An investigation of within-dimension stimuli in categorization and change detection

Anthea G. Blunden

https://orcid.org/0000-0001-8650-7010

Doctor of Philosophy

Melbourne School of Psychological Sciences

The University of Melbourne

In total fulfilment of the requirements of the Doctor of Philosophy

February 2020

Abstract

This thesis examined the way in which information is combined in order to make decisions. It focused specifically on within-dimension decision-making with spatially separate stimuli. The primary aim was to characterize the underlying organization of processing over time (i.e., identify the processing architecture; whether information processing proceeds in serial, parallel, or is pooled into a single decision-making channel), stopping rules (i.e., whether the process is exhaustive or self-terminating), and the efficiency of processing (i.e., workload capacity). These attributes were assessed through the analysis of response times (RTs) using Systems Factorial Technology (SFT), the Logical-Rules paradigm, and relevant computational modeling. In the first part of the thesis, results from the categorization experiments showed that for the majority of participants, processing occurs coactively (i.e., is pooled into a single decision process). Workload capacity, however, was shown to be limited. This suggests that a violation of context invariance may have occurred and several theories are considered as potential explanations for this finding.

In the second part of the thesis, a novel modeling framework for characterizing the time course of change detection based on information held in visual short term memory was presented. Specifically, we sought to answer whether change detection is better captured by a first-order integration model, in which information is pooled from each location, or a second-order integration model, in which each location is processed independently. We conducted two experiments with both disjunctive OR rules and conjunctive AND rules (across locations) using a double

factorial paradigm and a redundant target paradigm. These experiments showed that although capacity is generally limited in both tasks, architecture varies from parallel self-terminating in the OR task to serial self-terminating in the AND task. This novel framework allowed for model comparisons across a large set of models, ruling out several competing explanations of change detection.

As a whole, this thesis found both differences and similarities in decision-making using within-dimension stimuli across categorization and change detection tasks. The finding of differences in decision-making strategy across categorization and change detection tasks highlights that different perceptual operations can yield a variety of experimental results. It may be expected that other tasks such as visual search, identification, and detection might also diverge. A finding of limited capacity across both task types, however, points to a potential common bottle neck in processing efficiency at an earlier encoding stage.

Declaration

This is to certify that:

1. This thesis comprises only my work towards the Doctor of Philosophy, except where indicated in the preface;

2. Due acknowledgement has been made in text to all other material used; and

3. This thesis is fewer than 100,000 words, exclusive of tables, figures, footnotes, references, and appendices.

Anthea G. Blunden

Preface

Percentage contribution and publication status:

Blunden, A. G. & Howe, P. D. L., & Little, D. R. (2019). Evidence that within-dimension features are generally processed coactively. *Attention Perception & Psychophysics. https://doi.org/10.3758/s13414-019-01775-8.*

This paper was published by Attention, Perception, & Psychophysics on the 28th June, 2019.

Anthea Blunden: 80% contribution

Piers Howe: 5% contribution

Daniel Little: 15% contribution

Blunden, A. G. & Hammond, D. & Howe, P. D. L., & Little, D. R. (2019). Characterizing the time course of decision-making in change detection. *Manuscript under review: Psychological Review*

This paper is in revision following peer review by Psychological Review.

Anthea Blunden: 70% contribution

Dylan Hammond: 10% contribution

Piers Howe: 5% contribution

Daniel Little: 15% contribution

This work was supported by an Australian Government Research Training Program Scholarship awarded to Anthea Blunden and by an ARC grant (DP160102360) awarded to A/Prof Daniel R. Little.

Acknowledgments

This thesis was written on Wurundjeri land and I pay my respects to Elders past, present, and emerging.

I would like to express my deepest gratitude for my supervisors A/Prof Daniel Little and A/Prof Piers Howe. I thank Dan for his knowledge, passion, and unwavering guidance and support throughout the PhD. The considerable joy of cognitive science was an unexpected gift. I thank Piers for his thoughtful and supportive mentoring, not only with regards to my PhD, but also academic life more generally, and for always making me laugh, even on darker days. I cannot thank them both enough for their guidance, insight, and dedication. I would also like to thank my PhD advisory committee: Prof Philip Smith, A/Prof Meredith McKague, and Dr David Sewell for their ongoing encouragement, and Dr Robert De Lisle and Danièle Martinie, foremost for their friendship, but also for their assistance with some aspects of the programming and data collection for Chapter 4. I am also very thankful to all of the participants who kindly volunteered their time, and who spent many hours pushing buttons in the dark with good humour.

I would like to thank the members of the Knowlab: Dr Geoff Saw, Dr Xian Liew, Dr Maggie Webb, Jun Cheng, Sarah Moneer, Deb Lin, Tam Dennis, Amanda Shanks, Nicole Christie, Dave Griffiths, Ariel Goh, and Andrew Wang, and the members of the Visual Cognition Lab: Weijia Chen, Jess Marris, Campbell Pryor, Marcellin Martinie, Larson Landes, and Kathryn Hull. I would also like to thank my (many!) students for their perpetual enthusiasm and

curiosity; from this I draw much inspiration and happiness. In particular, I would like to thank my honours student, Lauren Fong, who has continued this scholarly journey with considerable flair, and who has, with great skill, humoured me in addressing some of the remaining questions posed by this thesis.

Thank you to my line manager, Dr Judi Humberstone, who has been an impressive mentor and advocate, and from whom I gather significant inspiration and strength. Thank you also to my friends and colleagues from the Melbourne School of Psychological Sciences and wider psychological sciences community, in particular, Dr Vanja Rozenblat, Damien Crone, Christina Van Heer, Dr Simon Lilburn, Hayley Jach, Blake Cavve, Caitlyn Gourlay, Michael Susman, Dr Benita Green, and the members of the Learning and Teaching Strategy sub-committee, whose presence have made this journey a pleasure.

Finally, I would like to extend my deepest thanks to my friends and family, without the support of whom this thesis would simply not exist. To my partner Dylan, thank you for lending a small part of your considerable ingenuity to some of this project. I am so grateful for your time and collaboration, and I thank you from the bottom of my heart for being a constant source of inspiration, love, and solace.

Contents

List of Tables

List of Figures

Chapter 1

Overview of Current Thesis

Decisions often require the combination of information from multiple sources. For example, when choosing a snack to take on a hike, one may consider a variety of factors; taste, nutritional value, portability etc. Decisions involving primarily visual stimuli typically also require the combination of information from multiple sources. For example, deciding whether an object lying across your path is a snake or a branch may involve assessments on a number of visual dimensions such as color, shape, texture, and movement. Many perceptual decisions not only require the integration of visual information from different sources but also across time points. For example, motorists may need to monitor changes in traffic conditions, doctors may need to decide whether a rash has worsened or improved, and so on. These kinds of decisions necessitate holding information in visual short term memory (VSTM) in order to make a relevant comparison. This kind of information integration is particularly important for a number of reasons. Not only does it allow us to make decisions about changes in the visual world, but it also allows for the comparison of information which cannot be simultaneously fixated. Even more fundamentally, our visual system must be able to form a correspondence between pre- and post-saccadic input in order to maintain a stable representation of the visual scene. The ability to integrate visual information is therefore crucial both to forming a basic visual

representation and to making higher level decisions about visual stimuli.

This thesis explores the way in which visual information is combined in order to make decisions. I focus specifically on decision-making utilizing features that comprise dimensions which are separated in space but are composed of differing levels of the *same* feature type, henceforth referred to as *within-dimension* features. I consider decision-making using within-dimension features across two cognitive tasks; in categorization, and in a one-shot change detection task.

The primary aim of this thesis is to characterize the underlying organization of processing over time (i.e., identify the processing architecture; whether information processing proceeds in serial, parallel, or is pooled into a single decision-making channel; Kantowitz, 1974; Sternberg, 1969; Schweickert, 1993; Townsend, 1984), stopping rules (i.e., whether the process is exhaustive or self-terminating), and the efficiency of processing (i.e. workload capacity; Townsend & Ashby, 1983; Wenger & Townsend, 2000). These attributes are assessed through the analysis of response times (RTs) using Systems Factorial Technology (SFT; Townsend & Nozawa, 1995; Little, Altieri, et al., 2017), the Logical-Rules paradigm (Little, Altieri, et al., 2017; Fifić et al., 2010) and relevant computational modeling (Fifić et al., 2010). While these methods are well established in the categorization literature (Blunden et al., 2015; Cheng et al., 2017; Fifić et al., 2010; Little et al., 2013, 2011; Moneer et al., 2016), analysis of the way in which information is integrated in visual short term memory (VSTM) in order to make a decision has been largely absent from the literature (although see Wilken & Ma, 2004). This thesis therefore presents an extension of the Logical-Rules models (Fifić et al., 2010) in order to characterize decision making based on information held in VSTM. This approach provides novel insight, not only as to how information is integrated in order to make a decision, but also how this process unfolds over time. This extension further provides a way to unify models of change detection with models of categorization decision making.

The present thesis comprises seven experiments across the two broader themes of categorization and change-detection. In Chapters 3 and 4 I explore categoriza-

tion decision-making utilizing within-dimension stimuli. Specifically, in Chapter 3, I present two categorization experiments characterizing the architecture of decision-making using luminance discs of opposite polarity separated in space. In the first experiment, the stimuli are separated by a boundary of luminance discs which is subsequently removed in the second experiment. In Chapter 4, I characterize the processing capacity of these stimuli using a redundant targets design. In Chapter 5, I present an extension of the Logical-Rules paradigm in order to diagnose processing in a one-shot multi-element change detection task. This chapter provides a theoretical overview of the novel extension of the Logical-Rules paradigm into VSTM, followed by four experiments. The first two experiments diagnose the processing architecture of change detection decision-making for two discs of opposite luminance polarity. In the first experiment, participants are required to use an "OR" decision rule (i.e., they are required to indicate whether *any* item has changed). In the second experiment, participants are required to use an "AND" decision rule (i.e. they must indicate whether *both* items have changed). The second two experiments also use AND and OR decision rules, but focus on processing capacity. Chapter 6 provides a summary and conclusion of the current work.

In the following chapter I provide a general literature review of categorization and change detection of visual information, focusing particularly on characterizing the decision-making process.

Chapter 2

General Literature Review

The way in which individuals form a mental representation of visual stimuli, and the processes that utilize this information in order to make categorization decisions, form a fundamental part of our everyday functioning. For example, categorization reduces the complexity of the environment, it allows for the generalisation of past learning when encountered with a novel stimulus or situation, and it assists in making decisions.

Historically, when categorization is studied in the laboratory, participants are presented with a set of training items which vary on multiple dimensions and participants are required to extrapolate from this training in order to classify new stimuli. For example, in their seminal study, Posner & Keele (1968) presented participants with a training set of dot patterns and investigated how their classification of new dot patterns varied as a function of their distance from a prototype pattern. Several competing theories of categorization explain performance in this type of task. For example, in exemplar-based memory models, individuals store a mental representation of items most typical to a given category ("exemplars") and classify new stimuli based on their similarity to these exemplars (Medin & Schaffer, 1978; Nosofsky, 1986; Estes, 1986). In prototype models, individuals form an abstract representation of each category, classifying new stimuli as belonging to the category with the most similar prototype (Reed,

1972). Finally, in decision-bound theory, items are classified based on their location in perceptual space. Individuals partition the perceptual space using a decision-boundary and learn to associate certain areas of perceptual space with certain responses (Ashby, 1992; Ashby et al., 1994; Maddox, 1992; Maddox & Ashby, 1996).

At the end of the 20th century, models of categorization decision-making using multidimensional stimuli provided detailed and complex quantitative accounts for accuracy in a wide range of categorization tasks (Anderson, 1991; Ashby, 1992; Estes, 1986; Kruschke, 1992; Nosofsky, 1992); however, a major limitation of these models was that they were unable to account for the time course of processing in categorization. The Exemplar-Based Random Walk model (EBRW; Nosofsky & Palmeri, 1997) was formulated to address this limitation. The EBRW combines the Generalized Context Model (GCM; Nosofsky, 1986), with the Instance-Based model of Automaticity (Logan, 1988) and the Random Walk model (Laming, 1968; Link & Heath, 1975; Luce, 1986; Ratcliff, 1978; Townsend & Ashby, 1983; Thornton & Gilden, 2007). In the EBRW model, a target stimulus is compared to a set of multidimensional exemplars previously stored in psychological space, along with their category label. A similarity value based on the distance in the psychological space between target and exemplars is computed for each category. This similarity value is calculated by taking the summed similarity across all items within a category. The summed similarity value is then converted to a sampling probability by dividing the total similarity score with the total summed similarity of all exemplars across categories which drives a sequential-sampling (random-walk) process in which information is sampled until sufficient evidence (as determined by a decision-criterion) has been sampled in order to make a decision.

The stochastic GRT model (Ashby, 2000), is another example of a model which assumes that a pooling of information drives RT. Like the decision-bound models described earlier, stochastic GRT assumes that decisions are made by establishing decision boundaries in perceptual space. In this model, each

stimulus is represented by a distribution of multidimensional perceptual effects. Individuals establish one or more decision boundaries, separating the perceptual space into one or more category regions. At each step of the process a sample is taken from the perceptual distribution. If the sample lies in a particular category region, the random-walk process will take a step towards that category. Samples are accumulated until a particular decision criterion is reached.

While the pooling of information is one way in which a categorization decision may unfold over time, there are other alternate possibilities which warrant consideration. For example, dimensions may be processed one at a time in serial, or simultaneously but in separate decision-making channels (i.e., in parallel). Fifić et al. (2010) introduced the Logical-Rules models to provide scope for exploring these other accounts of processing architecture. This research combines the non-parametric Systems Factorial Technology (SFT; Townsend & Nozawa, 1995; Little, Altieri, et al., 2017) with parametric modeling providing powerful scope for diagnosing processing architecture in categorization tasks. The Logical-Rules models form the basis for analysis in this thesis and will be considered in detail in Section 2.2.

In VSTM decision-making, and in change detection research in particular, similar limitations as those seen in categorization exist. While the change detection literature has made significant progress in a number of domains (for example, characterizing of capacity limits in the visual short-term memory processes that underlie change detection; Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974, differentiating the use of discrete information from the use of continuous resources; Wilken & Ma, 2004; Bays & Husain, 2008; Donkin et al., 2016, and developing mixture models which encompass elements of guessing and information-based responding; Zhang & Luck, 2008; Cowan & Rouder, 2009), few accounts of the way in which information is combined in order to make a decision exist. While there are some notable exceptions (see e.g., Wilken & Ma, 2004), these models typically do not account for information processing over time, or alternatively, only investigate

single items in a single decision-making channel (Donkin et al., 2013; Lilburn, 2016; P. L. Smith, 2016). Providing a novel account of change detection decision-making which diagnoses the architecture of information integration across the entire time course of processing is one of the primary goals in the current thesis.

This chapter will provide an overview of some key themes and empirical research relevant to categorization and change detection decision-making. I first focus on the first major theme of the thesis: categorization decision-making, reviewing SFT and the Logical-Rules models and their contribution, not only to characterizing decision-making using a wide range of multidimensional stimuli, but also to answering more fundamental questions regarding perception, attention, and decision-making (Griffiths et al., 2017; Little et al., 2011; Fifić et al., 2010; Little et al., 2013; Moneer et al., 2016). I then consider the question of decision-making using within-dimension stimuli, a question not yet considered by the categorization literature, but one which has formed the basis of investigations in other areas such as visual attention (Duncan & Humphreys, 1989; Wolfe et al., 1990; Mordkoff & Yantis, 1993) and in VSTM research (Huang et al., 2007; Mance et al., 2012; Sewell et al., 2014). I then turn to the second major theme of the thesis, namely change detection decision-making. Here I briefly review change detection research more generally, before turning to the more detailed questions of information integration in change detection, as well as the time course of information processing in change detection decision-making.

2.1 Systems Factorial Technology

The methods and statistical tools of SFT (Townsend & Nozawa, 1995; Little, Altieri, et al., 2017) were formulated to answer two of the most fundamental questions in cognitive science. Firstly, what is the *mental architecture* of processing within a system (see Figure 3.1)? One way that this question applies to categorization is to ask, when categorizing an object for example, is each element processed one at a time in a *serial* manner? Or are the elements consid-

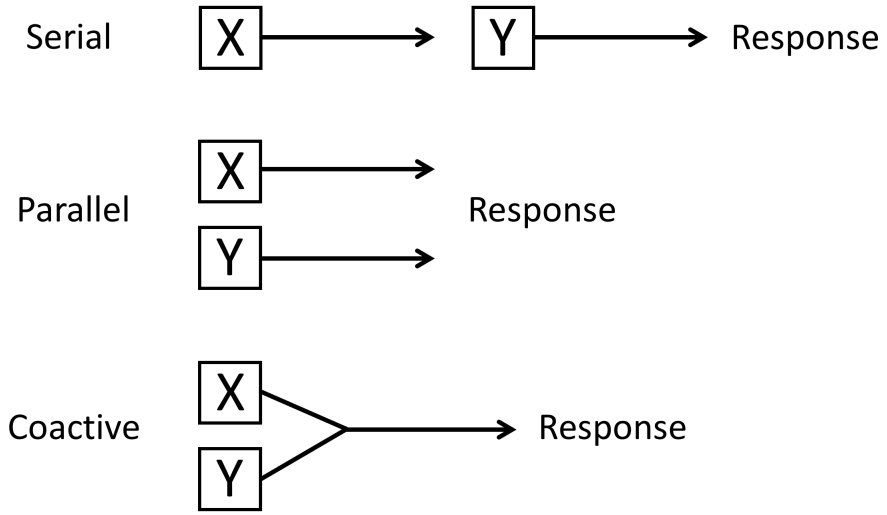

Figure 2.1: Schematic illustration of the three processing architectures: serial, parallel, and coactive.

ered simultaneously, but in separate decision-making channels (i.e. in *parallel*)? Alternatively, are the elements pooled together into a single decision making channel (i.e., processed *coactively*)? A related question is whether it is necessary to consider all of the available information before coming to a decision (so called *exhaustive processing*), or can a decision be reached without processing all of the information (*self-terminating processing*)? Finally, what is the *workload capacity* of a system? Does adding additional elements slow the processing of the system (i.e., is the system limited in capacity)? Or alternatively, does adding additional items speed up (i.e. super capacity) or not affect (i.e. unlimited capacity) total processing time? The following provides a detailed overview on how SFT can assist to answer these questions.

2.1.1 The double factorial paradigm

The double factorial paradigm (DFP; see Figure 2.2) builds on the classic redundant targets detection paradigm (see e.g. Egeth & Mordkoff, 1991; Snodgrass & Townsend, 1980) by combining two experimental manipulations factorially, one affecting workload and another affecting within-channel processing speed,

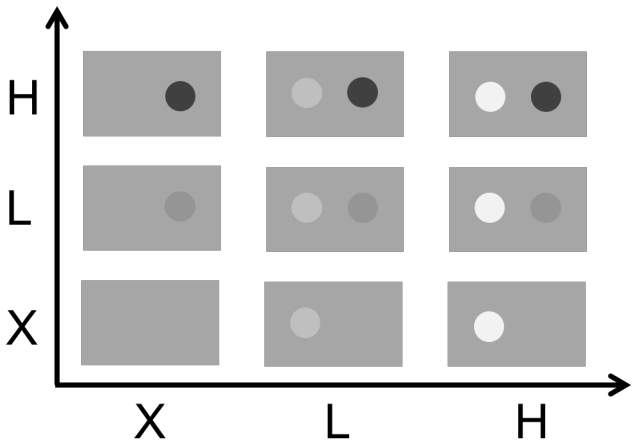

Figure 2.2: The double factorial paradigm design. H = high salience, L = low salience, X = no stimulus (single- and redundant-target items).

in order to make statistical inferences about processing architecture via the analysis of RT distributions. In this design, the participant is required to make speeded detection responses (usually via a button press). In a task employing an OR decision rule, participants must make a certain response when one or both targets are detected, and an alternate response if no targets are detected (i.e., a redundant non-target target trial). The OR decision rule is therefore disjunctive, in that the process can self-terminate once a single target is detected. Conversely, in a task employing an AND decision rule, participants are required to make a certain response only when both targets are detected, and an alternate response if one or no targets are detected. An AND decision rule requires exhaustive processing before a response can be made and is therefore a conjunctive decision rule.

Within-channel processing speed is systematically varied by factorially manipulating the salience of the items. For example, in Figure 2.2, salience is manipulated in terms of difference in luminance (lighter or darker) compared to the background. Workload is varied simply by the number of items present in the display (two, one, or zero items). This gives rise to a nine item stimulus space, with each stimulus having a unique identifier based on its salience and

workload. For instance, stimuli with a high salience item in both locations is referred to as *HH*, a stimulus with a high salience item in the first location and a low salience item in the second location is referred to as *HL*, the converse of this situation is referred to as *LH*, and finally the stimulus with a low salience item in both locations is referred to as *LL*. The single target trials combine low (L) or high (H) salience items with no item in the other location (X). Finally, the stimulus containing no items in either location (XX) is referred to as the *redundant* stimulus.

2.1.2 The mean interaction contrast

The mean interaction contrast (MIC) provides a useful summary of the RTs from the DFP. Using the double-target items, it is calculated by finding the difference between the low and high salience values on one dimension and then the difference between the low and high salience values on the other dimension:

$$MIC = (RT_{LL} - RT_{LH}) - (RT_{HL} - RT_{HH}) \tag{2.1}$$

An additive pattern of results ($MIC = 0$), is evidence for serial processing, regardless of stopping rule. An under-additive MIC ($MIC < 0$) is indicative of a parallel exhaustive model and an over-additive MIC ($MIC > 0$) is predicted by both a parallel self-terminating model and a coactive model.

2.1.3 The survivor interaction contrast

A more sensitive contrast than the MIC is the survivor interaction contrast (SIC). The SIC analyzes the functional form of the entire RT distribution and is calculated using the survivor function for each stimulus, at each time value, t:

$$SIC(t) = [S_{LL}(t) - S_{LH}(t)] - [S_{HL}(t) - S_{HH}(t)] \tag{2.2}$$

where the survivor function, $S(t)$, is the complement of the cumulative distribution function, $f(t)$, and represents the probability that a response has not been

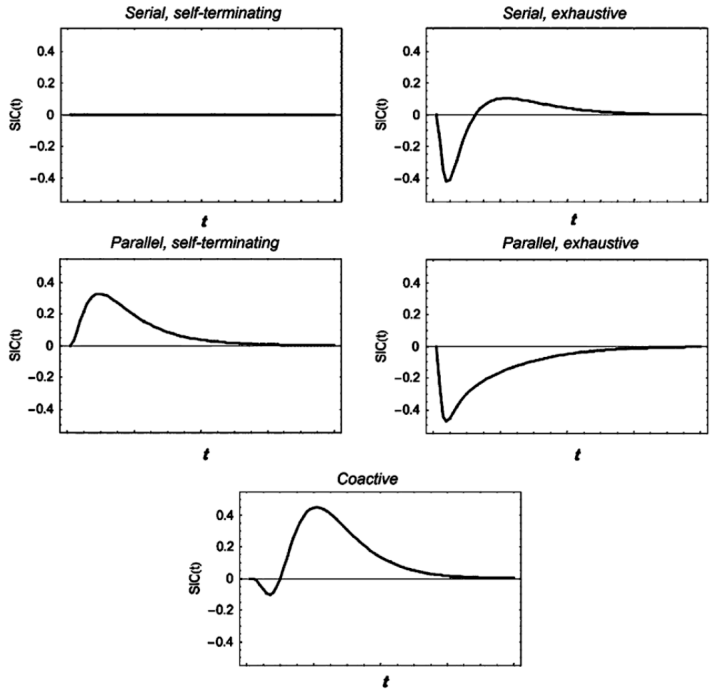

Figure 2.3: SIC predictions for the serial, parallel, and coactive models.

made by time, t.

