priors on the following trial.

#### Categorizing the Object

In each trial, the observer also needed to categorize an object as "A" or "B". Even though the Feature-Focused Bayesian Observer remained uncertain as to the actual distributions defining each category, it had already calculated the prior probability and likelihood for all the hypothesized distributions that could define each of the 2 categories (outlined above). Considering this information, the observer used the measured feature values for the sides and dots features to determine the category to which the object was more likely to belong (Figure 2.5).

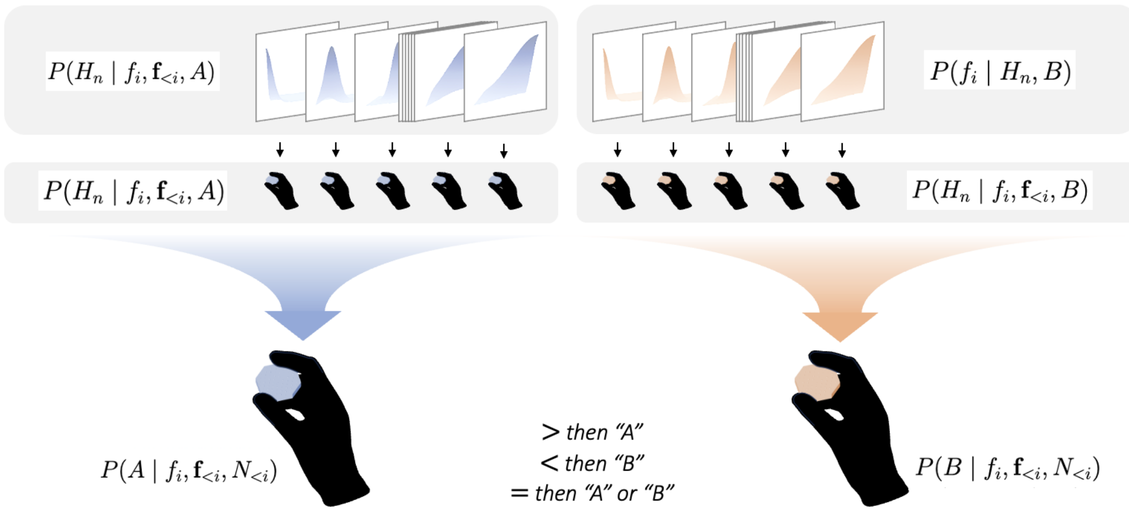



**Figure 2.5: Overview of Object Categorization** Outline of the Bayesian calculations used to simulate our Feature-Focused Bayesian Observer's object categorization process on a single trial of a haptic categorization task. The Observer considered all hypothesized categories and the likelihood of those categories correctly describing Categories A and B to identify the category from which the measured feature was more likely to be 'felt'.

We quantified the likelihood that the measured object features were sampled from Category A, considering all 1225 potential Category A distributions, using Equation 2.6. We quantified the corresponding likelihood for Category B using Equation 2.7.

$$P(f_i \mid A, \mathbf{f}_{<i}) = \sum_{H_n} [P(f_i \mid H_n, A) \times P(H_n \mid \mathbf{f}_{f_i}, \mathbf{f}_{<i}, A)] \tag{2.6}$$

$$P(f_i \mid B, \mathbf{f}_{<i}) = \sum_{H_n} [P(f_i \mid H_n, B) \times P(H_n \mid \mathbf{f}_{f_i}, \mathbf{f}_{<i}, B)] \tag{2.7}$$

Next, we used Bayes' formula (Equation 2.8) to calculate the posterior probability that the object belonged to Category A.

$$P(A \mid f_i, \mathbf{f}_{<i}, \theta) = \frac{P(f_i \mid A, \mathbf{f}_{<i}) \times \theta}{P(f_i \mid A, \mathbf{f}_{<i}) \times \theta + P(f_i \mid B, \mathbf{f}_{<i}) \times (1 - \theta)} \tag{2.8}$$

In Equation 2.8, $\theta$ denotes the observer's prior probability for (or the prevalence of) Category A. In some cases, we informed the observer that we planned to present equal proportions of objects from both categories. In this case, the prevalence of objects from each category was equal ($\theta = 0.5$).

If the model remained naive to the proportion of A's and B's that we planned to present, we used Equation 2.10 to estimate the prevalence of objects from Category A ($\theta$) based on the number of A's and B's presented on previous trials ($N_{<i}$) for 10 hypothesized $\theta$ values using Bayesian parameter estimation, a calculation further outlined in Section 2.2.4. Then, we calculated the posterior probability that the object belonged to Category A using each of the hypothesized $\theta$ values (Equation 2.8). Following that, we used Equation 2.9 to marginalize over $\theta$ and determine the probability that the object belonged to Category A, considering all of the potential $\theta$ values.

$$P(A \mid f_i, \mathbf{f}_{<i}, N_{<i}) = \sum_{\theta} P(A \mid f_i, \mathbf{f}_{<i}, \theta) \times P(\theta \mid N_{<i}) \tag{2.9}$$

If the posterior probability of Category A $> 0.5$, then our Feature-Focused Bayesian Observer categorized the object as "A". If the posterior probability of Category A $< 0.5$, then our Feature-Focused Bayesian Observer categorized the object as "B". If the posteriors for both categories were equal, then the Observer randomly categorized the object.

### 2.2.4 Non-Sensory Computational Observers

Despite the valuable information available from a stimulus' properties, individuals could dismiss this information and rely solely on the prevalence of stimuli from each group to make categorization decisions. We modelled 3 non-sensory reference observers to simulate such strategies. We later used these observers to verify that participants reasonably attempted our haptic categorization task.

#### Null Observer

We modelled a Null Observer that employed the least optimal approach of those considered. It received no sensory input about an object and randomly categorized it as "A" or "B" with equal chance.

#### Prevalence MAP Estimate Observer

We further modelled a Prevalence MAP Estimate Observer, which considered the proportion of A's and B's presented on past trials to determine whether the current object belonged to Category A or B. This observer considered 10 values as possible proportions of all objects belonging to Category A (or prevalence, $\theta$, values), ranging from 0.05 to 0.95 in increments of 0.1. It used the number of A's and B's presented over past trials ($N_{<i}$) to quantify the probability that each $\theta$ value gave rise to the proportion of each category presented. We express these likelihoods using Equation 2.10.

$$P(N_{<i} \mid \theta) \propto (\theta^{N_{A,<i}}) \times (1 - \theta^{N_{B,<i}}) \tag{2.10}$$

We considered all $\theta$ hypotheses equally with uniform priors. As a result, the posterior probability for each $\theta$ hypothesis, $P(\theta \mid N_{<i})$, equaled its likelihood divided by the sum of the likelihoods for all hypothesis.

Using these posteriors, we calculated the probability that the current trial's object belonged to category A, $P(A \mid N_{<i})$, based exclusively on the prevalence of the categories from all previous trials (Equation 2.11).

$$P(A \mid N_{<i}) = \sum_{\theta} [\theta \times P(\theta \mid N_{<i})] \tag{2.11}$$

Because the object must belong to either Category A or B, $P(B \mid N_{<i}) = 1 - P(A \mid N_{<i})$. This Prevalence MAP Estimate categorized the object according to the larger probability, $P(A \mid N_{<i})$ or $P(B \mid N_{<i})$.

If $P(A \mid N_{<i}) = P(B \mid N_{<i})$, then the observer randomly categorized the object as "A" or "B" with equal chance.

#### Prevalence Probability Matching Observer

This final non-sensory observer performed the same calculations to approximate the prevalence of Category A (Equation 2.10) and probability that the current trial's object belonged to Category A (Equation 2.11) as the Prevalence MAP Estimate Observer. It then flipped a coin with two sides, "A" and "B", to categorize the object, where the probability of landing on "A" equalled $P(A \mid N_{<i})$, determined by Equation 2.11.

### 2.2.5 Comparing our Computational Observers to a Participant using Bayesian Model Comparison

We further compared participants to these computational observers using Bayesian model comparison. This process quantified the degree to which each observer can simulate participant performance, relative to our other computational observers.

In this chapter, we compared our computational observers to simulated participants. To simulate a participant, we assigned them sensory noise values ($\sigma_{sensS}$ and $\sigma_{sensD}$) and ran them through 810 trials of the experiment. On each trial, the simulated participant received an object from one of the categories and drew a sensory measurement of the features based on their sensory noise sigma. Using these feature measurements, the simulated participant learned the categories and categorized objects according to the calculations previously outlined for our computational observers. While we ran each simulated participant once through the experiment, we ran 10 individual repetitions to implement each computational observer in order to account for variation in model performance resulting from sensory noise. We compared each participant to a unique set of computational models, generated using the same sequence of objects, to prevent performance differences resulting from object sequence.

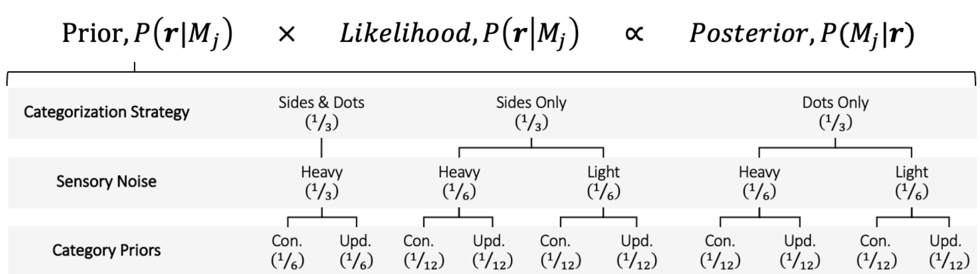



**Figure 2.6: Overview of Model Comparison** To identify the computational observer that best simulated a participant's performance, we used Bayesian model comparison to quantify the probability that a given Observer produced the same response sequence as the participant (i.e. its posterior probability). The Observer most likely to produce the same response sequence best simulated the participant. This calculation considered a prior and likelihood for each computational observer, or model, to quantify its posterior probability. The model priors depended on the computational observers considered, and are specified in parentheses next to the corresponding observer. For example, if only 3 models (SD, SO, and DO) were considered, then the prior for each model was $\frac{1}{3}$. However, if we considered a heavy and light sensory noise model for the SO and DO Observers, then the priors were $\frac{1}{3}$ for the SD Observer and $\frac{1}{6}$ for the SO and DO Observers. We followed the same structure for observers that assumed equal category prevalence (Con.) and those that did not make this assumption (Upd.).

**Prior Probabilities** The prior probabilities for each computational observer, or model ($M_j$), depended on the number and type of observers considered. Figure 2.6 outlines the model priors, which assumed that a participant was equally likely to employ an SD, an SO, or a DO categorization strategy. If we considered only these 3 Observers, then each model had a prior of $\frac{1}{3}$. This changed if we considered that a participant's sensory noise may differ, choosing to model the SO and DO Observers with both heavy and light sensory noise. In this case, the SD Observer had a prior of $\frac{1}{3}$, while each SO and DO Observer had a prior of $\frac{1}{6}$. We could further divide each Observer based on prevalence assumptions, where one Observer assumed equal (constant) prevalence of Categories A and B (Con.) and a second estimated (updates) the prevalence of each category (Upd.).

For our reference Null and Prevalence-Focused Observers, rather than assign a prior to these improbable models, we instead simply calculated a Bayes factor to compare each of these observers to the observer with the highest likelihood (Equation 2.12).

$$Bayes\ Factor = \frac{maximum\ liklihood\ across\ all\ observers}{likelihood\ of\ Null\ or\ Prevalence{-}Focused\ Observer} \tag{2.12}$$

**Likelihoods** We denoted the actual object sampled on a given trial, i, as $F_i$. By the end of the experiment, each participant produced a sequence of 810 categorization responses, which we denoted as **r**. On a given trial, $x = 1$ if the participant responded “A” and $x = 0$ if the participant responded “B”. Note that “A” represented a participant response, which may or may not align with the actual object category. We aimed to determine the probability that each model, $M_j$, would produce the same response sequence (Equation 2.13). Equation 2.13 used a given model’s unique probability of responding “A” or “B” when presented with $F_i$ on each trial of a given repetition ($P(\text{“}A\text{”}_{w,i} \mid F_i)$ and $P(\text{“}B\text{”}_{w,i} \mid F_i)$) to express the probability that this model would produce the participant’s response sequence. We included a lapse rate of 0.01, assuming that on 1% of trials, participants randomly responded “A” or “B”. Note that because the participant must categorize the object as A or B, we defined $P(\text{“}B\text{”}_{w,i} \mid F_i)$ as $1 - P(\text{“}A\text{”}_{w,i} \mid F_i)$.

We further averaged this probability across $w = 10$ repetitions using Equation 2.14, which obtained unique sensory measurements of an object’s features on each trial to account for a range of sensory measurements that could be made by a participant completing the same task. See Sections 2.2.5: Feature-Focused Bayesian Observer, 2.2.5: Null Observer, and 2.2.5: Prevalence-Focused Observer for observer specific $P(\text{“}A\text{”}_{w,i} \mid F_i)$ values.

$$P(\mathbf{r} \mid M_{j,w}) = \prod_{i=1}^{810} [(0.01)(0.5) + (0.99)(P(\text{“}A\text{”}_{w,i} \mid F_i)^{x_i} \times P(\text{“}B\text{”}_{w,i} \mid F_i)^{1-x_i})] \tag{2.13}$$

$$P(\mathbf{r} \mid M_j) = \frac{1}{10} \sum_{w=1}^{10} P(\mathbf{r} \mid M_{j,w}) \tag{2.14}$$

**Posterior Probabilities** We used the priors and likelihoods calculated for each model to determine the posterior probability that it simulated a participant's categorization strategy, given their response sequence (Equation 2.15). We classified the observer with the largest posterior as the one that most accurately simulated a participant's categorization strategy.

$$P(M_j \mid \mathbf{r}) = \frac{P(\mathbf{r} \mid M_j) \times P(M_j)}{\sum_{M_j} [P(\mathbf{r} \mid M_j) \times P(M_j)]} \tag{2.15}$$

**Calculating $\mathbf{P(\text{“}A\text{”}_{w,i} \mid F_i)}$ for a Feature-Focused Bayesian Observer**

Our Feature-Focused Bayesian Observer considered sensory noise, category noise, and category learning to calculate $P(\text{“}A\text{”}_{w,i} \mid F_i)$. To start, we considered $k$ possible measurements ($f_k$) that the observer may have 'felt' when given the actual object, $F_i$ (Equation 2.16). We calculated the probability of 'feeling' each $f_k$ using a 2D Gaussian distribution centred at the actual feature measurements of $F_i$ with a standard deviation equal to the observer's sensory noise. Each $f_k$ resulted in either an "A" or a "B" categorization. The sum of all $P(f_k \mid F_i, w)$ where the observer 'felt' $f_k$ and categorized the object as "A" sum to $P(\text{“}A\text{”}_{w,i} \mid F_i)$. We performed the same calculation for Category B.

$$P(\text{“}A\text{”}_{w,i} \mid F_i) = \sum_k P(f_k \mid F_i, w) \times 1 \; if \; P(A \mid f_{k,i}, w) > 0.5, else, 0 \tag{2.16}$$

If we specified to the model at the outset that we planned to present equal numbers of A's and B's during the task (i.e. $\theta = 0.5$), then we determined which $f_k$ values resulted in an "A" categorization response using Equation 2.17. A model responded "A" when the probability that an object was sampled from Category A, considering the feature measurement ($P(A \mid f_{k,i}, \theta)$), was greater than 0.5. Equation 2.17 used Bayes' theorem to calculate $P(A \mid f_{k,i}, \theta)$ based on the likelihood of obtaining $f_k$ from objects of either category.

$$P(A \mid f_{k,i}, \theta) = \frac{P(f_{k,i} \mid A) \times \theta}{P(f_{k,i} \mid A) \times \theta + P(f_{k,i} \mid B) \times (1 - \theta)} \tag{2.17}$$

If we did not specify the prevalence of Category A or B, then we considered 10 $\theta$ hypotheses ranging from 0.05 to 0.95 and used Equation 2.17 to calculate $P(A \mid f_{k,i}, \theta)$ for each $\theta$. Using Equation 2.18, we calculated $P(A \mid f_{k,i})$, where we weighted each $P(A \mid f_{k,i}, \theta, \mathbf{f}_{<i})$ value using the probability that $\theta$ correctly defined the category prevalence. These weights were previously calculated using the likelihoods from Equation 2.10.

$$P(A \mid f_{k,i}) = \sum_{\theta} [P(A \mid f_{k,i}, \theta, \mathbf{f}_{<i}) \times P(\theta \mid N_{<i})] \tag{2.18}$$

Equation 2.17 considered the probability of 'feeling' $f_k$ from an object belonging to Category A or B, values that we calculated using Equation 2.19. Because any object could belong to either category and the model did not know the actual feature values of $F_i$, we considered that $F_i$ could be any one of the 25 objects ($g = 25$). We calculated the probability that any object, $F_{g,i}$, could give rise to the perceived measurement ($P(f_{k,i} \mid F_{g,i})$) and the probability of sampling $F_{g,i}$ from Category A ($P(F_{g,i} \mid A)$. We quantified $P(f_{k,i} \mid F_{g,i})$ using a 2D Gaussian distribution centred at the actual feature measurements of $F_{g,i}$ with a standard deviation equal to sensory noise. We quantified $P(F_{g,i} \mid A)$ using Equation 2.20, which considered many hypothesized

2D Gaussian distributions that could describe Category A. This calculation considered both the probability of randomly sampling $F_{g,i}$ from the hypothesized category and the probability of that hypothesis correctly describing the actual category, based on what the observer has learned about Category A on completed trials ($\mathbf{f}_{<i}$ and $N_{<i}$). Because our model learned new information about the categories on each trial, these values changed on each trial. We used equations 2.19 and 2.20 to calculate $P(f_{k,i} \mid B)$ and $P(F_{g,i} \mid B)$ as well.

$$P(f_{k,i} \mid A) = \sum_{g=1}^{25} P(f_{k,i} \mid F_{g,i}) \times P(F_{g,i} \mid A) \tag{2.19}$$

$$P(F_{g,i} \mid A) = \sum_{H_n} P(F_{g,i} \mid H_n, A) \times P(H_n \mid \mathbf{f}_{<i}, N_{<i}, A) \tag{2.20}$$

Calculations for SD Observers considered both feature measurements throughout these calculations. Conversely, SO and DO Observers considered only the corresponding feature in their calculations.

#### Calculating P(“A”$_{\mathbf{w,i}}$ | F$_{\mathbf{i}}$) for a Null Observer

Our Null Observer categorized objects as “A” or “B” on trials with equal probability. As a result, $P(\text{“}A\text{”}_{w,i} \mid F_i) = P(\text{“}B\text{”}_{w,i} \mid F_i) = 0.5$.

#### Calculating P(“A”$_{\mathbf{w,i}}$ | F$_{\mathbf{i}}$) for a Prevalence MAP Estimate Observer

Our Prevalence MAP Estimate Observer based $P(\text{“}A\text{”}_{w,i} \mid F_i)$ on the magnitude of $P(A \mid N_{<i})$ calculated using Equation 2.11. If $P(A \mid N_{<i}) > 0.5$, then $P(\text{“}A\text{”}_{w,i} \mid F_i) = 1$. Similarly, if $P(A \mid N_{<i}) < 0.5$, then $P(\text{“}A\text{”}_{w,i} \mid F_i) = 0$.

#### Calculating P(“A”$_{\mathbf{w,i}}$ | F$_{\mathbf{i}}$) for a Prevalence Probability Matching Observer

Our Prevalence Probability Matching Observer used Equation 2.11 to calculate the probability that a given object belonged to Category A based on the proportion of A’s and B’s previously presented, $P(A \mid N_{<i})$. On each trial, this $P(A \mid N_{<i})$ value equalled $P(\text{“}A\text{”}_{w,i} \mid F_i)$.

## 2.3 Results

We produced computational observers for two variations of our haptic categorization task.

### 2.3.1 Manipulate Category Variation

In a first variation of our haptic categorization task, we manipulated the variation of the 2D Gaussian distributions defining Categories A and B. We created 4 experimental conditions to test low, moderate, and high category overlap. We define these conditions in Figure 2.7.

In all conditions, we specified at the outset that we were equally likely to present an object from each category on a given trial during the experiment.

$d'$ values for low and high overlap, or for the 0.75/0.75 and 1.25/1.25 conditions, using an SD strategy with heavy sensory noise equalled 1.81 and 1.51, respectively. $d'$ values for each degree of overlap using a DO strategy with light sensory noise equalled 1.85 and 1.36, respectively. Note that $d'$ for a DO (Light) strategy is greater than for an SD strategy in the 0.75/0.75 condition, while an SD strategy has a larger $d'$ in the 1.25/1.25 condition.

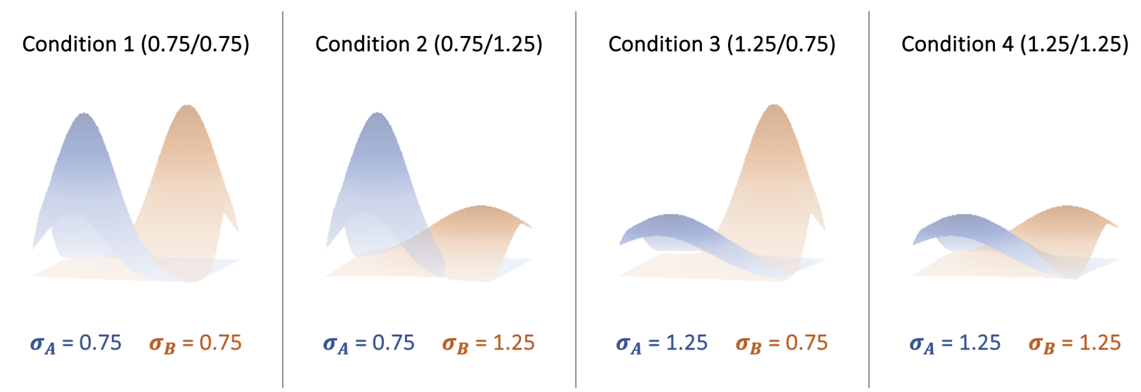



**Figure 2.7: Category Variation Conditions** Summary of the four experimental conditions with unique category standard deviation of Categories A and B, each described by 2D Gaussian distributions. These standard deviation ($\sigma$) values apply to both dimensions of the category distributions.

We produced 6 computational observers: Null Observer, SD Observer (heavy sensory noise), SO Observer (heavy sensory noise), SO Observer (light sensory noise), DO Observer (heavy sensory noise), and

DO Observer (light sensory noise). Each observer produced a learning curve that illustrates the cumulative percent correct (PC) achieved on each trial of the experiment. Figure 2.8 shows sample learning curves of performance on this haptic categorization task. Learning curves for our Null Observer maintained a consistent PC at approximately 50% correct on all trials. Learning curves for all variations of our Feature-Focused Bayesian Observer increased in PC during early trials and approached asymptotic performance in later trials.

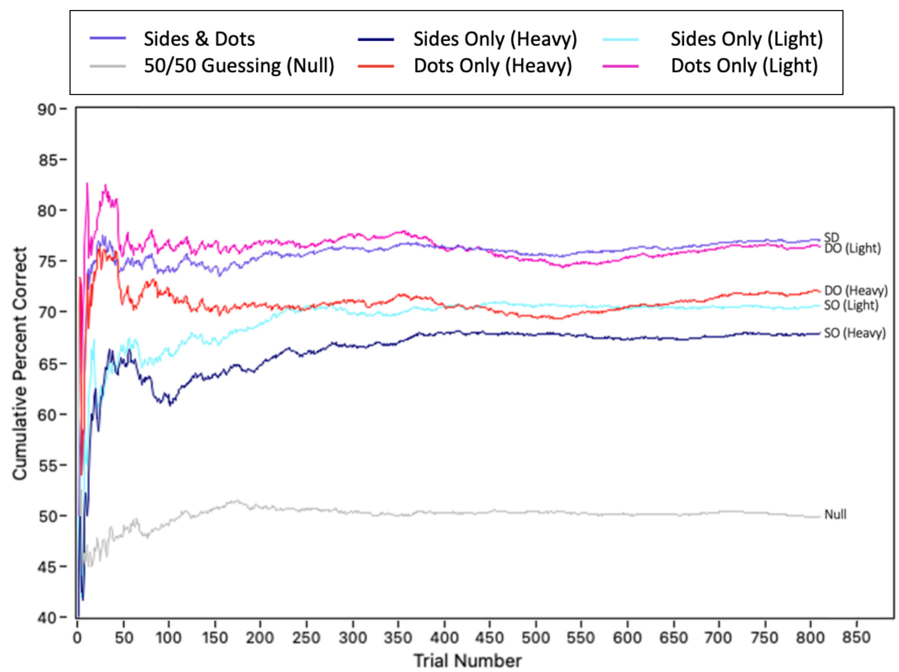



**Figure 2.8: Sample Learning Curves** Sample learning curves for 6 computational observers, each of which is the average Percent Correct (PC) of 10 repetitions on 810 trials of a haptic categorization task. These observers were run in the 1.25/0.75 experimental condition. The Null Observer maintains a PC at approximately 50% correct. All 5 variations of the Feature-Focused Bayesian Observer improve in PC as trial number increases to an asymptotic performance.

We calculated approximate asymptotic performance for each of our computational observers using 200 repetitions, estimating the asymptote as the average cumulative PC after 100 000 trials (Table 2.2). When calculating asymptotes, we input the actual category mean and standard deviation of both categories, meaning that the observer knew the actual category distributions from the outset and did not need to learn them. In all conditions, asymptotic performance of the Null Observer equalled 50.0%. The DO (Light) Observer had the highest asymptote compared to other observers in the 0.75/0.75 condition. The SD Observer had the highest asymptote compared to other observers in the 1.25/1.25 condition. The DO (Light) and SD Observers had similar asymtotes in the 0.75/1.25 and 1.25/0.75 conditions. All Feature-Focused Bayesian Observers had their highest asymptotic performance in the 0.75/0.75 condition and their lowest asymptotic performance in the 1.25/1.25 condition. Asymptotic performance was almost identical in the 0.75/1.25 and 1.25/0.75 conditions for all observers.