The predictions for serial, parallel, and coactive models as well as exhaustive and self-terminating stopping rules are shown in Figure 5.4. While both serial exhaustive and self-terminating models predict $MIC = 0$, the SIC functions are distinctly shaped. The serial self-terminating function is flat for all times t, whereas the serial exhaustive model predicts an S-shaped curve with an early negative region and later positive region which integrates to zero (i.e. the MIC).

The intuitive reasoning for these forms is as follows: For a serial self-terminating model, only one item need be processed before a decision can be made. This means that only one dimension contributes to the overall RT (e.g., LL becomes L, LH becomes L, HL becomes H and HH becomes H). When referring to Equation 5.4 above, we can see that the functions cancel out leaving a flat SIC curve. For serial exhaustive models, an additive pattern arises as both LH and HL items will show some slowing relative the to HH item due to their lower salience on one of the dimensions. The increase of RT for the LL item

compared to the HH item is therefore simply the sum of the individual sources of slowing.

The parallel self-terminating model predicts an entirely positive SIC function. This is because the RT is determined by the faster of the two decisions (i.e., the minimum processing time). The RT for the HH, HL, and LH items will therefore be shorter compared to the LL item. For parallel exhaustive models the converse is true. RT is determined by the slower of the two decisions (i.e. the maximum processing time). The RTs for the LL, LH, and HL items are therefore longer than the HH item, predicting an entirely negative function.

The coactive model predicts an initial negative portion (sometimes referred to as a negative "blip"), followed by a positive function. This function therefore integrates to a positive value. This over-additive pattern has been shown by Townsend and Nozawa's (1995) original mathematical proof and also later in simulations by Fifić, Nosofsky, & Townsend (2008).

These predictions do not depend on the particular forms of the RT distributions. This means that they are non-parametric and can apply to entire classes of serial, parallel, and coactive models, making them a powerful tool for assessing processing architecture.

2.1.4 Workload capacity

Assessing a system's workload capacity (sometimes termed *information capacity*, or *processing capacity* in the literature) can provide additional information on the efficiency of that system, in particular, whether there is a cost, benefit, or no effect to the system when workload is varied (Townsend & Ashby, 1983; Wenger & Townsend, 2000). Importantly, it is distinct from the concept of an *item capacity* which generally refers to an absolute limit on the number of items which can be stored in memory (see e.g. Luck & Vogel, 1997; Vogel et al., 2001). While interesting in its own regard, a measure of workload capacity can also offer additional diagnostic power, particularly when distinct architectures and stopping predict the same SIC signatures (this is particularly an issue for

interactive parallel models whose processing channels are dependent on each other in some way, see e.g., work by Eidels et al., 2011).

In the double factorial paradigm, the number of targets is varied from one to two targets, and consequently, the change in the rate of processing can be computed as the number of targets increases. Hence, workload capacity can be calculated by comparing the pattern of RTs for the single target items to the double target item. To interpret the capacity value, this calculation is compared to the expected performance of an unlimited capacity system, whose predictions are derived from the single target trials.

If adding additional signals slows the processing rate beyond that expected from the baseline model, then processing is of limited capacity. If the processing rate is unaffected, processing is of unlimited capacity. Finally, if the processing rate increases with additional signals, the system is operating at super capacity.

The capacity coefficient is computed using the integrated hazard functions, $H(t)$, of the single and double targets:

$$H(t) = \int \frac{f(t)}{S(t)} dt = -\log(S(t)) \tag{2.3}$$

To compute capacity, the integrated hazard function must be calculated for the two workload conditions. Hence, the integrated hazard function for the condition where the participant is presented with redundant information (i.e., two targets; H_{AB}(t)) must be divided by the sum of the two integrated hazard functions from the single target conditions (H_A(t) and H_B(t)):

$$C_{or}(t) = \frac{H_{AB}(t)}{H_A(t) + H_B(t)} \tag{2.4}$$

Note that we use the notation $C_{or}(t)$ above to distinguish the capacity coefficient in an OR task to the capacity coefficient in an AND task. This is discussed in further detail below.

The capacity coefficient $C(t)$ can be interpreted by comparing it to the predicted capacity coefficient of an unlimited capacity independent parallel

(UCIP) model. This model assumes that the redundant target condition is determined by the minimum time of either channel in isolation. The UCIP model therefore predicts that the rate of processing will be unaffected by the workload and therefore will satisfy the equality:

$$H_{AB}(t) = H_A(t) + H_B(t) \tag{2.5}$$

A system with unlimited capacity will therefore produce a capacity coefficient equal to one across the entire time course of processing; $C_{OR}(t) = 1$. This provides a benchmark for measuring workload capacity. A limited capacity system predicts $C_{OR}(t) < 1$. An unlimited capacity system predicts $C_{OR}(t) > 1$.

It is important to note that an RT advantage for detecting the redundant target could arise solely due to statistical facilitation in an independent parallel race model (i.e., the minimum time expected for the detection of two possible targets is smaller than the detection of any single target alone; Raab, 1962). J. Miller (1982) formulated the *race model inequality* (RMI) to provide an upper bound on the speeding of responses which can be accounted for by statistical facilitation alone. Conversely, a lower bound on the level of limited capacity provided by the UCIP model is provided by the *Grice bound* (Grice et al., 1984). This bound assumes that the fastest of the two single channel items is slower than the response to the redundant target item.

The ways of calculating and interpreting the capacity coefficient described above assume that processing can self-terminate (i.e., it assumes an OR decision rule). In an AND task, however, both target locations must be processed exhaustively. In order to maintain the same interpretation, an exhaustive parallel model must now be used as a baseline model. The $C_{AND}(t)$ therefore uses the reverse integrated hazard function; $K(t)$ (Townsend & Wenger, 2004). $K(t)$ can be defined as the log of $f(t)$:

$$K(t) = \log(f(t)) \tag{2.6}$$

Capacity for an AND decision rule is therefore defined as the ratio of the sum of the reverse integrated hazard functions for the single targets, K_A(t) and K_B(t), over the double target, K_{AB}(t):

$$C(t)_{AND} = \frac{K_A(t) + K_B(t)}{K_{AB}(t)} \tag{2.7}$$

This formulation of the coefficient for the AND case allows for the same interpretation as $C_{OR}(t)$. That is, $C_{AND}(t) = 1$ indicates unlimited capacity, $C_{AND}(t) < 1$ indicates limited capacity, and $C_{AND}(t) > 1$ indicates super capacity.

2.1.5 SFT in categorization

Fifić, Nosofsky, & Townsend (2008) were the first researchers to apply the SFT methodology to a multidimensional categorization task. Their primary aim was to use SFT to address a foundational question in multidimensional categorization research; namely, the difference between integral and separable stimuli. Integral dimensions are dimensions which cannot be easily attended to in isolation (Garner, 1974; Shepard, 1964). For example it is very difficult to attend to only the hue, saturation, or brightness of a color. These dimensions are generally thought to be processed holistically and therefore accord well with the concept of coactivity. Separable dimensions, on the other hand, can be easily attended to independently, even when in combination (Garner & Felfoldy, 1970; Garner, 1974). Separability is therefore more akin to the two types of independent processing architectures; serial and parallel.

Fifić, Nosofsky, & Townsend (2008) tested participants on a series of integral stimuli (Munsell colors varying in saturation and brightness) and separable stimuli (two rectangles separated or overlapped in space, varying in the hue of one rectangle and the position of a line contained within the other rectangle). As predicted, they found that integral stimuli were best described by a coactive processing strategy, whereas there was no evidence for coactivity for the separable

stimuli, even when they overlapped in space. This research forms the immediate precursor to the Logical-Rules models (Fifić et al., 2010) which, in turn, form the basis of the current thesis. An overview of the Logical-Rules models is provided in the following section.

2.2 The Logical-Rules models

The Logical-Rules models (Fifić et al., 2010) expanded on the work of Fifić, Nosofsky, & Townsend (2008) in a number of key ways. Firstly, they provide a full process account of categorization over the entire time course of decision-making. Secondly, they provide additional diagnostic power for identifying processing architecture through their *contrast category* analysis. Thirdly, they complement non-parametric SFT analysis with computational modeling.

The Logical-Rules models synthesize several main approaches from both categorization and the modeling of RT data. From a categorization perspective, they follow early rule-based theories of categorization, which assumed that categories were represented as a series of logical rules (Bourne, 1970; Levine, 1975; Trabasso & Bower, 1968). Specifically, the Logical-Rules models assume that individuals make individual decisions regarding the value of a stimulus's dimensions and then combine these using logical operations (e.g. "AND", "OR", "NOT"; Nosofsky, 1989; Feldman, 2000; Ashby & Gott, 1988) in order to reach a final decision. This is combined with a decision-bound/ GRT approach (Ashby & Townsend, 1986; Ashby & Gott, 1988; Ashby & Lee, 1991; Ashby et al., 1994; Maddox, 1992; Maddox & Ashby, 1996). To determine RT, the Logical-Rules models combine evidence accumulation models (Brown & Heathcote, 2008; Busemeyer, 1985; Luce, 1986; Ratcliff, 1978) with the mental architectures approach.

An example of the typical logical rules decision space, accompanied by the stimuli used in the current thesis, is shown in Figure 2.4. In the logical rules design, the top right quadrant is identical to the DFP and so can be used for SIC

analyses such as the MIC and SIC as described earlier. Unlike the DFP, however, the logical rules design comprises two continuous dimensions (dimensions X and Y in Figure 2.4), each varying on *three* levels. This gives a nine item stimulus space, with each individual stimulus comprising a single value from each dimension.

The space is divided into two categories via a decision boundary (the dotted line in Figure 2.4). The upper right quadrant is referred to as the *target* category (category A) whereas the remaining items belong to the *contrast* category (category B). In order to correctly classify a stimulus as belonging to category A, a conjunctive rule must be satisfied (i.e., both items must be processed before a correct decision can be made). With reference to Figure 2.4 we can see that, in order to belong to category A, the stimulus must have a value of at least X_1 on the X dimensions AND Y_1 on the Y dimension. Conversely, to correctly classify a stimulus as belonging to category B, only a disjunctive rule need be satisfied (i.e. processing is able to self-terminate). With reference to Figure 2.4 we can see that, in order to belong to category B, the stimulus have a value of less than X_1 on the X dimension OR a value of less than Y_1 on the Y dimension.

These logical rules are instantiated using decision-bound theory, a GRT approach (Ashby & Townsend, 1986; Ashby & Gott, 1988; Ashby & Lee, 1991; Ashby et al., 1994; Maddox, 1992; Maddox & Ashby, 1996). Here, participants are assumed to establish two orthogonal decision bounds (the dotted line in Figure 2.4) giving a fixed criterion along each dimension. Hence, in order for a stimulus to be correctly classified as belonging to category A, it must lie above AND to the left of the decision boundary. Conversely, to be correctly classified as belonging to category B, a stimulus must lie below OR to the left of the decision boundary. As specified by GRT, each stimulus value is represented as a normal distribution of perceptual effects. Due to perceptual noise, these stimuli overlap, resulting in errors in the sequential-sampling process.

Earlier research utilizing the Logical-Rules models utilized the Random-Walk model (Luce, 1986) to account for RT. Rather than combining perceptual

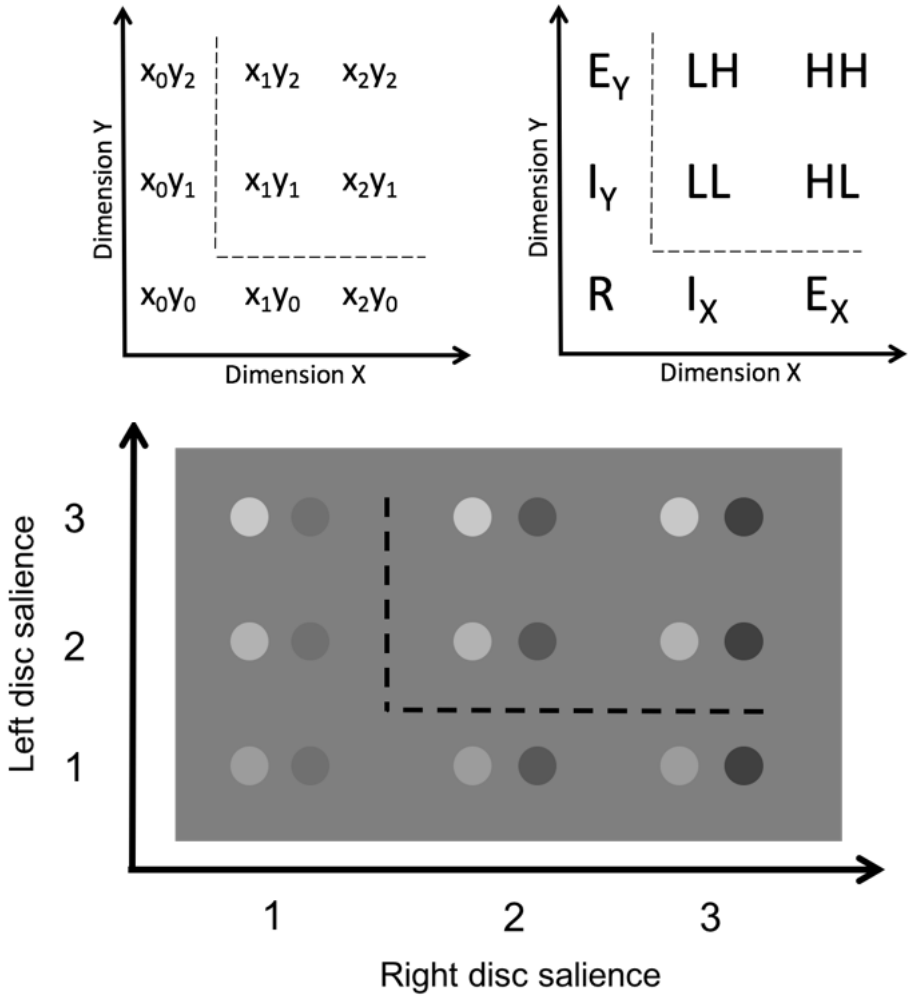

Figure 2.4: The logical rules decision space. Top left panel: Schematic diagram of the stimulus space. Stimuli are comprised of two dimensions (dimension X, the right disk, and dimension Y, the left disk). Each dimension varies on three levels which combine orthogonally to form a nine-item stimulus space. Top right panel: Schematic diagram showing discriminability and category membership. The target category (A) includes high (H) and low (L) salience dimensions and is defined by a conjunctive ("AND") rule. The contrast category (B) is defined by a disjunctive ("OR") rule and includes internal (I), external (E), and redundant (R), stimuli. The dotted line represents the decision boundary. Bottom panel: Stimulus space for experiments presented in Chapter 3, shown here as a point of comparison to the double factorial paradigm. Discs are comprised of three different salience levels (1 = Low Luminance, 2 = Medium Luminance, and 3 = High Luminance), for both black (darker than the background) and white (lighter than the background) levels of luminance. Note: each pair of discs forms one stimulus. Stimuli to the right and above of the decision boundary (indicated by the dotted line) belong to category A. All other stimuli belong to category B.

information together to feed a single random-walk process (as seen in previous models such as the EBRW; Nosofsky & Palmeri, 1997) the Logical-Rules models assume that a separate and independent random-walk process occurs for each dimension. The final decision is therefore dependent on the relevant logical rule for each category. For example, a stimulus can only be categorized as belonging to category A if both individual random-walk processes reach the category A decision criterion. Conversely, only one random-walk process need reach a category B criterion to be classified as belonging to category B.

Using this framework each of the candidate architectures can be instantiated. For example, a serial exhaustive process is simply the sum of time taken for the random-walk to complete on dimension X and dimension Y. For serial self-terminating models, if the process is able to terminate for the first-processed dimension, then RT is determined by the time taken for the single random-walk process on that dimension. Conversely, if a decision cannot be logically made, RT is determined by the sum of the two random-walk processes. For the parallel exhaustive model, both random-walks must complete, and therefore RT is determined by the slower of the two random-walks. Conversely, for a parallel self-terminating model RT is determined by the fastest random-walk process. For the coactive model, it is assumed that each sample drives a single, pooled random-walk.

While the random-walk process formed the basis for predicting RTs in earlier research using the Logical-Rules models, later research (see e.g. Blunden et al., 2020; Cheng et al., 2018) utilizing the Logical-Rules models utilized the Linear Ballistic Accumulator (LBA; Brown & Heathcote, 2008). The LBA approximates the random-walk process and provides an efficient method for predicting the decision time for each dimension. Drift rates for the LBA are generated by integrating the perceptual distributions with respect to the decision boundary within each category region. The decision times for each accumulator are then either, for example, summed for the serial exhaustive model or used to find the maximum time prediction for the parallel model. For the coactive model, rather

than modeling the perceptual distributions independently, the variability of the perception of both dimensions is modeled as a bivariate normal distribution.

In addition to the target category predictions afforded by SFT, the Logical-Rules models provide two additional measures for diagnosing processing architecture. Firstly, the models can be instantiated parametrically and can therefore be subject to model comparison processes. Secondly, they provide additional qualitative predictions to that of SFT via the contrast category. These predictions are the focus of the following section.

2.2.1 Contrast Category Predictions

In addition to the target category, the contrast category also provides qualitative predictions (for a summary of the Logical-Rules models mean RT predictions across both target and contrast categories, see Figure 3.4). In particular, as the contrast category is defined by a disjunctive decision rule, the order in which dimensions are processed provides a useful point for distinguishing each of the models. For example, in a serial self-terminating model, observers may process the dimensions in a fixed order (e.g. they may always process dimension X first, followed by dimension Y if required). Referring again to Figure 2.4, if an observer who is adopting this X then-if-necessary Y decision strategy is presented with a stimulus which has a value of 0 on the X dimension, then the disjunctive rule is satisfied and processing can terminate. If however, they are presented with x_1y_0 or x_2y_0 processing must continue to the Y dimension in order to make a correct categorization decision. This pattern of processing leads to the following predictions. Firstly, RTs for the first-processed dimension are approximately equivalent whereas RTs for the second-processed dimension are comparatively slower (as they must wait until processing is finished on the first-processed dimension). Secondly, the exterior item, x_2y_0, is faster than the interior item, x_1y_0. This is because the exterior item is further from the decision boundary and is therefore easier to determine that it is *not* unique to the contrast category.

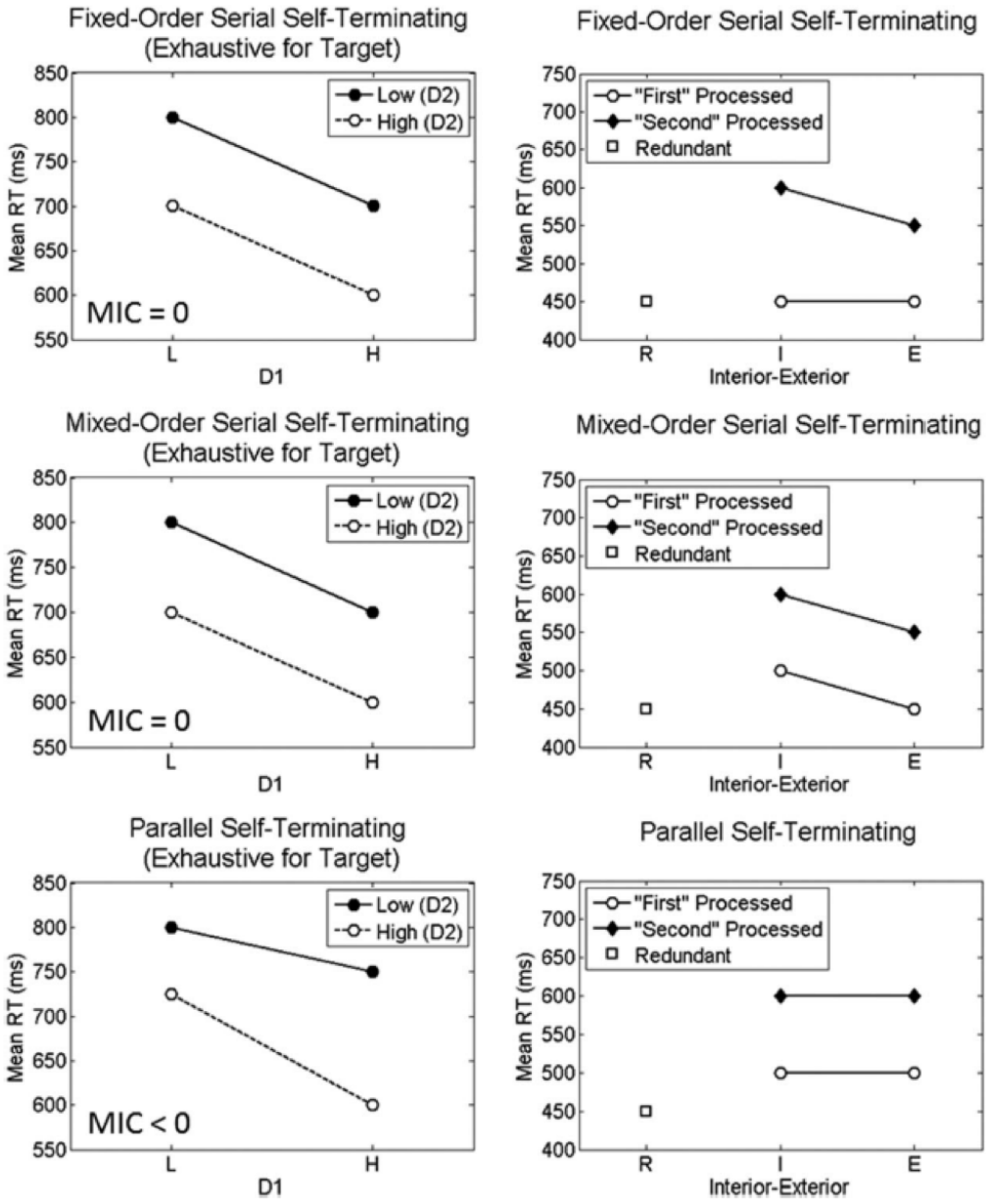

Figure 2.5: RT predictions for each mental architecture. These predictions were generated from simulations but are proven to hold under mild assumptions (Townsend & Nozawa, 1995). Left panel: predictions for target category, category A. Right panel: predictions for contrast category, category B. Each row represents one of the candidate architectures. D1 = First processed dimension. D2 = Second processed dimension. EBRW = Exemplar-Based Random Walk. Continues on next page.

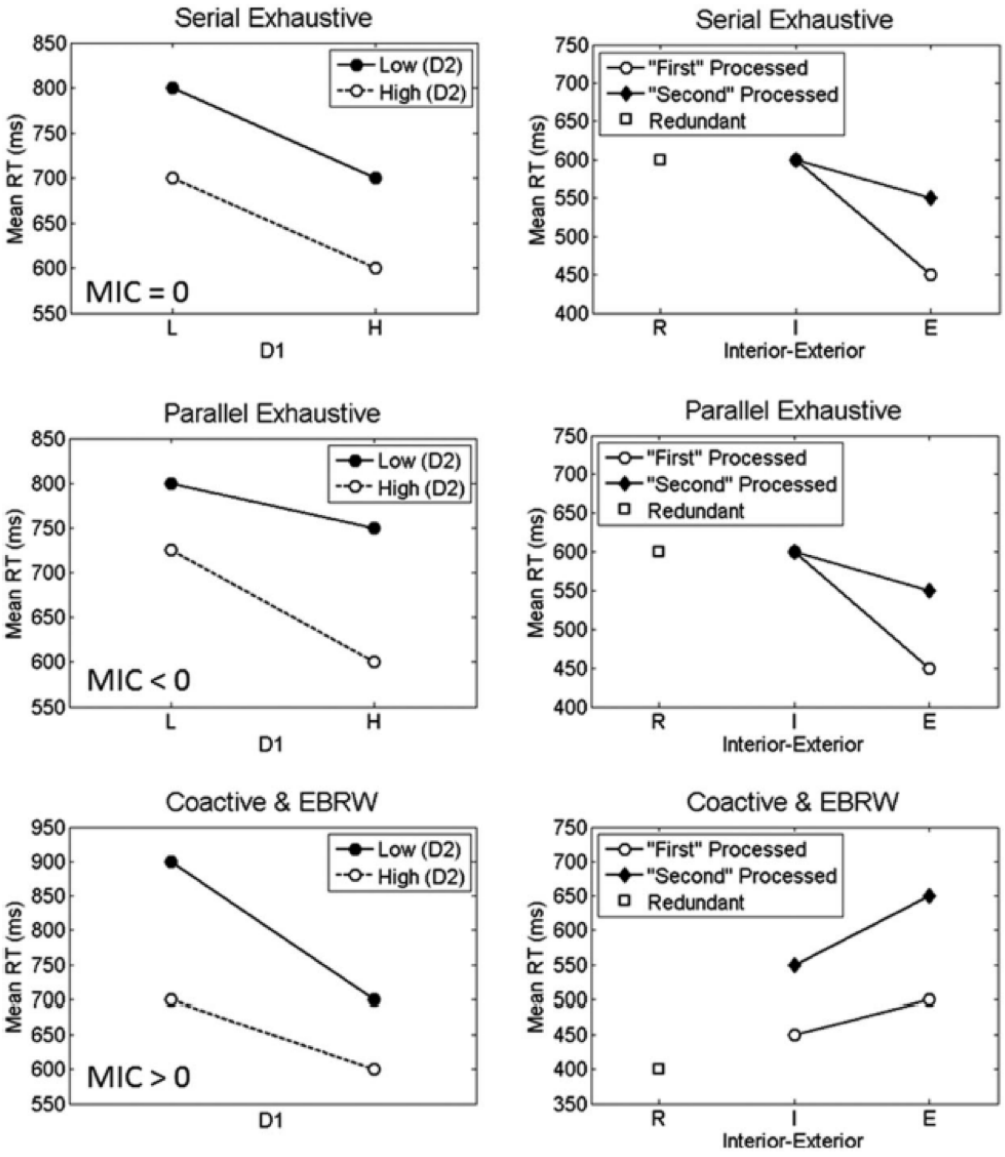

Figure 2.6: *Figure 3.4* continued.

In addition, to fixed-order self-terminating models, processing can also occur in a mixed order. That is, on some trials observers may process dimension X first, followed by dimension Y if required, while on other trials they may process dimension Y first, followed by dimension X. In a mixed-order self-terminating model, the redundant stimulus therefore has the greatest advantage as it satisfies the disjunctive rule on both dimensions. The exterior item is again faster than the interior item on both dimensions due to its position relative to the decision boundary. Put another way, on average, the time taken to switch to the other dimension is shorter for exterior items.

For the parallel self-terminating model, the redundant stimulus again has the greatest advantage. As both dimensions yield a contrast category response, statistical facilitation can occur between the dimensions (Raab, 1962) (i.e. there are more chances to self-terminate). For the remaining interior and exterior items, processing time is roughly equivalent for contrast category classifications (as processing time is determined by the minimum RT from either dimension).

A serial exhaustive model predicts that (depending on the relative rates of processing) the redundant stimulus will be the slowest item. This is because the overall RT is determined by the sum of both dimensions and the redundant stimulus lies closest to both decision boundaries, making it most difficult to distinguish on both dimensions with interior items being slower than exterior items. In a parallel exhaustive model RT is determined by the maximum processing time. This again means that the redundant stimulus is slowest, with interior items being slower than exterior items.

In the coactive model, the perception of both dimensions is considered to be jointly distributed as a bivariate normal. As one moves from the exterior stimuli to the interior stimuli, evidence becomes more consistent with a category B response. This is because a greater proportion of the distribution will lie in the contrast category region and consequently there is a higher probability of samples being drawn from the contrast category region. This means that the interior stimuli are faster than the exterior stimuli, with the redundant stimulus

having the greatest RT advantage. In the following section I will consider how these predictions have contributed to recent advances in categorization decision-making using the Logical-Rules models. In particular, I will consider how recent work has utilized rule-based theories of categorization, not only to characterize decision-making which necessitates the integration of a range of different features (e.g., size, color, shape) and their configurations, but also to answer more fundamental questions regarding perception, attention, and decision-making.

2.3 Processing architecture of multidimensional stimuli

In their seminal paper, Fifić et al. (2010) varied saturation and line position of two rectangles separated in space and participants were instructed to process the items using a fixed-order serial self-terminating strategy. The results showed that participants did indeed process the items in a fixed-order serial-self terminating manner, validating the ability of the models to diagnose decision-making strategies. A further test of the models utilized lamp stimuli which varied the width of the base and the curvature of the top piece. Although these dimensions formed part of a single object, they were still processed in a serial manner. Stemming from a validation of the models, (Little et al., 2011) sought to diagnose participant's chosen processing strategy when no explicit instructions were given. Using schematic drawings of lamp stimuli, Little et al. (2011) showed that the participants' chosen processing strategy was still serial when the dimensions were spatially separated. However, when they overlapped the same rectangular stimuli as used in Fifić et al. (2010) (Experiment 2, Little et al., 2011), they found that processing was best described as a mixture of serial and parallel processing.

These findings are interesting as they highlight the importance of spatial configuration in determining processing architecture. In Fifić et al. (2010) and

Little et al.'s (2011) Experiment 1, the dimensions were *separable*. That is, they could be easily attended to in isolation (Garner & Felfoldy, 1970; Garner, 1974). Importantly, they were also presented separately in space, meaning that selective attention was likely needed to resolve the feature values at each location independently, leading to serial processing. However, in Little et al.'s (2011) Experiment 2, the dimensions were separable but instead *co-located* in space. This lead to processing which was best described as a mixture between parallel and serial processing. From these results, it seems that spatial location plays a key role in determining processing strategy.

However, spatial location is not the only factor determining categorization strategy. For example, *integral* stimuli are thought to be processed holistically or configurally (Lockhead & King, 1977; Nosofsky, 1988). This means that the dimensions which comprise these stimuli cannot be attended to independently and selectively in isolation (Garner, 1974; Shepard, 1987; Nosofsky, 1988). Fifić, Nosofsky, & Townsend (2008) found that Munsell colors (a hallmark integral stimuli) varying on the dimensions of brightness and saturation were processed coactively (see also Little et al., 2013; Blunden et al., 2015). As participants were unable to attend to each dimensions separately, they instead pooled the visual representation into a single coactive processing channel. In contrast to Fifić et al. (2010) and Little et al. (2011) where spatial information seemed to determine processing strategy, the type of dimension in Fifić, Nosofsky, & Townsend (2008) - that is, the fact that the dimensions were integral rather than separable - seemed to be the main determinant of processing architecture. Of course, while Little et al. (2011) found differences in processing strategy depending on location for their separable stimuli, importantly, processing was always independent. Clearly both spatial location and stimulus type are both important considerations.

Although integral dimensions must necessarily occupy the same location, separable dimensions may either be located in different parts of the visual scene, or may occupy the same location. A unique type of separable stimulus,

however, is composed of so-called *whole-object* dimensions. These dimensions comprise the entire object (e.g., color, size, and shape) and therefore must necessarily be co-located. Although the dimensions themselves are separable, these stimuli have characteristically been considered to be integral (Biederman & Checkosky, 1970; L. B. Smith & Kilroy, 1979). Moneer et al. (2016) sought to test this using the Logical-Rules framework to provide a rigorous formal definition of integral processing (operationalized as coactivity) and separable processing (operationalized as serial or parallel processing). Moneer et al. (2016) showed that whole-object dimensions are not best described as integral, but that they actually elicit independent, multi-channel processing (i.e., serial or parallel processing). Similar results have been found with composite faces, which have also been traditionally treated as holistic (Cheng et al., 2017). Hence, the processing of different feature types not only in the same spatial location but comprising the whole object, depends on whether those features are separable or integral.

A naturally occurring question which stems from this body of work concerns stimuli which are located separately in space, but which do not comprise different dimensions. While past research in categorization has focused on stimuli which comprise different feature types, henceforth referred to as *between-dimension* features (e.g., shape vs. color), *within-dimension* features which comprise differing levels of the *same* type of feature (e.g., gray scale discs varying in luminance, triangles varying in size etc.) have largely been overlooked (although see Fitousi, 2019; discussed in further in Chapter 3, for an exception). As such the architecture underlying processing of within-dimension features is unclear. It is possible that because within-dimension features are presented in separate locations, selective attention will be required to resolve the feature values for decision-making. Consequently, processing may proceed serially. Alternatively, within-dimension features may be processed in a parallel or coactive manner. Further it remains unclear what the relevant stopping-rule and processing capacity may be. These questions forms the basis of the first major investigation of the

current thesis, presented in Chapters 3 and 4. Although the categorization literature has not yet considered within-dimension features, the distinction between within- and between-dimension features forms a key theoretical difference in other areas of the visual attention literature. As such, it is useful to consider studies which utilize within-dimension features in other domains and how they may cohere with potential categorization decision-making architectures.

2.4 Within-dimension stimuli

2.4.1 Theories of Visual Attention and Visual Search

A primary focus of theories of visual attention such as Feature Integration Theory (FIT; A. Treisman & Gelade, 1980) and Guided Search (Wolfe, 1994a, 2007) is how visual features present in the visual scene may alter the attentional mechanisms used to complete visual search tasks. Although they do not speak explicitly to decision-making architecture, they nonetheless provide some useful starting hypotheses regarding how attention may influence decision-making using within-dimension stimuli. For example, Guided Search proposes that focused visual attention is driven by an early pre-attentive parallel processing stage. At this early stage, different featural dimensions such as luminance, orientation, and spatial frequency are represented separately in different *feature-maps*. These individual feature maps are then summed together to form a master *salience map* which subsequently guides attention in order to complete the search process. Whether the dimensions which comprise the objects in the visual search display are between-dimension or within-dimension is critical to the efficiency of the search. For example, visual search for an item containing a conjunction of two between-dimensions features (e.g., the target is a red vertical line and distractors are red horizontal lines and green vertical lines) can still be guided efficiently in parallel. However, when search involves items whose features are represented on the same feature map and are therefore within-dimension features (e.g., search for a red-green target among red-blue and blue-green distractors) it is no longer

efficient (Wolfe et al., 1990).

Given that the attentional mechanisms for between- and within-dimensions are different for visual search tasks, a difference may also be expected in the categorization experiments presented in the current thesis. For example, as within-dimension stimuli are represented at an early stage as a single feature-map we may expect a similar holistic representation in categorization, leading to a finding of coactive processing. However, Guided Search is primarily concerned with guiding attention to different locations in the visual scene. As such, it is less clear what role visual attention plays in the processing of features at a specific location since feature processing is confounded with the visual search process itself. That is, although the search may proceed serially, the feature dimensions of any individual object may proceed in parallel, or even coactively. Wolfe et al. (1990) provided a possible step towards clarifying this question by having participants complete a "search" with a set size of one (i.e., a simple identification task). Wolfe found that there was no difference in RT for identifying whether a color $\times$ color (within-dimension) conjunction was a target versus a color $\times$ orientation (between-dimension) conjunction was a target. This suggests that there is no cost to identifying whether or not a conjunction is a target for within-dimension feature conjunctions versus between-dimension feature conjunctions. However, this result alone does not address whether the colors in a color $\times$ color target, which are of interest in the present work, are processed independently or not. I return to this discussion in Chapter 3.

2.4.2 Consolidation into VSTM

Another area which has utilized within-dimension stimuli is the literature investigating VSTM consolidation. For example, Mance et al. (2012) investigated the sensory encoding of color stimuli into VSTM using the simultaneous-sequential paradigm. In this paradigm, to-be-processed items are either presented to the participant simultaneously or one at a time in a sequential manner. If a processing advantage is found for the sequential condition over the simultaneous

condition, this shows that the workload capacity is limited. However, if no such advantage is found, this provides evidence that items are able to be processed concurrently without any limit on workload capacity. In their task, Mance and colleagues (2012) presented participants with 1-4 colored squares either simultaneously or sequentially and found no performance advantage for the sequential condition over the simultaneous condition for up to two items. They suggested that encoding of color features into VSTM occurs without any limit in capacity and in parallel. However, it is important to note that the simultaneous-sequential paradigm cannot be used to make inferences about serial and parallel processing since a sequential advantage can reflect either serial or certain types of limited capacity parallel processing (Townsend, 1990a). Nonetheless, this finding does suggest that encoding of within-dimension features is unlimited in capacity. While this does not necessarily indicate parallel encoding, it is at least consistent with it. Using an analogous design, this result has also been shown to extend to orientation (Sewell et al., 2014).

In addition to within-dimension features, J. R. Miller et al. (2014) also investigated the processing capacity of the sensory encoding of between-dimension features. Using the simultaneous-sequential paradigm, Miller and colleagues presented participants with two discs, one of which was one of four candidate colors and the other, one of four possible orientations of sinusoidal gratings. Unlike the processing of within-dimension features, a sequential advantage was found for encoding the simultaneously presented orientation and color items. This suggests that while within-dimension features can be encoded with unlimited capacity, the same does not hold for between-dimension features. These findings, in conjunction with those of the visual search literature, are of interest to the current thesis as they imply that different stages of processing using visual information may have different limits on processing that are dependent on the combination of features being processed.

An alternative explanation provided by the authors is that it is possible that the grating stimuli used were more difficult to encode, which may have therefore

exceeded the system's capacity limits in a way that colour did not[1]. While the possibility that different capacity limits may exist for different combinations of features is important to consider, it is also important to note that these findings do not give an indication of the processing architecture of the decision-making stage. The main subject of the first part of the thesis will therefore be a complete characterization of the decision-making architecture, stopping rule, and capacity of within-dimension stimuli in a categorization task. This is the focus of the investigation presented in Chapter 3.

Although decision-making using visual elements currently present in the scene is important, arguably the vast proportion of decision-making involves integrating information, not only currently present in the visual scene, but also from memory. In the second part of the thesis I therefore present an extension of the Logical-Rules paradigm in order to diagnose processing in a one-shot multi-element change detection task. I now turn to a review of the relevant change detection literature, focusing on the integration of information is visual short term memory (VSTM).

2.5 The one-shot change detection task

Early evidence for a specific short-term memory store specifically for visual information was presented by Posner & Keele (1967). In this experiment participants were asked to identify whether two successively presented letter stimuli matched. In one set of trials the case of the letter was switched (i.e., a lower case "a" would be correctly identified as matching an uppercase "A") and in another set the case was consistent (i.e., the visual information was identical between displays). Posner & Keele (1967) found an advantage in mean RT for items which were visually identical as opposed to simply maintaining the identity of the stimuli. Drawing on this work Phillips & Baddeley (1971) presented participants with a 5×5 array of randomly presented black and white squares which could not

[1]See also Becker et al. (2013), although note this is not found to be the case in other research using orientation stimuli (Sewell et al., 2014)

be verbally coded. These were interceded with a retention interval and random interrupting mask. They found that change detection performance was well above chance, although this performance decreased substantially after retention intervals of more than three seconds, further showing evidence for a specific visual short term memory which was not lexically mediated.

These studies form the basis of all modern investigations using the *one-shot change detection task*. This task involves the presentation of a memory array followed by a probe array after a short period of time has passed (i.e., the retention interval). Participants must then identify whether the probe array differs from the memory array on a given dimension (e.g., by reporting whether the color of any disc has changed; Phillips, 1974; Vogel et al., 2006). Generally, experimental work using this task has provided insight into the series of processes necessary for change detection. A memory representation of the visual scene must first be formed (Sperling, 1960) and this representation must then be consolidated into visual short-term memory (VSTM; Jolicoeur & Dell'Acqua, 1998; Mance et al., 2012; Sewell et al., 2014; Phillips, 1974; Vogel et al., 2006). The information must then be retained (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997) until it is retrieved (Hollingworth, 2003; Mitroff et al., 2004) and compared to the current scene in order to make a response (Hyun et al., 2009; C.-T. Yang, 2011; C.-T. Yang et al., 2011). These mechanisms are inferred from results supporting the limited capacity of visual memory (Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974).

A key result of the modern change detection literature for the current thesis is the finding that capacity limitations apply to discrete objects but not necessarily to the features which comprise them. This object-based account is supported by the finding that accuracy decreases as a function of the number of items in the display and that performance is improved by binding features into single objects (Duncan, 1984; Luck & Vogel, 1997; Vogel et al., 2001). On the other hand, some features that appear as conjunctions within a single object can seem to be treated as independent items (Wheeler & Treisman, 2002). Further, Fougnie & Alvarez

(2011) found that when participants were asked to recall the individual features of an object, errors were independent of one another (i.e., even if participants could not recall one feature of an object, they were often still able to recall the other feature). This is counter to an object-based account, which would predict that an inability to recall one feature should also lead to an inability to recall the other. While these findings implicate the possibility that not only objects but also features may compete for access into VSTM, these results do not imply either a pooled or independent decision-making architecture. That is, decision making can be both pooled or independent at the level of features or the level of objects. In short, the explanations mimic each other.

The limited capacity of visual short-term memory has prompted theorizing of a discrete storage limit of around four objects; items are either encoded into these "slots", or not encoded at all (Cowan, 2001), in which case that participant resorts to guessing. These models are, in essence, high threshold models (Luce, 1963). High threshold models, and their equivalent discrete object models in a change detection task, predict distinctly linear ROCs. This is because, while high threshold models may fail to encode an item into memory, they will never mistake a distractor for a target. These models are generally not supported by ROC data, which is characteristically curved (e.g., Swets, 1961; Swets et al., 1961; Tanner & Swets, 1954a,b; see Wilken & Ma, 2004 for an example). Instead, ROCs appear to support a theory in which a noisy signal is compared to an internal criterion to generate a response (e.g., signal detection theory; Green & Swets, 1966; Peterson et al., 1954). This type of model is best described as utilizing flexibly allocated resources, whereby memory can be flexibly allocated among items with no necessary limit on the number of items able to be stored (Wilken & Ma, 2004). One challenge to this result has been investigations by Rouder et al. (2008). Using a standard change detection task, Rouder et al. (2008) showed that a high threshold version of a slots-based model outperformed a signal detection model, with straight ROC curves best accounting for human performance. This result was later replicated by Donkin et al. (2014).

However, Donkin et al. (2016) provided evidence that empirical differences in how change detection is measured may lead to different outcomes, supporting either discrete object theories or resource models. They found that when the number of to-be-remembered items was consistent over a session, participants were able to flexibly allocate their attention. However, when set size varied from trial-to-trial participants were more likely to engage in "slot-like" encoding of information. That is, when the environment was unpredictable, participants were more likely to focus on a smaller subset of items. Keshvari et al. (2013) found that change detection requiring low magnitude changes rather than highly distinct and categorically different changes (e.g., changes between distinct and nameable colors) was also best described by a resource-based account. Evidently, there are strategic differences in how attention is employed that depend on task constraints.

2.6 Integration of information in VSTM

Although the change detection literature has made great advances in (a) developing of models that differentiate "slot-based" accounts (i.e., the use of discrete information) from the use of continuous resources (see e.g., Wilken & Ma, 2004; Bays & Husain, 2008), (b) developing mixture models which encompass elements of guessing and information-based responding (Zhang & Luck, 2008; Cowan & Rouder, 2009), and (c) characterizing capacity limits in the VSTM processes that underlie change detection (Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974), the way in which evidence is integrated in order to make change detection decisions has largely been overlooked.

A notable exception to this is work done by Wilken & Ma (2004). Wilken & Ma (2004) contrasted a high threshold model with two signal detection models: a first-order signal detection model and a second-order signal detection model. This distinction between first- and second-order models comes from seminal work done by Shaw (1980, 1982) regarding the integration of information to

make simple target identification and visual search decisions, such as monitoring multiple locations for the presence of light, cross modal monitoring of auditory and visual information, and letter detection. In first-order accounts, relevant information is pooled and a single decision is made based on this combined information. In second-order accounts a separate decision is made based on each piece of relevant information. These separate decisions are then aggregated to form a final decision.

Wilken & Ma (2004) contrasted both first- and second-order accounts of the effect of set size on change detection accuracy. In their Experiments 1 to 3, testing changes in color, orientation, and spatial frequency, respectively, set size was varied from 2 to 8 items presented in a circular array around fixation. After a briefly displayed memory set (100 ms) and blank interval (1500 ms), a test display was briefly presented (100 ms) after which participants indicated the presence of a change in any of the locations. Hit rates decreased and false alarms increased across set size in all three experiments. The ROCs had a curvilinear shape ruling out a high threshold model. However, their ROC results were unable to differentiate their first- and second-order integration models. Although the minimum evidence model quantitatively fit better in some instances (e.g., using χ^2), the pooled model was quantitatively better in other conditions.

2.7 RT accounts of VSTM decision-making

While response time analysis forms the basis for answering such questions in categorization, investigation of response time has largely been missing from the change detection literature. The majority of work has instead inferred properties of the memory store using accuracy under varying memory load. However, unique insights into the decision component of change detection as distinct from the memory store can be gleaned using a response time account. For example, Donkin et al. (2013) replicated the results of Rouder et al. (n.d.) (discussed previously) using two class variants of the LBA model, one to represent a slots-based account

and one to represent a continuous resources account. In the continuous resources account, the quality of the visual representations varied based on the number of to-be-remembered items, with all items stored in memory. Evidence for a decision-making process was then accumulated from these representations. In the slots-based account, responses were accumulated either from a limited number of precise representations, or alternatively, a guessing process was utilized. They found support for the slots-based account, although there may be small changes to memory quality when the upper limit of VSTM item capacity was not met. However, as discussed earlier, later work by Donkin et al. (2016) showed that this was highly dependent on the parameters of the task (i.e., whether the task was predictable or unpredictable).

Lilburn (2016) also investigated the effect of task type on decision-making. He compared a change detection task with an orientation discrimination task, controlling for decision complexity using a post-stimulus probe, and obtained both accuracy and response time data. When modelling participant sensitivity he found that a sample-size relationship accounted for performance as a function of memory load in the orientation discrimination task. In the change detection task the same relationship held, however, an additional item representing the probe needed to be accounted for. This account was supported by response time modeling using the diffusion model. For change detection decisions, an additional time constant was needed to account for encoding the probe array as well as a comparison process between the probe and memory item. A process which accounted for intrusions from non-target information into the decision making process was also found.