**Table 2.2: Manipulate Category Variation – Observer Asymptotes** Summary of approximate asymptotic performance for 6 Computational observers in each of our 4 experimental conditions. Asymptotes were calculated as the average performance of 200 simulated observer repetitions after 100 000 trials.

| | Experimental Condition | | | |
|---|---|---|---|---|
| **Computational Observer** | 0.75/0.75 | 0.75/1.25 | 1.25/0.75 | 1.25/1.25 |
| SD | 81.2 | 77.3 | 77.2 | 73.1 |
| SO (Heavy) | 70.5 | 67.8 | 67.8 | 65.2 |
| SO (Light) | 74.8 | 71.4 | 71.4 | 68.1 |
| DO (Heavy) | 75.9 | 72.4 | 72.3 | 68.7 |
| DO (Light) | 82.0 | 77.2 | 77.2 | 72.2 |
| Null | 50.0 | 50.0 | 50.0 | 50.0 |

#### Comparison to Simulated Participants

We generated 100 simulated participants corresponding to the categorization strategies of each of our computational observers to evaluate the accuracy of our model comparison process. We omitted the Null Observer from this comparison, as the Bayes factor comparing the Null Observer and most likely observer did not exceed a values of $10^{-31}$. In the resulting model comparison, considering the Feature-Focused Bayesian Observers, we simulated observers and participants using the same sequence of objects selected from categories defined by 2D Gaussian distributions with standard deviation equal to 0.75 in both dimensions. Table 2.3 summarizes the number of participants correctly and incorrectly classified by our Bayesian model comparison.

**Table 2.3: Manipulate Category Variation – Confusion Matrix** We generated 100 simulated participants using each of following categorization strategies: SD, SO (Heavy), DO (Heavy), SO (Light), and DO (Light). We compared each simulated participant to the computational observers and identified the observer that best modelled the participant's performance according to our Bayesian model comparison. All participants and observers were simulated in the 0.75/0.75 experimental condition.

| **Model Comparison Classification** | **Actual Categorization Strategy** SD | SO (Heavy) | DO (Heavy) | SO (Light) | DO (Light) |
|---|---|---|---|---|---|
| SD | 100 | 0 | 3 | 0 | 0 |
| SO (Heavy) | 0 | 97 | 0 | 44 | 0 |
| DO (Heavy) | 0 | 0 | 91 | 0 | 0 |
| SO (Light) | 0 | 3 | 0 | 56 | 0 |
| DO (Light) | 0 | 0 | 6 | 0 | 100 |

All simulated participants using an SD or a DO (Light) categorization strategy were correctly classified. We correctly classified 97% and

91% of SO (Heavy) and DO (Heavy) participants, respectively. All misclassified SO (Heavy) participants were classified as SO (Light). Misclassified DO (Heavy) participants were classified as either SD or DO (Light). 56% or SO (Light) participants were correctly classified, while 44% were misclassified as SO (Heavy).

We further compared the same simulated participants to the same computational models using a mean squared error (MSE) analysis, rather than our Bayesian model comparison (Table 2.4). We classified the best matched model as the one with the smallest MSE.

**Table 2.4: Manipulate Category Variation – Confusion Matrix for MSE Analysis** We generated 100 simulated participants using each of following categorization strategies: SD, SO (Heavy), DO (Heavy), SO (Light), and DO (Light). These were the same simulated participants considered in table 2.3. We compared each simulated participant to the computational observers and identified the observer that best modelled the participant's performance as the model with the smallest MSE. All participants and observers were simulated in the 0.75/0.75 experimental condition.

| | **Actual Categorization Strategy** | | | | |
|---|---|---|---|---|---|
| **MSE Classification** | SD | SO (Heavy) | DO (Heavy) | SO (Light) | DO (Light) |
| SD | 54 | 0 | 5 | 7 | 43 |
| SO (Heavy) | 0 | 76 | 6 | 10 | 0 |
| DO (Heavy) | 5 | 15 | 54 | 49 | 1 |
| SO (Light) | 3 | 9 | 34 | 34 | 1 |
| DO (Light) | 38 | 0 | 1 | 0 | 55 |

All sets of simulated participants had more participants misclassified by an MSE comparison than by our Bayesian model comparison. We correctly classified 76% of simulated SO (Heavy) participants and 34% of simulated SO (Light) participants. We correctly classified 54% to 55% of participants in all other sets of simulated participants.

### 2.3.2 Manipulate Prior Probabilities of Categories

In a second variation of our haptic categorization task, we manipulated the proportion of objects presented from each category. We created 4 experimental conditions, where we specified that we would sample equal proportions of objects from Categories A and B on the first 405 trials, but that proportions could differ for the last 405 trials. In all conditions, category standard deviation equalled 0.75 in both dimensions. We illustrate these conditions in Figure 2.9.

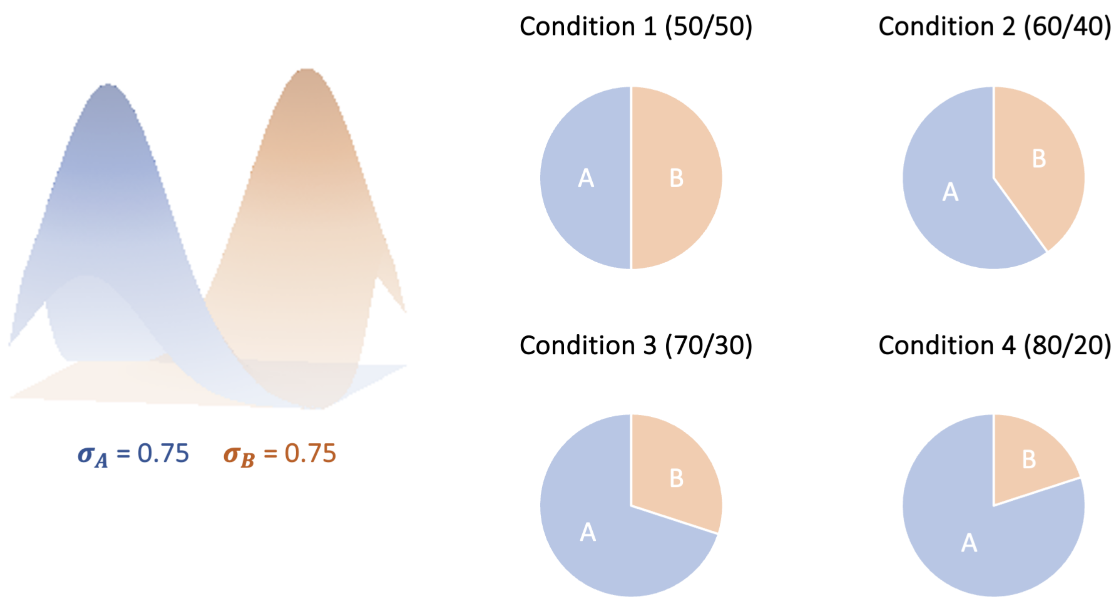



**Figure 2.9: Category Prior Conditions** Summary of the four experimental conditions tested in Experiment 2, where we manipulated the category proportion of objects presented from the two categories on each day of the experiment. On Day 1, participants in all conditions received the same proportion of objects from A and B.

We produced 13 computational observers to simulate performance on this task variation. 5 of these 12 observers were identical to the

Feature-Focused Bayesian Observers generated in the previous section: SD Observer (heavy sensory noise), SO Observer (heavy sensory noise), SO Observer (light sensory noise), DO Observer (heavy sensory noise), and DO Observer (light sensory noise). In this section, we refer to these as constant category prior observers, as they assumed that objects from both categories are equally likely to be presented on both days of the experiment. We also generated 5 updating category prior observers, which considered the same features as the constant category prior observers, but made no assumptions regarding category priors on the last 405 trials. The constant and updating category prior observers performed identical calculations on the first 405 trials, and differed only in the sensory noise that affects all observer repetitions. Finally, we generated a Null Observer, a Prevalence MAP Estimate Observer, and a Prevalence Probability Matching Observer on this task variation.

Similar to the previous set of observers, cumulative PC of Feature-Focused Bayesian Observers improved in early trials and approached an asymptote in later trials. This pattern held true for the Prevalence MAP Estimate and Prevalence Probability Matching Observers. Table 2.5 summarizes asymptotic performance for each observer in each condition. Constant and updating category prior observers approached equal asymptotes in the 50/50 condition, while updating category prior observers approached higher asymptotes in the 60/40, 70/30, and 80/20 conditions. In these 3 conditions, the Prevalence MAP Estimate and Prevalence Probability Matching Observers approached an asymptotic performance lower than than that of the Feature-Focused Bayesian Observers, but higher than that of the Null Observer.

**Table 2.5: Manipulate Category Priors – Observer Asymptotes** Summary of approximate asymptotic performance for 6 Computational observers in each of our 4 experimental conditions. Asymptotes were calculated as the average performance of 200 simulated observer repetitions after 100 000 trials.

| | | **Experimental Condition** | | | |
|---|---|---|---|---|---|
| | **Computational Observer** | 50/50 | 60/40 | 70/30 | 80/20 |
| **Updating Category Prior** | SD | 81.2 | 81.8 | 83.7 | 87.0 |
| | SO (Heavy) | 70.5 | 71.8 | 75.6 | 81.8 |
| | SO (Light) | 74.8 | 75.9 | 78.7 | 83.6 |
| | DO (Heavy) | 75.9 | 76.8 | 79.5 | 84.0 |
| | DO (Light) | 81.9 | 82.5 | 84.3 | 87.3 |
| **Constant Category Prior** | SD | 81.2 | 81.3 | 81.2 | 81.2 |
| | SO (Heavy) | 70.5 | 70.8 | 70.6 | 70.6 |
| | SO (Light) | 74.8 | 75.0 | 74.7 | 74.9 |
| | DO (Heavy) | 75.9 | 75.9 | 75.9 | 75.9 |
| | DO (Light) | 81.9 | 82.0 | 81.9 | 81.9 |
| | Null | 50.0 | 50.0 | 50.0 | 50.0 |
| | Prevalence MAP Estimate | 50.0 | 60.0 | 70.0 | 80.0 |
| | Prevalence Probability Matching | 50.0 | 52.0 | 58.0 | 67.9 |

#### Comparison to Simulated Participants

We generated 100 simulated participants using the same strategy as each of our computational observers. We compared each simulated participant to our computational observers and classified it as using the strategy of the observer that best simulated participant performance. (Table 2.6). As with the first task variation, we omitted the Null, Prevalence MAP Estimate, and Prevalence Probability Matching Observers from our analysis, for which our Bayes factors did not exceed values of $10^{-59}$, $10^{-28}$, and $10^{-29}$, respectively.

**Table 2.6: Manipulate Category Priors – Confusion Matrix** We generated 100 simulated participants using each of following categorization strategies, considering both constant and updating category prior versions of the Feature-Focused Bayesian Observers: SD, SO (Heavy), DO (Heavy), SO (Light), and DO (Light). We compared each simulated participant to the computational Observers and identified the Observer that best modelled the participant's performance. All participants and observers were simulated in the 80/20 experimental condition.

| | | **Actual Categorization Strategy** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Updating Category Prior** | | | | | **Constant Category Prior** | | | | |
| | **Model Comparison Classification** | SD | SO (Heavy) | SO (Light) | DO (Heavy) | DO (Light) | SD | SO (Heavy) | SO (Light) | DO (Heavy) | DO (Light) |
| **Updating Category Prior** | SD | 100 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| | SO (Heavy) | 0 | 66 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | SO (Light) | 0 | 34 | 99 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | DO (Heavy) | 0 | 0 | 0 | 87 | 0 | 0 | 0 | 0 | 0 | 0 |
| | DO (Light) | 0 | 0 | 0 | 8 | 100 | 0 | 0 | 0 | 0 | 0 |
| **Constant Category Prior** | SD | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 1 | 0 |
| | SO (Heavy) | 0 | 0 | 0 | 0 | 0 | 0 | 64 | 1 | 0 | 0 |
| | SO (Light) | 0 | 0 | 0 | 0 | 0 | 0 | 36 | 99 | 0 | 0 |
| | DO (Heavy) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 88 | 0 |
| | DO (Light) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 100 |

We correctly classified all simulated participants generated using categorization strategies identical to the following observers: SD (constant category prior), SD (updating category prior), DO (Light) (constant category prior), and DO (Light) (updating category prior). We correctly classified 87% of DO (Heavy) (updating category prior) participants, with remaining participants misclassified as DO (Light) (updating category prior) or SD (updating category prior). We correctly classified remaining simulated participants with $> 50\%$ accuracy, where we misclassified participants as using the same strategy with a different level of sensory noise (i.e. heavy/light).

## 2.4 Discussion

### 2.4.1 Summary of Findings

In this chapter, we proposed 3 different computational observers: Feature-Focused Bayesian Observer, Prevalence MAP Estimate Observer, Prevalence Probability Matching Observer, and Null Observer. Our central observer, the Feature-Focused Bayesian Observer, learns novel categories and categorizes objects on each trial of a haptic categorization task. Its performance increases in PC toward an asymptote that increases when the observer's sensory noise decreases. The observer may attend to a single feature or multiple features on the haptic stimuli, which further impacts its asymptotic performance.

We further implemented a Bayesian model comparison to compare participant performance to several different computational observers. We compared simulated participants using a known categorization strategy to the observers and identified the observer that best simulated participant performance (Figure 2.3, 2.6). These results demonstrate the high accuracy of our model comparison in classifying participant categorization strategies.

### 2.4.2 The Challenge of Defining Optimality

Most research considers Bayesian models as optimal when they combine all available cues to increase reliability of a perceived stimulus (Alais and Burr, 2004; Ernst and Banks, 2002; Gepshtein and Banks, 2003; Bejjanki et al., 2011; Bankieris et al., 2017). For example, Alais and Burr (2004) demonstrated that participants combine a visual and an auditory cue to optimally locate the origin in space of a stimulus, rather than relying on a single cue. Initially, we expected our SD Observer to consistently achieve the highest PC on our task, as integrating information from both features provides more reliable information about the objects. Contrary to our expectations, the DO (Light) Observer achieved the highest PC in many cases, a pattern further observed amongst the asymptotes in our experimental conditions (tables 2.2 and 2.5). If we define our optimal observer as the one that achieves the highest PC on our task, then SD and DO (Light) Observers can both perform optimally, depending on the object sequence and experimental condition. Despite the benefit of minimizing sensory uncertainty by combining multiple sensory estimates in an SD categorization strategy, the higher cognitive demand of attending to two features may increase the sensory noise associated with each measurement. Using a DO strategy with sufficiently small sensory noise when interpreting the dots feature could provide more reliable information.

### 2.4.3 Assumptions of our Feature-Focused Bayesian Observer

This Feature-Focused Bayesian Observer ultimately aimed to simulate human performance. In doing this, it made assumptions about the state of the world, which may or may not align with human assumptions.

#### Gaussian Categories with Feature Independence

We considered a total of 2450 hypotheses for the categories, each of which defines a 2D Gaussian distribution with no correlation between features. In reality, people may consider any distribution to define the categories, such as Poisson, logarithmic, or uniform distributions. While we could have generated additional hypotheses for different distribution types, past research suggests that humans may make this Gaussian assumption when presented with uniformly distributed categories (Gifford et al., 2014). Accordingly, we accepted this as a reasonable assumption for the computational observer.

#### Set Range of Actual Feature Values

Like humans, this observer experienced sensory noise when measuring sides and dots feature values. When interpreting these measurements, our Feature-Focused Bayesian Observer considered several actual feature values that may have produced a measured feature value. This observer considered 8 possible values for the actual feature values, ranging $\pm 2$ $\sigma_{sens}$ (sensory noise variability), from the actual feature value. We implemented this 8 value range as a computational consideration to maintain reasonable run times to produce observers. While we would ideally consider a wider range of actual feature values, the probabilities for more extreme values would become very small, likely having negligible effects on our final results.

#### Constant Sensory Noise

This observer assumed that sensory noise remained constant across objects and trials. That is, the observer experienced the same amount of sensory noise with any of the 25 objects, as well as on any trial from 1 to 810. We acknowledge that this may differ from human experience, such as in a participant who experiences greater sensory noise in later trials as their focus diminishes.

### 2.4.4 Assumptions of our Bayesian Model Comparison

In the same vein, our Bayesian model comparison makes assumptions regarding the state of the world and the past experience of participants.

#### Exhaustive List Of Computational Models

In this study, we consider only Bayesian computational observers and 3 non-sensory reference Observers. While Bayes' theorem does provide a promising framework for modelling human sensory perception (Alais and Burr, 2004; Bejjanki et al., 2011; Bankieris et al., 2017), researchers have proposed alternative models. For example, Granato et al. (2022) trained a machine learning model on a visual categorization task, which could effectively accomplish the task and qualitatively simulate human performance. Our comparison between participants and observers assumed that we provided an exhaustive list of computational observers, overlooking the possibility that an alternative observers better simulates sensory perception. Future studies may compare Bayesian models to other computational observers.

We can better investigate the goodness-of-fit between a participant and winning model by comparing the performance of a participant to the 10 repetitions that make up that model. See Figure A.3.2 for illustrations of this comparison in participants tested on categories with manipulated category variation (note that these are the 24 participants tested in Chapter 3). See Figure A.4.2 for these illustrations for participants tested on categories with manipulated category priors (note that these are the 20 participants tested in Chapter 4).

### 2.4.5 Computational Limitations

Many of the previously mentioned assumptions help to address computational limitations. If we considered infinite categorical distributions to describe A and B, infinite hypotheses for the actual feature values, and infinite computational models of categorization, our model would exceed the processing capabilities of any existing computer. In the following sections, we outline these limitations.

#### Assigning Priors to Computational Observers

In comparing computational models to participants, we imposed hierarchical priors across all models. Our priors assumed that participants employ SD, SO, and DO categorization strategies with equal chance. While this may be appropriate for simulated participants, humans may favour certain strategies depending on past experiences. For example, a person who does more manual labour may have callused hands and in turn poorer tactile acuity than average. This individual may find the sides feature easier to measure than the dots feature, as they can discern the side length and vertex angles more easily than the raised dots. In this case, the individual may favour SO or SD strategies, where they rely more on measuring the sides feature.

Despite knowing that models may not be considered equally in all cases, it proves extremely difficult to quantify all of a person's life experience into a single numerical prior probability. Because our study exclusion criteria excluded individuals with poor tactile acuity, or notable factors that would cause them to perform uniquely, we consider it reasonable to assume that participants are approximately equally likely to employ these three categorization strategies. Accordingly, we interpret the results of this model comparison analysis as the posterior probability for a model if participants did begin the experiment with uniform priors for SD, SO, and DO strategies.

#### Sensory Variability Between Repetitions

The sensory variability associated with measuring object features during this task results in infinite possible sequences of measured values across the 810 trials. In this study, we generate and average across

10 repetitions of our Feature-Focused Bayesian Observer. Ideally, we could generate and average across infinite repetitions.

Observers with less sensory noise are less likely to measure feature values far from the actual value. With sufficiently small sensory noise, fewer repetitions likely encompass enough sequences of measured feature values to produce a representative observer of participant performance. To investigate this, we briefly compared the categorization strategy results from models with 10 and 20 repetitions. Qualitatively, we observed no notable differences between these results, suggesting that 10 repetitions are sufficient to model performance.

**Dividing Observer Repetitions**

This variability between repetitions becomes increasingly relevant in Section 2.3.3., where we model constant and updating category prior observers. All observers have equal category priors for the first 405 trials, then either learn the categorical priors using parameter estimation on the last 405 trials or continue to consider equal category priors for the entirety of the experiment. While these two models differ on the final 405 trials, they employ identical strategies on the first 405. Accordingly, observers simulating the same categorization strategy should be equally effective in simulating participant's performance on the first

405 trials. We distribute observer repetitions in a manner that makes these two models nearly equal as possible at simulating a participant's response after the first 405 trials. Because of variation between repetitions, these probabilities often differed somewhat, meaning that strategy classifications from the final 405 trials were biased at least slightly toward one or the other of of the two models.

#### Number of Feature Value Hypotheses

While our Feature-Focused Bayesian Observer considers a set number of actual values for an object's sides and dots features, one may want to model an Observer that considers more values. Because each additional value increases the processing time to simulate this observer, we were limited to our current ranges to maintain a reasonable processing time.

### 2.4.6 MSE Analysis vs. Bayesian model Comparison

We performed 2 comparisons between our simulated participants and computational models, one using MSE and another using Bayesian model comparison. In all groups of simulated participants, the Bayesian model comparison proved much more accurate in classifying participant strategies. This finding provides further evidence that Bayesian model comparison can correctly classify participant strategies.

## 2.5 Conclusion

Overall, we put forward a Bayesian model of haptic perception that successfully learns categories and categorizes stimuli on a haptic categorization task. We identify the model type that best simulates participants using known categorization strategies with high accuracy.

# Chapter 3

# Experiment 1: Manipulation of Category Variation

## 3.1 Introduction

In the previous chapter, we outlined a Feature-Focused Bayesian Observer that can effectively simulate human performance on a haptic categorization task. We further simulated participants on this task and classified their strategy with high accuracy.

In this chapter, we report a study in which we tested human participants on the same task to compare to our computational observers. We tested human participants on one of four variations of this task, where we manipulated the standard deviation of the 2D Gaussian distributions defining categories A and B to create four task variations with unique category overlap. Then, we compared human and Bayesian model performance on the task. We hypothesized that human performance would be best simulated by either (1) an optimal Bayesian model, (2) a sub-optimal Bayesian model, or (3) a non-Bayesian model.

## 3.2 Methods

### 3.2.1 Participants

We collected data from 24 participants aged 18 to 22 (average age 19 years, 22 F, 2 M) recruited from the McMaster undergraduate student population using SONA. All participants self-reported they were unaffected by any of the following conditions: diabetes, nervous system disorder or injury, learning disability, dyslexia, attention deficit disorder, cognitive impairment, carpal tunnel syndrome, arthritis of the hands, hyperhidrosis. This study was approved by the McMaster Ethics Research Board and all participants were compensated with cash or course credit.

### 3.2.2 Experimental Conditions

We assigned each participant to one of four experimental conditions, which we counterbalanced based on their order of entry into the experiment. In total, we tested 6 participants in each of the conditions. Each condition used a unique set of 2D Gaussian distributions to define categories A and B (Table 3.1).

**Table 3.1: Experiment 1 – Conditions** Summary of the four experimental conditions tested in Experiment 1, during which we manipulated the category standard deviation of categories A and B. Both categories were described by two dimensional Gaussian distributions. These standard deviation ($\sigma$) values applied to both dimensions of the category distributions. See Chapter 2, Figure 2.7 for additional details.

| | Category A Standard Deviation | Category B Standard Deviation |
|---|---|---|
| Condition 1 | 0.75 | 0.75 |
| Condition 2 | 0.75 | 1.25 |
| Condition 3 | 1.25 | 0.75 |
| Condition 4 | 1.25 | 1.25 |

In each of these conditions, the standard deviation remained constant across both dimensions of the 2D Gaussian distribution. Condition 1 produced the least overlap between categories, while condition 4 produced the most overlap between categories.

### 3.2.3 Experimental Procedure

We used an experimental procedure almost identical to that described in Chapter 2. See Chapter 2, Subsection 2.2.1 for details. We made two minor adjustments when testing this set of human participants. First, we had participants complete 405 trials on each of two days spaced 1 week apart, rather than 810 trials on a single testing day. This ensured that participants did not exceed 2.5 hours of testing on a single day, after which we deemed it unreasonable to expect focus and effort from participants. We further divided the trials on each day into 9 blocks of 45 trials, where participants received a 1 minute break after every block and a 5 minute break after every 3 blocks. Second, we named the two categories “Elyk” and “Noek”, counterbalancing the assignment of each name to either category A or B across participants. These names, with which participants had no past experience, prevented name associations from impacting performance.

### 3.2.4 Haptic Exploratory Procedures

Previous studies defined eight haptic EPs that participants employ to obtain tactile information about an unknown object (Lederman and Klatzky, 1987, 1993). We determined that four of these previously defined EPs describe hand movements that participants could employ to extract information about the 3D printed objects used as stimuli in our study. In Table 3.2, we list and define these four EPs as they relate to our study, as well as identify the tactile features (sides, dots, or both) participants could explore with the EP.

During the experiment, we used a GoPro Camera (Hero Session or Hero 5 Session) to record participants’ hands while they explored the haptic stimuli. The camera was mounted on a stand approximately 51cm above the participants’ hands and recorded them from a top-down view. We recorded 15 trials per block of testing: trials 1 to 5, trials 20 to 24, and 41 to 45. In total, we obtained video from 270

**Table 3.2: Exploratory Procedure definitions** Summary of the four exploratory procedures that we recognized as participant strategies used during our haptic categorization task. Each strategy focused on investigating one or both of the two object features, sides and dots. Enclosure obtained information about the sides and dots features simultaneously, while participants could use the other exploratory procedures to obtain information about one feature at a time.

| **Exploratory Procedure** | **Feature Under Investigation** | **Description** |
|---|---|---|
| Enclosure | Sides AND Dots | Participant has contact with as much of the object as possible for a period of time. This commonly involves making a fist around the object. |
| Static | Sides OR Dots | Participant has sustained contact between the object and one or more parts of the hand(s) with no movement between the skin and object feature. |
| Swipe | Dots | Participant moves one or more parts of the hand(s) to swipe across the dots feature of the object. |
| Contour Following | Sides | Participant maintains contact with the perimeter of the object and uses a smooth, coordinated movement to feel the side lengths and vertices. Contour following ends when the participant lifts their hand off of the object. |

of the 810 trials completed. We informed participants that we would record their hands during the experiment, but did not specify the total number of trials or the specific trials that we planned to record.

We reviewed videos offline to score EPs used during haptic exploration. Table 3.3 details the information collected from each video. Initially, we recorded additional data to that listed in Table 3.3. Because participant hand movements can be subtle, these additional data proved difficult to identify and record reliably. As a result, we decided to omit those data from our analysis.

To train on how to watch and score videos, all experimenters watched and scored sample data from a pilot study in which experimenters acted

**Table 3.3: Exploratory Procedure Scoring** Summary of EP information collected from video recordings. We recorded one or more details for each EP. Bolded text, followed by a more detailed description, defines individual pieces of information scored. Experimenters scored EP details using keywords, indicated by italicized text.

| Exploratory Procedure | Data Recorded |
|---|---|
| Enclosure | **EP Used:** Did the participant use this EP? [*true or false*] |
| Static Contact | **EP Used:** Did the participant use this EP? [*true or false*] |
| | **Hand(s) Used:** If the participant used this EP, which hand(s) did they use? [*left or right or both*] |
| Swipe | **EP Used:** Did the participant use this EP? [*true or false*] |
| Contour Following | **EP Used:** Did the participant use this EP? [*true or false*] |
| | **Hand(s) Used:** If the participant used this EP, which hand(s) did they use? [*left or right or both*] |
| General Information | **Table contact:** Did the participant lift the object off of the table or leave it on the table during haptic exploration? [*on or off or both*] |
| | **Hand(s) Used:** Throughout the entire trial, which hand(s) did the participant use to explore the object? [*left or right or both*] |
| | **Trial Time:** For how long did the participant explore the object? [*0, 1, 2, 3, 4, 5, 6, >6 seconds*] |

as participants. We completed rounds of training until at least 80% of information scored by each pair of experimenters matched. After four rounds of training, between 84% to 93% of scores from each pair of experimenters matched.