The results of Lilburn (2016) and Donkin et al. (2016) show that response time data can also provide meaningful conclusions about the change detection process, as well as give greater insight into how decision-making strategy may be altered according to task parameters. However, the architecture of the decision process utilized in both Donkin et al. (2013) and Lilburn (2016) was not explicitly examined. In both studies a single item was cued meaning that decision-making

only need be represented by a single processing channel (represented by a single pair of LBAs in Donkin et al., 2013 and a single diffusion process in Lilburn (2016)). While this was necessary to answer the particular questions posed in these works, it remains unclear how information from multiple sources might be integrated to drive decision-making which utilizes the VSTM store. This is a fundamental question which is yet to be addressed by the change-detection literature and forms the basis of the second part of this thesis.

However, Lilburn's (2016) results also bring to mind the question of whether there are different complexities in processes between perceptual categorization and change detection. For example, whether information can be directly accessed from working memory in a manner which is similar to perceptual information present in the visual scene and can therefore be easily subject to SFT analysis is an open question. If information from working memory must be selected into a single-item focus of attention for use in a decision-making process (see e.g., Oberauer & Hein, 2012; Sewell et al., 2016) then the basic architectural definitions of serial, parallel, and coactive may not hold. Further, while the salience manipulation of SFT is designed to isolate the decision-making stage specifically, if the salience manipulation also affects an earlier selection process, then this may confuse the SFT analyses. While coupling a measure of capacity with a measure of architecture can help to elucidate more complex models (Eidels et al., 2011), this is nonetheless an important limitation of the analysis method to keep in mind. These issues are considered in greater detail in the empirical chapters.

The current thesis proposes a novel way to diagnose evidence integration (i.e., whether it is best described as first- or second-order) by using a measure of response time. This is achieved through an extension of the Logical-Rules models which can not only distinguish pooled versus independent processing, but can also diagnose processing architecture and stopping rule. The first part of the thesis therefore focuses on characterizing the architecture, stopping rule, and capacity of decision-making utilizing two luminance discs separated in space.

This is then extended to novel modeling framework using a change detection task.

Chapter 3

Categorization decision-making using within-dimension stimuli

The majority of this chapter was published in the Attention, Perception, & Psychophysics special issue in honor of the contributions of Anne Treisman:

Blunden, A. G., Howe, P. D. L., & Little, D. R. (2019). Evidence that within-dimension features are generally processed coactively. *Attention, Perception, & Psychophysics. https://doi.org/10.3758/s13414-019-01775-8.*

Minor edits have been made for this Thesis

3.1 Evidence that within-dimension features are generally processed coactively

Every day we make decisions by identifying, discriminating, comparing, and categorizing objects that have different visual features which come from the same dimension (e.g., colors, sizes, shapes etc.). To give a few examples of this everyday decision making, imagine you wish to select a ripe banana from a bunch: How do you decide which banana is the most preferable? An obvious approach is to compare the color of the bananas, preferring the bright yellow bananas while avoiding the green and overripe brown bananas. Likewise, a food safety inspector may need to comply with federal guidelines on the color of meat in order to decide whether the rib eye is safe, and must therefore compare the color of the rib eye with a meat safety color chart. A navy ship office may need to identify and interpret the color combination of the signal flags of an approaching vessel. Understanding how we make categorization decisions using stimuli that vary on the *same* type of feature is a fundamental question in the psychology of perception and cognition. Our interest in these examples and the present paper, is not how we identify the color, but rather how that color satisfies some criterion that informs some decision: Is this the ripest banana? Is the meat fresh enough? What is the other ship's intention?

Recent work has utilized rule-based theories of categorization incorporating theories of response time (RT), not only to characterize decision-making, which necessitates the integration of a range of different features (e.g., size, color, shape) and their configurations, but also to answer more fundamental questions regarding perception, attention, and decision-making (Griffiths et al., 2017; Little et al., 2011; Fifić et al., 2010; Little et al., 2013; Moneer et al., 2016). One critical question regards the underlying architecture of decision-making. Here, architecture refers to the organization of mental processes, or the way in which we combine information (Kantowitz, 1974; Sternberg, 1969; Schweickert, 1993; Townsend, 1984). More technically, it refers to distinguishing between serial,

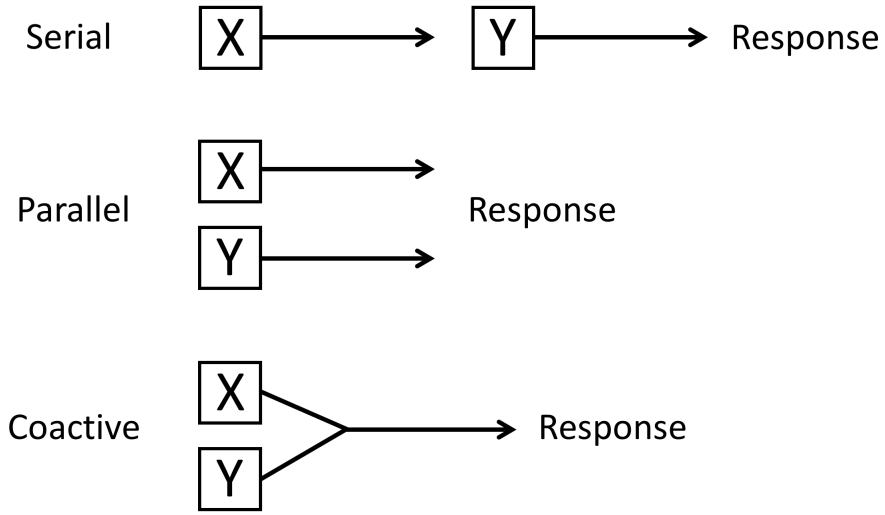

Figure 3.1: Schematic illustration of the the three processing architectures: serial, parallel, and coactive. This example uses a hypothetical stimulus with two dimensions: X and Y. In a serial model decisions about each dimension are made sequentially, one after the other. In a parallel model, decisions are made simultaneously, but in independent processing channels. In a coactive model the information from both stimuli is combine into a single processing channel from which a decision is made.

parallel, and coactive processing. For example, when selecting a banana do you make a decision on the color of each banana one at a time (i.e., serially)? Or alternatively, do you do so simultaneously, in separate decision-making channels (i.e., in parallel), or do you pool all your perceptual information into one single decision-making channel and make one overall decision (i.e., by processing the colors *coactively*; J. Miller, 1982; see Figure 3.1)?

Beyond identifying the processing architecture, which is not always straightforward due to pervasive model mimicry in many tasks (Townsend, 1990a), a further question is what factors affect the processing architecture of decision-making? For example, characteristics such as whether or not features or dimensions can be attended to in isolation, and the configurations of these dimensions, such as spatial separation, have been shown to play a key role in determining how information from different dimensions is integrated (Moneer et al., 2016; Little et al., 2011). For example, it is difficult to attend to hue, brightness, or saturation

independently. It would be therefore impossible to judge the ripeness of a banana based on the hue of its color alone, while ignoring saturation or brightness. Instead, information about these three dimensions must be pooled in order to make a categorization decision. On the other hand, judging which variety of bananas to buy (e.g., the Cavendish or the Lady Fingers, two popular Australian varieties) might require comparison of two separate dimensions which are easy to attend to in isolation: freshness and price. Such information is likely to be located in two different spatial locations requiring attention to be independently deployed to each, with categorization decisions made independently. The ability to diagnose the underlying processing architecture is therefore crucial as it provides information about the role of spatial attention and selective attention in the sequencing of dimensions for processing. Here we use spatial attention to refer specifically to the allocation of attention across space, while selective attention refers to the more general process of weighting specific stimulus attributes, potentially including distance. Many paradigms for studying attention focus on accuracy, choice data, or mean RT alone (for example, spatial cuing tasks, see e.g., Posner et al., 1980, or visual search, see e.g., A. Treisman & Gelade, 1980); however, these paradigms are often limited in their ability to differentiate serial, parallel, and coactive architectures from each other.

We seek to add to the growing body of work on visual processing in categorization by testing an important configuration of multidimensional stimuli such as those in the introductory examples: stimuli which are composed of dimensions separated in space but which comprise differing levels of the *same* type of feature, henceforth referred to as *within-dimension features* (see Figure 3.2 for some examples). This type of stimulus is interesting from the perspective of visual attention as it has features that would appear on the same feature map, for instance, in Treisman and Gelade's (1980) Feature Integration Theory (FIT). In our experiments, the features are separated spatially and take on different luminance values but can be continuously transformed from one to another. By contrast, this would not be the case for features conceived to be on different

feature maps (e.g., shape vs. color), which we term *between-dimension features* (see Figure 3.2). This is a key distinction in many theories of visual search, which find differences in performance between the two stimulus types (Wolfe et al., 1990). It is therefore possible that FIT may foreshadow a difference in categorization strategy for between and within-dimension stimuli.

Theories of visual attention such as FIT (A. Treisman & Gelade, 1980) and Guided Search (Wolfe, 1994a, 2007) provide useful insight into how attention operates as a function of which features are present in the visual scene. Although these theories do not explicitly address categorization, they propose that focused visual attention is driven by an early pre-attentive parallel processing stage. In this stage, different visual dimensions (e.g., luminance, orientation, spatial frequency, etc.) are registered separately and only later combined or bound to form a visual representation of an object or visual scene. Guided Search, for example, suggests that different dimensions are represented in separate maps which are summed together to form a master *salience map* that subsequently guides attention to relevant areas of a visual scene. Because of this guidance, when a target does not share any features with the distractors (e.g., a red square among green squares), search occurs efficiently, and purportedly in parallel. In these cases the target "pops out", and search time is generally fast and independent of the number of distractors. If, however, there is not a sufficient difference between the target and the distractors to cause "pop out", search is not efficient (see e.g., Duncan & Humphreys, 1989) and is instead driven by effortful, attentive processing which is purportedly more serial in nature (Wolfe, 1998, 1994a, 2007). Importantly, these theories emphasize attention as being driven by the specific features involved in the task, and therefore these tasks are often used to infer how features are processed. Given we are using within-dimensions features, we may expect that these are combined pre-attentively into the same feature map which then provides a single signal for decision making. This would mean decision making would occur coactively. Independent processing (e.g., serial or parallel) may be more consistent with between-dimension stimuli from

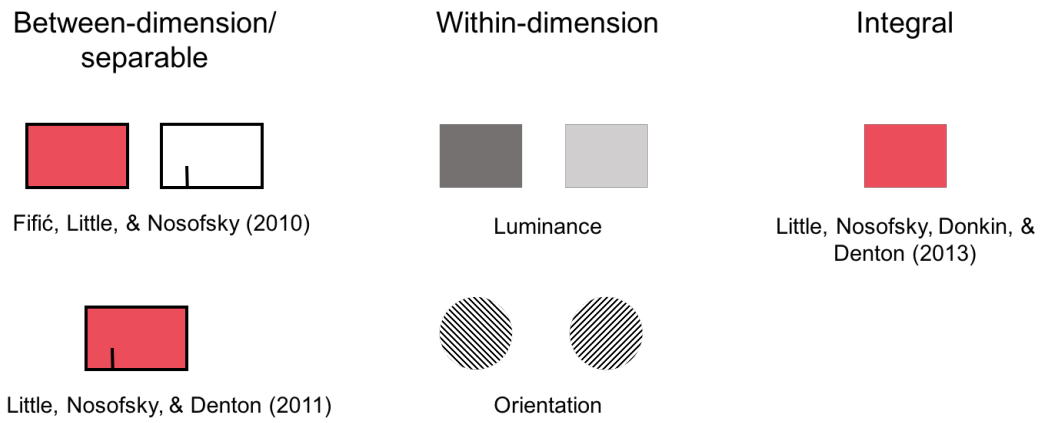

Figure 3.2: Examples of different feature types. Left panel: Examples of two experiments utilizing between-dimension stimuli which are separable. The stimuli are between-dimension as they incorporate features which would be located on separate feature maps (namely, color and line position). The stimuli are separable as each dimension can be easily identified in isolation from the other (i.e., the color is red, and the line is to the left). Both of these characteristics hold whether the stimuli are co-located (bottom), or separated in space (top). Middle panel: Examples of stimuli which are within-dimension. These stimuli are within-dimension as they incorporate features which would be located on the same feature map (in these examples, luminance and orientation). By definition these cannot be overlapped in two-dimensional space. Right panel: An example of an integral stimulus. Color is considered integral as its comprising dimensions (i.e., hue, saturation, and brightness) cannot be readily identified in isolation.

different feature maps (indeed, this is what we see in the categorization literature; see e.g., Fifić et al., 2010).

To assess the processing of within-dimension features in perceptual categorization, we utilized Systems Factorial Technology (SFT; Townsend & Nozawa, 1995; Little, Altieri, et al., 2017). Specifically, we investigated stimuli of opposite luminance polarity in separate locations of a visual display in a categorization decision-making task. Several previous studies of visual attention (Duncan & Humphreys, 1989; Wolfe et al., 1990; Mordkoff & Yantis, 1993) have also examined within-dimension features, and we return to these studies in our discussion; we first review related work in perceptual categorization and then present our current experiments.

3.1.1 Dimensional Processing in Categorization

The diagnosis of processing architecture during decision-making using visual information has been the focus of recent work in categorization (Blunden et al., 2015; Cheng et al., 2017; Fifić et al., 2010; Little et al., 2013, 2011; Moneer et al., 2016). This literature draws on theories of visual attention to investigate how information from multi-dimensional sources is integrated during decision-making. The goal of this work is to provide a quantitative and detailed diagnosis of mental architecture using the logical rules modeling framework (Fifić et al., 2010) which complements the SFT analyses (Townsend & Nozawa, 1995; Little, Altieri, et al., 2017) in order to yield patterns of RTs across the entire time course of information processing.

To date, this work has focused on the integration of information from between-dimension features and has been successful in characterizing the underlying architecture of this process using a variety of different dimension types and configurations. Most relevant to the current work is the seminal paper by Fifić et al. (2010) introducing the logical rule models. Fifić et al. (2010) presented participants with two rectangles separated in space, one of which varied in saturation, and the other of which varied in the position of a line contained within the rectangle. Because these dimensions could easily be attended to independently (i.e., because they are *separable* – see e.g., Garner & Felfoldy, 1970, Garner, 1974 – and physically separated), these dimensions could be processed in serial when participants were instructed to do so. Serial processing was also found for spatially separate dimensions when participants were allowed to adopt their own categorization strategies (Little et al., 2011). However, when two separable, between-dimension features were presented in a spatially overlapped fashion (Experiment 2, Little et al., 2011), processing was more parallel.

These findings are interesting as they highlight the importance of spatial configuration and attention in determining the architecture that underlies categorization. Before the introduction of the logical rule models, all models

of categorization RT, including the successful Exemplar-Based Random Walk Model (Nosofsky & Palmeri, 1997), stochastic General Recognition Theory (GRT; Ashby, 2000), and decision bound models (Ashby et al., 1994; Maddox, 1992; Maddox & Ashby, 1996), assumed that the features of objects were pooled together into single objects (i.e., what we term coactivity). In both of the cases described above, both used separable dimensions, but having features positioned in separate locations induced serial processing, whereas overlapping the features in space produced more parallel processing. This would suggest that feature type is not necessarily a key component affecting processing strategy in categorization decision-making tasks but rather that location of the features is primary. Following this logic, we would expect our within-dimension luminance stimuli to require selective attention to resolve the feature values at each location, and consequently, categorization decision should proceed serially with separate micro-decisions made at each location combined using logical rules to determine the final response.

On the other hand, there are likely additional factors that may determine how features are processed. For instance, not all features can easily be attended to independently or selectively (Garner, 1974; Shepard, 1987; Nosofsky, 1988); these types of features, termed integral features, are thought to be processed holistically or configurally (e.g., hue, saturation, and brightness of colors in the Munsell color space; Lockhead & King, 1977; Nosofsky, 1988). Using the same categorization methodology, Little et al. (2013) found that the categorization of colors varying in saturation and brightness was best described by a coactive process (see also Blunden et al., 2015). Instead of making decisions separately along the brightness dimension and saturation dimension, participants instead pooled information about these dimensions into a single decision-making channel. In contrast to Fifić et al. (2010) and Little et al. (2011), the type of dimension - that is, the fact that the dimensions were integral rather than separable - played a key role in determining the processing architecture.

Integral dimensions necessarily occupy the same spatial location. However,

Moneer et al. (2016) investigated categorization of *whole-object* features, which are features that comprise an entire object, and therefore spatially co-located, but are notionally separable, such as shape and size or shape and color. These whole-object features are between-dimension features in the sense that they would activate different feature maps[1], but, despite being comprised of separable features, these features have been traditionally been characterized as integral (Biederman & Checkosky, 1970; L. B. Smith & Kilroy, 1979). Using the logical rules paradigm, Moneer et al. (2016) showed that these dimensions elicit independent, multichannel processing (i.e., serial or parallel processing). Similar results have been found with composite faces, which have also been traditionally treated as holistic (Cheng et al., 2017). Hence, the processing of different feature types not only in the same spatial location but comprising the whole object, depends on whether those features are separable or integral.

A natural question arising from these experiments concerns the categorization of within-dimension but spatially separated features of the type which are often used in studies of visual search (e.g., a red pop-out target in a field of green distractors). In the present paper, we use the strong inferential methods to answer this question. Within-dimension stimuli, such as luminance discs, provide an important point of investigation. Likely, because these within-dimension features are presented in separate locations, selective attention will be required to resolve the feature values. Consequently, spatial configuration will play a stronger role in determining processing strategy, and we expect processing to proceed serially as a result. On the other hand, different these luminance values would be represented by the same feature map (A. Treisman & Gelade, 1980) and so it is possible that they would instead be processed coactively. While parallel processing seems somewhat unlikely, it is nonetheless worthy of consideration as a possible means for processing the within-dimension features. For example, in simple redundant target detection of two luminance targets, processing appears

[1]Technically, size could be considered within the same dimension as shape as it is an aspect of shape; when varied orthogonally in a restricted stimulus set, size and shape are nonetheless separable in the sense that they satisfy empirically observable markers of separable (e.g., Garner, 1974).

to proceed in parallel but with limited capacity (Townsend & Nozawa, 1995; C.-T. Yang et al., 2014). Visual attention, stimulus configuration, and dimension type, may all play key roles, and none of the candidate models (serial, parallel, and coactive) can be ruled out a priori. In the remainder of the introduction, we provide a detailed explanation of the categorization decision-making paradigm and logical rule models which we utilize to uncover how within-dimension features are processed.

3.1.2 Logical Rules Design

In order to differentiate the processing architectures, we utilize a design which provides strong diagnostic contrasts between the predictions of each of the candidate architectures. It is useful to describe these predictions with reference to Figure 3.3, which shows a variant of the double factorial design proposed by Fifić et al. (2010). In this design, individual stimuli are comprised of the orthogonal combination of two dimensions, each of which vary over three levels. This generates nine stimuli, each of which comprise different levels of the two dimensions. While the double factorial paradigm has previously been implemented using single items which comprise both dimensions (e.g., halves of a face; Cheng et al., 2017, or two parts of a lamp; Fifić et al., 2010) the within-dimension feature stimuli used in the present experiment have different values of the same feature at different respective locations in space. Hence, each stimulus comprises a pair of discs, with each disc varying on three levels of saliency with respect to the background.

The upper right quadrant comprises the target category (category A) stimuli. The dotted line represents the category boundary between the target category and the contrast category (category B). Items that lie closer to the category boundary should be more difficult to discriminate (Ashby & Gott, 1988; Nosofsky, 1986); hence, stimulus dimensional values of either high discriminability (H) or low discriminability (L) combine to form four stimuli which are defined by their relative difficulty in discriminability: HH, HL, LH, and LL.

The contrast category stimuli are also identified by their location in the category space. The redundant stimulus, R, satisfies both of the boundary decisions necessary to classify a stimulus as belonging to the contrast category (i.e., it is both to the left and below the decision boundary). The stimuli adjacent to R are termed the interior stimuli, I_X and I_Y, whereas the stimuli at the far edges are termed the exterior stimuli, E_X and E_Y.

In order to correctly classify a target category stimulus, a conjunctive rule on both dimensions must be satisfied. That is, a stimulus must have a value on both dimension X and Y that exceeds the horizontal and vertical decision boundaries, respectively. Specifically, the luminance of both disks must have a value of 2 (medium luminance) or higher. Hence, stimuli from the target category must be processed exhaustively (both dimensions must be processed before a correct decision can be made). Stimuli belonging to the contrast category can be correctly classified using a disjunctive rule (i.e., stimuli need only be below or to the left of the horizontal and vertical decision boundaries, respectively).

Note that these rules, used to instantiate the categories in the task, do not presume any sort of processing architecture. The fact that both dimensions must be processed exhaustively to correctly classify a target category stimulus, does not preclude this processing from being carried out one dimension at a time in serial, or simultaneously in parallel, or indeed pooled into a single integrated percept. The following section describes how the predictions from each of these models (and combinations of stopping rules) varies across both categories. The important point to note is that the model predicts that processing will be exhaustive for the target category when responding is correct, but processing may be self-terminating or exhaustive for the contrast category. Of course, a participant might self-terminate and still be accurate when responding to a contrast category item.

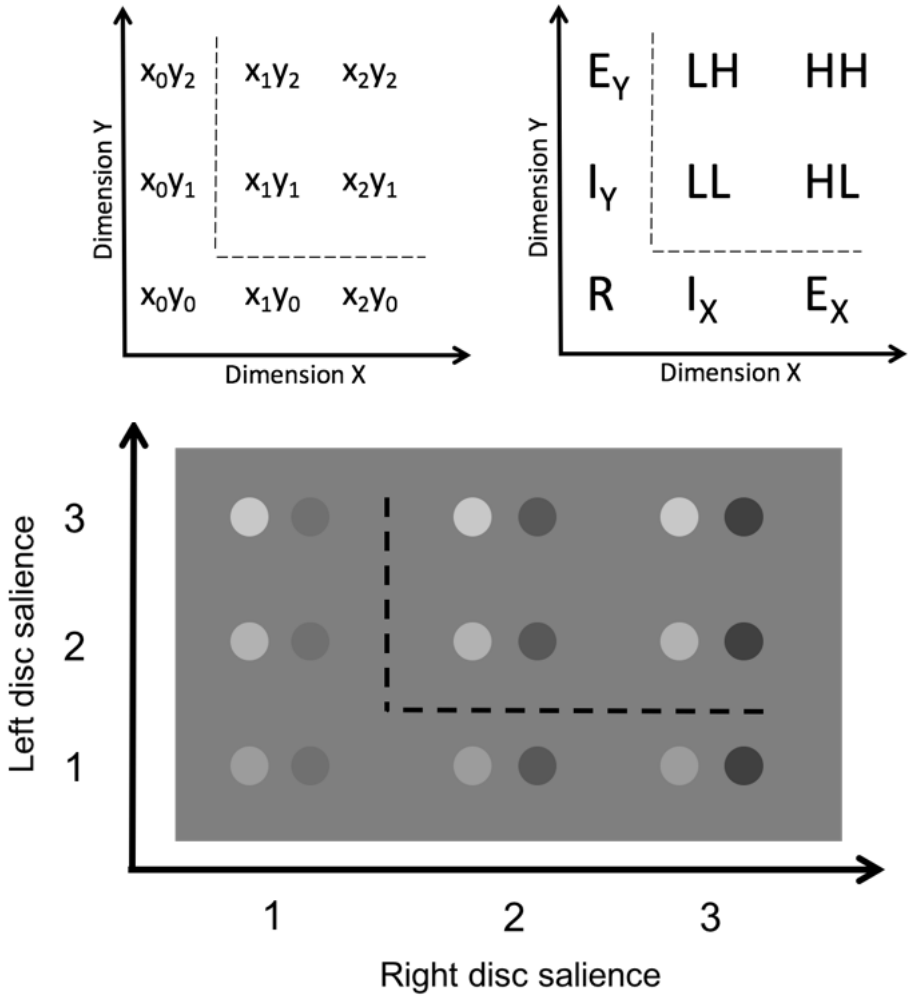

Figure 3.3: Top left panel: Schematic diagram of the stimulus space. Stimuli are comprised of two dimensions (dimension X, the right disk, and dimension Y, the left disk). Each dimension varies on three levels which combine orthogonally to form a nine-item stimulus space. Top right panel: Schematic diagram showing discriminability and category membership. The target category (A) includes positions of high (H) and low (L) salience items and is defined by a conjunctive ("AND") rule. The contrast category (B) is defined by a disjunctive ("OR") rule and includes internal (I), external (E), and redundant (R), stimuli. The dotted line represents the decision boundary. Bottom panel: Stimulus space showing example stimuli. Discs are comprised of three different salience levels (1 = Low Luminance, 2 = Medium Luminance, and 3 = High Luminance), for both black (darker than the background) and white (lighter than the background) levels of luminance. Note: each pair of discs forms one stimulus. Stimuli to the right and above of the decision boundary (indicated by the dotted line) belong to the target category. All other stimuli belong to the contrast category.

Target Category Predictions

Discriminating target category discs that are close to the decision boundary should be slower than discriminating discs that are further away (Ashby & Gott, 1988). The way in which these discriminations are combined varies for each model, and qualitatively different mean RTs are predicted by each model architecture for the target category stimuli (shown in the left panel of Figure 3.4). These RTs are readily summarized by the mean interaction contrast (MIC). The MIC is calculated by finding the difference between the difference of the low and high discriminability values on one dimension and the difference of the low and high discriminability values on the other dimension:

$$MIC = (RT_{LL} - RT_{LH}) - (RT_{HL} - RT_{HH}) \tag{3.1}$$

Serial models predict an additive pattern of mean RTs ($MIC = 0$), parallel models predict an under-additive pattern of mean RTs ($MIC < 0$), and coactive models predict an over-additive pattern of mean RTs ($MIC > 0$; Townsend & Nozawa, 1995). A brief explanation for these predictions is outlined below; however, for a more detailed outline, please see Fifić et al. (2010).

Serial models predict an additive pattern as both LH and HL items will show some slowing relative to the HH item due to their lower discriminability on one of the dimensions. The increase of RT for the LL item compared to the HH item is simply the sum of the individual sources of slowing. Parallel models predict an under-additive pattern because the RTs for the target category items are determined by the slower of the two decisions (i.e., the maximum processing time). The LH and HL stimuli will be therefore much slower than the HH stimuli. The LL stimulus will be only slightly slower than either the LH or HL stimuli. Finally, Townsend & Nozawa (1995) provide a mathematical proof demonstrating coactive models result in an over-additive pattern of results. This finding has been corroborated by simulations done by Fifić, Nosofsky, & Townsend (2008). The predictions outlined above are non-parametric in that they do not depend

on the particular forms of the RT distributions; hence, the qualitative contrasts apply to the entire class of serial, parallel, and coactive models.

Further diagnostic evidence for processing architecture from the target category can be found by calculating the survivor interaction contrast (SIC). The SIC is calculated using the survivor function for each stimulus, at each time value, t:

$$SIC(t) = [S_{LL}(t) - S_{LH}(t)] - [S_{HL}(t) - S_{HH}(t)] \tag{3.2}$$

where the survivor function, $S(t)$, is the complement of the cumulative distribution function, $F(t)$, and represents the probability that a response has not been made by time, t.

Different mental architectures also produce qualitatively distinct predictions for the SIC (Townsend & Nozawa, 1995; see Figure 5.4) when using an exhaustive stopping rule (as is necessary for correct responses in the target category). Serial models predict an initially negative function which becomes positive, with the entire function integrating to zero (i.e., the MIC equals zero). Parallel models predict an entirely negative function. Coactive models predict an initial negative blip, with the majority of the SIC being positive. Coactive models do not integrate to zero but rather integrate to a positive value (note that since the target category uses an AND rule, all of the serial and parallel models, including those with a self-terminating rule, must predict an exhaustive SIC as shown in Figure 5.4)[2].

Contrast Category Predictions

The contrast category stimuli can also be used to make diagnostic judgments using mean RTs (Fifić et al., 2010; see also Little et al., 2015, 2018). Illustrative

[2]Note that self-terminating serial and parallel models make different SIC predictions; however, since our factorial stimuli in the target category require exhaustive processing (i.e., both dimensions must be processed before a correct decision can be made), self-terminating strategies applied to the target category would be accompanied by higher error-rates. As described in the method and results, we encouraged highly accurate responding throughout. Therefore, we only consider self-terminating strategies when considering the contrast category results.

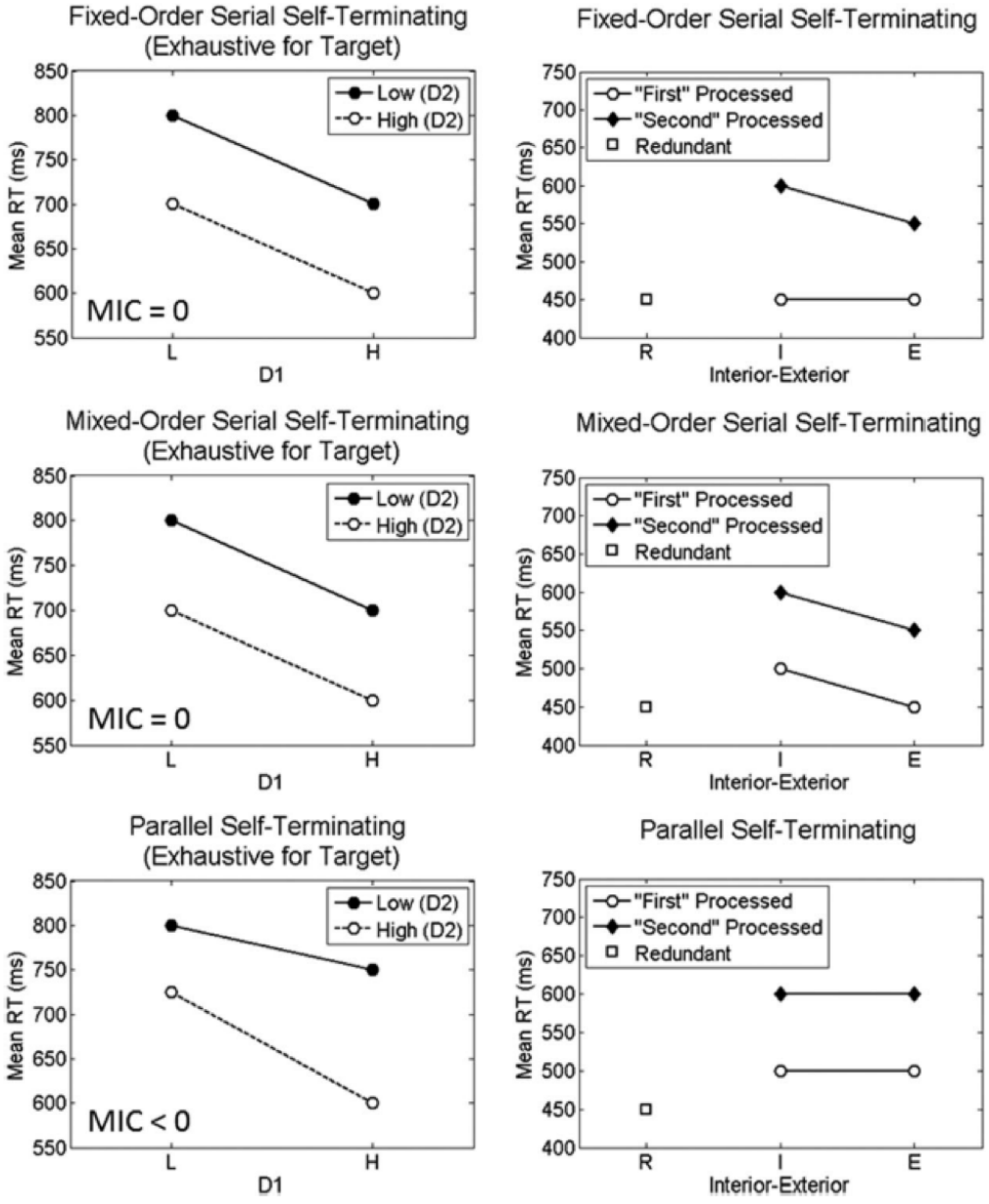

Figure 3.4: Illustrative RT predictions for each mental architecture. These predictions were generated from simulations but are proven to hold under mild assumptions (Townsend & Nozawa, 1995). Left panel: predictions for target category, category A. Right panel: predictions for contrast category, category B. Each row represents one of the candidate architectures. D1 = First processed dimension. D2 = Second processed dimension. Continues on next page.

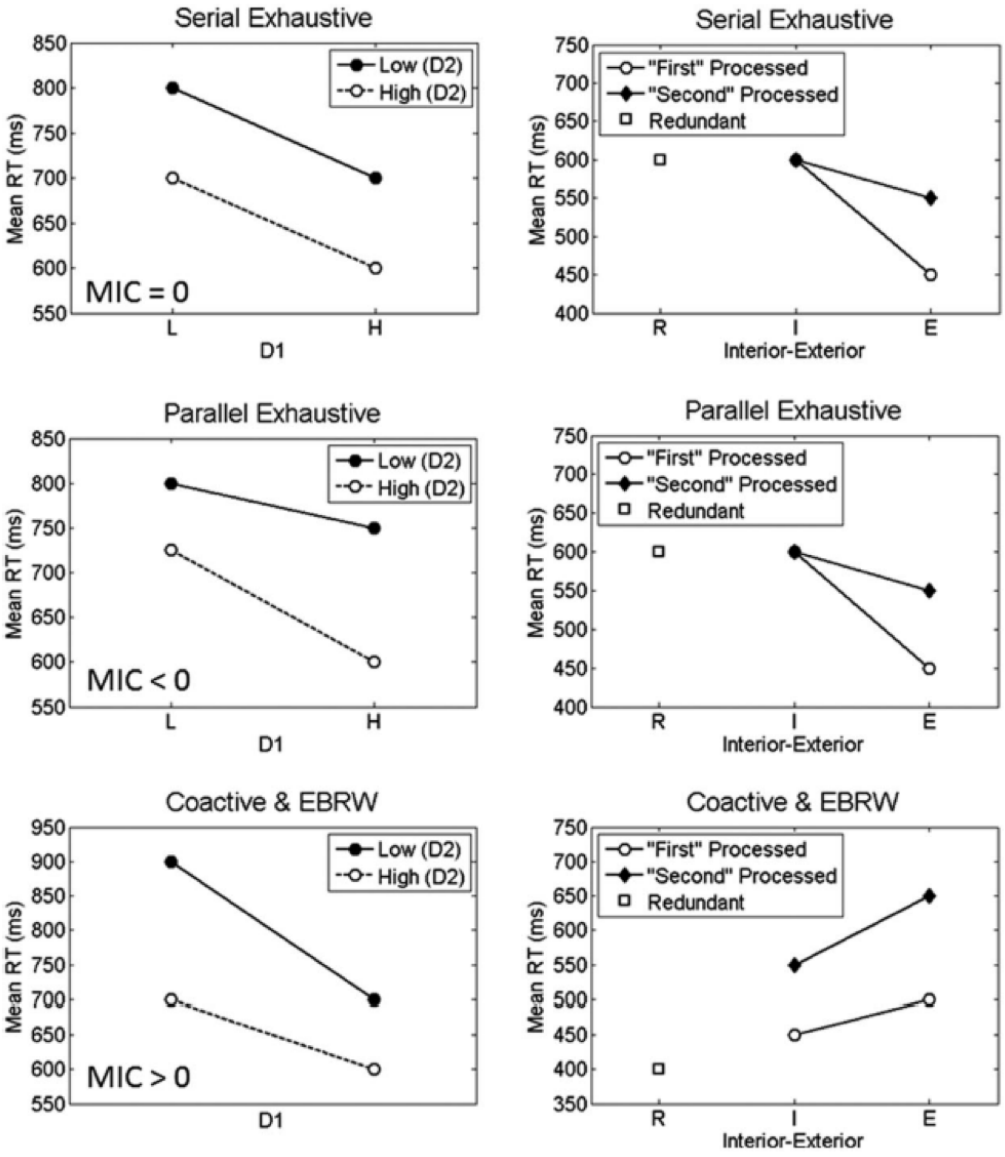

Figure 3.5: *Figure 3.4* continued.

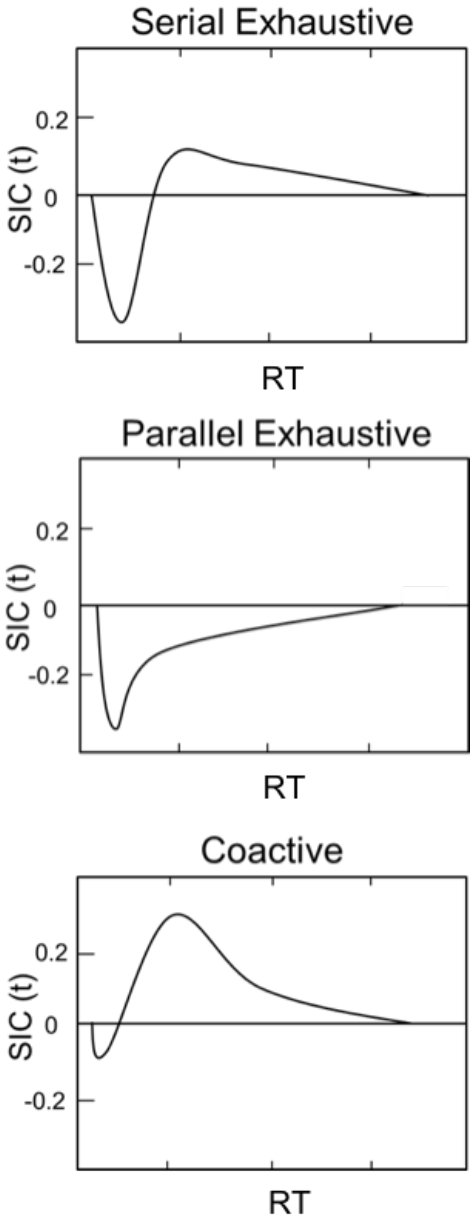

Figure 3.6: Schematic of the survivor interaction contrast (SIC) predictions for serial, parallel, and coactive architectures.

predictions for each model are shown in Figure 3.4. For example, consider the predictions of a fixed-order serial self-terminating model, in which dimension x is processed first followed by, if necessary, dimension y (for reference see Figure 3.3, left panel). The presentation of items which satisfy the disjunctive contrast category rule on dimension x (i.e., x_0y_0, x_0y_1, x_0y_2) will lead to a decision without further processing. If, however, x_1y_0 or x_2y_0 are presented and x is processed first, then further processing of dimension y is necessary in order to make a correct categorization decision. This leads to a general prediction that RTs for the first processed dimension will be approximately equivalent, whereas RTs for the second processed dimension are comparatively slower (as they need to wait for the completion of the first processed dimension). Further, the exterior item (x_2y_0) is processed faster compared to the interior item (x_1y_0) because the first processed x dimension for the exterior stimulus is further from the decision boundary and therefore easier to judge as not belonging to the contrast category.

For mixed-order self-terminating models, the first-processed dimension may be dimension x on some trials and dimension y on other trials. It follows that the redundant stimulus has the greatest processing advantage, as both dimensions satisfy the disjunctive decision rule. Further, these models predict that exterior items will be processed faster than interior items on both dimensions. Averaged across trials, where participants switch from one dimension to the other, the time preceding the switch is shorter for exterior items.

For a parallel self-terminating model, the RT for a contrast category item is determined by the minimum processing time needed to make a categorization decision. More specifically, in the current task, the processing time will be determined by the dimension which yields a contrast category response. The redundant stimulus therefore has a processing advantage in this case as both dimensions yield contrast category responses, allowing for statistical facilitation between the dimensions (Raab, 1962).

In the case of exhaustive models, both dimensions are processed regardless of whether the disjunctive rule is satisfied or not. For a serial exhaustive model,

this means that total RT comprises the sum of RTs on each single dimension. This leads to a prediction that, depending on the relative rates of processing, the redundant stimulus will have the slowest RT (as it lies closest to both decision boundaries), with interior items being processed slower than exterior items. The parallel exhaustive model predicts that the total RT is determined by the maximum processing time needed to make a categorization decision. Again, redundant and interior items are slower as they lie closer to the decision boundaries compared to the exterior items.

Finally, a general prediction of the coactive architecture is that the interior stimuli will be processed faster than the exterior stimuli. The intuition is that the interior items are located closer to the left-most corner of the decision space, and therefore when the dimensions are pooled, interior items pool more evidence for a contrast category response compared to exterior items (where the evidence for one of the pooled dimensions is comparatively more diagnostic of a target category response).

This design offers considerable diagnosticity as data from each item can be used to differentiate architectures and associated stopping rules. As the analysis is non-parametric it does not rely on assumptions regarding the underlying probability distributions of the RTs. Further inferences are provided by formally instantiating each of the candidate architectures in a parametric RT model. In this model, the representation of each stimulus is instantiated as a multivariate signal detection model (e.g., using General Recognition Theory; Ashby & Townsend, 1986). This representation is used to derive rates of accumulation which are passed to a sequential sampling model (or pair of models for the serial and parallel models) to model the RT. These models, described fully below, allow us to account for correct and error RTs across all of the items simultaneously.

3.2 Current Study

In the present study, we sought to determine the processing architecture underlying categorization decisions about luminance discs of different polarity in different spatial locations (e.g., lighter on the left, darker on the right). Requiring the discs to have different polarity ensured we would be able to manipulate the brightness of each disc independently without participants relying on the overall brightness of the display. However, due to the nature of the discriminability manipulation, it is possible that participants could make categorization decisions based on the overall *contrast* polarity (see Figure 3.3); that is, the difference between the left and right disc luminance is greater for the target category stimuli compared to the contrast category stimuli. Consequently, a correct target category decision could be made based on high contrast polarity and a correct contrast category decision could be made based on a low contrast polarity. As we were interested in how participants made decisions based on individual luminance levels, rather than an overall evaluation of the contrast of the presented stimuli, we included a set of catch trial stimuli. These stimuli have high contrast polarity (now lighter on the right, darker on the left; essentially the reversal of the four target category items) but are associated with a contrast category response. This manipulation ensured participants could not make categorization decisions based on contrast, and instead necessitated that participants use the individual luminance levels of each disc.

We ran two versions of this experiment. In Experiment 1, the screen was divided by a series of boundary discs whose luminance levels were randomly sampled from the possible luminance levels of all discs in the category space. This was done to reduce the probability that items were grouped into a single object. These discs were removed in Experiment 2. Experiment 2, therefore, provides an almost direct replication of Experiment 1 but with a single minor methodological difference.

3.3 Experiments

3.3.1 General Method

Participants

For Experiments 1 and 2, respectively, seven participants from the University of Melbourne community (14 in total, 10 females and four males, aged between 19 and 26 years) with normal or corrected-to-normal vision completed the study. Participants were recruited via advertising placed on notice boards within the Melbourne School of Psychological Sciences and through the school's online recruiting system. All participants were naïve to the purpose of the experiment. Participants provided informed consent and were reimbursed $10 per session, plus an extra $3 bonus for accuracy within a session greater than 90%. The participants from Experiment 1 are referred to as B1 - B7 (B denoting "boundary"), and participants from Experiment 2 are referred to as NB1 - NB7 (NB denoting "no boundary"). Testing was approved by the Melbourne Human Research Ethics Committee (Approval Number 1034866).

As we were interested in the individual-level decision mechanisms, we adopted an expert observer paradigm in which each observer acted as an independent replication of the experiment (see e.g., Normand, 2016; Little & Smith, 2018). Following relevant precedents, we collected a large number of trials for each individual item (N $\approx$ 300) in order to estimate the RT distribution for each item. As demonstrated in P. L. Smith & Little (2018), this approach has considerable advantages over traditional group designs which tend to be underpowered. Our goal is therefore not to estimate a population level parameter but is to test the predictions of each of the models.

Stimuli and Apparatus

Illustrative examples of the stimuli used in Experiments 1 and 2 are shown in Figure 3.3. Stimuli were presented at a monitor resolution of 1280 $\times$ 1024 and participants viewed the screen at a distance of approximately 60 cm. Stimuli

were nine sets of two discs of different luminance levels presented on a gray background (RGB color space values [128 128 128]). The discs subtended a visual angle of 1.91° with centers 11.34° of visual angle to the left and right of fixation (the center of the screen).

For each stimulus, there was a white disc on the left and a black disc on the right. The specific level of luminance was varied. The set of stimuli was created by orthogonally combining the luminance level of the left disc and the luminance level of the right disc. These follow the logical rules design introduced by Fifić et al. (2010), whereby the discriminability manipulation was achieved by varying the luminance level of both discs by three possible increments in comparison to the background [3]. RGB coordinates were as follows for each salience manipulation: High salience black: [64 64 64], mid salience black: [88 88 88], low salience black: [112 112 112], high salience white: [200 200 200], mid salience white: [178 178 178], and low salience white: [156 156 156]. An additional four pairs of discs were added to act as catch trial stimuli. For these discs, the contrast was an orthogonal combination of high or mid salience black and white; however, the left disc was now darker than the background, and the right disc was lighter than the background (i.e., the contrasts were reversed compared to the primary experimental stimuli).

In Experiment 1, the screen was divided by a boundary of 29 discs (also subtended at a visual angle of 1.91° at 60 cm viewing distance) presented as a central vertical column. The luminance values of the boundary discs were randomized from trial to trial using six possible RGB color space values (drawn from the values used to implement the salience manipulation). This boundary was removed in Experiment 2. All other aspects were the same. RTs for categorization were collected using a calibrated RT box (Li et al., 2010).

[3]The luminance values were determined from pilot testing which aimed to find values which resulted in high and low salience manipulations which had the expected ordering of RTs when combined into pairs.

Procedure

For Experiment 1, two participants completed five one-hour sessions of categorization on consecutive or near consecutive days[4]. The remaining participants from Experiments 1 and 2 completed six sessions. At the beginning of the task, participants were shown experimental instructions, as well as an example of the stimuli. Each session consisted of 867 trials (17 practice trials and 850 experimental trials). The contrast category stimuli (nine in total, including the catch trials) were presented five times per block, and the target category items (four in total) were presented ten times per block (i.e., 85 trials per block). This was done to minimize the development of a response bias for contrast category items. All stimuli were presented in randomized order within blocks. In between each block, participants were shown their percent correct on the current block and given the option to take a short break. During each trial, a fixation cross was presented for 1500 ms. A stimulus was then presented and participants were asked to decide whether the stimulus belonged to either category A or category B. Stimuli were presented for 5000 ms or until a response was made. Feedback was presented for incorrect responses. For responses greater than 5000 ms, the feedback "Too Slow" was presented and the trial was removed from the analysis. An example of a single trial for both experiments (with and without boundary discs) is shown in Figure 3.7.