We watched and scored the last trial of each five trial recording, i.e. trials 5, 24, and 45 of each testing block, totalling 54 trials for each participant. Different pairs of experimenters, selected from our 4 available experimenters, watched and scored video for each of our 24 participants. Experimenters remained blind to the computational model that best simulated performance throughout the scoring process. They watched videos on 1x speed as many times as necessary to confidently record all information included in Table 3.3. A third experimenter compiled scores from the two experimenters watching each video. In case of disagreement between experimenters scoring participant data, responses were compiled according to the guidelines in Table A.1.

## 3.3 Results

### 3.3.1 Human Participant Results

Each participant completed two days of testing on our haptic categorization task. We calculated the average PC across participants for each experimental condition and each study day (Figure 3.1). Average PCs on Day 2 were higher than on Day 1 for all conditions. On both days, condition 1 (0.75/0.75) had the highest average PC, while condition 4 (1.25/1.25) had the lowest average PC on both days. Average PC from participants in conditions 2 and 3 remained between conditions 1 and 4 on both days.

Figure 3.1 illustrates average participant learning curves in each of the four experimental conditions. These learning curves compare average PC across all participants in a given condition for sets of 3 experimental blocks (referred to as block sections). Block sections incorporate

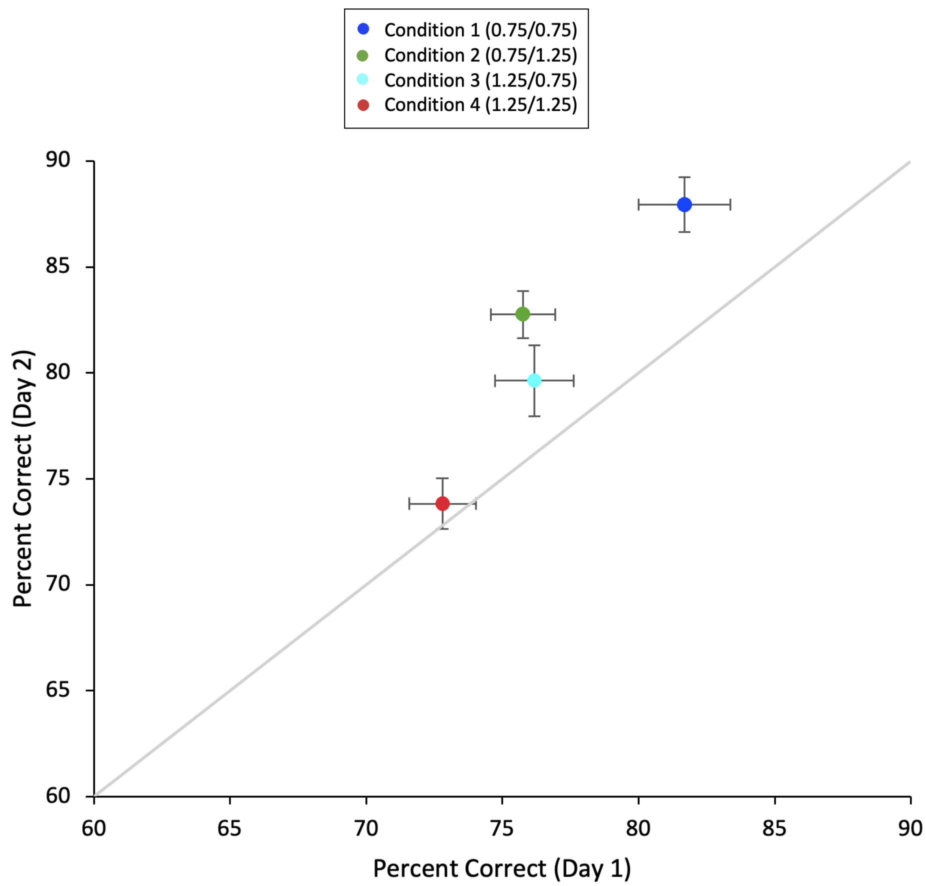
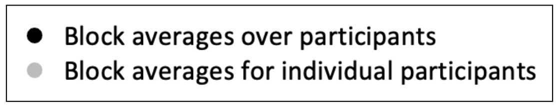


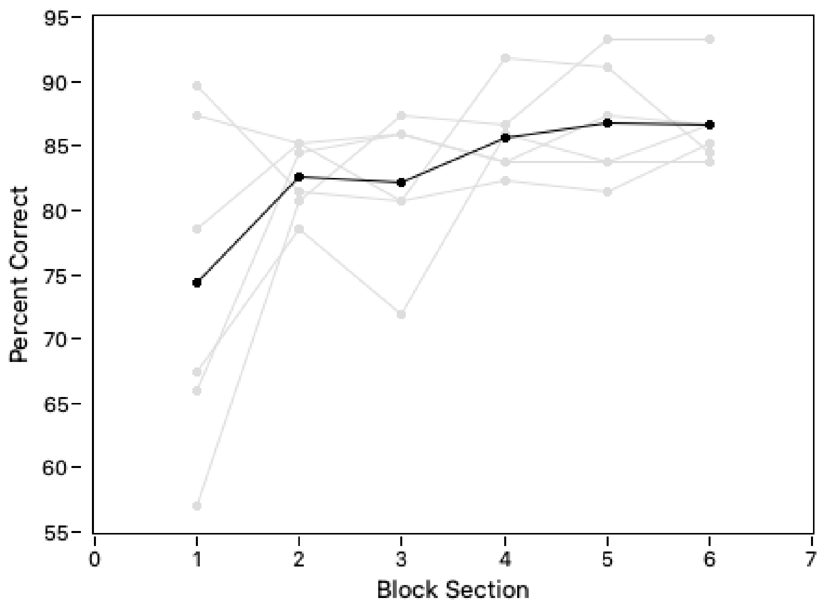

**Figure 3.1: Experiment 1 – Human Performance on Haptic Categorization Task** Average PC of participants in each of the four experimental conditions study Days 1 and 2. The light grey line represents points at which Day 1 PC equals Day 2 PC. The error bars show ±1 Standard Error (SE). n = 24 (n = 6 within each condition).

3 blocks of 45 trials each. For example, block section 1 includes Day 1 blocks 1 to 3. Block sections 1, 2, and 3 include blocks from Day 1, while block sections 4, 5, and 6 include blocks from Day 2. In all conditions, average participant PC increased from block section 1 to 6. For conditions 0.75/0.75, 0.75/1.25, 1.25/0.75, and 1.25/1.25, average PC increased by 12.4%, 9.0%, 7.3%, and 1.1% from the first to last block section, respectively (Figure 3.1). In conditions 3 and 4, PC decreased

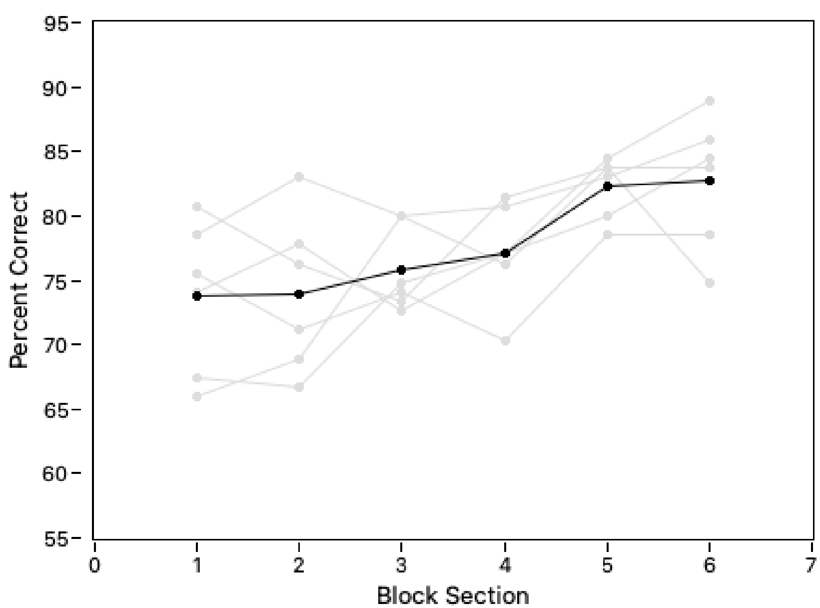
by 3.58% between block sections 3 and 4 (Figure 3.1c,d). Any other decreases in average PC were $\leq 0.5\%$.



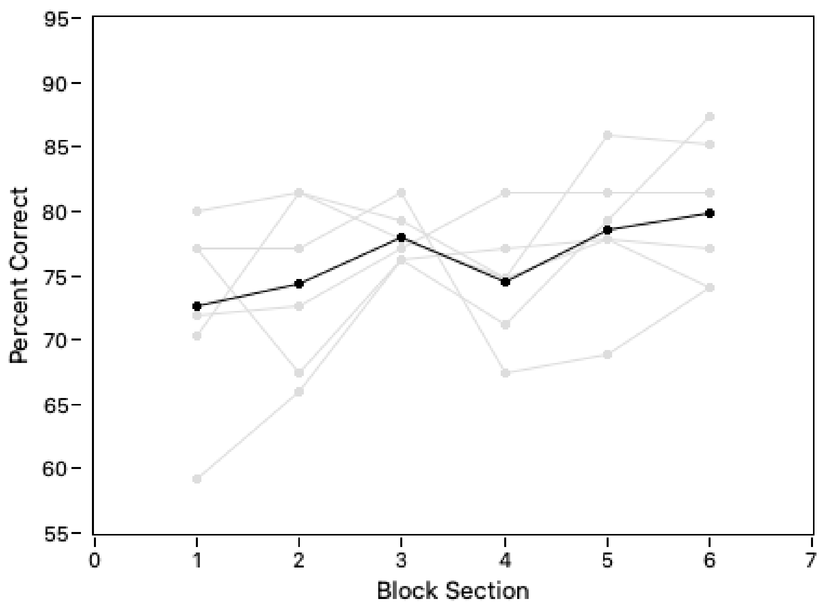

**(a)** Condition 1 (0.75/0.75)




**(b)** Condition 2 (0.75/1.25)



**(c)** Condition 3 (1.25/0.75)




**(d)** Condition 4 (1.25/1.25)

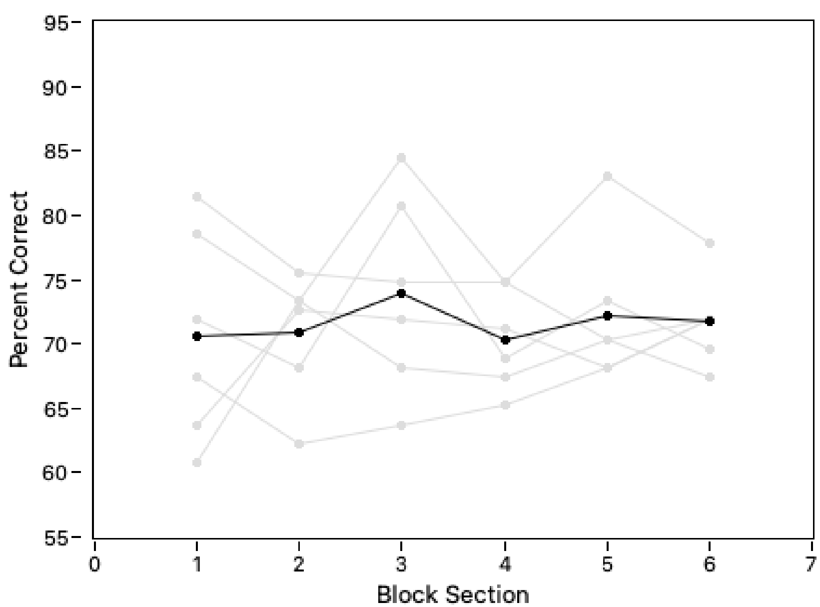



**Figure 3.1: Experiment 1 – Learning on Haptic Categorization Task** Participant PC averaged over windows of 135 trials (block sections). Black lines illustrate the average learning curves across all participants in a given condition. Grey lines illustrate individual participant PC on each block section. n = 20 (n = 5 within each condition).

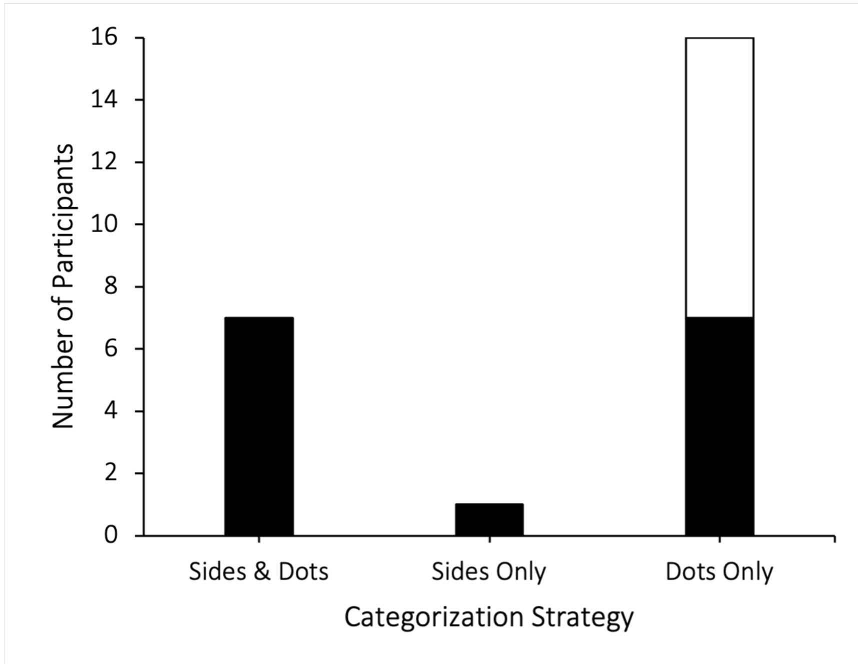
### 3.3.2 Comparison to Bayesian Observers

We compared human performance on our haptic categorization task to that of 5 different computational models: SD, DO (Heavy), DO (Light), SO (Heavy), and SO (Light). As in Chapter 2, we omitted the Null Observer from this comparison, as the Bayes' factor comparing the Null and most likely observers for each participant did not exceed $10^{-13}$.

We generated a unique set of models for each participant according the specific object sequence they received during the experiment. We compared Day 1 and Day 2 performance to the computational models separately to classify the strategy that each participant used on each study day. On Day 1, we classified 9 participants as using a DO (Light) strategy, 7 as using a DO (Heavy) strategy, 7 as using an SD strategy, and 1 as using an SO (Heavy) strategy (Figure 3.2a). We did not classify any participants as using an SO (Light) strategy. On Day 2, we classified 12 participants as using an SD strategy, 11 as using a DO (Light) strategy, and 1 as using a DO (Heavy) strategy (Figure 3.2b). We did not classify any participants as using an SO (Heavy) or SO (Light) strategy.

We could, alternatively, compare participant performance to the computational models using a mean squared error calculation, rather

**(a)** Day 1 Strategy Classification




**(b)** Day 2 Strategy Classification

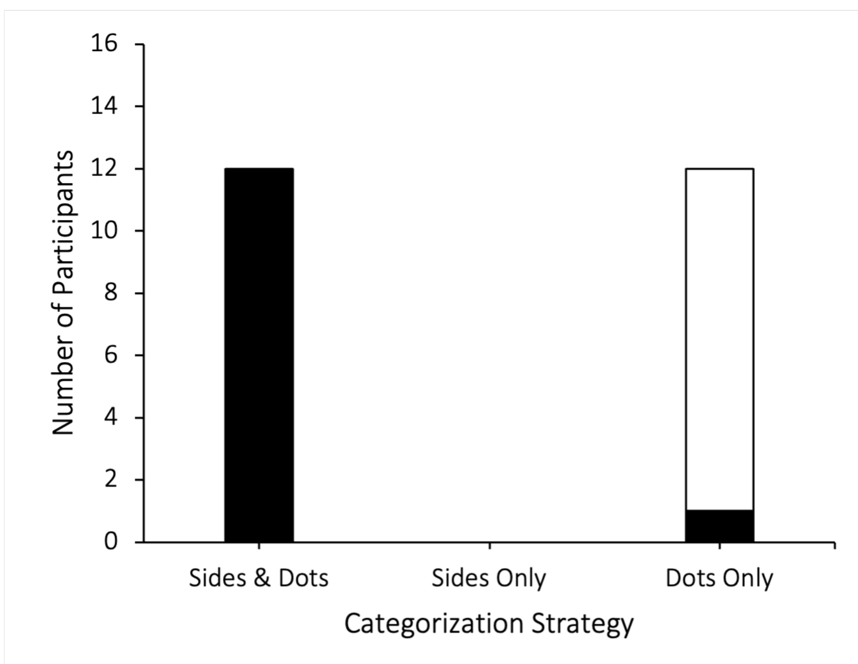



**Figure 3.2: Experiment 1 – Strategy Classification** Number of participants classified as using each categorization strategy on Day 1 and Day 2. Black Bars represent observers simulated with heavy sensory noise, while white bars represent observers simulated with light sensory noise. n = 24.

than using Bayesian model comparison. This comparison identifies the model that produces a learning curve closest to that of the participant as the best fitting model. If we use this comparison with the same set of 24 participant, 17 and 12 participants are classified differently on Days 1 and 2, respectively.

In total, our model comparison classified 15 participants as using the same strategy on both days, while 9 participants changed strategies between study days (Figure 3.3). Of the 15 participants classified as maintaining the same strategy on both days, 8 participants used DO (Light), 6 used SD, and 1 used DO (Heavy). We classified the following strategy changes in participants between study days: DO (Heavy) to DO (Light), DO (Light) to SD, DO (Heavy) to SD, SO (Heavy) to SD, and SD to DO (Light). For participants who changed strategy, PC increased by an average of 3.8% (std dev 1.68%) (See Table A.3 for individual participant values). For participants who maintained the same strategy, PC increased by an average of 1.17% (std dev 2.52%) (See Table A.2 for individual participant values).

Individual learning curves and posterior probabilities for each observer model quantify the relative probability of all computational models in best simulating human performance. Figure 3.4a illustrates the learning curve for a participant classified as SD on both days, as well as

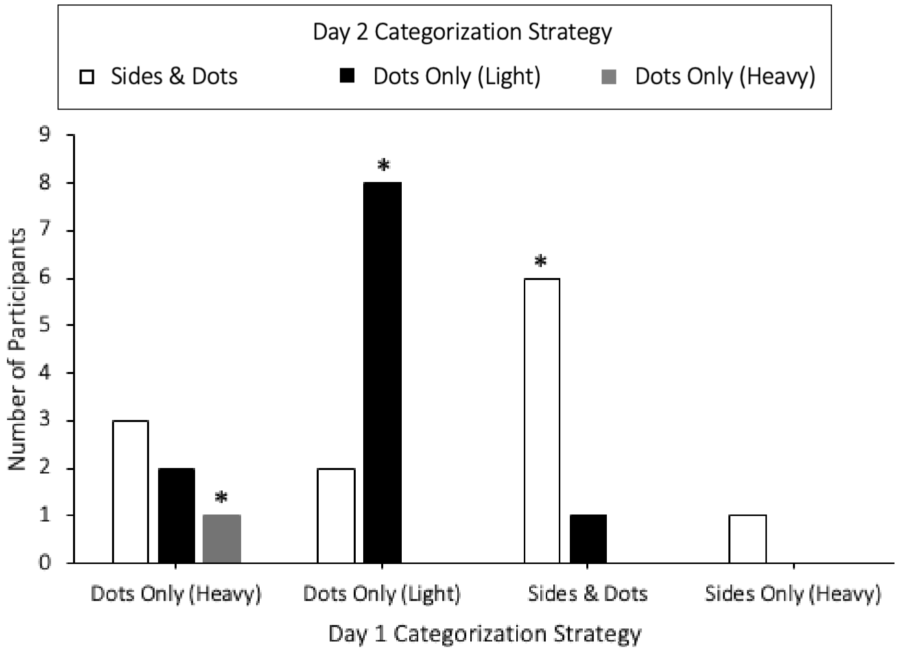



**Figure 3.3: Experiment 1 – Changes in Strategy Classification Between Experiment Days** Count of participant classification changes between Days 1 and 2. The x-axis indicates a participant's Day 1 categorization strategy, while the bar colour indicates their Day 2 categorization strategy. *15 participants maintained the same strategy on both study days, while the remaining 9 participants changed strategies between study days. n = 24.

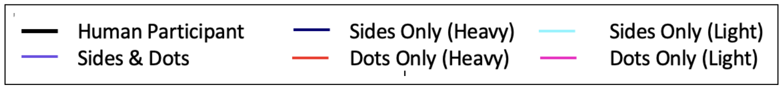
the computational observers to which we compared their performance. Figure 3.4c,e illustrate the most probable model, determined by our Bayesian model comparison, on each trial of study Days 1 and 2. On both days, the Bayesian model comparison recognizes SD as the most probable model relatively quickly.

Our Bayesian model comparison recognized SO (Heavy) on Day 1 to SD on Day 2 as one of the strategy changes performed by a participant in the study. Figure 3.4b illustrates the learning curve for one sample participant who made this strategy change, as well as the computational observers to which we compared their performance. On Day 1, our

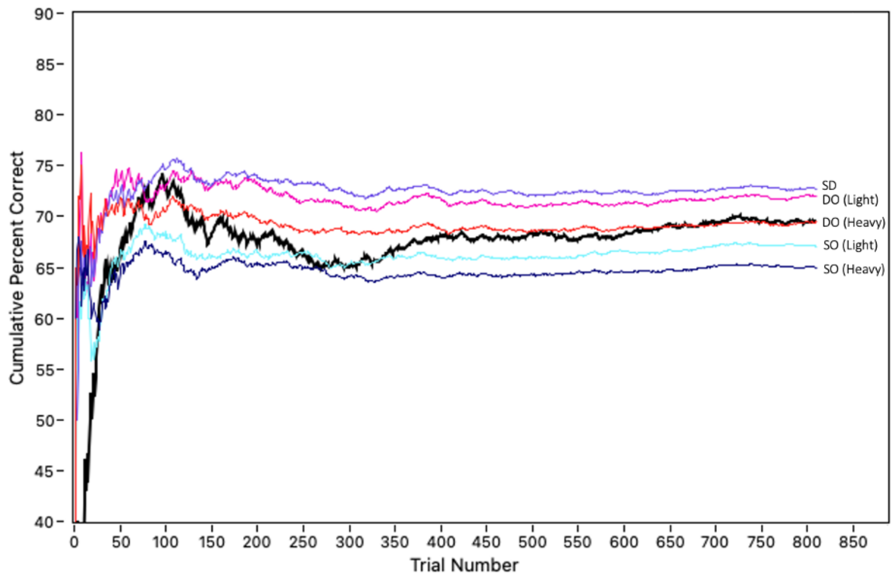
Bayesian model comparison classified the participant as using an SO (Heavy) strategy on most trials (Figure 3.4d). On Day 2, the Bayesian model comparison classified the participant as using an SD strategy on most trials (Figure 3.4f).

We compared both sample participants in Figure 3.4 to 5 Feature-Focused Bayesian Observers. In one comparison, the DO (Light) Observer achieved the highest PC by the end of trial 810, while the SD observer achieved the highest PC in the other comparison. Overall, the DO (Light) Observer achieves the highest PC for 10 of 24 sets of participant models. The SD Observer achieves the highest PC for 13 of 24 sets of participant models. For one participant's set of computational observers, the DO (Light) and SD Observers end with an equal PC at trial 810.

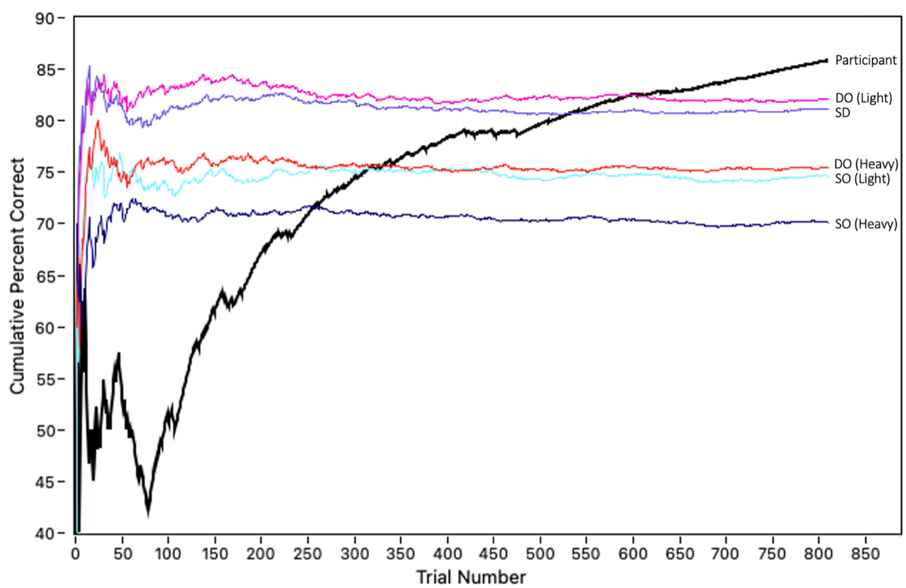
Upon comparing each participant to the Feature-Focused Bayesian Observers across 810 trials, our model comparison analysis classified all participants but 1 as either an SD or a DO (Light) Observer. Of the 6 participants in each experimental condition, 1 was classified as SD in each of the 0.75/0.75 and 0.75/1.25 conditions, while 2 were classified as SD in each of the 1.25/0.75 and 1.25/1.25 conditions. In the 0.75/0.75, 0.75/1.25, 1.25/0.75, and 1.25/1.25 conditions, 5, 5, 4, and 3 participants were classified as DO (Light), respectively.