Data Analysis

To analyze the target category, we focused on individual participant ANOVAs, using the interaction effect in a 2×2 factor design to assess the MIC. We used a series of planned t-tests to assess the pattern of contrast category RTs. These analyses follow relevant precedents (Fifić et al., 2010; Little et al., 2011, 2013), and allow us to make inferences about the specific processing patterns for each individual participant. Two shortcomings of this method are apparent. First, each analysis only considers a subset of the data. That is, even though

[4]After the first two participants, we elected to expand the experiment to six sessions.

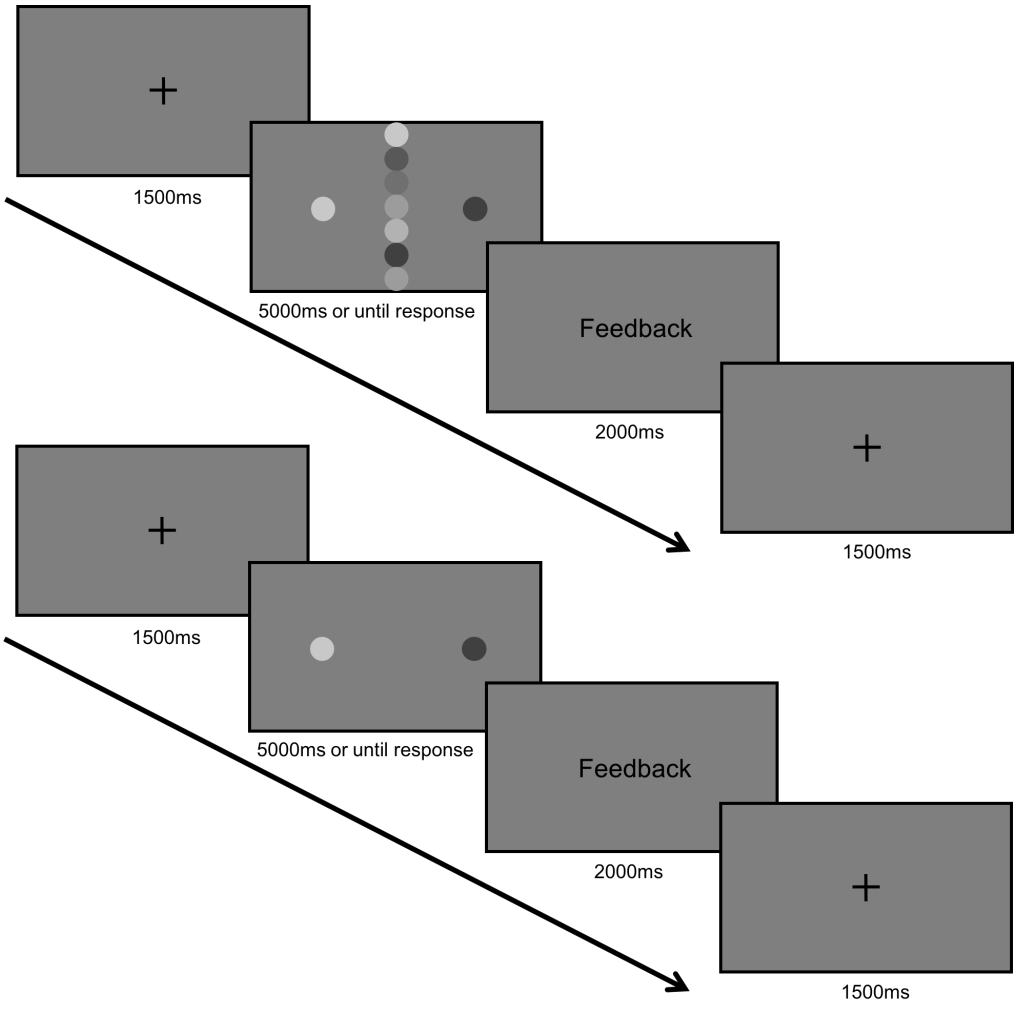

Figure 3.7: First panel: Trial order for Experiment 1. Second panel: Trial order for Experiment 2. The repeated fixation cross indicates the start of the subsequent trial.

the models make predictions across all of the items, an ANOVA across all nine items would be unwieldy and difficult to interpret. Second, each analysis only considers correct RTs. Although accuracy is high for most participants, a more complete analysis would also take into account patterns of error RTs. In order to deal with these issues, we complement our statistical analyses with computational model fitting in which we fit parametric instantiations of each of the models of interest (and relevant extensions) to the correct and error RT distributions for all of the items simultaneously. We then use model selection (i.e., the Deviance Information Criterion, DIC; Gelman et al., 2014) to select the model that provides the best explanation for our data. In summary, our analysis proceeds in two passes. We first use the non-parametric SFT analyses coupled with statistical tests to rule out specific models; we then instantiate the remaining models parametrically and fit them to the data comparing how well each fits the data taking into account the complexity of the model.

3.3.2 Results

Experiment 1

For all participants, the first session was considered practice and was excluded from further analysis. This was done to ensure participants had appropriately learned the categories and had developed a stable categorization strategy. Additionally, RTs less than 200 ms or greater than 3000 ms were excluded. These cut-offs are commonly used in the RT literature (see e.g., Donkin et al., 2011). Less than 1% of trials in total were removed using this method. Mean correct RTs, mean error RTs, and error rates are presented in Table 3.1. Error rates tended to be low across all participants excepting in some cases for the LL, E_X, E_Y, and I_Y stimuli. While the analyses of SFT assume perfect accuracy (Townsend & Nozawa, 1995), Townsend & Wenger (2004) support the robustness of the SIC functions up to an error rate of 30%, which is much higher than that usually seen in studies utilizing the double factorial paradigm. The simulations

of Fifić, Nosofsky, & Townsend (2008) further demonstrate that estimates of the SIC are robust to violations of this assumption.

Table 3.1: Observed Mean Correct and Error RTs (ms), and Error Rates for Individual Stimuli for each Participant in Experiment 1

		Items												
Participant	Variable	HH	HL	LH	LL	EX	IX	EY	IY	R	CHH	CHL	CLH	CLL
B1	RT correct	914.12	1059.6	986.51	1136.2	1075.2	1053.6	1155.9	1191	947.91	747.87	703.29	738.37	804.64
	RT error	982.51	1030.6	1223	1273.8	1202.6	1032.3	1081.6	1225.7	1248.4	*	*	*	*
	p(error)	0.004	0.033	0.041	0.112	0.082	0.057	0.251	0.158	0.004	*	*	*	*
B2	RT correct	618.08	650.13	678.17	716.02	653.44	618	770.04	718.56	589.62	541.74	496.78	530.68	545.46
	RT error	514.77	463.84	834.93	1020.8	1384.1	1131.8	1017.6	1240.8	*	769.58	410.05	836.56	384.74
	p(error)	0.016	0.016	0.034	0.076	0.012	0.024	0.076	0.041	*	0.004	0.008	0.016	0.004
B3	RT correct	509.15	542.44	534.14	574.46	538.19	508.03	548.51	538.91	449.16	445.17	457.87	464.59	452.46
	RT error	522.27	392.8	598.75	548.59	701.73	566.27	1017.4	628.66	689.94	558.7	376.23	*	464.45
	p(error)	0.006	0.014	0.016	0.032	0.012	0.012	0.016	0.02	0.004	0.004	0.004	*	0.004
B4	RT correct	619.74	681.55	673.79	759.13	655.61	651.61	690.3	713.51	580.97	606.12	621.71	646.85	594.5
	RT error	854.57	841.25	1009.6	974.05	962.25	825.5	605.29	695.71	*	640.12	*	*	*
	p(error)	0.004	0.023	0.006	0.027	0.008	0.004	0.037	0.033	*	0.008	*	*	*
B5	RT correct	1004.6	1113.5	1079.6	1200.2	1004.1	1077.5	1197.1	1310.1	957.71	981.65	1020.3	1029.6	1044.9
	RT error	*	715.08	1124.5	1194.6	661.44	*	1442.4	1167.5	*	*	871.24	*	*
	p(error)	*	0.006	0.019	0.017	0.004	*	0.033	0.025	*	*	0.008	*	*
B6	RT correct	538.98	596.49	681.48	793.83	699.7	693.81	785.15	747.92	535.82	469.89	479.81	493.83	508.92
	RT error	467.23	580.77	1253.3	1029	967.27	1008.7	1083	1038.2	*	429.6	415.05	*	*
	p(error)	0.003	0.014	0.034	0.058	0.122	0.061	0.155	0.109	*	<.001	<.001	*	*
B7	RT correct	579.23	644.07	618.43	688.69	620.21	612.14	653.75	605.4	553.83	438.83	443.22	448.57	452.21
	RT error	574.71	689.15	660.49	749.3	581.55	597.31	853.48	786.51	*	*	*	368.26	*
	p(error)	0.008	0.015	0.018	0.051	0.015	0.015	0.101	0.025	*	*	*	0.005	*

Note: * indicates error free performance; B1 = Boundary participant 1; CHH = catch trial stimulus which is high salience black on the left and high salience white on the right.

In order to interpret the SIC functions, it is necessary that the Survivor functions are ordered such that $S_{HH}(t) \leq S_{HL}(t) \approx S_{LH}(t) \leq S_{LL}(t)$ with the strict inequality holding for at least one time point (Townsend & Nozawa, 1995). A series of Kolmogorov-Smirnov (KS) tests (Houpt et al., 2013) were used to check that each participant's survivor functions followed this ordering. If the assumption of stochastic dominance holds, the first four columns of Table 3.11 should be significant, whereas the last four should not. No violations of stochastic dominance were found for this experiment (see Appendix A, Table 3.11). Survivor functions are shown in Appendix A, Figure 3.15.

Target Category

Figure 3.8 shows the mean RTs and corresponding MICs. All the MICs are near zero but in the positive direction.

To analyze the target category RTs, we conducted a series of 5 (sessions: 2-6) $\times$ 2 (left disc: L or H) $\times$ 2 (right disc: L or H) ANOVAs on the Target Category RTs for each individual participant (see Table 3.2)[5].

[5]As two participants completed only five sessions a 4 $\times$ 2 $\times$ 2 ANOVA was used for these participants. In the case of participant B6, the initial ANOVA indicated a significant three-way interaction effect, suggesting that the processing architecture may have changed across sessions.

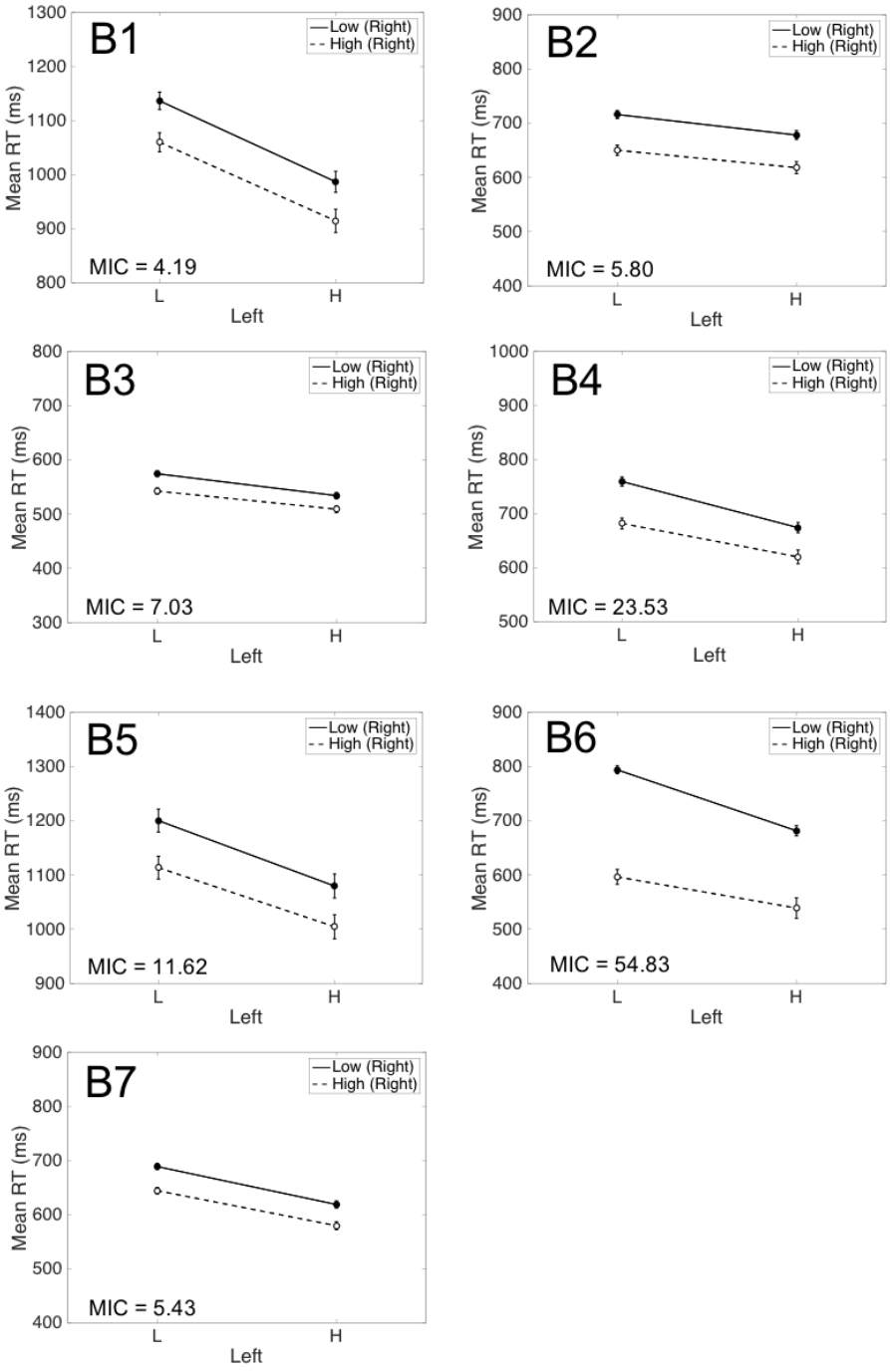

Figure 3.8: Observed target category mean RTs and MICs for individual participants in Experiment 1. The two left-hand points represent low discriminability on the left disc and the two right-hand points represent high discriminability on the left disc. The solid line represents low discriminability on the right disc, and the dotted line represents high discriminability on the right disc. Error bars represent one standard error.

We first summarize the results which were common across all or most participants:

1. There was a main effect of session, indicating that RTs become faster over the course of the experiment.

2. There was a significant main effect of disc discriminability for both discs indicating that the discriminability manipulation was effective.

3. For some participants, session interacted with one or both of the dimensions indicating for some sessions the left disc was processed faster than the right and vice versa.

4. The three-way interaction was not significant, indicating a stable relationship between target category items across sessions. That is, participants were not changing processing strategy from session to session.

5. With the exception of participant B6, the Left Disc × Right Disc interaction (see Figure 3.8) was not significantly different from zero. Although the MIC was positive, a non-significant interaction is consistent with serial processing. The test of the interaction in the present case is a test of the *point prediction* of the serial model, which predicts that the interaction term should equal zero (cf. Sternberg, 1969). This presents a different goal to the typical null hypothesis significance testing case, where the goal of the significance cut-off is to place some criteria on the false positive rate. In the present case, an alpha criterion of .05 is biased toward the serial model (Fox & Houpt, 2016). Consequently, caution must be taken when interpreting a non-significant result in this context.

Figure 3.9 shows the SICs. When considering the model predictions presented in Figure 5.4, it can be seen that in all cases the SICs have a large positive portion, ruling out parallel processing for target category stimuli. Generally, the

After excluding the first two sessions, the three-way interaction was no longer significant, suggesting a stable strategy after the first two sessions. Hence, we removed the second session as well, and a 3 × 2 × 2 ANOVA was conducted for this participant

Table 3.2: Target Category Statistical Results for Individual Participants in Experiment 1

Variable	*df*	*F*	p	*df*	*F*	p	*df*	*F*	p	*df*	*F*	p
	B1			B2			B3			B4		
Session	4	17.28	<.001	4	18.31	<.001	4	47.95	<.001	4	11.18	<.001
Left	1	16.31	<.001	1	49.15	<.001	1	27.92	<.001	1	41.98	<.001
Right	1	66.56	<.001	1	15.48	<.001	1	44.07	<.001	1	52.9	<.001
Session x L	4	1.46	0.211	4	0.36	0.840	4	3.11	0.015	4	0.63	0.640
Session x R	4	4.55	0.001	4	0.31	0.868	4	0.63	0.644	4	2.11	0.078
Left x Right	1	0.05	0.830	1	0.17	0.683	1	0.46	0.499	1	1.33	0.250
Sess x L x R	4	1.92	0.105	4	0.26	0.903	4	0.19	0.945	4	1.37	0.243
Error	1838			1864			1909			1908		
	B5			B6			B7					
Session	4	85.60	<.001	2	151.35	<.001	3	8.44	<.001			
Left	1	15.92	<.001	1	237.52	<.001	1	42.2	<.001			
Right	1	32	<.001	1	57.41	<.001	1	109.12	<.001			
Session x L	4	2.57	0.040	2	21.23	<.001	3	0.34	0.793			
Session x R	4	6.36	<.001	2	5.02	0.007	3	0.67	0.57			
Left x Right	1	0.03	0.854	1	6.17	0.013	1	0.16	0.686			
Sess x L x R	4	0.67	0.614	2	1.50	0.225	3	0.25	0.862			
Error	1889			1132			1526					

SICs also appear to have a greater positive region than negative region. This is most consistent with a coactive pattern of results.

Using two one-sided KS-Tests from Houpt's (2013) SFT analysis package, we also sought to determine whether the positive and negative portions of the SICs were significantly different to zero. Two null-hypothesis tests were performed: one which determines whether the largest value of the SIC is significantly greater than zero (D+) and one which determines whether the lowest value of the SIC is significantly lower than zero (D-; see Houpt & Townsend, 2010). Like the MIC, the null hypothesis for the Houpt-Townsend statistic is $SIC(t) = 0$ for all times t, a conservative significance level biases the test toward retaining the null hypothesis (i.e., a serial model). We therefore adopted a less conservative cut-off of $\alpha = .33$. This value has been shown to work well in model recovery tests using this statistic (Fox & Houpt, 2016). Both positive and negative D-tests are displayed in Table 3.3.

For most participants, the positive deflection in the SIC is significantly greater than zero, whereas the negative deflection was not. This provides support for the coactive processing architecture. For B3, however, both positive and negative values were significant, which is indicative of either serial processing or coactive processing. For B5 neither value was significant.

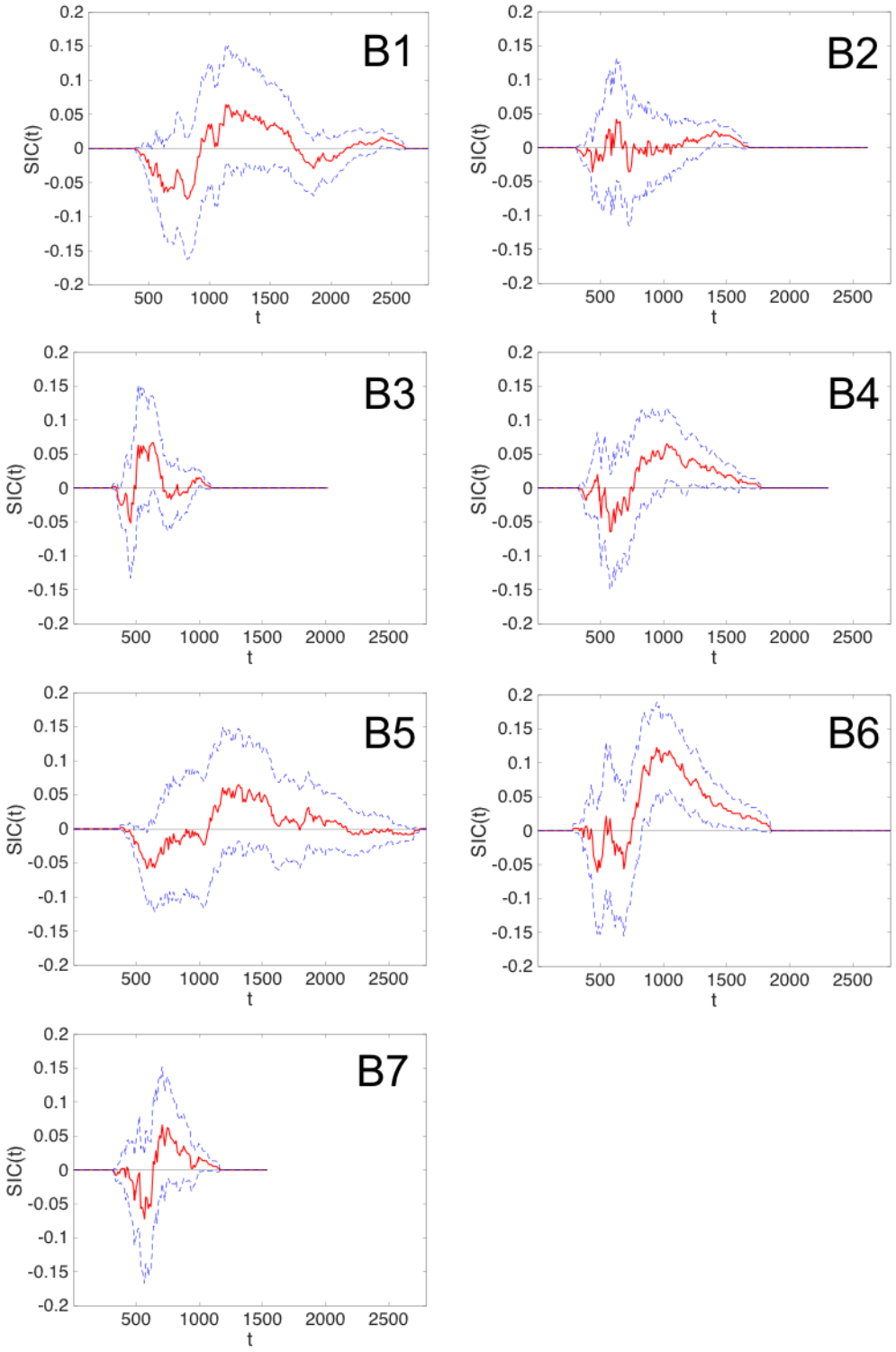

Figure 3.9: Observed target category SICs (red line) for individual participants in Experiment 1. Blue lines represent 95% bootstrapped confidence intervals.

Table 3.3: Directional KS-tests for individual participants in Experiment 1

Participant	*D+*	*p*	*D-*	*p*
B1	0.073	0.270*	0.06	0.419
B2	0.079	0.216*	0.023	0.882
B3	0.070	0.300*	0.081	0.198*
B4	0.069	0.312*	0.063	0.378
B5	0.067	0.334	0.060	0.423
B6	0.117	0.140*	0.078	0.407
B7	0.076	0.319*	0.074	0.341

Note: * indicates a significant difference with an alpha level of .33. D+ tests whether the SIC is significantly greater than zero. D- tests whether the SIC is significantly lower than zero.