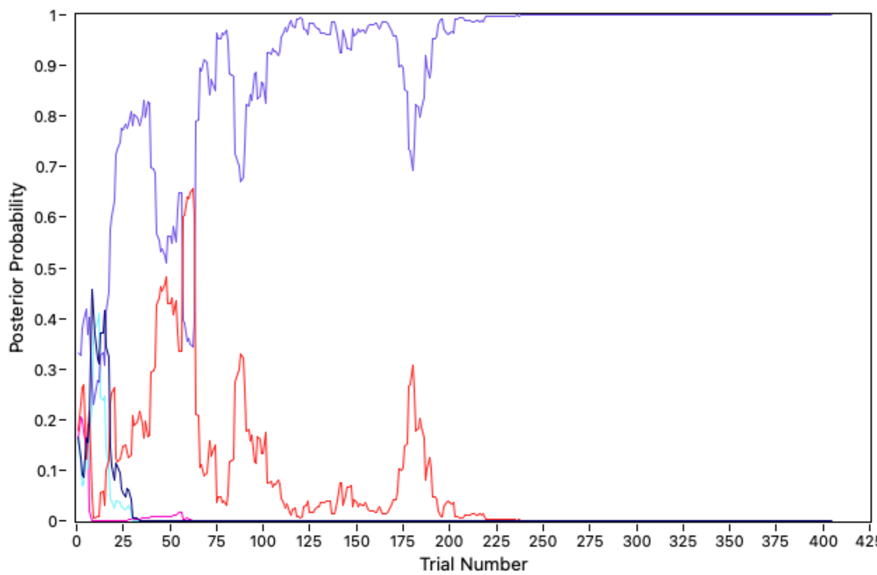

**(a)** Sample Participant 1: Learning Curves for Human Participant and Computational Models



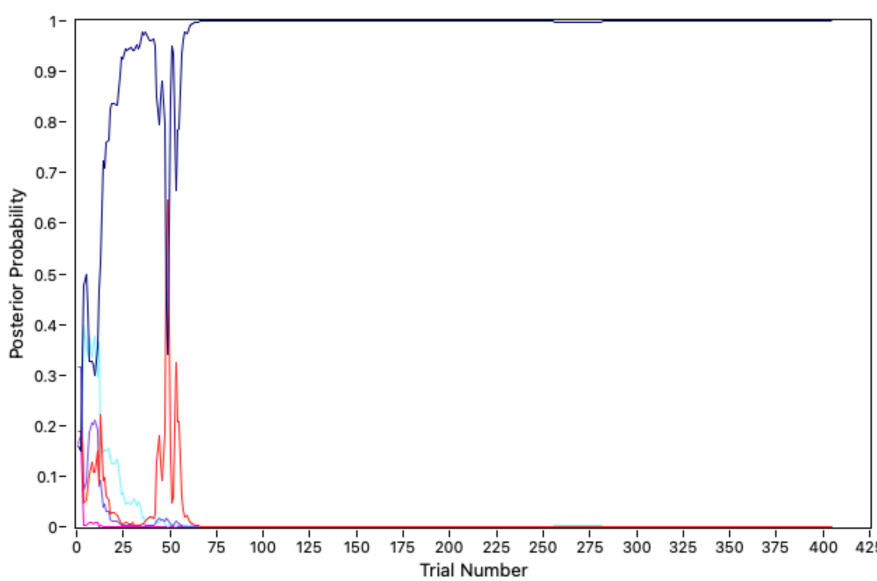

**(b)** Sample Participant 2: Learning Curves for Human Participant and Computational Models



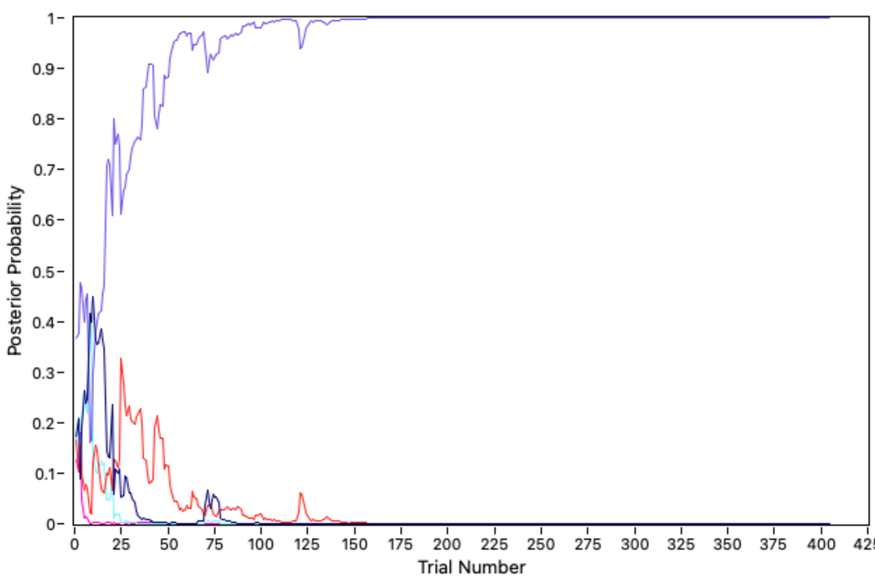
**(c)** Sample Participant 1: Day 1 Strategy Classification




**(d)** Sample Participant 2: Day 1 Strategy Classification



**(e)** Sample Participant 1: Day 2 Strategy Classification




**(f)** Sample Participant 2: Day 2 Strategy Classification

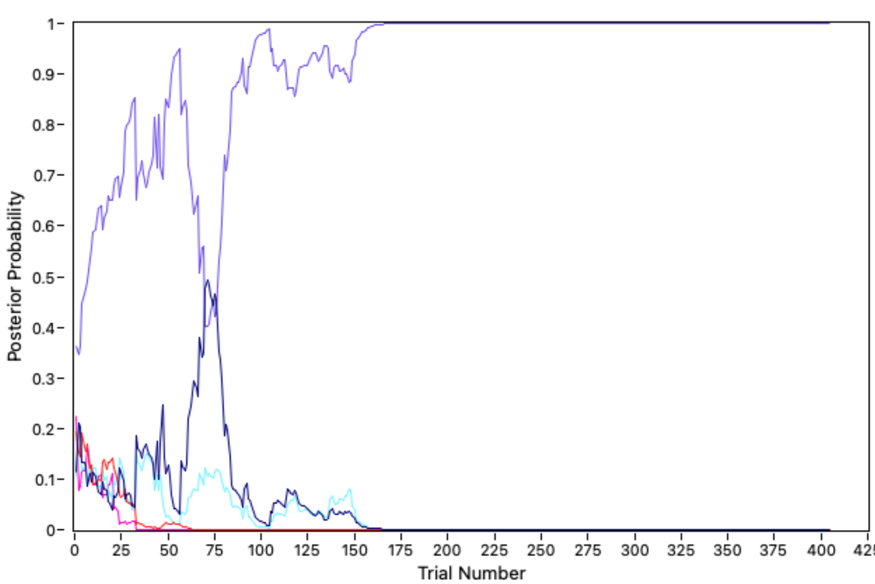


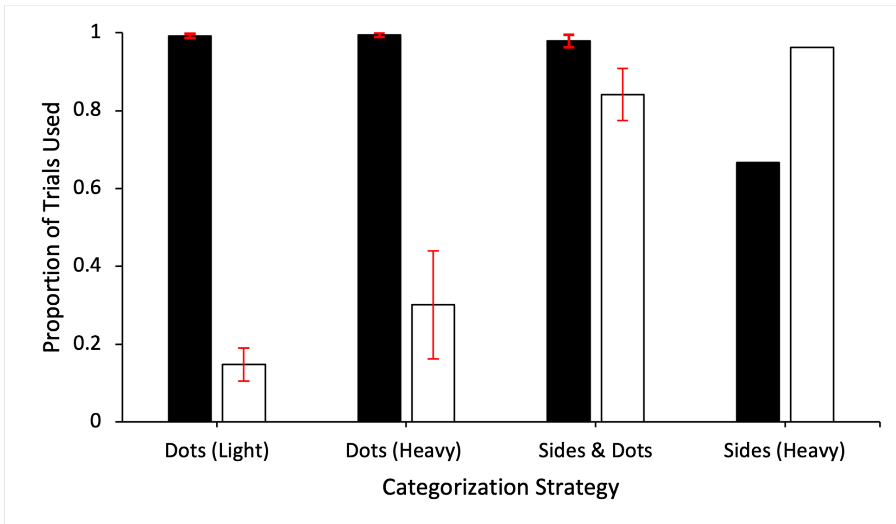

**Figure 3.4: Bayesian Classification of a Sample Participant** (a,b) Learning curves for two different human participants and 5 computational models of their performance, each made up of 10 repetitions of performance on 810 trials of the haptic categorization task. (c – f) Model posteriors for each computational model on all trials of the experiment. The participant and observers illustrated in a,c,e were tested in the 1.25/1.25 experimental condition. The participant and observers illustrated in b,d,f were tested in the 0.75/0.75 experimental condition.

### 3.3.3 Exploratory Procedures

We recorded EP use from each participant on 54 trials of the experiment and calculated the proportion of trials on which participants used swipe and contour following during haptic exploration (Figure 3.5). We averaged EP use over all participants classified as using a given categorization strategy. The 9 participants classified as DO (Light) on Day 1 swiped on 98.9% and contour followed on 23.0% of trials. The 11 participants classified as DO (Light) on Day 2 swiped on 99.0% and contour followed on 38.7% trials. The 7 participants classified as DO (Heavy) on Day 1 used swipe and contour following on an average of 99.5% and 30.2% of trials, respectively. The single participant classified as DO (Heavy) on Day 2 used both EPs on 100% of trials. The 7 participants classified as SD on Day 1 used swipe on 98.1% of trials and contour following on 82.1% of trials. On Day 2, the 12 participants classified as SD used swipe on 95.7% of trials and contour following on 89.2% of trials. One participant was classified as SO (Heavy) on Day 1, during which they used swipe on 66.7% of trials and contour following on 96.3% of trials.

Table 3.4 summarizes the average use of all 4 EPs on recorded trials.


**(a)** EP use on Day 1




**(b)** EP use on Day 2

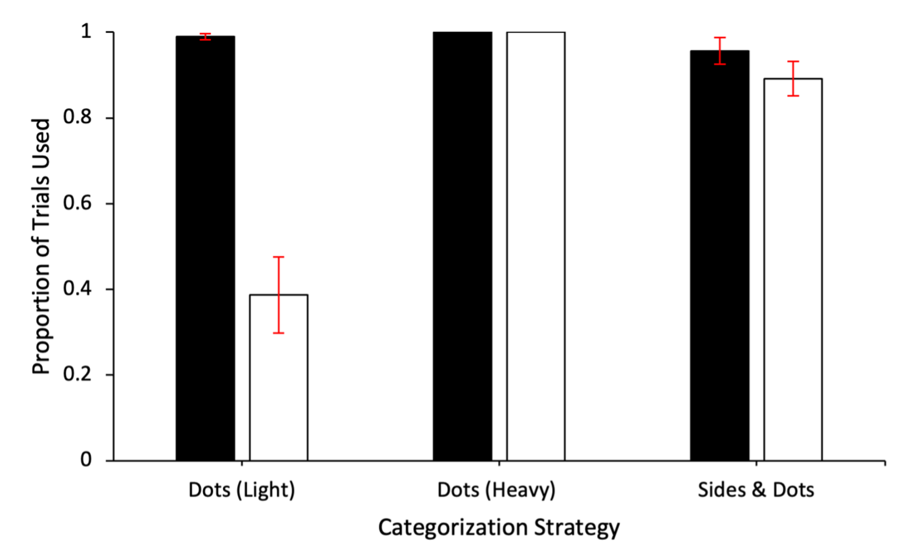



**Figure 3.5: Participant use of Swipe and Contour Following Exploratory Procedures** Average proportion of trials analyzed (each proportion is calculated with a total of 27 trials) in which the swipe and contour following strategies are employed by participants. Red error bars show ±1 SE. Proportions are specified for day (a) 1 and day (b) 2 individually. Number of participants averaged for each strategy corresponds to the number of participants classified as using that strategy. Number of participants is as follows: (a) DO (Light): n = 9 (b) DO (Light): n = 11 (a) DO (Heavy): n = 7 (b) DO (Heavy): n = 1 (a) SD: n = 7 (b) SD: n = 12 (a) SO (Heavy): n = 1.

**Table 3.4: Summary of Exploratory Procedure Use** Percent of trials (out of 27) on which participants classified as using each of the strategies used each of our five EPs. Each cell includes a percent that is averaged over all participants classified as using that strategy (number indicated in the rightmost column).

| | | Exploratory Procedure | | | | | |
|---|---|---|---|---|---|---|---|
| **Day** | **Strategy Classification** | **Static Contact (Dots)** | **Swipe** | **Static Contact (Sides)** | **Contour Following** | **Enclosure** | **Number of Participants** |
| **1** | **Null** | 0 | 0 | 0 | 0 | 0 | 0 |
| **1** | **SD** | 66.67 | 98.15 | 96.30 | 82.10 | 0.62 | 7 |
| **1** | **SO (Heavy)** | 100.00 | 66.67 | 100.00 | 96.30 | 0 | 1 |
| **1** | **DO (Heavy)** | 56.35 | 100.00 | 98.94 | 30.18 | 1.06 | 7 |
| **1** | **Sides (Light)** | 0 | 0 | 0 | 0 | 0 | 0 |
| **1** | **DO (Light)** | 68.15 | 98.89 | 86.67 | 22.96 | 0.74 | 9 |
| **2** | **Null** | 0 | 0 | 0 | 0 | 0 | 0 |
| **2** | **SD** | 67.28 | 95.68 | 83.02 | 89.20 | 0 | 12 |
| **2** | **SO (Heavy)** | 0 | 0 | 0 | 0 | 0 | 0 |
| **2** | **DO (Heavy)** | 0 | 100.00 | 100.00 | 100.00 | 0 | 1 |
| **2** | **Sides (Light)** | 0 | 0 | 0 | 0 | 0 | 0 |
| **2** | **DO (Light)** | 49.61 | 99.31 | 95.29 | 39.00 | 0 | 11 |

## 3.4 Discussion

### 3.4.1 Summary of Findings

Upon comparing participants to our computational observers, we found that our Feature-Focused Bayesian Observer effectively simulated all 24 participants' performance. Of the observers we considered, our Null and Prevalence-Focused Observers did not effectively simulate any of our participants' performance. This finding aligns with past studies, which suggest that Bayesian models accurately simulate human performance on a number of sensory categorization tasks (Bejjanki et al., 2011; Bankieris et al., 2017). Our model comparison further recognized a change in categorization strategy for 9 of our 24 participants between

Days 1 and 2, suggesting that they may change strategies to optimize performance.

#### Exploratory Procedure Analysis Supports the Bayesian Model Comparison Results

Lederman and Klatzky (1993) determined the relative accuracy of each EP in interpreting different haptic properties. Swipe proved most reliable in interpreting texture compared to other EPs, while contour following proved most reliable in interpreting exact shape compared to other EPs. Texture and exact shape are the primary informative properties associated with the dots and sides features of our objects, respectively. As a result, we expected these EPs to correlate with the strategy classification of each participant made by our Bayesian model comparison.

Figure 3.5a illustrates the proportion of trials on which participants using each categorization strategy used swipe and contour following on Day 1. Remarkably, our findings align closely with expectations. DO (Light) and DO (Heavy) categorization strategies require participants to interpret the texture of the object. As predicted, participants classified as DO (Light) and DO (Heavy) used swipe on almost all trials, while they use contour following on less than half of trials. An SO (Heavy) categorization strategy requires participants to interpret the exact shape of the object. As predicted, the participant classified as SO (Heavy) used contour following on almost trials, while they use swipe on far fewer trials. Our results further show that participants classified as using an SD strategy used both EPs on a large proportion of trials.

Figure 3.5b illustrates the same information for Day 2. Average EP use by participants classified as using DO (Light) and SD categorization strategies follow the same pattern as Day 1. In contrast to our expectations, DO (Heavy) has 100% average use for both swipe and contour following on Day 2. Because we only classified one participant as DO (Heavy) on Day 2, this could result from individual variation in performance. Alternatively, this may indicate that the participant was in the process of transitioning to an SD strategy, requiring them to use both EPs.

These results provide evidence to validate the accuracy of our Bayesian model comparison. They also suggest that we can reliably track EPs in our video recordings.

### 3.4.2 The Role of Memory

One considerable difference between our Feature-Focused Bayesian Observer and humans is memory abilities. We know that sleep consolidates learning into long term memory, and also that humans can forget information over time (Hennevin et al., 2007). Because Days 1 and 2 of this study were separated by 1 week, both of these factors likely influenced human performance on Day 2, but are not considered by our computational observers. The decreases in performance between Days 1 and 2 observed in Figure 3.1c,d suggest that human performance worsened at the beginning of Day 2, but improved rapidly as participants completed trials.

### 3.4.3 Defining Optimality for Individual Participants

In the previous chapter, we outlined the challenge of defining optimality (Section 2.4.2). Here, we highlight the same concept for individual human participants. As previously mentioned, the experimental condition, object sequence, and an individual's sensory noise all aid in determining an optimal strategy. In Chapter 2, we determined that the DO (Light) Observer has the highest asymptote in the 0.75/0.75 condition, while the SD Observer has the highest asymptote in the 1.25/1.25 condition. In accordance with this finding, we classified 5 participants as DO (Light) and 1 as SD in the 0.75/0.75 condition, while we classified 3 participants as DO (Light) and 2 as SD in the 1.25/1.25 condition. This suggests that optimality – and the strategy that participants choose to follow – are not generalizable, but rather dependent on the categories under investigation.

### 3.4.4 Time Dependence

We consider the possibility that performance on our haptic categorization task varies as a function of time provided during haptic exploration. In the current study, we allowed participants 5 seconds for haptic exploration. With enough time, participants' sensory noise associated with measuring the sides and dots features would likely decrease considerably. If the sensory noise decreased to a negligible value with prolonged exploration time, we could modify our Feature-Focused Bayesian Observer to consider only category noise to model the same haptic categorization task with unlimited time provided for haptic exploration. In this scenario, our SD and DO Observers would have the same (negligable) amount sensory noise, where we previously considered a DO observer with light sensory noise to reflect a participant attending to only the one feature and therefore having less sensory noise. While this sensory noise difference allowed a DO categorization strategy to prove optimal in many cases, increasing the time to essentially remove sensory noise as a variable would remove this occurrence and an SD observer would consistently obtain the most sensory information and perform optimally on the task, compared to the other models.

### 3.4.5 Rationale Behind Categorization Strategy Changes

Figure 3.3 summarizes the categorization strategy changes made by 9 of 24 participants between study days. While it proves difficult to determine whether these participants move toward more optimal strategies, we can reasonably explain the 5 strategy changes observed as a shift toward optimality. The 6 participants using dots (light/heavy) or SO (Heavy) on Day 1 and SD on Day 2 may attend to a second feature to make more reliable interpretations of the objects on Day 2. The 3 participants using DO (Heavy) or SD on Day 1 and DO (Light) on Day 2 minimize their sensory noise, possibly with experience or increased attention, to make more reliable interpretations of the objects.

Tables A.3 and A.2 further support the prediction that participants become more optimal on Day 2. They show that all 9 participants who change strategy achieve a higher PC by the end of Day 2 than they achieved at the end of Day 1. That said, we also consider the possibility that this improvement on Day 2 arises from a better understanding of the categories, rather than a change in categorization strategy.

Future studies may investigate and model the process by which individuals determine whether or not to change strategy. This process likely considers factors to maximize performance on the task, as well as quantify the amount of feedback required to identify the best strategy.

#### Number of Categorization Strategy Changes

In this chapter, we consider strategy changes between study Days 1 and 2. In Figure A.3.1, we narrow the window of trials considered to classify participant strategies. In reality, participants can make infinite strategy changes during the experiment. For our Bayesian model comparison to classify a participant's strategy, it must compare their performance on a range of trials to that of the computational models. Providing the Bayesian model comparison with a smaller range of trials to compare results in a less reliable classification. Providing it with a larger range of trials increases the risk of overlooking a strategy change within that trial range. We are limited in our ability to recognize strategy changes because we need to balance these two factors.

Figure A.3.1 suggests that only 5 participants maintained the same categorization strategy throughout all block sections – 4 less than when comparing Day 1 and 2 categorization strategies. We can consider the possibility that all participants use multiple strategies at some point, but our model comparison lacks the sensitivity to recognize these strategy changes.

### 3.4.6 Exploratory Procedure Functionality

While reviewing videos, we recorded information about participant use of 4 different EPs: static contact, enclosure, swipe, and contour following. During the experiment, most participants lifted objects off of the table to explore them. As a result of holding these objects, almost every participant consistently used static contact during haptic exploration. While participants can obtain information using static contact, its almost constant use made it challenging to determine whether a participant used it as a functional strategy (i.e. used to acquire sensory information) or a non-functional strategy (i.e. simply used to support the object, providing no information about the sides or dots feature). Table 3.4 summarizes the percent of trials on which participants of each strategy classification use static contact. These percentages consider only trials in which static was used through the middle of the trial, as almost all participants use static contact in a non-functional manner at some point to lift objects off of the table. We cannot know for certain, from a video, whether a strategy is functional or not. This makes static contact an unreliable indicator of categorization strategy on our haptic categorization task.

Unlike static contact, participants use enclosure on an average of less than 1.5% of trials. We expected minimal use of enclosure, as past studies concluded that enclosure is minimally reliable in interpreting object texture and shape (Lederman and Klatzky, 1993).

In future studies, we could further consider the possibility that the combination of EPs affects participant performance.

#### 3.4.7 Exploratory Procedure Ambiguity

Some hand movements used to feel our objects can be brief and subtle. This proved to be a common challenge when tracking EP use by each participant. In addition, the top down camera angle used to record video in our experiment produced some ambiguity, as there were frames in which we struggled to determine whether participants made contact with the object. We attempted to minimize errors by requiring two different experimenters to record information from each video. Future studies may add a second camera to capture video of hand movements from a second angle to minimize uncertainty in scoring EPs.

## 3.5 Conclusion

Overall, we provide evidence that humans employ Bayesian strategies to interpret and categorize stimuli, as our Feature-Focused Bayesian observer effectively simulated human sensory categorization in the haptic modality. We further validated these results with an analysis of each participant's EP use, which strongly supported the classification results of our Bayesian model comparison.

# Chapter 4

# Experiment 2: Manipulation of Category Prior Probabilities

## 4.1 Introduction

In Chapter 3, we outlined a study in which participants group objects into one of two categories with different levels of variation in the two categories. While this provides insight into the processing behind sensory categorization in humans, we can alter our categories further to better reflect naturally occurring categories. In reality, categories rarely occur in equal proportions. In most cases, one category appears more frequently than another.

In this chapter, we report a study in which we tested participants on a second variation of our haptic categorization task, where participants could receive objects from Category A more frequently on study Day 2 (i.e. modified prior probabilities for each category). We considered two hypotheses: (1) participants can learn and integrate category priors in their categorization decisions or (2) participants cannot learn and integrate category priors in their categorization decisions.

## 4.2 Methods

### 4.2.1 Participants

We collected data from 20 participants aged 17 to 22 (average age 19 years, 18 F, 2 M), recruited from the McMaster undergraduate student population using SONA. All participants self-reported they were unaffected by any of the following conditions: diabetes, nervous system disorder or injury, learning disability, dyslexia, attention deficit disorder, cognitive impairment, carpal tunnel syndrome, arthritis of the hands, hyperhidrosis. This study was approved by the McMaster Ethics Research Board and all participants were compensated with cash or course credit.

### 4.2.2 Experimental Conditions

We assigned each participant to 1 of 4 experimental conditions outlined in Table 4.1 based on their entry number into the experiment. In total, we assigned 5 participants to each experimental condition. In all conditions, we informed participants that we were equally likely to select an object from Category A or B to present to the participant (i.e. uniform priors for categories) on Day 1. On Day 2, we told participants that the prevalence of each category may change. Depending on the experimental condition in which participants were tested, category priors either remained equal, or shifted to increase the Category A prior, making us more likely to present an object from Category A on a given trial. In all conditions, objects were drawn from Category A and B distributions with a category sigma value of 0.75 in both feature dimensions. Note that these conditions were identical to those in Chapter 2: Table 2.9, on which we tested our computational observers.

**Table 4.1: Experiment 2 – Conditions** Summary of the four experimental conditions tested in Experiment 2, where we manipulated the category proportion of objects presented from the two categories on each day of the experiment. On Day 1, participants in all conditions received the same proportion of objects from A and B. See Chapter 2, Figure 2.9 for additional details.

| | **Day 1** | **Day 2** |
|---|---|---|
| Condition 1 | $P(A) = 0.5, P(B) = 0.5$ | $P(A) = 0.5, P(B) = 0.5$ |
| Condition 2 | $P(A) = 0.5, P(B) = 0.6$ | $P(A) = 0.6, P(B) = 0.4$ |
| Condition 3 | $P(A) = 0.5, P(B) = 0.7$ | $P(A) = 0.7, P(B) = 0.3$ |
| Condition 4 | $P(A) = 0.5, P(B) = 0.8$ | $P(A) = 0.8, P(B) = 0.2$ |

### 4.2.3 Experimental Procedure

We used an object and experimental procedure identical to that described for Experiment 1 (Chapter 3), with the exception of the experimental conditions in which we manipulate category priors rather than category variation and the constant category sigma value of 0.75.

## 4.3 Results

### 4.3.1 Human Participant Results

We calculated the average PC of participants in each experimental condition and plot the average PC on Day 1 against the average PC on Day 2 (Figure 4.1). On average, participants in all conditions achieved a higher PC on Day 2 than on Day 1. Participants in the 80/20 condition achieved the highest average PC on Day 1, while participants in the 70/30 condition achieved the lowest average PC. On Day 2, participants in the 80/20 condition achieved the highest average PC, followed by participants in the 70/30 condition. Participants in the 50/50 and 60/40 conditions achieved a similar PC on Day 2. Note that category prevalence was identical for all conditions on Day 1, so we attribute any performance differences to variability in participant responses and/or individual differences between participants assigned to each condition. In a similar vein, category prevalence remained constant across Days 1 & 2 in the 50/50 condition, so we attribute performance differences between days to response variability and/or additional practice on the task during Day 2 trials. While these factors remain relevant in the other 3 conditions, we consider the added factor of modified category priors as a rationale for changes in average PC as well.

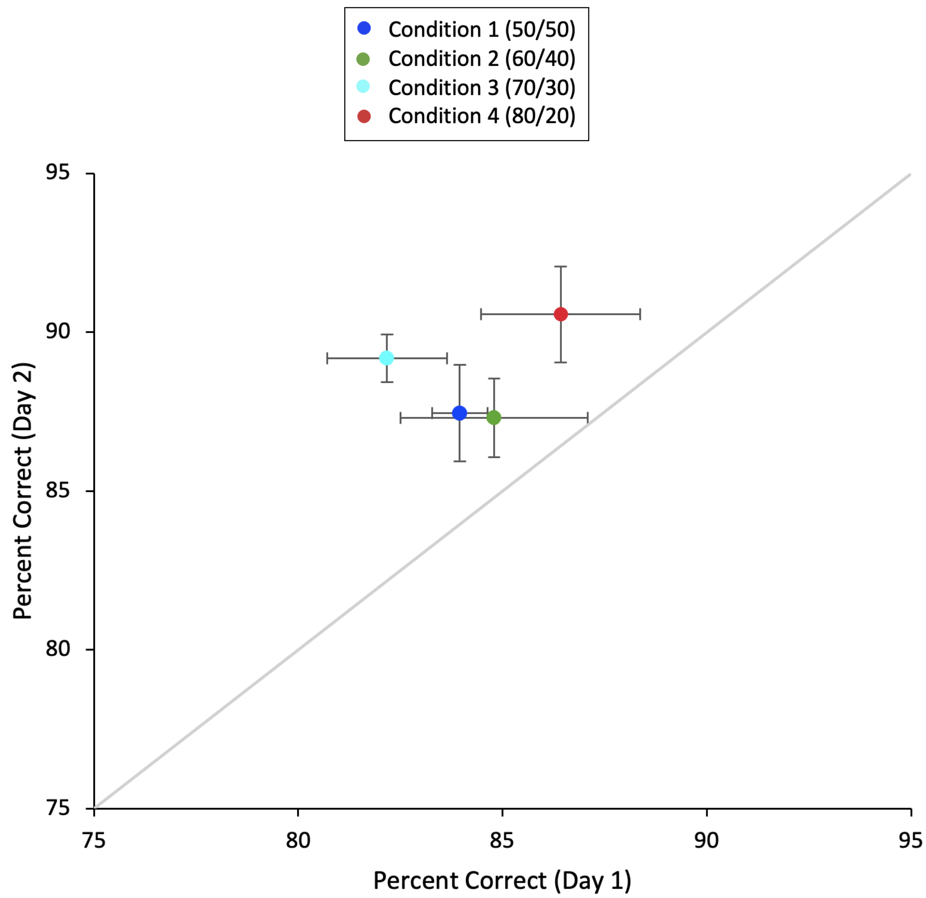


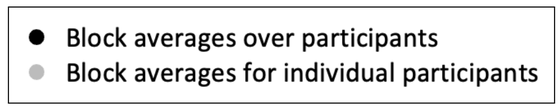

**Figure 4.1: Experiment 2 – Human Performance on Haptic Categorization Task** Average PC of participants in each of the four experimental conditions study Days 1 and 2. The light grey line represents points at which Day 1 PC equals Day 2 PC. The error bars show ±1 SE. n = 20 (n = 5 within each condition).