Contrast Category

The mean RTs for the contrast category are displayed in Figure 3.10. For the majority of participants the interior stimulus was faster than the exterior stimulus, on at least one of the dimensions, which is suggestive of coactivity. No other model predicts a faster interior item compared to an exterior item on any dimension. Consequently, coactive processing provides a potential explanation for the contrast category items for most of the participants. Nonetheless, participant B1 despite having a faster interior compared to exterior item on one dimension, shows the reverse pattern on the other, which could be indicative of serial processing. For B4, the RTs seem more consistent with a fixed-order serial model, with the interior item being slower on one dimension, and the exterior and interior being approximately equal on the other. Further, for B5 both interior items were slower than exterior items, suggesting mixed-order serial self-terminating processing.

For contrast category items, we conducted a series of planned t-tests comparing interior and exterior items on both dimensions and comparing the redundant stimulus to the other contrast category items (see Table 3.4). Except for two instances, the redundant stimulus was processed significantly faster than the other stimuli, providing evidence against an exhaustive stopping rule for the contrast category[6].

Although a smaller mean RT was recorded for the interior items compared to the exterior items for the majority of participants, as expected under coactive

[6]The left interior stimulus vs redundant stimulus comparison for B2 and the left exterior stimulus vs redundant stimulus comparison for B5 were not significant.

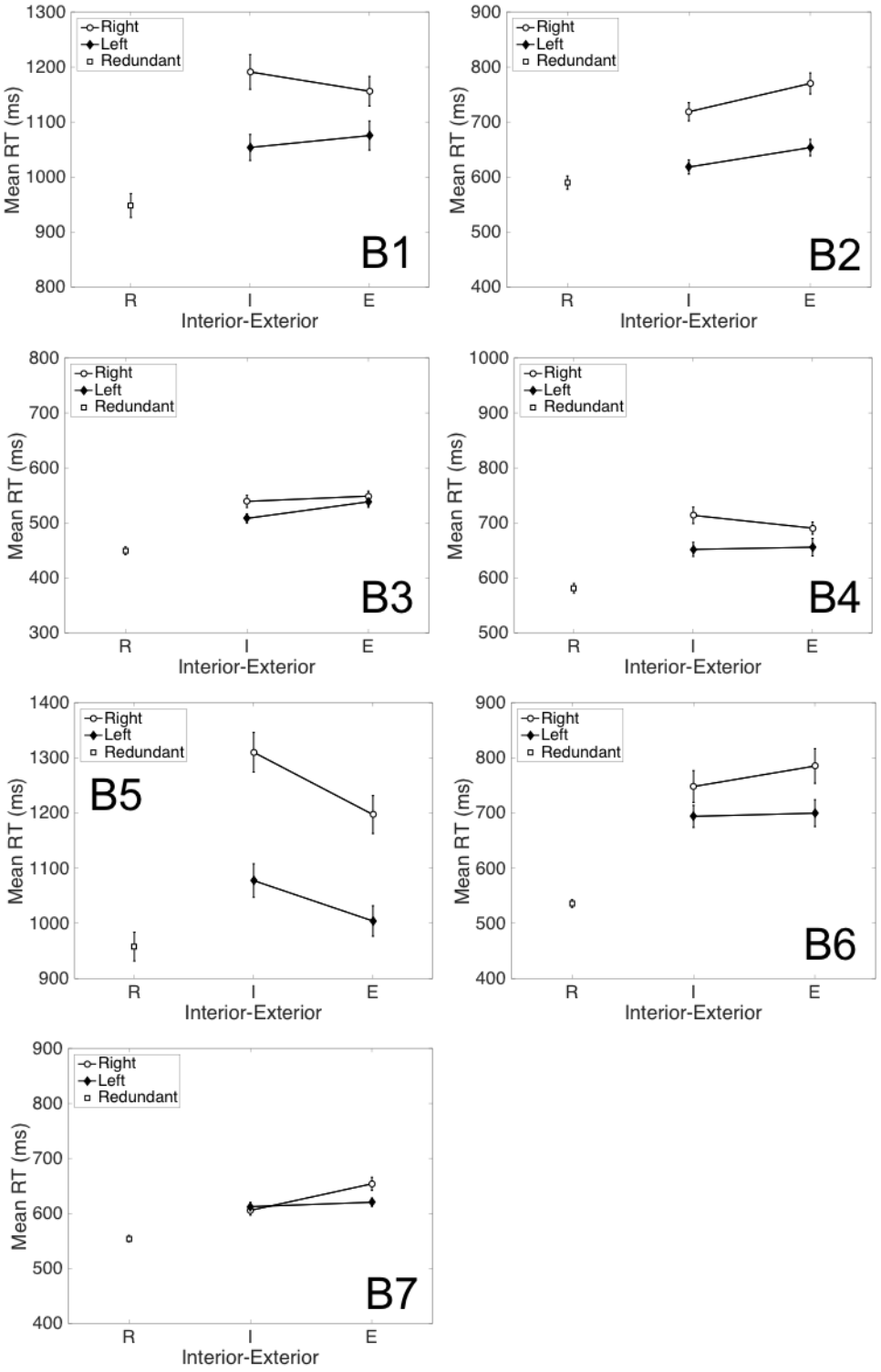

Figure 3.10: Observed contrast category mean RTs for individual participants in Experiment 1. Error bars represent one standard error. R = redundant stimulus, I = interior stimulus, E = exterior stimulus.

processing, this pattern was only significant for B2 and B7 for the right dimension and B3 for the left dimension. Nonetheless, this pattern of faster interior than exterior items is not predicted by any other model. However, parallel processing cannot be ruled out since that model predicts no significant difference between the interior and exterior items. For B4 and B5, the exterior items were faster than the interior items, suggesting serial processing. However, again, this pattern was only significant for B5 on the right dimension.

Table 3.4: Contrast Category Statistical Results for Individual Participants in Experiment 1

Stimulus Pair	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*
	B1				B2				B3				B4			
ELeft - ILeft	21.64	0.61	451	0.542	35.44	1.80	479	0.073	30.15	2.31	482	0.021	4.00	0.20	481	0.845
ERight - IRight	-35.10	-0.84	391	0.404	51.48	2.04	464	0.042	9.60	0.67	481	0.500	-23.21	-1.24	469	0.215
ELeft - R	127.33	3.74	461	<.001	63.82	3.31	481	0.001	89.03	7.28	485	<.001	74.64	4.14	485	<.001
ILeft - R	105.69	3.29	468	0.001	28.38	1.61	478	0.108	58.88	5.40	483	<.001	70.64	4.58	488	<.001
ERight - R	208.02	6.07	423	<.001	180.41	8.04	469	<.001	99.36	8.75	483	<.001	109.33	7.79	481	<.001
IRight - R	243.12	6.47	446	<.001	128.94	6.30	475	<.001	89.76	6.85	484	<.001	132.54	7.71	480	<.001

Stimulus Pair	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*
	B5				B6				B7			
ELeft - ILeft	-73.33	-1.78	478	0.075	5.89	0.19	266	0.851	8.07	0.75	384	0.455
ERight - IRight	-112.96	-2.27	467	0.024	37.23	0.87	254	0.383	48.35	3.30	369	0.001
ELeft - R	46.43	1.22	472	0.222	163.88	6.86	275	<.001	66.38	6.95	389	<.001
ILeft - R	119.75	2.98	478	0.003	157.99	7.62	285	<.001	58.31	5.91	387	<.001
ERight - R	239.40	5.57	466	<.001	249.33	8.37	271	<.001	99.92	7.70	374	<.001
IRight - R	352.36	7.94	473	<.001	212.10	7.53	277	<.001	51.57	4.86	387	<.001

3.3.3 Discussion

Taken together, the target category results for the majority of participants tend towards coactivity. While the MICs are positive for all participants except B5, the non-significant interaction between left and right dimensions in the target category is consistent with serial processing for all participants. When coupled with the SICs, however, a clearer pattern of coactivity emerges. First, the SICs all have a greater positive portion than negative portion which rules out parallel processing. Further, the positive portion of the SIC appears greater than the negative portion, suggesting coactive, rather than serial processing. This interpretation is supported by the directional KS-tests for most participants (excluding B3 and B5). Nonetheless, the target category data do not clearly rule-out serial processing.

Generally, the contrast category results also somewhat point to coactivity with

the interior item being faster than the exterior item at least on one dimension for four of the seven participants. For three participants (B1, B4 and B5), an interior item was slower than an exterior item, which could indicate serial processing (although B1 also shows the opposite pattern on the other dimension, which is more indicative of coactive processing). We defer further discussion of these results until after the presentation of Experiment 2.

3.3.4 Experiment 2

In Experiment 2, we removed the column of discs with randomly varying luminance that separated the left and right targets. Again, for all participants, the first session was considered practice and was excluded from further analysis. Additionally, RTs less than 200 ms or greater than 3000 ms were excluded. Less than 1% of trials in total were removed using this method. Mean correct RTs, mean error RTs, and error rates are presented in Table 4.3. Error rates tended to be low across all participants except for the LL and E_Y stimuli in some cases.

Table 3.5: Observed Mean Correct and Error RT (ms) and Error Rate for Individual Stimuli for Each Participant in Experiment 2

		Items												
Participant	Variable	HH	HL	LH	LL	EX	IX	EY	IY	R	CHH	CHL	CLLH	CLLL
NB1	RT correct	740.49	861.43	784.64	935.58	935.67	907.25	937.47	874.33	806.23	632.97	628.12	633.25	631.79
	RT error	728.76	622.56	938.09	1153.5	1035.8	684.4	935.9	1085	*	500	449.96	964.24	*
	p(error)	0.008	0.006	0.021	0.012	0.037	0.004	0.041	0.033	*	0.004	0.004	0.004	*
NB2	RT correct	594.99	635.64	640.03	719.75	717.56	726.74	900.22	709.05	558.42	464.34	451.73	477.62	480.25
	RT error	439.95	903.1	769.94	875.98	1001	1083.3	1188	1034	*	475.66	*	1666.7	1401.1
	p(error)	0.022	0.030	0.04	0.126	0.028	0.024	0.284	0.081	*	0.004	*	0.004	0.004
NB3	RT correct	687.46	698.89	706.04	790.53	797.15	759.14	824.41	780.84	676.54	671.82	684.33	694.02	700.09
	RT error	880.57	*	1363.20	1233.90	980.04	578.05	1224.60	852.62	*	*	*	*	1156.50
	p(error)	0.004	*	0.002	0.012	0.012	0.004	0.064	0.004	*	*	*	*	0.004
NB4	RT correct	697.95	724.16	726.24	809.64	844.63	850.66	735.48	805.03	687.01	694.07	684.39	720.82	701.15
	RT error	585.16	*	1085.10	1036.10	918.16	819.77	1958.2	1153.00	*	*	659.86	*	*
	p(error)	0.002	*	0.012	0.014	0.033	0.008	0.008	0.020	*	*	0.004	*	*
NB5	RT correct	784.01	853.99	873.08	928.82	886.01	806.02	917.59	768.38	628.53	626.35	641.94	653.49	678.92
	RT error	1572.4	956.1	1276.3	1342.1	1634.8	1239.9	1407.7	1010.7	739.94	518.28	561.4	*	1825.1
	p(error)	0.010	0.016	0.029	0.100	0.078	0.033	0.065	0.061	0.008	0.008	0.004	*	0.004
NB6	RT correct	647.66	752.24	701.59	812.44	494.56	476.38	657.41	720.21	505.82	526.86	536.71	558.59	542.4
	RT error	596.39	848.22	772.93	664.62	*	*	946.48	1251.50	*	*	1640.70	*	*
	p(error)	0.010	0.027	0.025	0.056	*	*	0.110	0.074	*	*	0.004	*	*
NB7	RT correct	698.78	751.91	850.35	827.91	950.92	954.24	943.63	914.95	732.31	689.45	678.72	656.07	655.54
	RT error	1295.40	888.18	1323.50	1157.30	1541.80	1304.70	1031.10	1113.80	1025.10	389.66	331.35	947.02	1083.50
	p(error)	0.008	0.014	0.0371	0.054	0.086	0.045	0.062	0.045	0.008	0.004	0.004	0.008	0.012

Note: * indicates error free performance; NB1 = No boundary participant 1; CHH = catch trial stimulus which is high salience black on the left and high salience white on the right.

A series of KS-tests (Houpt et al., 2013) were used to check that each participant's survivor functions were ordered to allow interpretation of the SFT analyses (Townsend & Nozawa, 1995, see Table 3.6). The survivor functions are shown in Appendix B, Table 3.16. Although the assumption held for most participants, there are some notable violations. Namely, for NB3, the HH

stimulus is not significantly higher than the HL and LH stimuli, and for NB7, the LH stimulus is not significantly greater than the LL stimulus. Consequently, we omitted these participants from further analysis.

Table 3.6: P-Values from KS-tests of stochastic dominance in Experiment 2. Violations of stochastic dominance are indicated in bold.

Participant	Dominance Test for Selective Influence							
	HH > HL	HH > LH	HL > LL	LH > LL	HH < HL	HH < LH	HL < LL	LH < LL
NB1	<.001	0.017	<.001	<.001	1.000	0.998	0.967	1.000
NB2	0.003	<.001	<.001	<.001	0.992	0.968	0.998	0.998
NB3	**0.264**	**0.115**	<.001	<.001	0.739	0.751	0.982	0.951
NB4	<.001	0.004	<.001	<.001	0.905	0.930	0.998	1.000
NB5	<.001	<.001	<.001	<.001	0.935	0.999	1.000	9.320
NB6	<.001	<.001	<.001	<.001	1.000	0.812	0.966	0.967
NB7	**0.060**	<.001	<.001	**0.393**	0.260	0.968	0.980	0.233

Target Category

Figure 3.11 shows the mean RTs and corresponding MICs. Except for NB5, all the MICs are positive which is indicative of coactive processing.

To analyze the target category RTs, we again conducted a series of 6 (sessions: 2-6) × 2 (left disc: L or H) × 2 (right disc: L or H) ANOVAs on the Target Category RTs for each individual participant (see Table 3.7).

Table 3.7: Target Category Statistical Results for Individual Participants in Experiment 2.

Variable	*df*	*F*	p	*df*	*F*	p	*df*	*F*	p
	NB1			NB2			NB4		
Session	4	33.61	<.001	4	15.69	<.001	4	25.65	<.001
Left	1	23.26	<.001	1	56.26	<.001	1	28.97	<.001
Right	1	124.46	<.001	1	48.34	<.001	1	27.01	<.001
Session x L	4	0.08	0.989	4	1.89	0.109	4	0.96	0.430
Session x R	4	3.42	0.009	4	0.74	0.567	4	2.74	0.027
Left x Right	1	1.39	0.238	1	5.24	0.022	1	7.59	0.006
Sess x L x R	4	0.30	0.879	4	1.93	0.104	4	0.23	0.923
Error	1908			1845			1925		
	NB5			NB6					
Session	4	22.79	<.001	4	27.53	<.001			
Left	1	23.30	<.001	1	22.41	<.001			
Right	1	14.19	<.001	1	81.52	<.001			
Session x L	4	2.16	0.071	4	0.31	0.872			
Session x R	4	2.96	0.019	4	4.71	0.001			
Left x Right	1	0.06	0.806	1	0.06	0.810			
Sess x L x R	4	0.84	0.500	4	0.15	0.963			
Error	1857			1875					

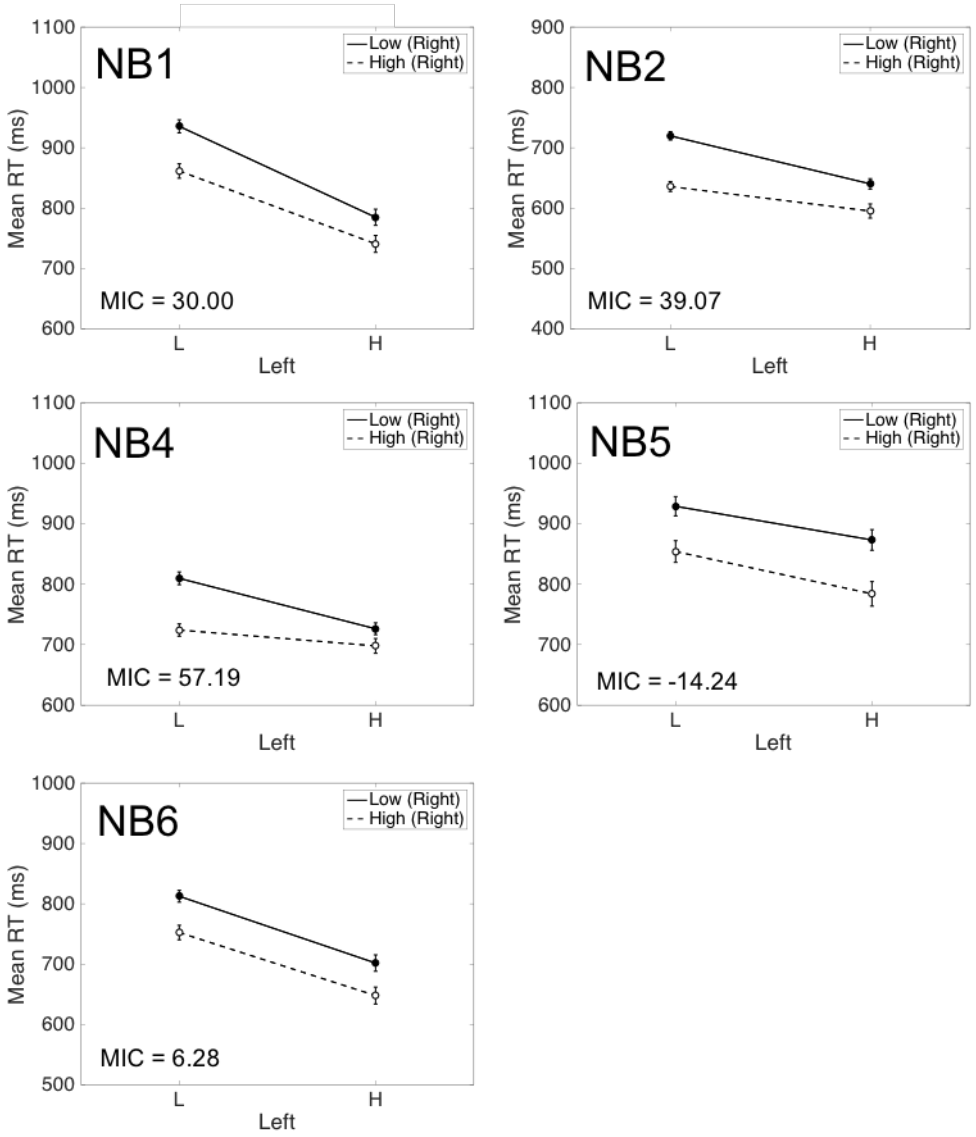

Figure 3.11: Observed target category mean RTs and MICs for individual participants in Experiment 2. The two left-hand points represent low discriminability on the left disc and the two right-hand points represent high discriminability on the left disc. The solid line represents low discriminability on the right disc, and the dotted line represents high discriminability on the right disc. Error bars represent standard error.

For all participants the results indicated that:

1. There was a main effect of session, indicating RTs became faster across sessions.

2. There was a significant main effect of disc discriminability for both discs across all participants.

3. For some participants, session interacted with one or both of the dimensions indicating for some sessions the left disc was processed faster than the right and vice versa.

4. The three-way interaction was not significant, indicating a stable relationship between target category items across sessions.

5. The Left $\times$ Right interaction was significant for participants NB2, NB4 indicating that the MIC was significantly positive, which supports the inference of coactivity. For all other participants, the non-significant interaction is consistent with serial processing, although again caution must be exercised when interpreting this result (Fox & Houpt, 2016) .

Figure 3.12 shows the SICs. When referenced to the model predictions shown in Figure 5.4, the majority of the SICs in Experiment 2 have a large positive area, ruling out parallel processing for the target category stimuli. Further, for these participants, the SICs appear to have a greater positive region than negative region. This is most consistent with a coactive pattern of results.

Using two one-sided KS-Tests from Houpt et al. (2013)'s SFT analysis package, we also sought to determine whether the positive and negative portions of the SICs were significantly different to zero. We again adopted a less conservative cut-off of $\alpha = .33$ to avoid bias toward the serial model; most participants showed significant differences at this level. Both positive and negative D-tests for Experiment 2 are displayed in Table 3.8.

For most participants, the positive deflection was significantly higher than zero, but the negative deflection was not. This provides further support for

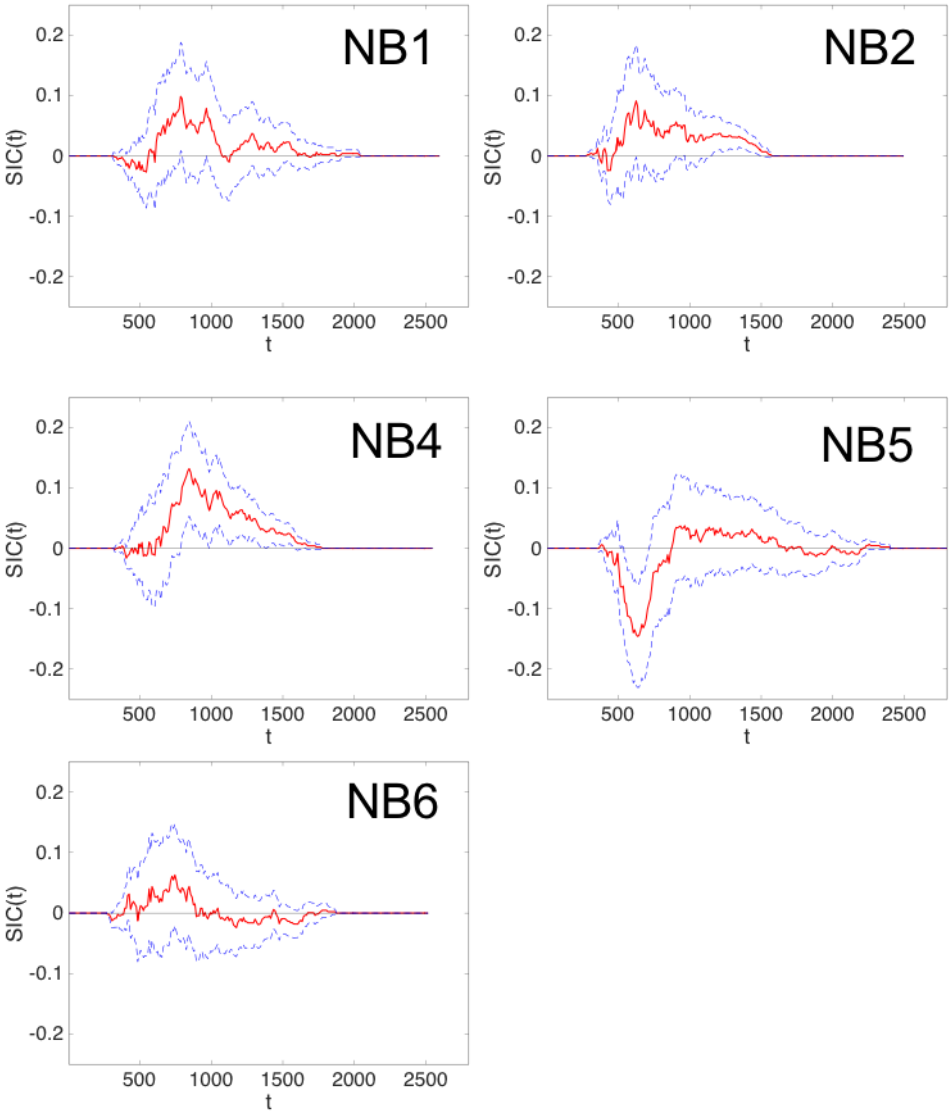

Figure 3.12: Observed target category SICs (red line) for individual participants in Experiment 2. Blue lines represent 95% bootstrapped confidence intervals.