In Figure 4.2, we plot the PC of participants in 3-block intervals throughout the experiment, referred to as block sections. Block section 1 includes Day 1 blocks 1 to 3; block section 2 includes Day 1 blocks 4 to 6, and so on. Black lines illustrate the average PC across participants in a condition, while grey lines illustrate PC of individual participants. Average PC increased between block sections 1 and 6 in all experimental conditions. Average PC in the 50/50, 60/40, 70/30, and 80/20 conditions increased by 5.93%, 7.56%, 8.59%, and 3.70%, respectively. In 3 of 4 experimental conditions, category priors changed between the end of block section 3 (end of Day 1) and the beginning of block section 4 (beginning of Day 2). Changes in average PC between the end of Day 1 and beginning of Day 2 in the 50/50, 60/40, 70/30, and 80/20 conditions changed by the following amounts, respectively: 2.22%, -0.59%, 4.30%, and 6.81%.

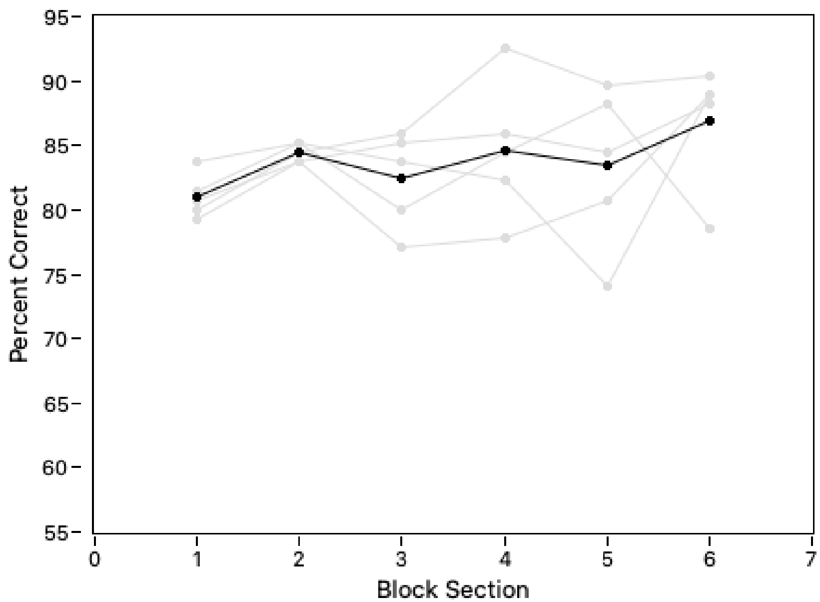
### 4.3.2 Comparison to Bayesian Observers

We compared each participant's performance to that of our Feature-Focused Bayesian Observers. As in Chapters 2 and 3, we omitted the Null, Prevalence MAP Estimate, and Prevalence Probability Matching Observers from these analyses, as none of their Bayes' factor exceeded $10^{-28}$ when compared to any of the participants' performance.

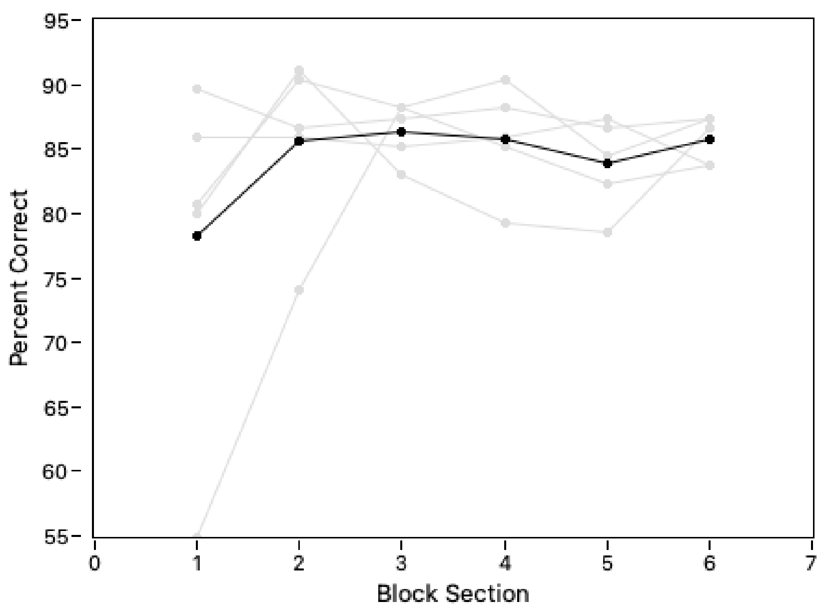
We compared participants to computational observers simulated using the same object sequence. To classify categorization strategy, we considered the posterior for both the constant and updating category prior observers that employed that strategy. For example, the posterior probability for an SD Observer equalled the sum of the posteriors for



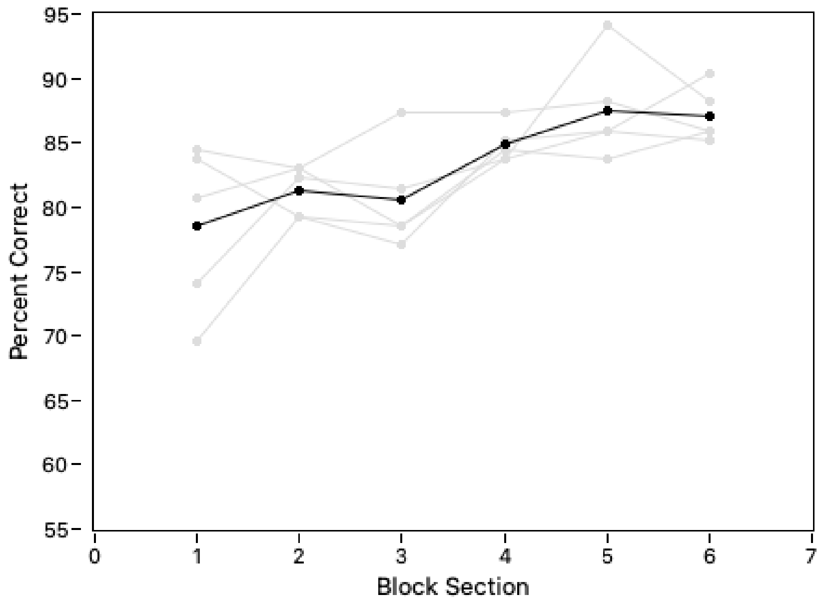

**(a)** Condition 1 (50/50)




**(b)** Condition 2 (60/40)



**(c)** Condition 3 (70/30)




**(d)** Condition 4 (80/20)

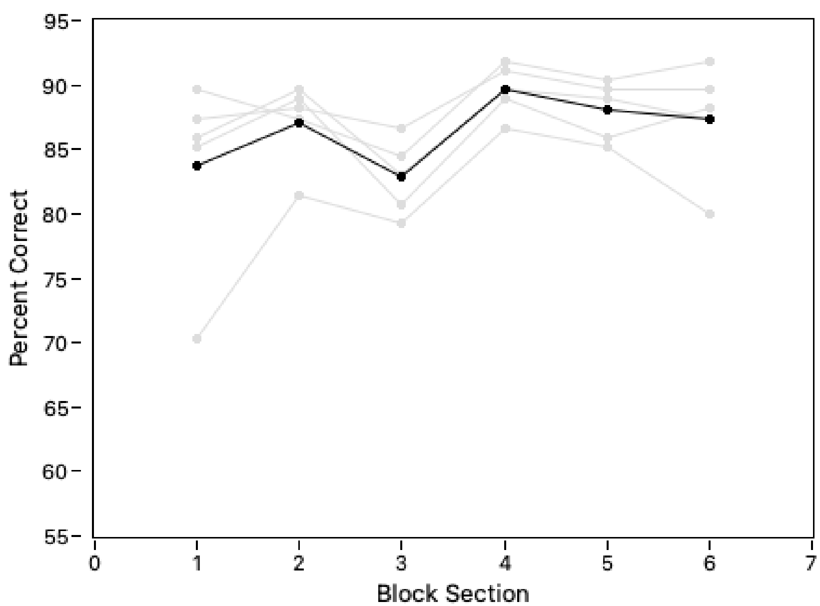



**Figure 4.2: Experiment 2 – Learning on Haptic Categorization Task** Participant PC averaged over windows of 135 trials (block sections). Black lines illustrate the average learning curves across all participants in a given condition. Grey lines illustrate individual participant PC on each block section. n = 20 (n = 5 within each condition).

the constant and the updating SD Observers. On Day 1, we classified 18 of 20 participants as using a DO (Light) categorization strategy, 1 participant as using an SD categorization strategy, and 1 participant as using an SO (Light) categorization strategy (Figure 4.3). On Day 2, we classifies 12 participants as DO (Light), 7 participants as SD, and 1 participants as DO (Heavy) (Figure 4.3). Figure 4.3 further illustrates changes in participant categorization strategies between Days 1 and 2, where we classified 11 participants as maintaining the same strategy and 9 as changing strategies. Of the 18 participants classified as DO (Light) on Day 1, 11 remained DO (Light) on Day 2, 6 changed to an SD categorization strategy, and 1 changed to a DO (Heavy) categorization strategy. The participant classified as SD on Day 1 changed to DO (Light) on Day 2, while the participant classified as SO (Light) on Day 1 changed to SD on Day 2.

For each participant, we quantified the posterior probability for constant and updating category prior observers as the sum of the posteriors for observers with the same category prior assumption using any of the categorization strategies. On Day 2, updating category prior observers best simulated the performance of 15 participants, while constant category prior observers best simulated the performance of the remaining 5 participants (Figure 4.4). Of these 5 participants, 2 were tested in the

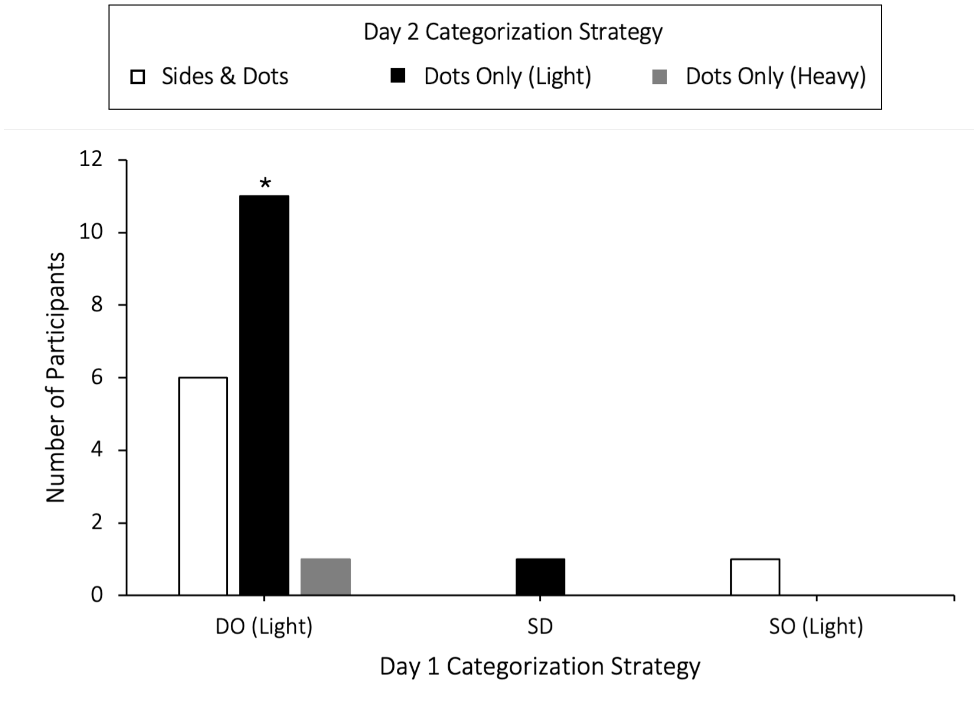



**Figure 4.3: Experiment 2 – Changes in Strategy Classification Between Experiment Days** Count of participant classification changes between Days 1 and 2 of the experiment. The x-axis represents a participant's Day 1 categorization strategy, while the bar colour represents a participant's Day 2 categorization strategy. *11 DO (Light) participants maintained the same strategy across both study days, while the remaining 9 participants switched categorization strategies between Days 1 and 2. n = 20.

50/50 condition, 2 in the 60/40 condition, and 1 in the 70/30 condition. Updating category prior observers best simulated all participants' performance in the 80/20 condition.

## 4.4 Discussion

### 4.4.1 Summary of Findings

Results from this study further support the hypothesis that Bayesian models can effectively simulate human categorization strategies in our haptic categorization experiment. We determined that Bayesian ob-

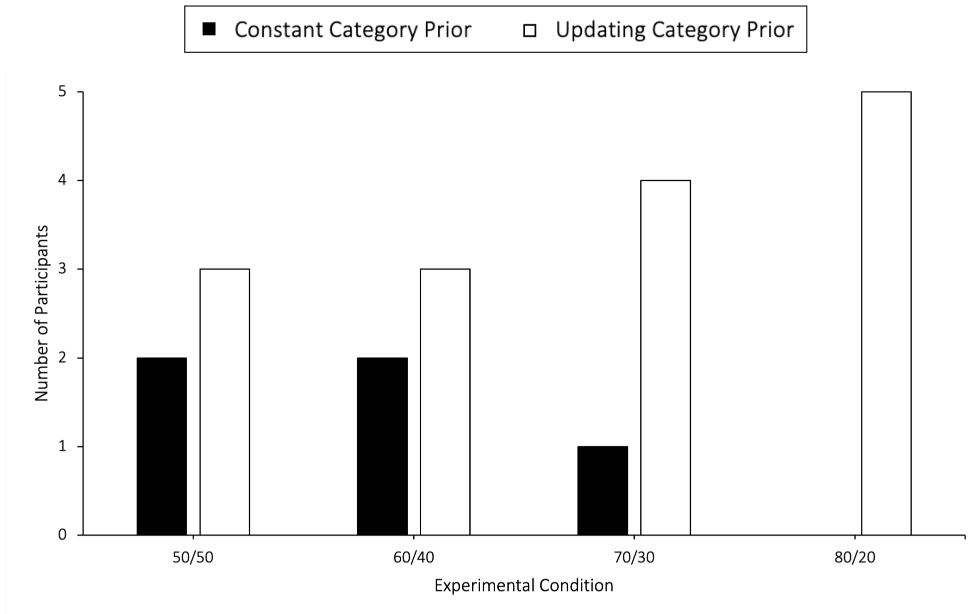



**Figure 4.4: Number of Constant and Updating Prior Strategy Classifications** Count of the number of participants in each experimental condition best simulated by constant and updating category prior observer on study Day 2. Black bars illustrate the number of participants best simulated by a constant category prior observer, while white bars illustrate the number of participants best simulated by an updating category prior observer. n = 20 (n = 5 within each condition).

servers best simulate human performance when they update their beliefs of category prevalence. These findings build upon our results from Experiment 1 to suggest that participants incorporate changes in category prevalence, as well as changes in category variation, into their categorization decisions.

### 4.4.2 Classifying Participants in the 50/50 Condition

As category priors become more extreme, performance of the constant and updating category prior observers becomes increasingly different. Consequently, these two observer types perform quite similarly when simulated in the 50/50 experimental condition. In this study we classified 2 of 5 participants tested in the 50/50 condition as using a constant category prior categorization strategy on Day 2, while the other 3 were classified as using an updating category prior strategy. In theory, participants should employ an updating category prior strategy on Day 2 because we do not inform them of the category priors in our experiment instructions. One possible explanation for the constant category prior strategies is that people may favour 50/50 category priors until they are sufficiently different from 50/50.

#### 4.4.3 Rationale Behind Categorization Strategy Changes

On Day 1 of this experiment, we classified 18 of 20 participants as using a DO (Light) categorization strategy. According to Table 2.5, the DO (Light) Observer approached the highest asymptote on Day 1, compared to all other observers. This provides a reasonable rationale for this finding, that a DO (Light) categorization strategy is optimal.

We determined that 9 participants changed categorization strategies between Days 1 and 2, while 11 participants maintained the same strategy. Table 2.5 indicates that the difference between the SD and DO (Light) Observer asymptotes decreases as category priors get more extreme. This further explains the change of categorization strategy from

DO (Light) to SD observed in 6 participants.

### 4.4.4 Learning the Category Priors

Past studies suggest humans can learn and incorporate category priors into their decisions during categorization tasks (Berniker et al., 2010; Nagai et al., 2012). Our results, illustrated in Figure 4.4, further support this claim. In all experimental conditions, updating category prior observers best simulated performance more commonly than constant category prior observers. As the category priors became more extreme and the updating and constant observers varied more in performance, updating category prior observers best simulated a larger proportion of participants' performance. This suggests that human participants can, in fact, learn and incorporate category priors into their categorization decisions.

### 4.4.5 Human Implications

In the previous chapter, we outlined several considerations that become relevant when testing human participants. These remain relevant for the participants discussed in the current chapter. See Chapter 3, Sections 3.4.2, 3.4.3, 3.4.3, and 3.4.4 for details.

## 4.5 Conclusion

These findings provide further evidence to support the hypothesis that Bayesian models effectively simulate sensory, and more specifically tactile, perception. They further support the hypothesis that people consider and integrate environmental priors, namely the prevalence of categories, in the decision making process.

# Chapter 5

# General Discussion

## 5.1 Summary

This project aimed to deepen our understanding of the haptic modality, specifically the processes behind haptic categorization. We built upon research from Gauder (2024), which suggested that Bayesian Observers that account for sensory measurement noise experienced by humans can effectively model human haptic categorization. We used computational observers to model human performance on two variations of a haptic categorization task, during which participants categorize haptic stimuli and use feedback to learn two novel Gaussian categories, labelled A and B. We used haptic stimuli that differ in number of sides and dot spacing, 2 features that informed category identity.

### 5.1.1 Computational Observers

We produced computational observers that use unique computational approaches to model this categorization task. Our central model, referred to as our Feature-Focused Bayesian Observer, weighted the available sensory features (sides and dots) based on their reliability to 'feel' and categorize haptic stimuli. We successfully modelled 3 versions of this Feature-Focused Bayesian Observer, each of which learned the categories and categorized the stimuli using one or both of the available sensory features (sides, dots). We compared participants to our computational models using Bayesian model comparison and identified the observer that best simulated participant performance (i.e. best simulated their categorization strategy) with high accuracy.

### 5.1.2 Experiment 1: Modification of Category Variation

We tested 24 human participants in 1 of 4 experimental conditions, where each condition had unique overlap between the Gaussian categories A and B. Participants tested in the condition with the least category overlap achieved the highest average PC, while participants tested in the condition with the most category overlap achieved the lowest average PC. By study Day 2, all but 1 of the 24 participants' performance was best simulated by either an SD or a DO (light) Observer. In all cases, one of these two Observers produced the highest percent correct on the task, suggesting that participants employed Bayesian-optimal categorization strategies during the experiment. These categorization strategy classifications are further validated by our exploration of EP use, where participant EPs closely aligned with the categorization strategy identified.

### 5.1.3 Experiment 2: Modification of Category Prior Probabilities

We tested 20 human participants in 1 of 4 experimental conditions, where each condition received a different ratio of Category A and B objects on study Day 2. Bayesian models that did not assume equal prevalence of A's and B's and instead estimated the prevalence of each category simulated the performance of 15 participants better than models that did make this assumption. These results support our hypothesis that humans can estimate and integrate category prevalence into their categorization decisions.

## 5.2 Future Directions

### 5.2.1 Quantify Sensory Noise for Individual Participants

Individual differences pose a challenge for researchers during behavioural studies. While humans share many behaviours, each individual may have a distinct process and physical compositions (ex. tactile receptor density, skin compliance, etc.) for considering and interpreting relevant variables when completing these studies. The average values for sensory noise considered in this study were calculated using data collected from participants tested previously on the same object set. Future research may test participants on that initial task to quantify their individual sensory noise prior to testing them on a categorization task, allowing the researchers to produce Bayesian models with their specific sensory noise.

Another factor that makes an individual's sensory noise difficult to quantify is that it may differ across trials. For example, participant focus and attention span influence sensory noise and almost certainly vary throughout the experiment. Because these factors can vary meaningfully and prove almost impossible to measure, we rely on our experimental design to maximize participant focus and minimize the impact of this noise on performance. We implement breaks and provide participants with their scores throughout the experiment to maximize focus and motivation.

#### 5.2.2 Quantify Sensory Noise for Individual Stimulus Levels

The sensory noise associated with a stimulus commonly changes with stimulus level, where its magnitude is a constant fraction of the stimulus level, rather than remaining constant across all stimulus levels. This phenomenon is described by Weber's law (Ekman, 1959). Future studies may conduct a two interval forced choice task to determine the sensory noise associated with discriminating between specific feature levels (i.e. 4mm dot spacing, 5mm dot spacing, etc.), rather than assuming a constant value of sensory noise across feature levels.

#### 5.2.3 Quantify Sensory Noise based on Alternative Sensory Cues

The Bayesian observers considered in these studies assume that all participants interpret the sides and dots features as number of sides and dot spacing. While all participants feel objects with these same features, they don't necessarily attend to the same sensory cues. For instance, rather than interpreting number of sides, participants may consider side length or vertex angle. In a similar manner, they may interpret dot spacing as number of dots or vibration frequency when swiping across dots. While such alternative sensory cues provide information about the same object feature, they may not have a linear relationship with the number of sides and dot spacing cues and corresponding sensory noise that we considered.

### 5.2.4 Expand Object Set

Future studies may extend the object set in the current study to greater than 25 objects, in turn creating objects with smaller increments between feature values. The current object set uses increments of 1 side and 1mm for the sides and dots features, respectively. These increments may exceed the magnitude of some participants' sensory noise, making it difficult to quantify their sensory noise using the current object set. A smaller increment would allow us to better measure an individual's sensory noise and model their performance.

### 5.2.5 Modify Categories Further

In the current study, we test participants on 2 variations of our haptic categorization task. Analogous studies in other sensory modalities test participants on sensory categorization tasks with further manipulations applied to Gaussian categories, such as adding correlations between dimensions of the category (Bankieris et al., 2017). In a similar manner, future studies may further manipulate the categories in this haptic categorization task to determine whether Bayesian observers remain effective in simulating human performance. For example, we could further modify our categories to correlate the sides and dots features within each category, in turn adding a level of complexity for participants to learn and integrate into their decision-making process.

#### 5.2.6 Automate Exploratory Procedure Scoring

Agreement between our model comparison and EP scoring results dramatically increased our confidence in the ability of our Feature-Focused Bayesian Observer and model comparison process to effectively simulate and classify human performance and categorization strategy. Put differently, our EP analysis strongly supported the accuracy of our model comparison results. Ideally, we could score EP use across all trials and participants. However, watching and scoring trial videos proves extremely time consuming and tedious. An undergraduate student in the lab began preliminary exploration of machine learning algorithms that measure contact time between an object and a hand on a given trial. The surgical field has had success in a similar area, where they designed deep learning models that can identify surgical hand gestures and procedure types with accuracy significantly above chance (Luongo et al., 2021; Khalid et al., 2020). This emerging field of deep learning video analysis provides a promising foundation for automating our EP scoring process.

## 5.3 Broader Applications

As the theory that Bayes' Theorem effectively models sensory perception gains traction and evidence-based support, we can begin to broaden its applications. Past studies extend this research to special populations, such as autistic spectrum disorder (ASD), to propose the theory that ASD individuals establish unique priors compared to control individuals, resulting in varied perception of their environment (Angeletos Chrysaitis and Seriès, 2023). Despite this theory's popularity, studies provide weak evidence to support its claims. Sapey-Triomphe et al. (2023) recently used Bayesian models to find evidence that ASD populations do not have unique priors, but rather a unique learning process that results in varied perception. Our Feature-Focused Bayesian Observer models this learning phase and could further extend this research to haptic categorization tasks.

We can further extend this research to robotics that aim to simulate human tactile behaviours. Yuan et al. (2017) investigate the ability of a robot to perceive object hardness, a feature commonly explored using pressure. Solak and Jamone (2023) further explore hand movements for robotic haptic exploration, attempting to incorporate movements that allow an autonomous robotic hand to obtain information about an unknown object in a safe and effective manner. Our findings associate hand movements with optimal and sub-optimal strategies, which may help inform these findings and possibly aid in optimizing robotic Exploratory Procedure (EP)s.

# Appendix A

# Supplemental Information

## A.1 Participant Instructions

Below are the experiment instructions recorded and played for participants at the beginning of each study day. We repeated instructions for participants until they could provide a sufficient summary to the experimenter.

### A.1.1 Experiment 1 – Manipulation of Category Sigma: Participant Instructions

**Days 1 & 2** "For this experiment, your task will be to try to learn two different categories of objects using only touch. These two object categories are called Elyk and Noek and they differ by both shape and the density of dots on top. To be as good as you can, you will need to pay attention to both of these. Half of the objects are Elyk and half are Noek. There are many different Elyks and many different Noeks, and some of them look like each other. For example, say we measured the height of a group of 5-year-old's and 7-year-old's. Generally older children will be taller, but some younger children will be as tall or taller than some older children.

During the experiment, you will sit behind this sheet and reach your hands through where I can give you an object to explore. During a trial, I will reach into a box and pull out an object at random. I will place the object on the table and give you five seconds to explore it. The computer will beep when the trial is over; as soon as you hear this sound, immediately put the piece down. I will then put your answer into the computer and it will respond with whether you were correct or incorrect. This sound will play when you are correct, and this sound when you are incorrect. The first trial may seem strange as you will need to answer without any knowledge of the piece, but as the test proceeds you'll start acquiring knowledge.

This experiment is divided into blocks. There are 45 trials in each block. There will be a one-minute break after every block, and a five-minute break after every three blocks. At the end of each block, I will reveal your score. The categories will not change throughout the experiment; every block will follow the same procedures, so you'll find out if you're improving over time. We will also be recording your hands as you explore the objects. Do you have any questions?"

### A.1.2 Experiment 2 – Manipulation of Category Priors: Participant Instructions

**Day 1** "Imagine that you spent the morning collecting seashells on a beach. You now have a large bin of shells. Scientists recently discovered two new types of shells on this beach, called Elyks and Noeks. Your task is to learn to identify Elyks and Noeks, which differ by both shape and the density of dots on top, using only touch. To be as good as you can, you will need to pay attention to both of these features. Half of the shells on this beach are Elyk and half are Noek. There are many different Elyks and many different Noeks, and some of them look like each other. For example, say we measured the height of a group of 5-year-old's and 7-year-old's. Generally older children will be taller, but some younger children will be as tall or taller than some older children.

During the experiment, you will sit behind this sheet and reach your hands through. During a trial, I will select a shell from your collection, place it on the table, and give you five seconds to explore it. The computer will beep when the trial is over; as soon as you hear this sound, immediately put the shell down. I will then put your answer into the computer and it will respond with whether you were correct or incorrect. This sound will play when you are correct, and this sound when you are incorrect. At the end of the trial, I will put the shell back into your bin of seashells. The first trial may seem strange as you will need to answer without any knowledge of the shell, but as the test proceeds, you'll start acquiring knowledge.

This experiment is divided into blocks. There are 45 trials in each block. There will be a one-minute break after every block, and a five-minute break after every three blocks. At the end of each block, I will reveal your score. The categories will not change throughout the experiment; every block will follow the same procedures, so you'll find out if you're improving over time. We will also be recording your hands as you explore the objects. Do you have any questions?"

**Day 2** "Last week, we asked you to imagine collecting seashells on a beach with two new types of shells, Elyks and Noeks. That beach had the same amount of Elyks and Noeks. This week you have moved to a new beach. Scientists have also discovered Elyks and Noeks at this new beach, but do not know whether half of the shells on the beach are Elyks and half are Noeks, or if one shell type is more common than the other.

Imagine now that you collected a bin of shells from this new beach and must learn to identify Elyks and Noeks, which differ by both shape and the density of dots on top, using only touch. To be as good as you can, you will need to pay attention to both of these features. There are many different Elyks and many different Noeks, and some of them look like each other. For example, say we measured the height of a group of 5-year-old's and 7-year-old's. Generally older children will be taller, but some younger children will be as tall or taller than some older children.

During the experiment, you will sit behind this sheet and reach your hands through. During a trial, I will select a shell from your collection, place it on the table, and give you five seconds to explore it. The computer will beep when the trial is over; as soon as you hear this sound, immediately put the shell down. I will then put your answer into the computer and it will respond with whether you were correct or incorrect. This sound will play when you are correct, and this sound when you are incorrect. At the end of the trial, I will put the shell back into your bin of seashells.

This experiment is divided into blocks. There are 45 trials in each block. There will be a one-minute break after every block, and a five-minute break after every three blocks. At the end of each block, I will reveal your score. The categories will not change throughout the experiment; every block will follow the same procedures, so you'll find out if you're improving over time. We will also be recording your hands as you explore the objects. Do you have any questions?"

## A.2 Experimenter Instructions for Compiling EP Data

**Table A.1: Exploratory Procedure Compilation Guidelines** Different pairs of experimenters (from the 4 available experimenters) reviewed and scored video recordings of each participant's hand movements during this experiment. If the pair of experimenters scored any movements differently, then a third experimenter followed these guidelines to compile information from the pair of experimenters. Note that experimenters were instructed to highlight information in yellow if they were uncertain of their response.

| **Case** | **Resolution** |
|---|---|
| Experimenter responses match | Record matching response |
| One experimenter highlighted information in yellow (unsure of response) and the other did not | Record non-highlighted response |
| Experimenter comments provide reasoning and resolution for a disagreement | Use comments to determine the appropriate information |
| Trial times differ by less than 4 seconds | Average the two times |
| One experimenter records timing in trial as 'middle' and the other records 'both' | Record 'both' |
| One experimenter records timing in trial as 'Beginning & End' and the other records 'both' | Record 'Beginning & End' |
| One experimenter records 'TRUE' for an EP and the other records 'FALSE' | Watch the trial video. If the procedure was not used, record 'FALSE'. If the procedure was used, record 'TRUE' and record the remaining information as done by the experimenter that initially entered 'TRUE'. |
| The procedure is static. One experimenter recorded 'TRUE' at 'beginning/end' and the other recorded 'FALSE' | Record 'TRUE' at 'beginning/end' |
| None of the above rules apply | Watch the trial video. Record the the response of the experimenter with whom you agree. |
| All 3 experimenters disagree | All three experimenters rewatch the trial video together and discuss to determine the correct information to record |

## A.3 Chapter 3: Supplemental Information

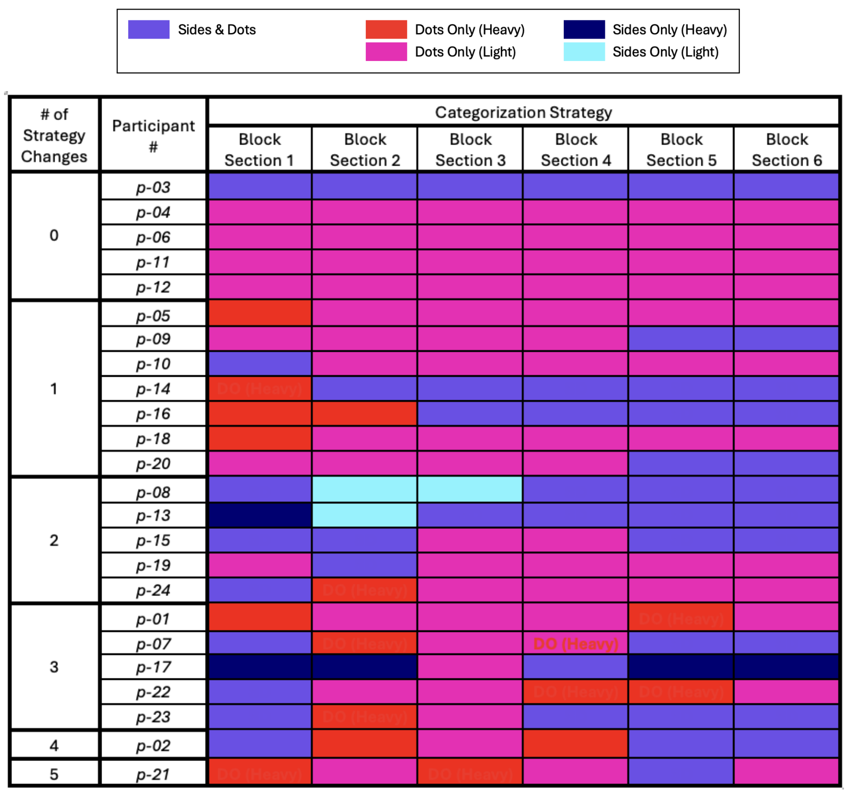
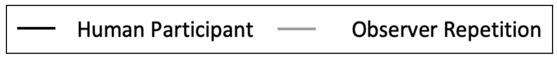
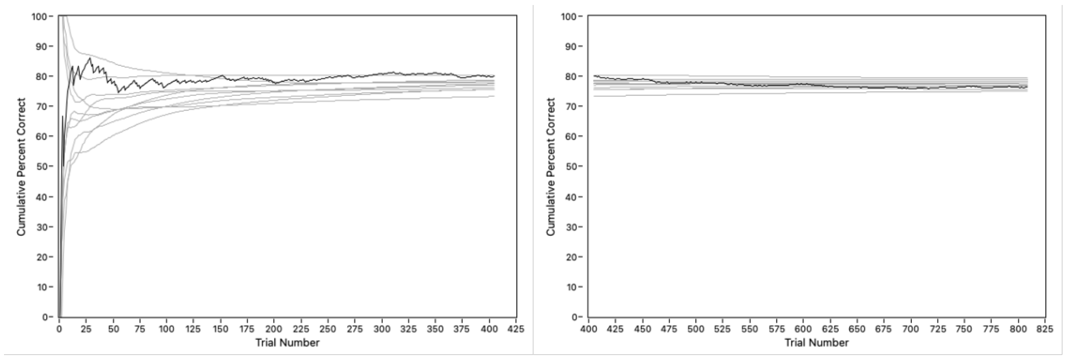


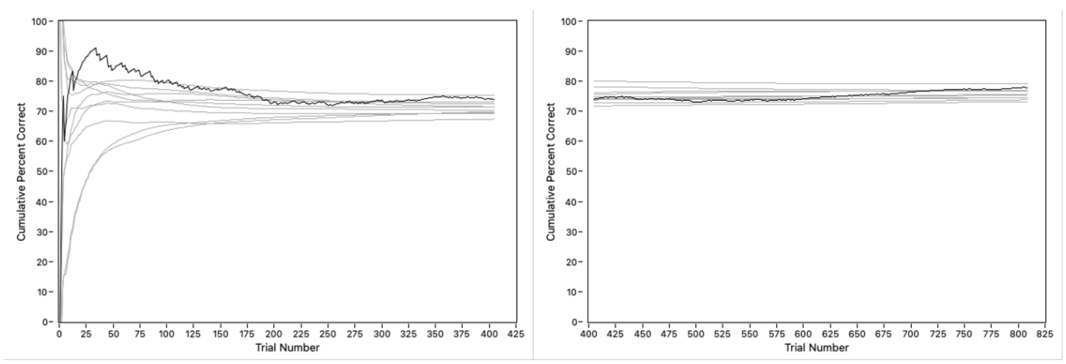

**Figure A.3.1: Experiment 1 – Categorization Strategy Classification for Individual Participants** For each of the 24 participants in Experiment 1, we compared windows of 135 trials (block sections) through the 810 trials total to 5 Feature-Focused Bayesian Observers to identify the observer that best simulated participant performance. Participants are listed according to the number of times that we recorded a change in strategy. We identified 5 participants as maintaining the same categorization strategy throughout all block sections, 7 participants as making 1 strategy change, 5 participants as making 2 strategy changes, 5 participants as making 3 strategy changes, 1 participant as making 4 strategy changes, and 1 participant as making 5 strategy changes.



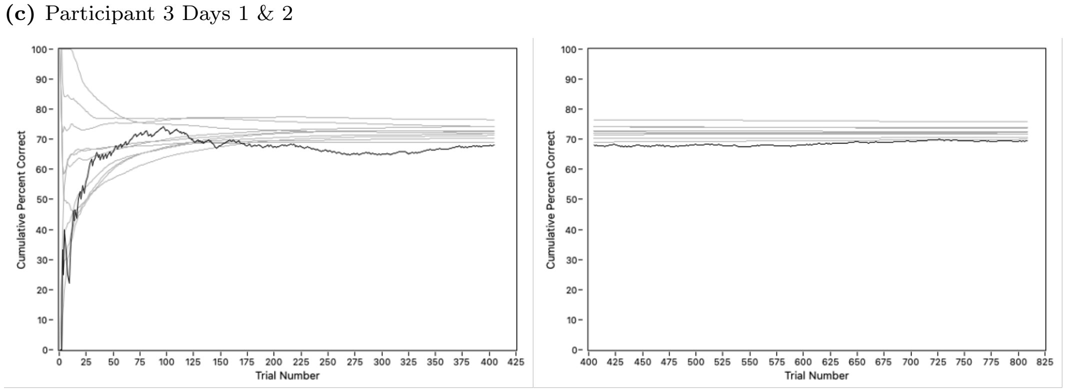

**(a)** Participant 1 Days 1 & 2



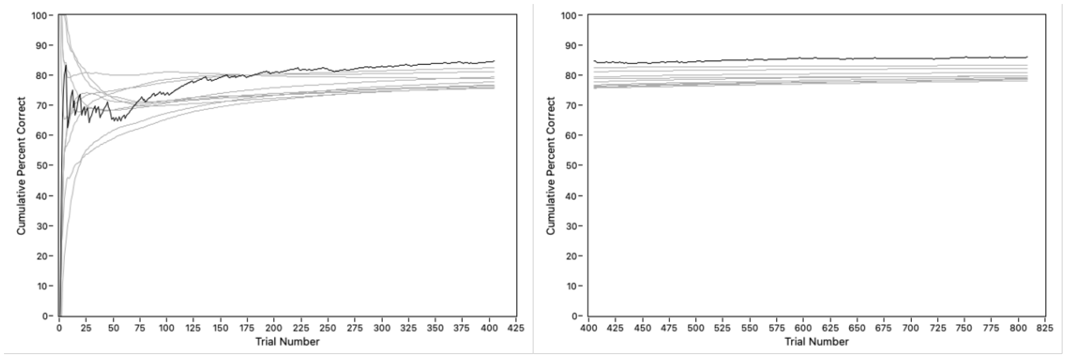

**(b)** Participant 2 Days 1 & 2



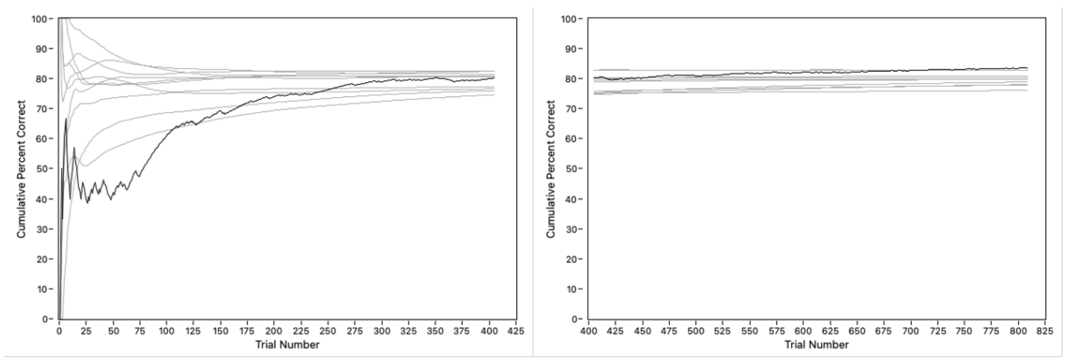

**(c)** Participant 3 Days 1 & 2



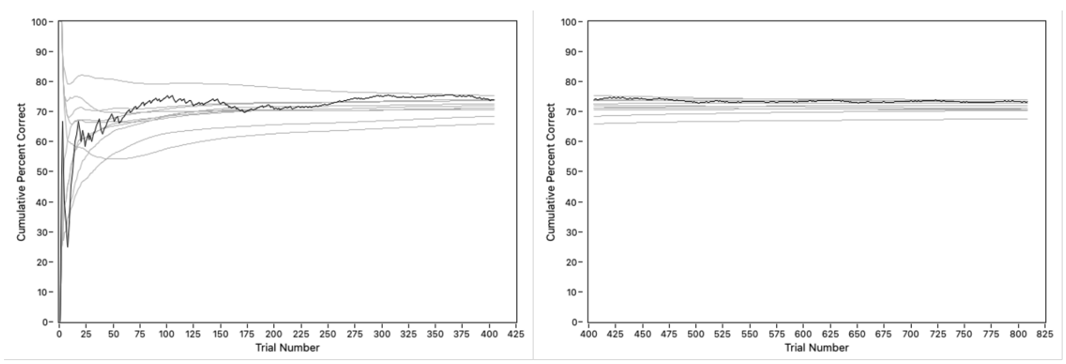
**(d)** Participant 4 Days 1 & 2



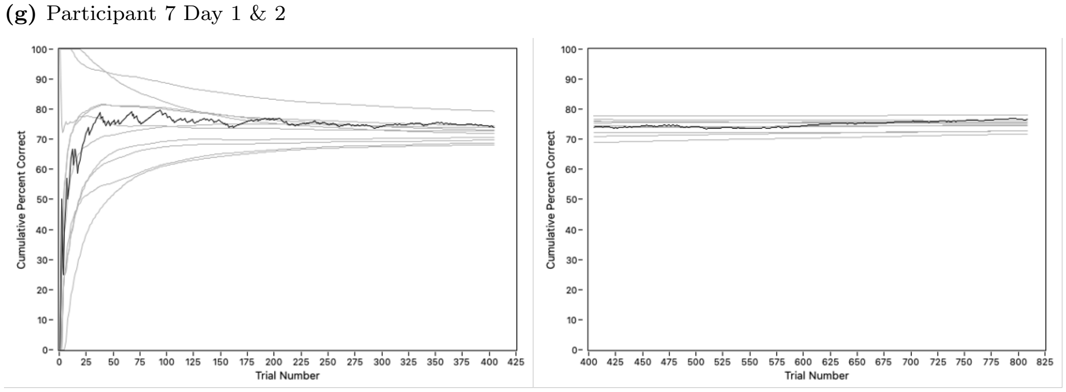

**(e)** Participant 5 Days 1 & 2



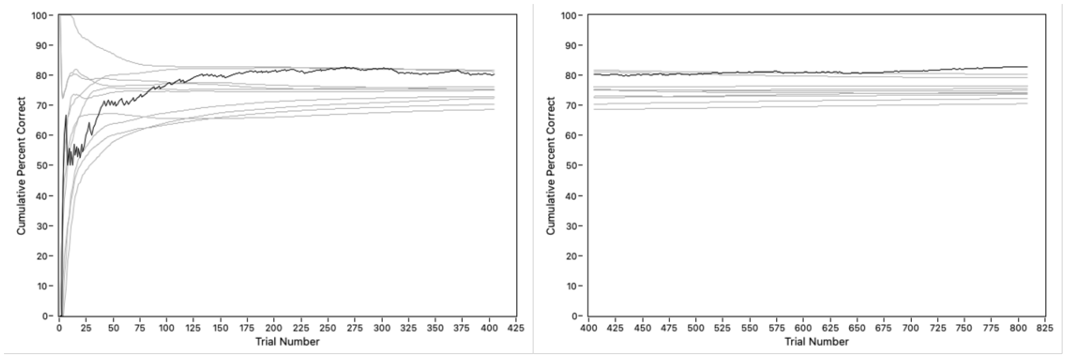

**(f)** Participant 6 Days 1 & 2



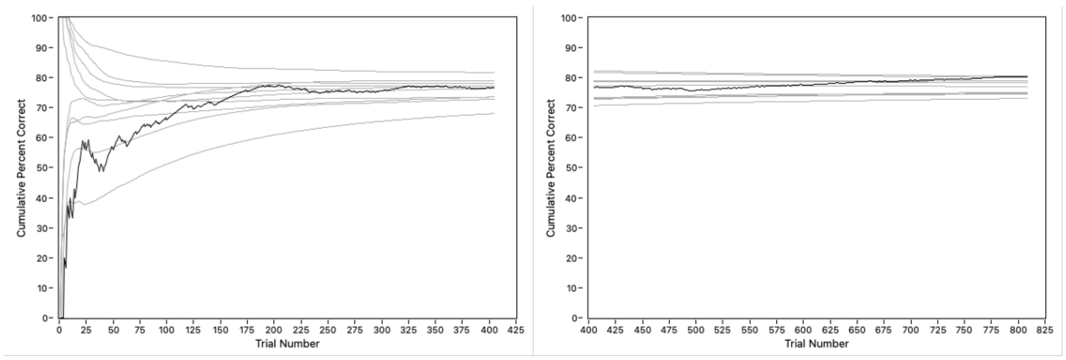

**(g)** Participant 7 Day 1 & 2



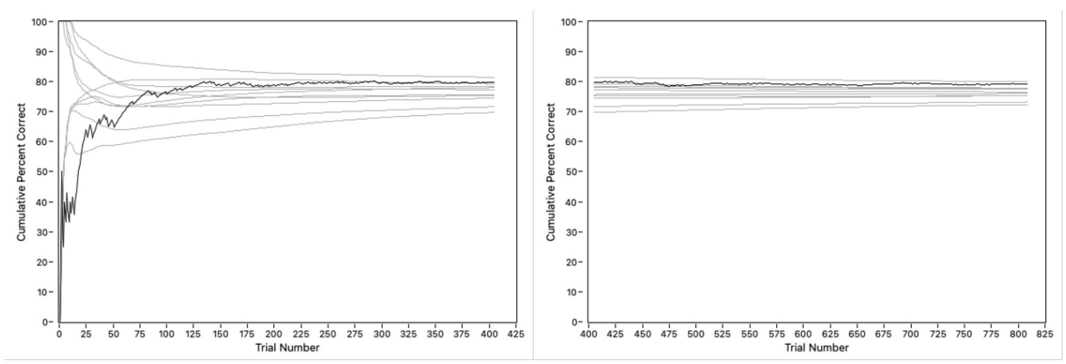
**(h)** Participant 8 Day 1 & 2



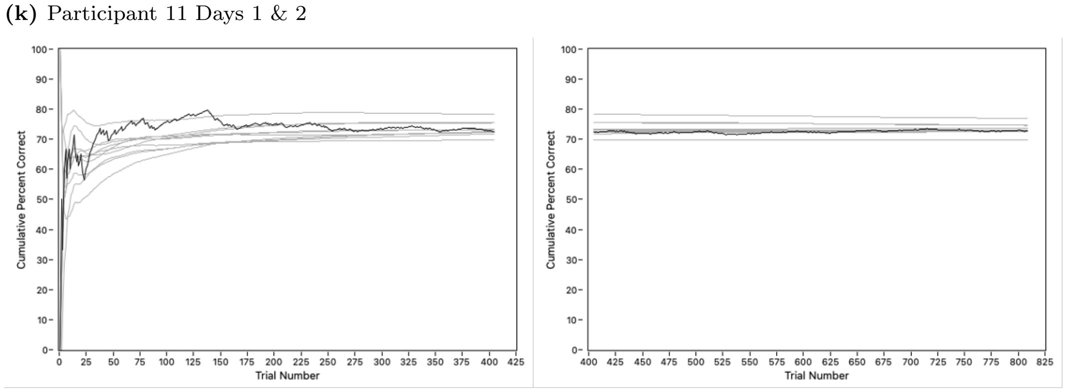

**(i)** Participant 9 Day 1 & 2



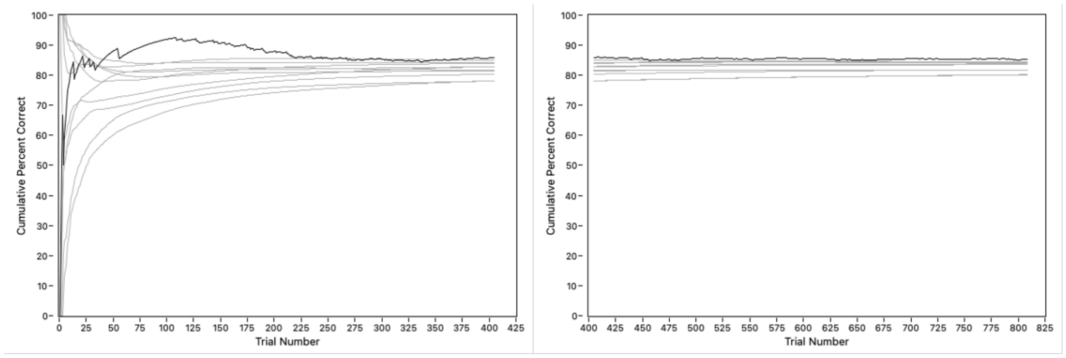

**(j)** Participant 10 Days 1 & 2



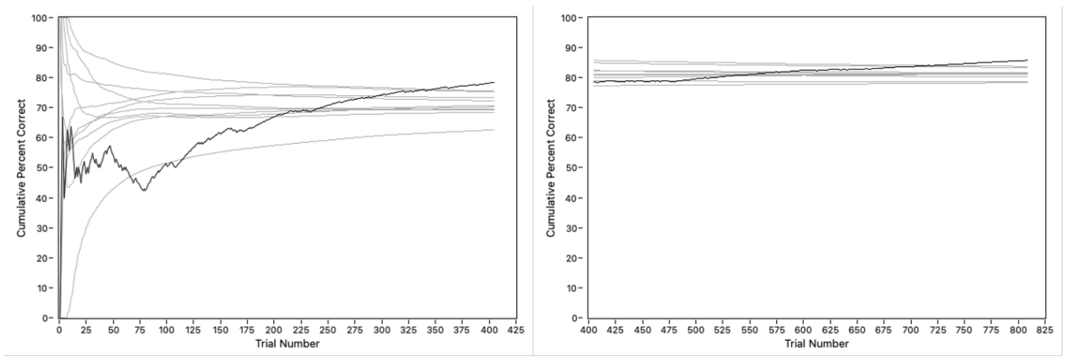

**(k)** Participant 11 Days 1 & 2



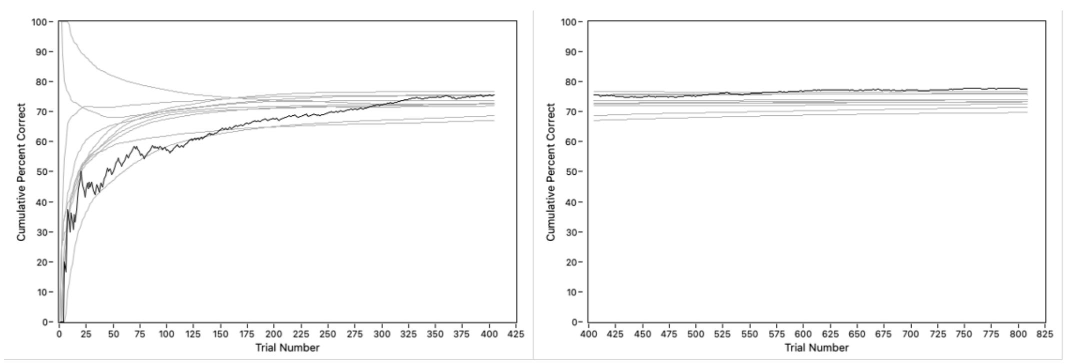
**(l)** Participant 12 Days 1 & 2



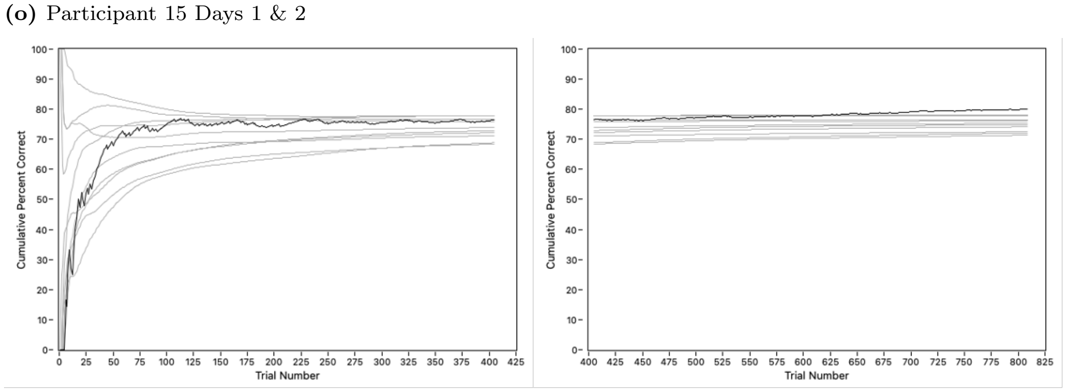

**(m)** Participant 13 Days 1 & 2



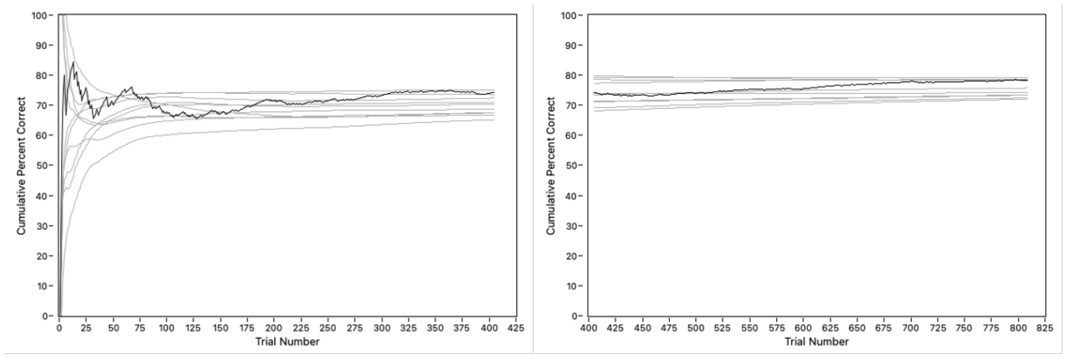

**(n)** Participant 14 Days 1 & 2



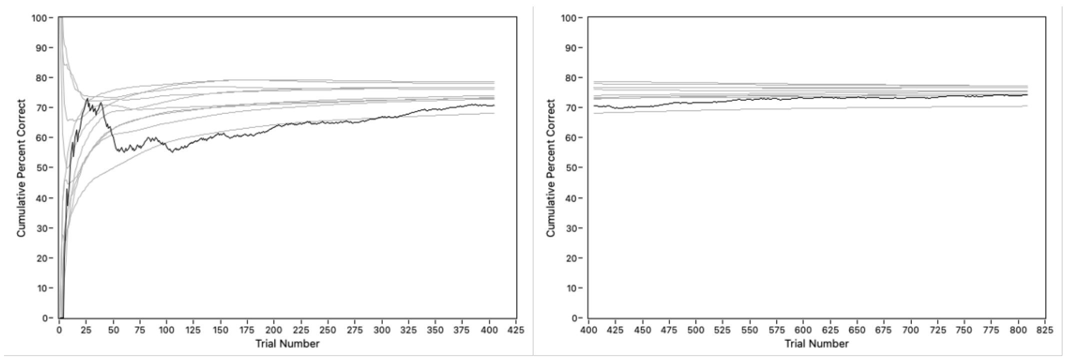

**(o)** Participant 15 Days 1 & 2



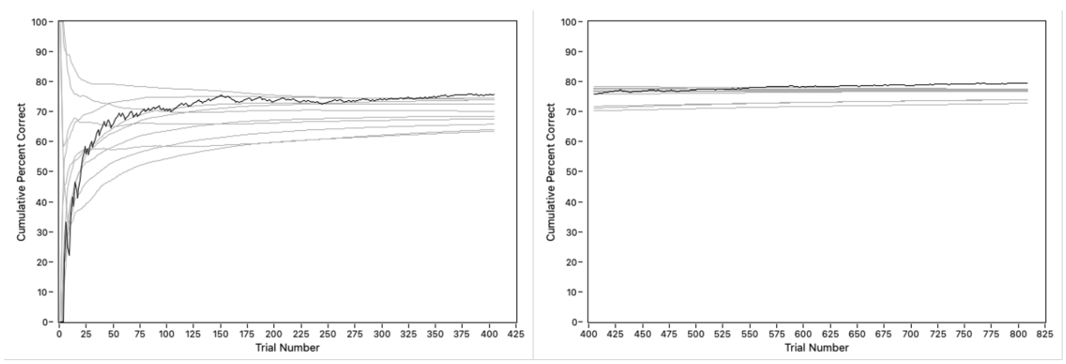
**(p)** Participant 16 Days 1 & 2



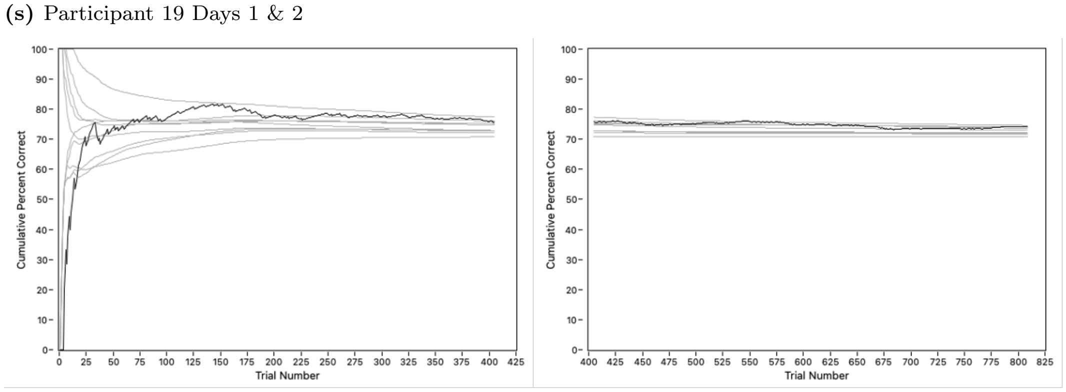

**(q)** Participant 17 Days 1 & 2



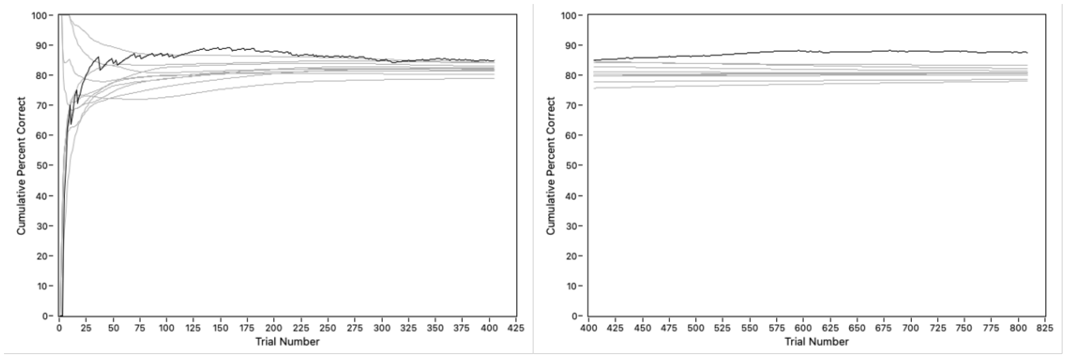

**(r)** Participant 18 Days 1 & 2



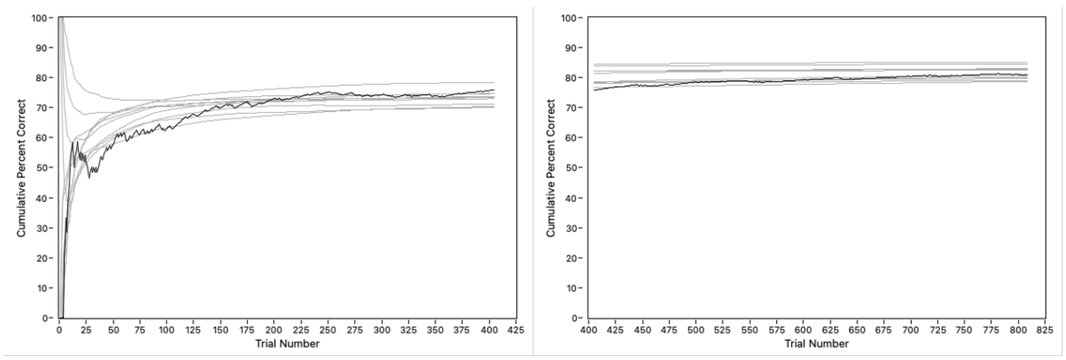

**(s)** Participant 19 Days 1 & 2



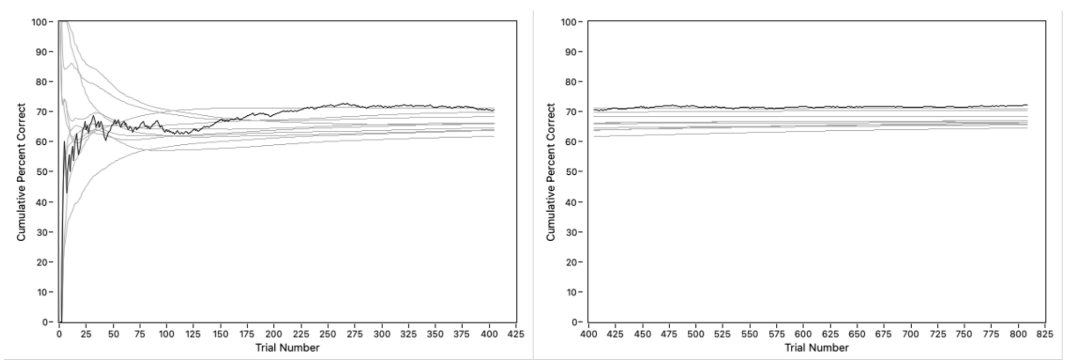
**(t)** Participant 20 Days 1 & 2



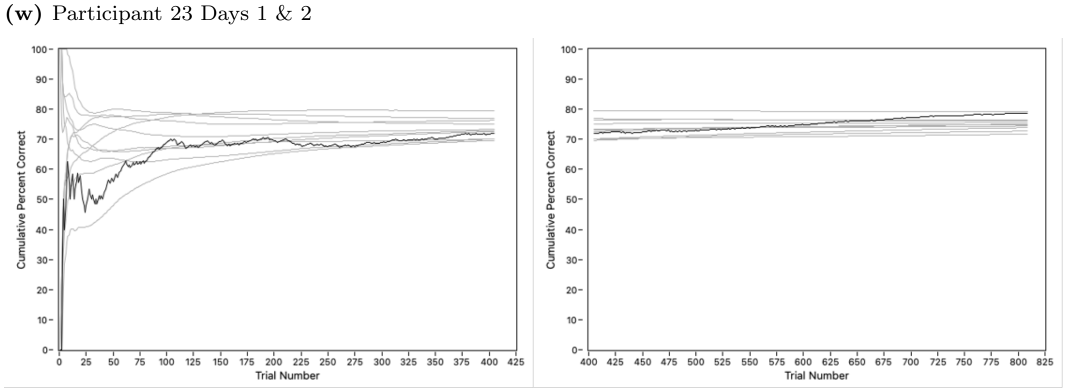

**(u)** Participant 21 Days 1 & 2




**(v)** Participant 22 Days 1 & 2




**(w)** Participant 23 Days 1 & 2



**(x)** Participant 24 Days 1 & 2

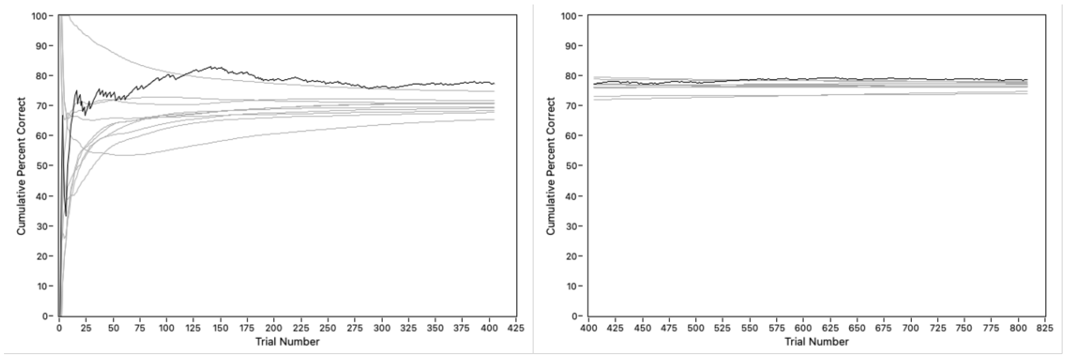



**Figure A.3.2: Experiment 1 – Participant Performance Compared to Winning Model Repetitions** We plot the learning curve for each of the 24 participants tested in Experiment 1 on Day 1 (left) and Day 2 (right). On each day, we classified participant categorization strategies using our model comparison analysis, which compared human performance to that of computational observers. Each computational observer was run on 10 simulated repetitions of the experiment. In grey, we plot each repetition of the most probable observer for each participant on Days 1 and 2. Figures a to x show results for participants 1 through 24.

**Table A.2: Experiment 1 – Performance Differences Between Days** Summary of the difference in PC of participants between the end of Day 1 and the end of Day 2. We classified the included participants as individuals who maintained the same categorization strategy between study days. On average, PC increased by 1.17%.

| Participant Number | $PC_{Day\ 2} - PC_{Day\ 1}$ |
|---|---|
| 1 | -3.61 |
| 3 | 1.32 |
| 4 | 1.34 |
| 5 | 3.19 |
| 6 | -0.77 |
| 8 | 2.70 |
| 10 | -0.40 |
| 11 | 0.21 |
| 12 | -0.64 |
| 14 | 0.34 |
| 15 | 3.56 |
| 17 | 3.67 |
| 19 | -1.64 |
| 22 | 1.69 |
| 23 | 6.52 |
| Average | 1.17 |

## A.4 Chapter 4: Supplemental Information

**Table A.3: Experiment 1 – Performance Differences Between Days** Summary of the difference in percent correct scores of participants between the end of Day 1 and the end of Day 2. We classified the included participants as individuals who change categorization strategies between study days. On average, PC increased by 3.79%.

| **Participant Number** | $\boldsymbol{PC_{Day\ 2} - PC_{Day\ 1}}$ |
|---|---|
| 2 | 3.80 |
| 7 | 2.56 |
| 9 | 3.56 |
| 13 | 7.27 |
| 16 | 4.17 |
| 18 | 3.68 |
| 20 | 2.70 |
| 21 | 5.04 |
| 24 | 1.33 |
| Average | 3.79 |

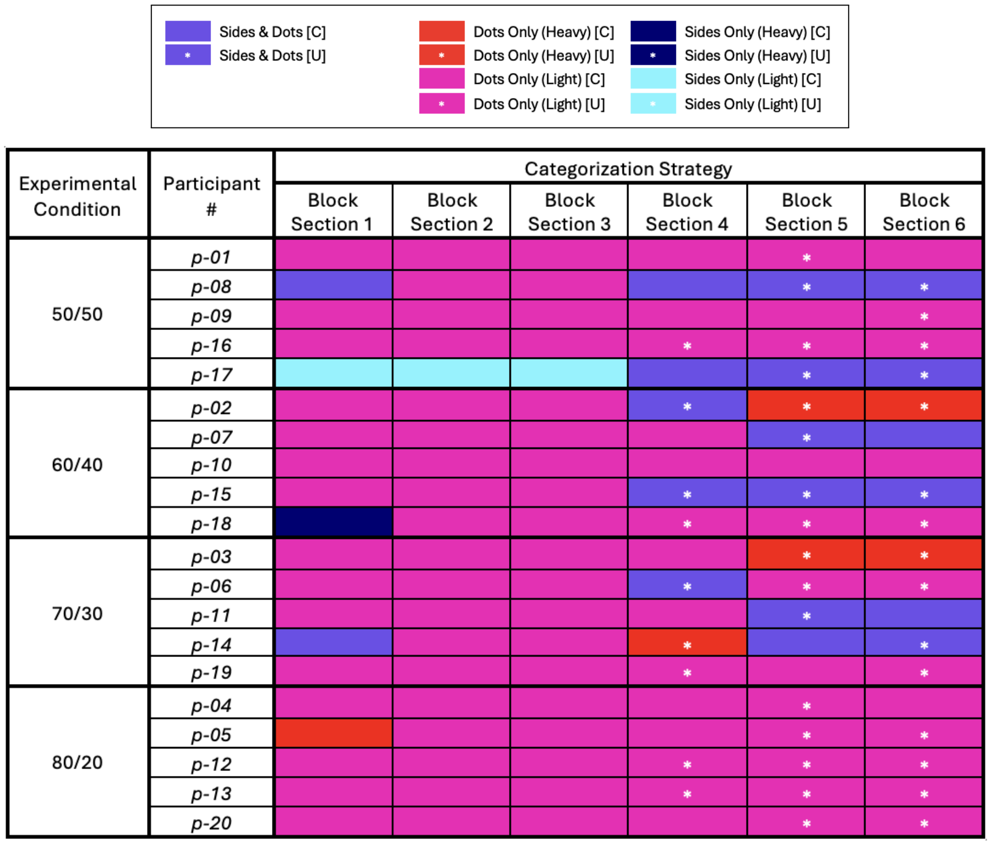
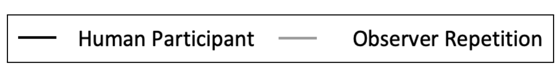
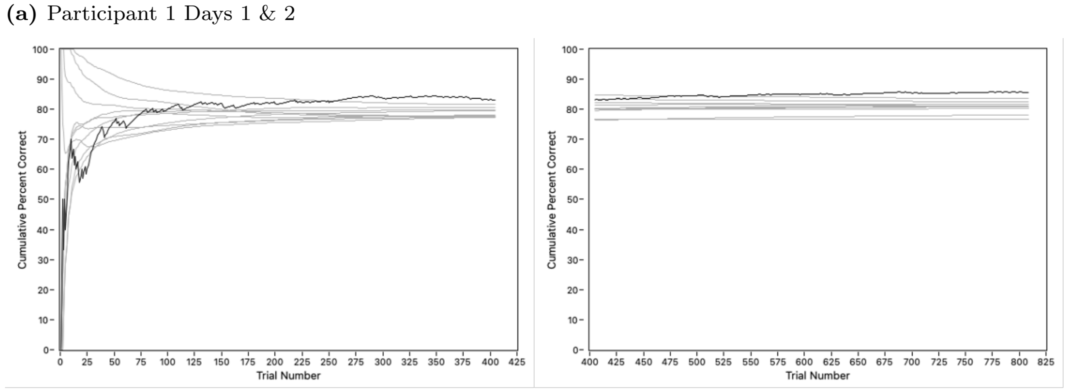


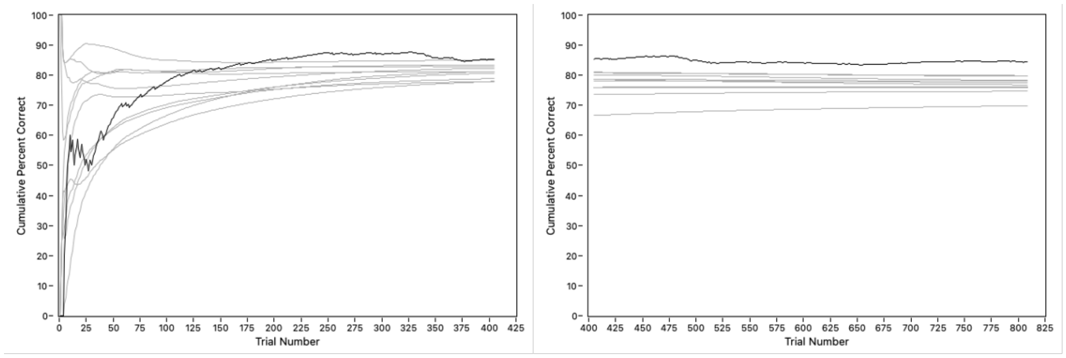

**Figure A.4.1: Experiment 2 – Categorization Strategy Classification for Individual Participants** For each of the 20 participants in Experiment 2, we compared windows of 135 trials (block sections) through the 810 trials total to 10 Feature-Focused Bayesian Observers to identify the observer that best simulated participant performance. A white * indicates that the best matched observer estimated the prevalence of Category A and B (updating category prior observers, or [U]), rather than assuming they appeared in equal proportions (constant category prior observers, or [C]). Participants are listed according to the experimental condition in which they were tested. An updating category prior observer best simulated participant performance 40 of 60 total block sections on Day 2.



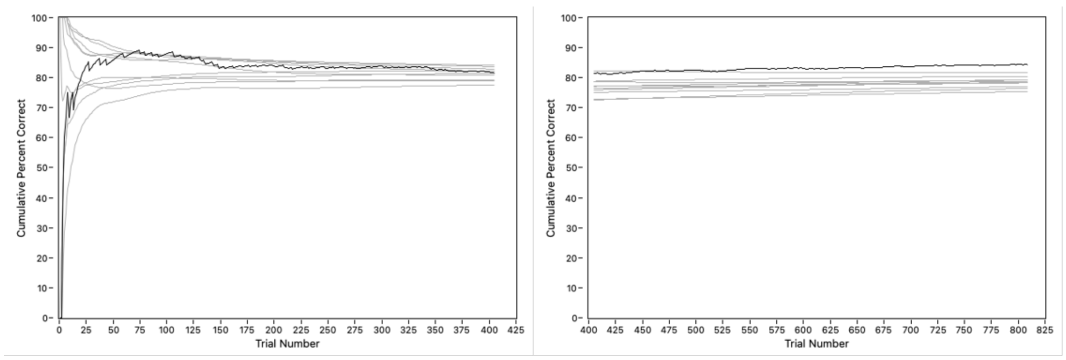

**(a)** Participant 1 Days 1 & 2



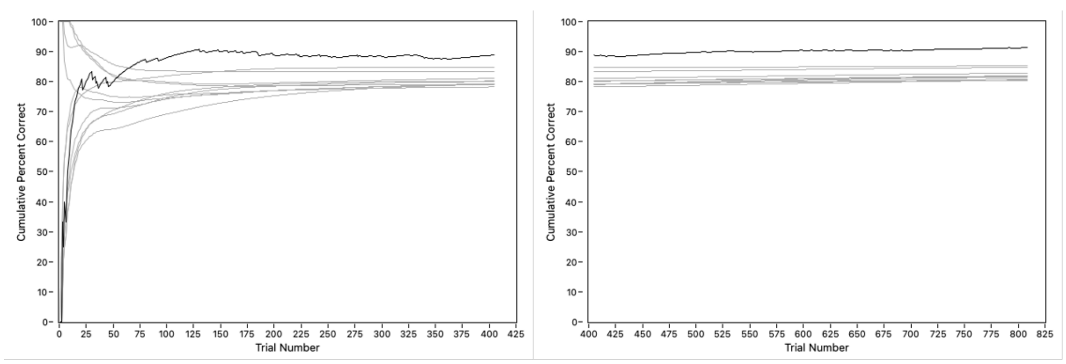
**(b)** Participant 2 Days 1 & 2



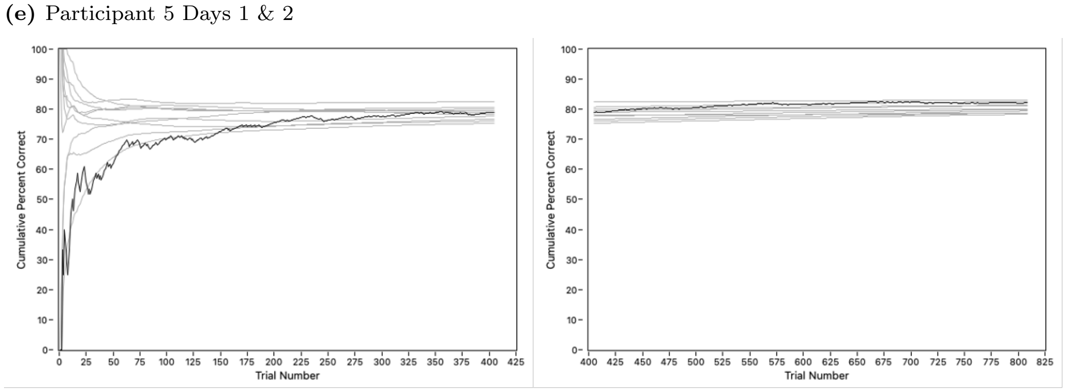

**(c)** Participant 3 Days 1 & 2



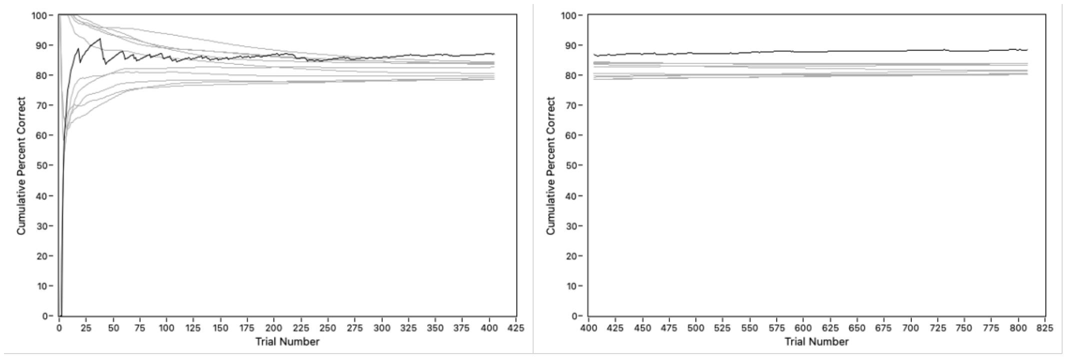

**(d)** Participant 4 Days 1 & 2



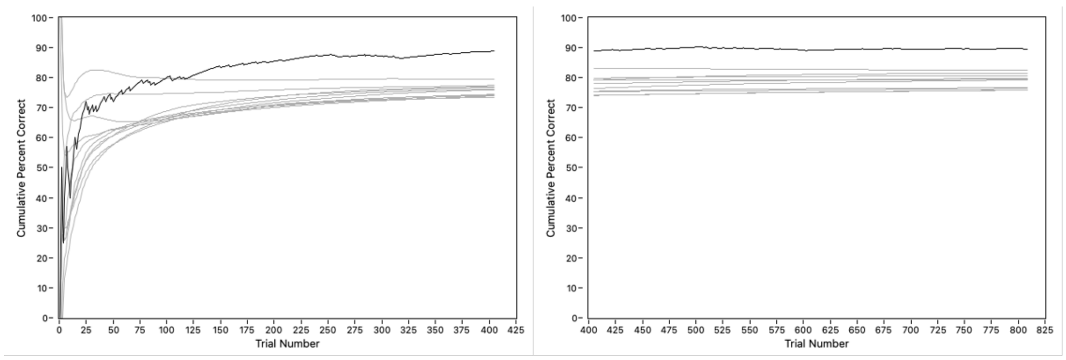

**(e)** Participant 5 Days 1 & 2



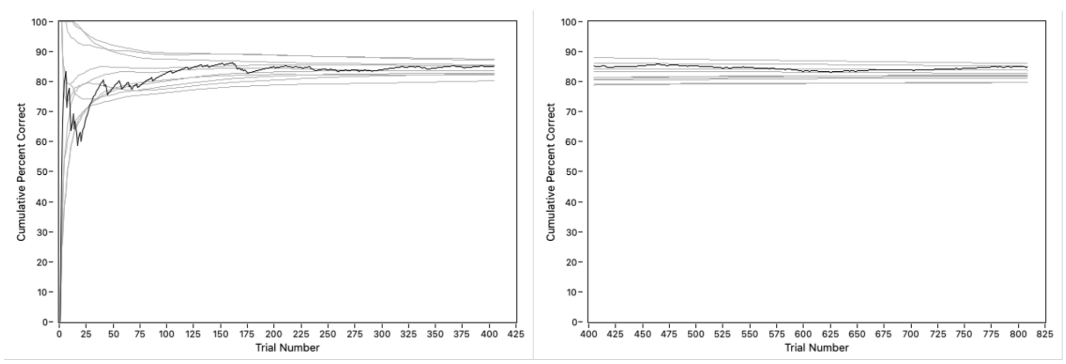
**(f)** Participant 6 Days 1 & 2



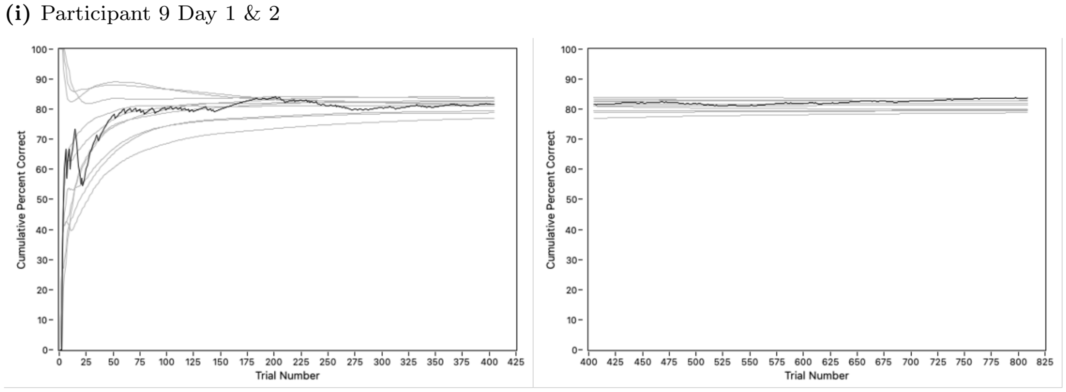

**(g)** Participant 7 Day 1 & 2



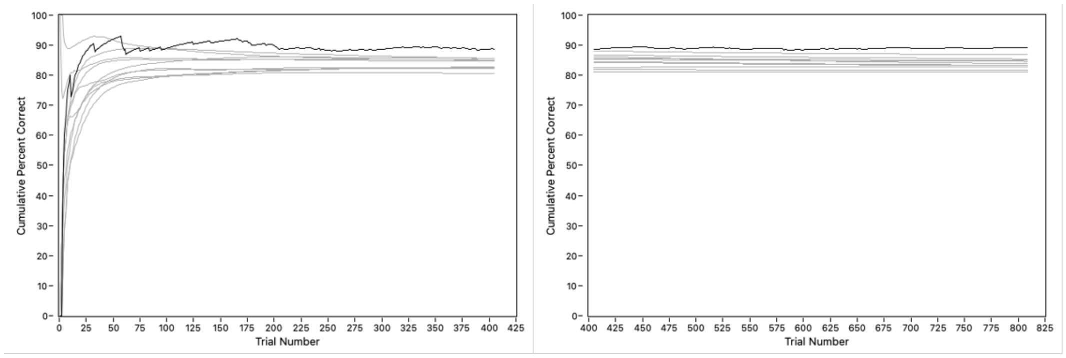

**(h)** Participant 8 Day 1 & 2



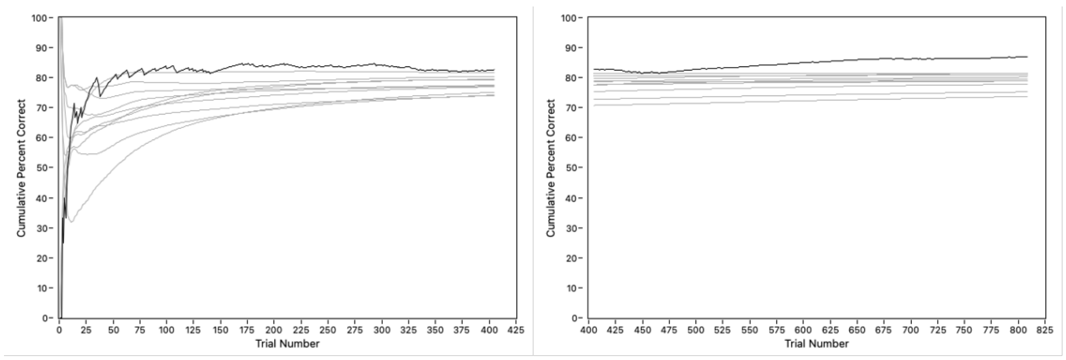

**(i)** Participant 9 Day 1 & 2



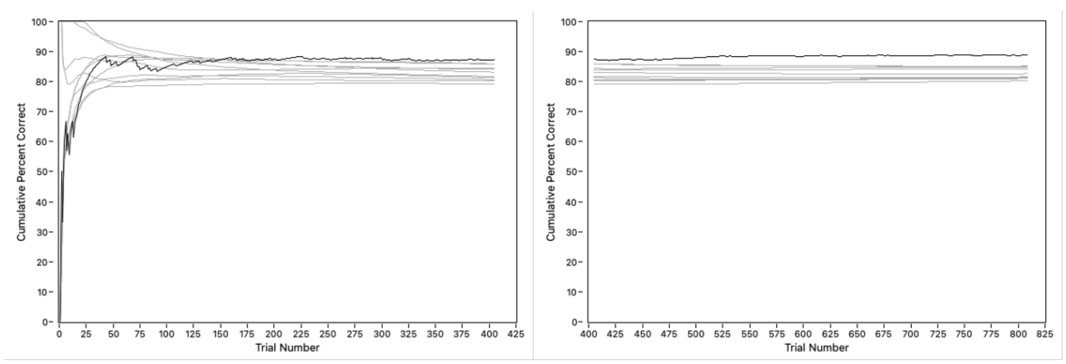
**(j)** Participant 10 Days 1 & 2



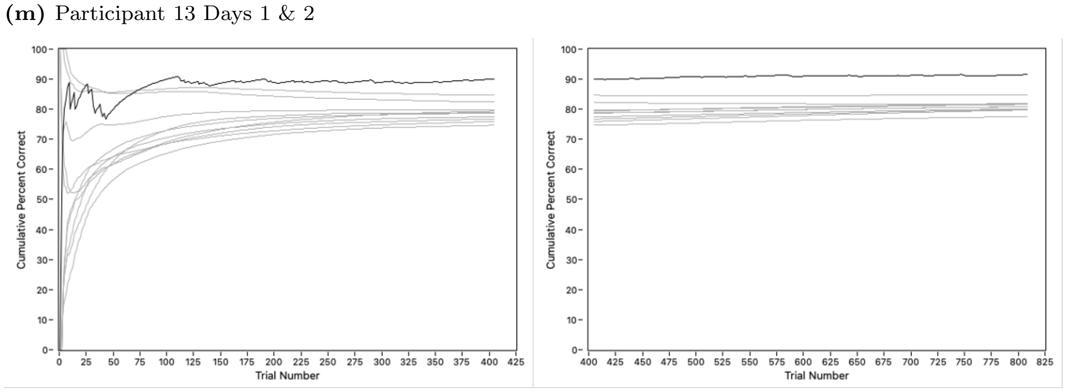

**(k)** Participant 11 Days 1 & 2



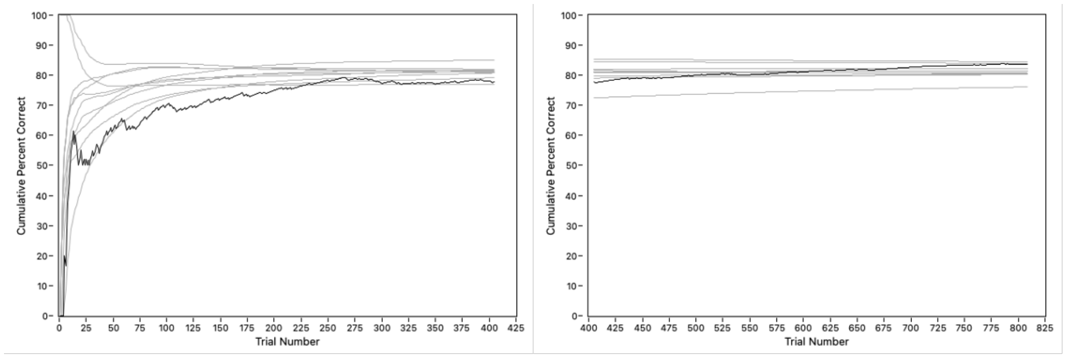

**(l)** Participant 12 Days 1 & 2



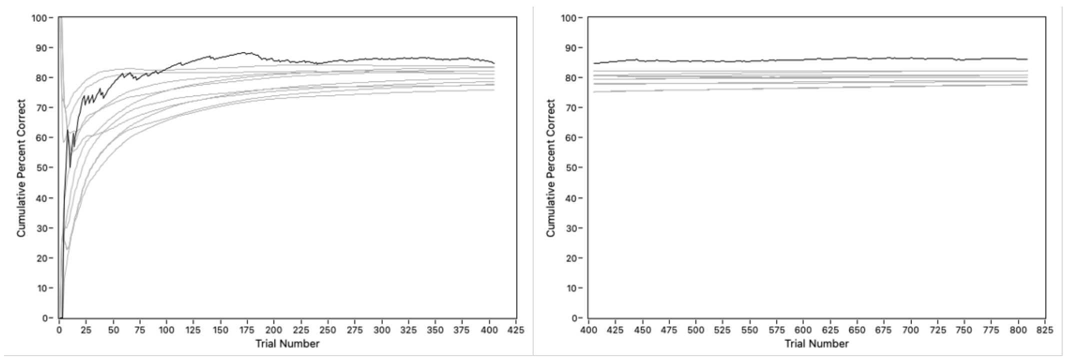

**(m)** Participant 13 Days 1 & 2



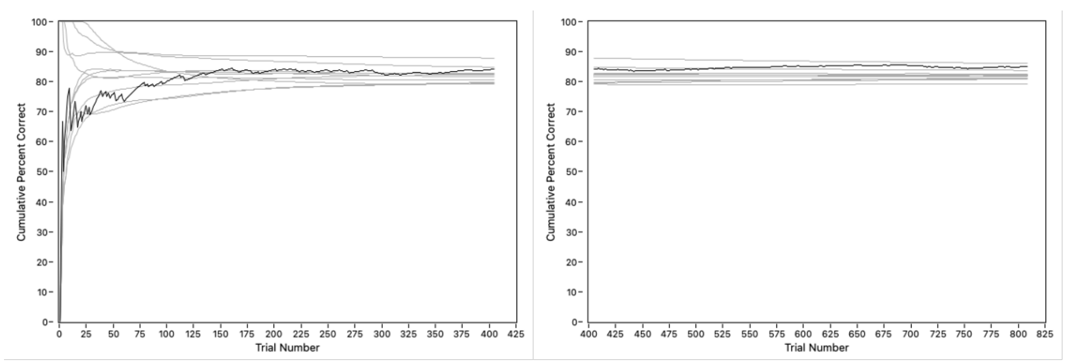
**(n)** Participant 14 Days 1 & 2



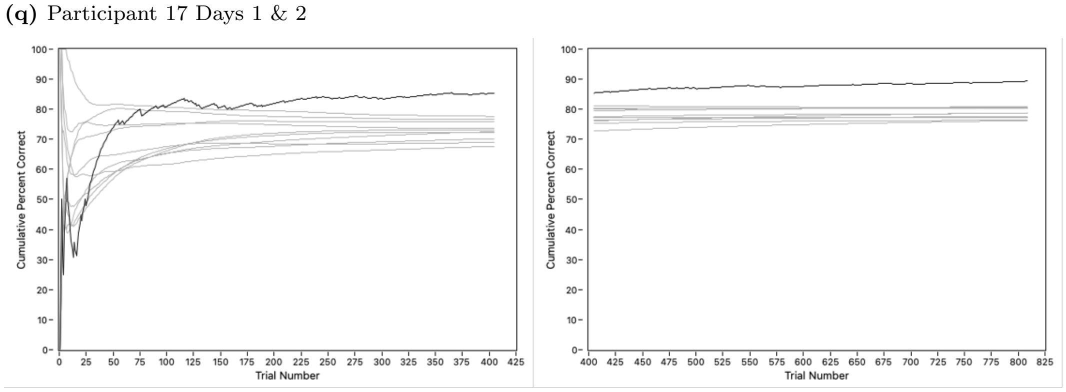

**(o)** Participant 15 Days 1 & 2



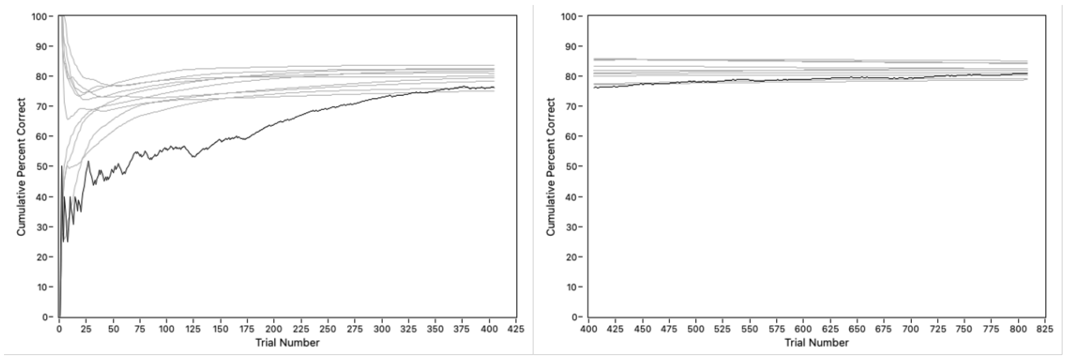

**(p)** Participant 16 Days 1 & 2



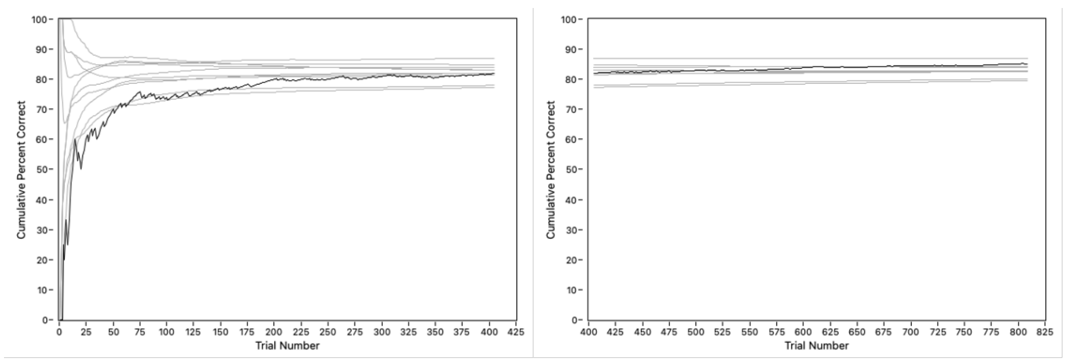

**(q)** Participant 17 Days 1 & 2



**(r)** Participant 18 Days 1 & 2




**(s)** Participant 19 Days 1 & 2




**(t)** Participant 20 Days 1 & 2

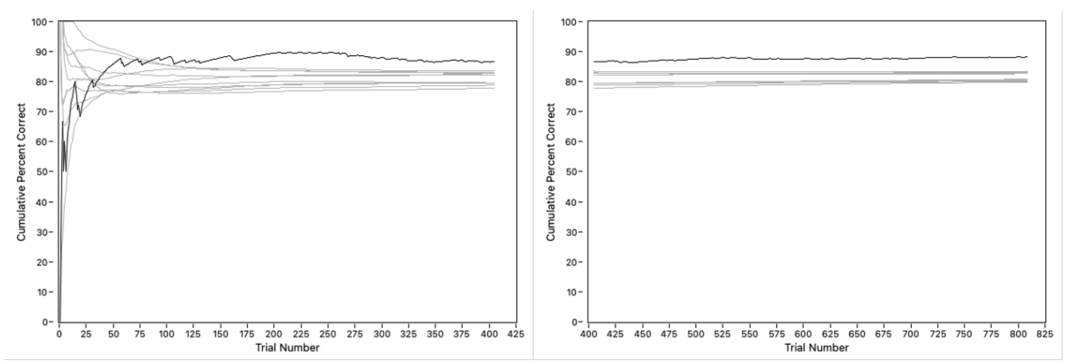



**Figure A.4.2: Experiment 2 – Participant Performance Compared to Winning Model Repetitions** We plot the learning curve for each of the 20 participants tested in Experiment 2 on Day 1 (left) and Day 2 (right). On each day, we classified participant categorization strategies using our model comparison analysis, which compared human performance to that of computational observers. Each computational observer was run on 10 simulated repetitions of the experiment. In grey, we plot each repetition of the most probable observer for each participant on Days 1 and 2. Figures a to t show results for participants 1 through 20.

**Table A.4: Experiment 2 – Performance Differences Between Days** Summary of the difference in percent correct scores of participants between the end of Day 1 and the end of Day 2. We classified the included participants as individuals who change categorization strategies between study days. On average, PC increased by 5.57%.

| Participant Number | $PC_{Day\ 2} - PC_{Day\ 1}$ |
|---|---|
| 2 | 0.49 |
| 3 | 4.20 |
| 7 | -1.98 |
| 8 | 8.89 |
| 11 | 8.40 |
| 12 | 2.96 |
| 13 | 2.96 |
| 14 | 11.60 |
| 15 | 2.96 |
| 17 | 7.65 |
| 18 | 13.09 |
| Average | 5.57 |

**Table A.5: Experiment 2 – Performance Differences Between Days** Summary of the difference in PC of participants between the end of Day 1 and the end of Day 2. We classified the included participants as individuals who maintained the same categorization strategy between study days. On average, PC increased by 3.32%.

| Participant Number | $PC_{Day\ 2} - PC_{Day\ 1}$ |
|---|---|
| 1 | 4.69 |
| 4 | -0.74 |
| 5 | 4.44 |
| 6 | 2.96 |
| 9 | 3.70 |
| 10 | 0.99 |
| 16 | 1.73 |
| 19 | 5.68 |
| 20 | 6.42 |
| Average | 3.32 |

# Bibliography

Alais, D. and Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3):257–262.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3):409–429. Num Pages: 409-429 Place: Washington, US Publisher: American Psychological Association (US).

Angeletos Chrysaitis, N. and Seriès, P. (2023). 10 years of Bayesian theories of autism: A comprehensive review. *Neuroscience & Biobehavioral Reviews*, 145:105022.

Arnold, D. H., Petrie, K., Murray, C., and Johnston, A. (2019). Suboptimal human multisensory cue combination. *Scientific Reports*, 9(1):5155.

Ashby, F. G. and Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experi-*

*mental Psychology: Learning, Memory, and Cognition*, 14(1):33–53. Num Pages: 33-53 Place: Washington, US Publisher: American Psychological Association (US).

Bankieris, K. R., Bejjanki, V. R., and Aslin, R. N. (2017). Sensory cue-combination in the context of newly learned categories. *Scientific Reports*, 7(1):10890. Publisher: Nature Publishing Group UK London.

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

Bejjanki, V. R., Clayards, M., Knill, D. C., and Aslin, R. N. (2011). Cue Integration in Categorical Tasks: Insights from Audio-Visual Speech Perception. *PLOS ONE*, 6(5):e19812. Publisher: Public Library of Science.

Berniker, M., Voss, M., and Kording, K. (2010). Learning Priors for Bayesian Computations in the Nervous System. *PLOS ONE*, 5(9):e12686. Publisher: Public Library of Science.

Ekman, G. (1959). Weber's Law and Related Functions. *The Journal of Psychology*. Publisher: Taylor & Francis Group.

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

Gauder, K. (2024). *Exploration of a Bayesian probabilistic model for categorization in the sense of touch*. PhD thesis, McMaster University.

Gepshtein, S. and Banks, M. S. (2003). Viewing Geometry Determines How Vision and Haptics Combine in Size Perception. *Current Biology*, 13(6):483–488.

Gifford, A. M., Cohen, Y. E., and Stocker, A. A. (2014). Characterizing the Impact of Category Uncertainty on Human Auditory Categorization Behavior. *PLOS Computational Biology*, 10(7):e1003715. Publisher: Public Library of Science.

Granato, G., Cartoni, E., Rold, F. D., Mattera, A., and Baldassarre, G. (2022). Integrating unsupervised and reinforcement learning in human categorical perception: A computational model. *PLOS ONE*, 17(5):e0267838. Publisher: Public Library of Science.

Gredin, N. V., Bishop, D. T., Williams, A. M., and Broadbent, D. P. (2021). Integrating explicit contextual priors and kinematic information during anticipation. *Journal of*

*Sports Sciences*, 39(7):783–791. Publisher: Routledge _eprint: https://doi.org/10.1080/02640414.2020.1845494.

Guo, K., Nevado, A., Robertson, R. G., Pulgarin, M., Thiele, A., and Young, M. P. (2004). Effects on orientation perception of manipulating the spatio–temporal prior probability of stimuli. *Vision Research*, 44(20):2349–2358.

Hansen, K., Hillenbrand, S., and Ungerleider, L. (2012). Effects of Prior Knowledge on Decisions Made Under Perceptual vs. Categorical Uncertainty. *Frontiers in Neuroscience*, 6. Publisher: Frontiers.

Hennevin, E., Huetz, C., and Edeline, J.-M. (2007). Neural representations during sleep: From sensory processing to memory traces. *Neurobiology of Learning and Memory*, 87(3):416–440.

Khalid, S., Goldenberg, M., Grantcharov, T., Taati, B., and Rudzicz, F. (2020). Evaluation of Deep Learning Models for Identifying Surgical Actions and Measuring Performance. *JAMA Network Open*, 3(3):e201664.

Klatzky, R. L., Lederman, S. J., and Metzger, V. A. (1985). Identifying objects by touch: An "expert system". *Perception & Psychophysics*, 37(4):299–302.

Knill, D. C. and Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43(24):2539–2558.

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

Lebaz, S., Jouffrais, C., and Picard, D. (2012). Haptic identification of raised-line drawings: high visuospatial imagers outperform low visuospatial imagers. *Psychological Research*, 76(5):667–675.

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

Lederman, S. J. and Klatzky, R. L. (1990). Haptic classification of common objects: Knowledge-driven exploration. *Cognitive Psychology*, 22(4):421–459.

Lederman, S. J. and Klatzky, R. L. (1993). Extracting object properties through haptic exploration. *Acta Psychologica*, 84(1):29–40.

Lederman, S. J., Klatzky, R. L., Chataway, C., and Summers, C. D. (1990). Visual mediation and the haptic recognition of two-dimensional pictures of common objects. *Perception & Psychophysics*, 47(1):54–64.

Li, Y., Seger, C., Chen, Q., and Mo, L. (2020). Left Inferior Frontal Gyrus Integrates Multisensory Information in Category Learning. *Cerebral Cortex*, 30(8):4410–4423. Publisher: Oxford University Press.

Lim, S.-J., Fiez, J. A., and Holt, L. L. (2019). Role of the striatum in incidental learning of sound categories. *Proceedings of the National Academy of Sciences*, 116(10):4671–4680. Publisher: Proceedings of the National Academy of Sciences.

Luongo, F., Hakim, R., Nguyen, J. H., Anandkumar, A., and Hung, A. J. (2021). Deep learning-based computer vision to recognize and classify suturing gestures in robot-assisted surgery. *Surgery*, 169(5):1240–1244.

Massaro, D. W. and Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9(5):753–771. Num Pages: 753-771 Place: Washington, US Publisher: American Psychological Association (US).

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

Miyazaki, M., Nozaki, D., and Nakajima, Y. (2005). Testing Bayesian Models of Human Coincidence Timing. *Journal of Neurophysiology*, 94(1):395–399. Publisher: American Physiological Society.

Nagai, Y., Suzuki, M., Miyazaki, M., and Kitazawa, S. (2012). Acquisition of Multiple Prior Distributions in Tactile Temporal Order Judgment. *Frontiers in Psychology*, 3.

Petrini, K., Remark, A., Smith, L., and Nardini, M. (2014). When vision is not an option: children's integration of auditory and haptic information is suboptimal. *Developmental Science*, 17(3):376–387. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/desc.12127.

Petzschner, F., Maier, P., and Glasauer, S. (2012). Combining Symbolic Cues with Sensory Input and Prior Experience in an Iterative Bayesian Framework. *Frontiers in Integrative Neuroscience*, 6.

Rohe, T. and Noppeney, U. (2015). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision*, 15(5):22.

Rosch, E. and Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4):573–605.

Sapey-Triomphe, L.-A., Sanchez, G., Hénaff, M.-A., Sonié, S., Schmitz, C., and Mattout, J. (2023). Disentangling sensory precision and prior expectation of change in autism during tactile discrimination. *npj Science of Learning*, 8(1):1–16. Publisher: Nature Publishing Group.

Schwarzer, G., KÜfer, I., and Wilkening, F. (1999). Learning categories by touch: On the development of holistic and analytic processing. *Memory & Cognition*, 27(5):868–877.

Solak, G. and Jamone, L. (2023). Haptic Exploration of Unknown Objects for Robust in-hand Manipulation. *IEEE Transactions on Haptics*, PP(99):1–12. Publisher: IEEE.

Tassinari, H., Hudson, T. E., and Landy, M. S. (2006). Combining Priors and Noisy Visual Cues in a Rapid Pointing Task. *Journal of Neuroscience*, 26(40):10154–10163. Publisher: Society for Neuroscience Section: Articles.

Withagen, A., Kappers, A. M. L., Vervloed, M. P. J., Knoors, H., and Verhoeven, L. (2013). The use of exploratory procedures by blind and sighted adults and children. *Attention, Perception, & Psychophysics*, 75(7):1451–1464.

Yuan, W., Zhu, C., Owens, A., Srinivasan, M. A., and Adelson, E. H. (2017). Shape-independent hardness estimation using deep learning and a GelSight tactile sensor. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 951–958.