Table 3.8: Directional KS-tests for individual participants in Experiment 2

Participant	*D+*	*p*	*D-*	*p*
NB1	0.101	0.083*	0.026	0.849
NB2	0.109	0.053*	0.040	0.677
NB4	0.131	0.015*	0.026	0.847
NB5	0.068	0.323*	0.138	0.009*
NB6	0.045	0.604	0.039	0.687

Note: * indicates a significant difference with an alpha level of .33. D+ tests whether the SIC is significantly greater than zero. D- tests whether the SIC is significantly lower than zero.

coactive processing. For NB5, however, both deflections were significant, suggesting serial processing or coactive processing. For NB6 neither the positive nor negative deflection was significant.

Contrast Category

The mean RTs for the contrast category are displayed in Figure 3.13. For most participants, the interior stimulus is faster than the exterior stimulus on at least one dimension, which supports an inference of coactivity. For NB4 and NB6, the contrast category RTs seem more consistent with a fixed-order serial model, with the interior item being slower on one dimension, and the exterior and interior being approximately equal on the other.

For contrast category items, we conducted a series of planned t-tests comparing interior and exterior items on both dimensions, and comparing the redundant stimulus to the other contrast category items (see Table 3.9). With the exception of one comparison[7], the redundant stimulus was processed significantly faster than the other stimuli providing evidence against an exhaustive model for the contrast category.

The interior item was significantly faster than the exterior item on the left dimension for NB5, and on the right dimension for NB1, NB2, and NB5 which is suggestive of coactive processing. For NB4 and NB6, the exterior item was significantly faster than the interior item for the right dimension which is indicative of serial processing for these participants.

3.3.5 Discussion

As was the case with Experiment 1, the target and contrast category results for Experiment 2 support a tentative inference of coactivity for most participants. However, there were differences across participants that make it difficult to clearly infer the architecture based on the non-parametric results. Although

[7]For NB6 the redundant stimulus was significantly *slower* than the left interior item. There was no significant difference in RT between the left exterior item and the redundant stimulus.

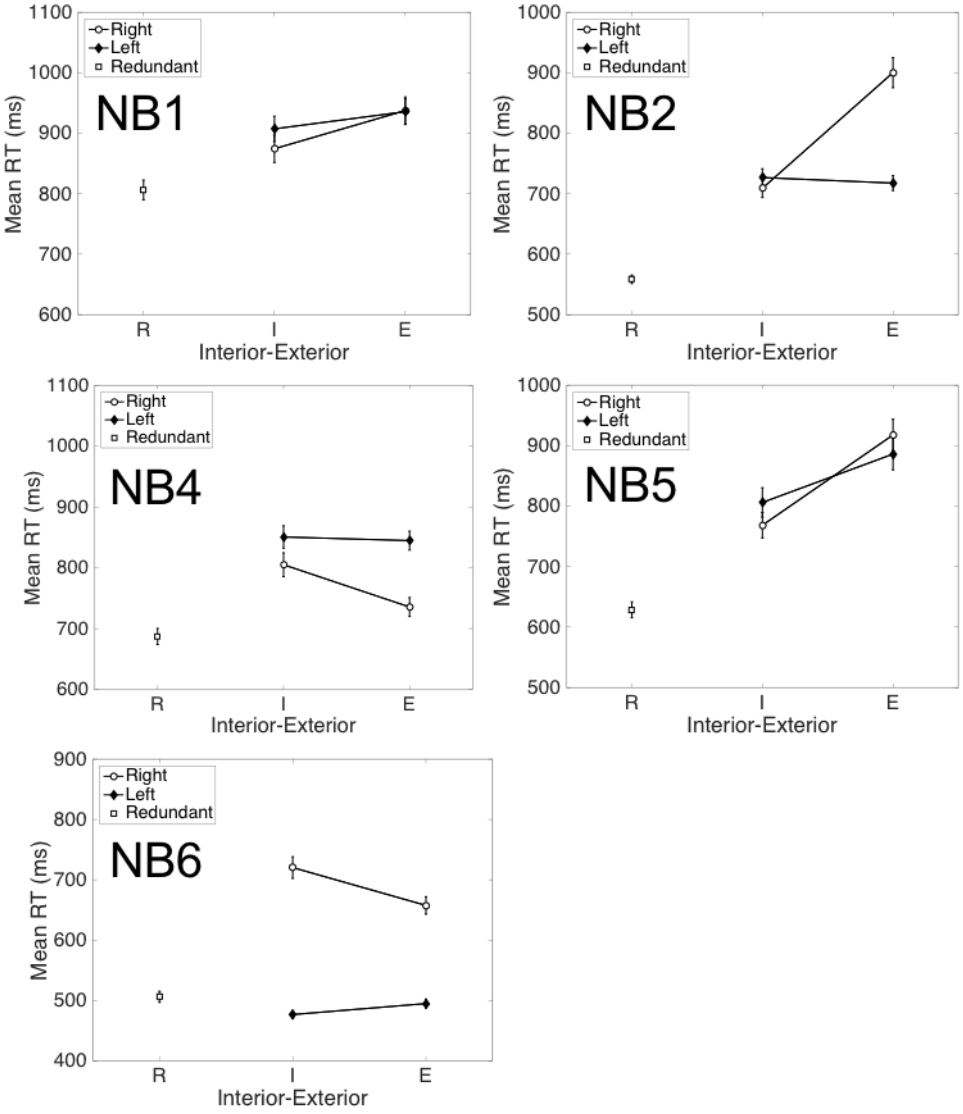

Figure 3.13: Observed contrast category mean RTs for individual participants in Experiment 2. Error bars represent standard error. R = redundant stimulus, I = interior stimulus, E = exterior stimulus.

Table 3.9: Contrast Category Statistical Results for Individual Participants in Experiment 2.

Stimulus Pair	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*
	NB1				NB2				NB4			
ELeft - ILeft	28.41	0.95	474	0.341	-9.19	-0.48	477	0.630	-6.03	-0.25	480	0.803
ERight - IRight	63.14	1.97	468	0.049	191.17	6.85	405	<.001	-69.55	-2.78	480	0.006
ELeft - R	129.44	4.86	479	<.001	159.14	11.34	484	<.001	157.62	7.72	479	<.001
ILeft - R	101.02	3.82	485	<.001	168.32	10.52	485	<.001	163.65	7.20	485	<.001
ERight - R	131.24	4.75	479	<.001	341.80	15.06	424	<.001	48.47	2.37	482	0.018
IRight - R	68.10	2.45	479	0.015	150.63	9.22	473	<.001	118.03	5.00	482	<.001

Stimulus Pair	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*
	NB5				NB6			
ELeft - ILeft	79.99	2.25	460	0.025	18.18	1.83	485	0.068
ERight - IRight	149.22	4.46	457	<.001	-62.80	-2.72	442	0.007
ELeft - R	257.48	8.99	462	<.001	-11.26	-0.96	486	0.339
ILeft - R	177.49	6.44	474	<.001	-29.44	-2.60	485	0.010
ERight - R	289.06	10.04	466	<.001	151.58	9.06	460	<.001
IRight - R	139.85	5.67	467	<.001	214.39	10.90	468	<.001

the SICs, for the most part tended to be positive and looked like the coactive prediction, the MICs were typically not significant. However, as indicated above, the typical NHST cutoff of .05 is biased toward the serial model, and this result should be interpreted cautiously. For the contrast category, there was substantial variability in the pattern of mean RTs, although we note that the predictions shown in Figure 3.4 are only illustrative based on simulations from one set of model parameters. In general, however, a coactive model is the only model under consideration that can predict faster interior than exterior processing (Little et al., 2015; Little & Smith, 2018)[8]. Nevertheless, the nonparametric analyses only take into account a subset of the data. They do not, for instance, account for error rates or error RT distributions, nor do they consider data from both the target category and the contrast category simultaneously. The SFT analyses further require meeting an assumption of stochastic dominance of the target category RTs. For these reasons, we also fit a set of computational models that parametrically instantiate our assumptions for the serial self-terminating, parallel self-terminating, and coactive models.

[8]This prediction relies on the assumption that the processing of each dimension is equivalent across levels of the other dimension. If this assumption is violated then other models may also make this prediction (Cheng et al., 2017).

Computational Modeling

We fit the models using Differential Evolution Markov Chain Monte Carlo (Turner et al., 2013). We then compared the models using the Deviance Information Criterion (DIC) which provides an estimate of model fit with a penalty for model complexity (Gelman et al., 2014). The DIC can be thought of as a Bayesian version of a maximum likelihood-based fit statistic like the Akaike Information Criteria (AIC; Akaike, 1974). A detailed discussion on the implementation of the computational models and calculation of the DIC is reported in Appendix C. As the contrast category RTs conclusively rule out exhaustive processing, we focused on fitting the self-terminating version of both serial and parallel models as well as the coactive model. We included the parallel model since, although this model is not supported by the SIC analyses, it cannot be ruled out on the basis of the contrast category results. Because the models do not rely on the assumption of stochastic dominance, we also fit the model to observers NB3 and NB5, who were omitted in our previous analyses. The computational modeling thus provides the strongest test of architecture since it utilizes all of the data including both correct and error RTs across all items simultaneously.

To summarize the models, the serial and parallel models are based on the assumption that observers make independent decisions about stimulus values along each dimension and then combine these using logical operations such as OR and AND; hence, we term these models the logical rule models. Decisions about the values of each of the discs are modeled as independent evidence accumulation processes. In the past, we have used random walk processes to model the RTs (Fifić et al., 2010; Luce, 1986; Ratcliff, 1978). Here we utilize the linear ballistic accumulator (LBA; Brown & Heathcote, 2008). To generate the drift rate for each of the LBA channels, we used GRT (Ashby & Townsend, 1986), which is a multivariate generalization of signal detection theory, and decision-bound theory (Ashby & Gott, 1988). Each stimulus is represented by a bivariate normal distribution representing the variability in the perception of the stimulus from moment to moment. The assumption is that, from moment

to moment, samples are drawn from this distribution and used to drive the sequential sampling process. When a sample falls in the target category region, the evidence increases in the accumulator for the target category, and likewise for the contrast category. The LBA approximates this process and provides an efficient method for predicting the decision time for each disc. Following Cheng et al. (2018) we generate drift rates for the LBA by integrating the perceptual distributions with respect to the decision boundary within each category region. The decision times for each accumulator are then either, for example, summed for the serial model or used to find the maximum time prediction for the parallel model. For the coactive model, rather than modeling the perceptual distributions independently, the variability of the perception of both discs is modeled as a bivariate normal distribution.

In the coactive model, as for the serial and parallel models, we assume that the means for the left and the right discs are aligned to a grid (as shown, e.g., in Figure 1). That is, we assume that the technical property of perceptual separability holds (Ashby & Townsend, 1986). To minimize reliance on this assumption, we also fit a highly flexible model in which we freely estimated a drift rate for each stimulus. This model can, for instance, capture coactive patterns which are accompanied by violations of perceptual separability. However, the free drift model should incur a high penalty for complexity. The free drift model is therefore an important comparison model because it includes many existing single-channel categorization models as special cases including the Exemplar-Based Random Walk model (Nosofsky & Palmeri, 1997) and stochastic GRT (Ashby, 2000). The DICs for each individual participant and model are shown in Table 3.10. The preferred model is the model with the lowest DIC.

In both experiments, the coactive model or the more flexible free drift model provided the best fit for most participants. This supports the general conclusion that processing of both discs tended to be pooled into a single coactive channel. The model-based analysis of each individual also provides some characterization of the differences in processing between each individual. Observers B4, NB5, and

Table 3.10: DIC values for each individual participant and candidate model across both Experiments 1 and 2. Lower values indicate better fit.

	Model			
Subject	Serial ST	Parallel ST	Coactive	Free Drift
B1	4071.5	4151.5	**3894.7**	3922.7
B2	-971.27	-1076.6	**-1113.7**	-1054.8
B3	-4245	-4289.4	**-4362.9**	-4307.5
B4	-1340.7	**-1517.5**	-1483.9	-1438.4
B5	3948.6	4035.1	**3796.2**	3826.8
B6	-258.99	-335.52	-388.92	**-396.36**
B7	-2772.1	-2832.3	-2810.3	**-2845.5**
NB1	1146.6	1169.8	**1111.7**	1162.4
NB2	98.664	-46.4	-142.7	**-197.78**
NB3	-161.47	-313.41	**-339.39**	-319.86
NB4	-391.83	-436.68	**-468.81**	-453.93
NB5	2558	**2417.3**	2468.9	2489.9
NB6	**-342.17**	-316.69	-323.22	-315.85
NB7	2513.7	**2386.4**	2424.9	2492.9

Note: Best fitting models indicated in bold

NB7 were best fit by the parallel self-terminating model while participant NB6 was best fit by the serial self-terminating model. Along with model comparison estimates, it is essential to provide an estimate of how well a model fits the data to ensure that the model is capturing the data well (Heathcote et al., 2015). In the interest of space, we present the posterior predictions of the coactive model and for the best fitting models (when this was not the coactive model) for each item along with posterior parameter estimates of the parameters in our supplementary material. In Figure 3.14, we show the model predictions for one participant (observer B2) to show that the model does provide a good fit to the data. The supplementary material shows that the fits to the other participants are comparable.

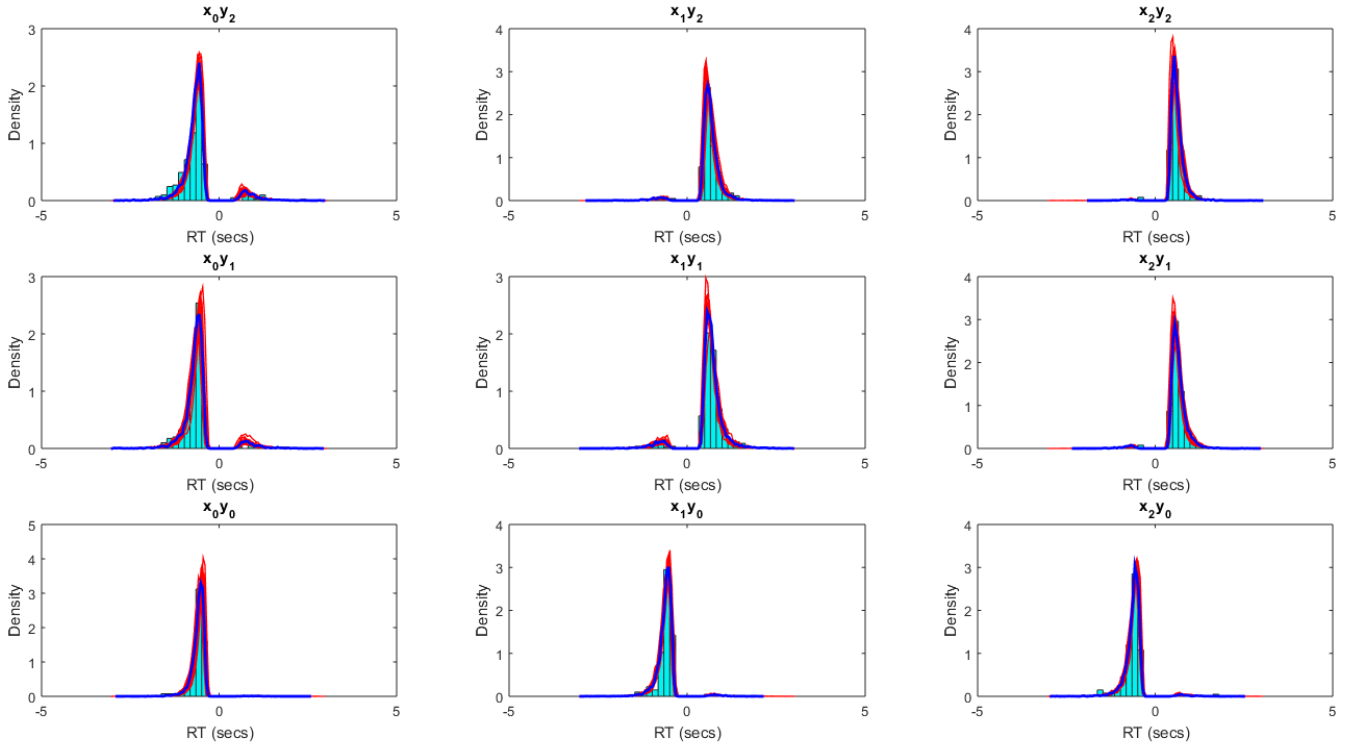

Figure 3.14: Posterior predictions from the coactive model for Observer B2 from Experiment 1. The data are shown as bars. The solid blue line is the posterior mean prediction, and the red lines are draws from the posterior predictive distribution.

3.4 General Discussion

In this paper, we have shown that decisions about spatially separated luminance discs appear to be processed coactively for most individuals (10 of 14). This pattern of coactivity was consistent with non-parametric analyses of MIC and SIC results for the target category in both Experiments 1 and 2 (see Figures 3.8 and 3.9, and Figures 3.11 and 3.12, respectively). This interpretation is supported statistically via the directional KS-Test results for both Experiments 1 and 2 (Houpt & Townsend, 2012). Further, the mean RT results for the contrast category also tended towards a pattern of coactivity (with some individual variation) in both Experiments 1 and 2 (see Figures 3.10 and 3.13, respectively). Finally, although there is individual variation in the data, the computational modeling provides strong evidence that the majority of participants favored a strategy in which information was pooled into a single channel.

3.4.1 Implications for RT theories of Categorization

This finding has implications for current RT theories of categorization and, in particular, the growing body of work utilizing the logical rule-based models. First, we add to the literature examining categorization decisions using visual stimuli by demonstrating that within-dimension luminance features are pooled together into a single decision channel. This finding is novel because it is not commensurate with the idea that spatially separated dimensions need to be resolved serially due to limits on the spatial aspects of visual attention as one might conclude from the results of Little et al. (2011). However, a limitation of the current study is the inability to characterize the specific effects of spatial attention and in particular, distinguish between space-based and object-based attention. Further investigation into spatial configuration is therefore an interesting avenue for future research.

It should also be noted, that since the stimuli were displayed until a response, at least one participant may have adopted a serial strategy. With brief presenta-

tions, we would expect a stronger tendency toward coactivity since this would limit the ability to make several eye saccades which could induce serial processing. We therefore also highlight that pooling of information, while commonly used, may not be mandatory at the time scale of presentation that we used. Given the possible effect of time-scale on processing architecture, investigating decisions which are time limited as well as those which are time unlimited is therefore a worthwhile pursuit for future research

The key point of difference from earlier perceptual categorization work is that features that belong to the same feature type *can* be pooled across space whereas features of different dimension types need to be attended to independently. In terms of our banana example, this suggests that you can select the ripest banana by considering the bunch as a whole, but in order to compare both ripeness and price, you would need to attend separately to the price tag and the bunch. However, an important caveat to this conclusion is that other within-dimension feature types have not been investigated here. For example, it is possible that the finding of coactivity in the current work is due to a strategy whereby the participants utilize the difference, or relative difference, in luminance between the left and right disc, rather than the values of the discs specifically. Such a strategy may be not exist for other within-dimension stimuli. Replicating the current work using other within-dimension features is therefore a worthy avenue for future work.

3.4.2 Implications for theories of visual attention

Visual Search

In visual search, the difference between within-dimension and between-dimension features is central to the efficiency of the search. For instance, visual search for items comprising the conjunction of two between-dimension features (e.g., searching for a red vertical line among distractors which are red horizontal lines and green vertical lines) can still be guided via parallel feature guidance

modules. However, guidance in search for feature conjunctions which comprise two within-dimension features (e.g., search for a red-green target among red-blue and blue-green distractors) is not as efficient. While a guidance module may be able to preferentially direct search to all red items, thus reducing the number of items to be searched, this is still not as efficient as search involving conjunctions of different feature types. In a series of experiments, Wolfe et al. (1990) showed that searches for conjunctions which were within-dimension (e.g., color $\times$ color and orientation $\times$ orientation) in a field of within-dimension distractors, were significantly less efficient than between-dimension searches (e.g., color $\times$ orientation) in a field of between-dimension distractors.

It is tempting to link the efficiency (i.e., speed and accuracy) with which the target is found, to the underlying architecture of processing. Indeed there are strong indications that an invariance of RT across set sizes is indicative of parallel processing (Wolfe, 2016; Townsend, 2016). On the other hand, it is well known that the increase in mean RT with increasing set size cannot be taken as an indicator of serial processing. The issue is that a limited capacity parallel model can yield identical inefficient set size functions to the serial model (see Townsend, 1971). Changes in processing due to a change in set size therefore give an indication of capacity (i.e., the efficiency at which a system can process information given varying workloads) which should be considered independently to that of questions of architecture (this concept is covered in further detail in the following section). Our present work uses a factorial combination of item difficulty in order to address the question of architecture directly thereby circumventing methodological issues with other methods (Little, Eidels, et al., 2017). Although our focus is on perceptual categorization, our results can provide additional insight into other perceptual tasks like simple visual search, where the primary question is whether a target is present or not, instead of is this target a type A or a type B.

A limitation of the current methods, however, is that we are unable to locate the specific stage (i.e., pre-attentively or attentively) in which the pooling of

information occurs. Guided Search (Wolfe, 1994a, 2007) proposes that feature maps are created and combined into a master salience map at a pre-attentive parallel processing stage and that this map subsequently guides attention. Commensurate with this theory, it seems likely that, given they belong to the same feature-map, within-dimension features could be pooled pre-attentively forming a single signal which drives the decision making process. However, it could also be that pooling occurs at an attentive stage. Indeed, given that highly similar perceptual operations can yield a variety of experimental results, it may be expected that different tasks such as visual search, identification, categorization, detection might also diverge. Investigating the locus of pooling of information, and more broadly investigating different task types utilizing the same experimental stimuli and ideally, the same participant pool, would be a worthwhile pursuit for developing a unified view of visual perception and cognition. This could be tested within the methodology of SFT but would require manipulations of salience thought to operate solely at the pre-attentive stage.

While in simple visual search the focus is on target presence or absence, for complex visual search tasks, target type may actually be of vital importance. For example, a radiologist may need to conduct a visual search of an x-ray to search for a potential cancerous tumor. This process may involve not only a search for a potential target, but also a categorization decision regarding the tumor. This decision may form a part of the search (does this part of the image constitute a malignant tumor or distracting information) or be somewhat independent from the search process itself (is this tumor, once found, malignant or benign?).

Fifić, Townsend, & Eidels (2008) have successfully used SFT to investigate the processing architecture of the search process, and whether processing architecture changes under varying experimental manipulations. They investigated search for letters versus non-letters, while also accounting for how feature complexity (number of features per item) and target-distractor similarity may additionally affect processing architecture. Overall, they found that target-distractor similarity (as modulated by feature complexity) rather than linguistic composition

changed the processing architecture of the search. They proposed this was likely achieved through positively interacting parallel channels.

The use of SFT in this instance is interesting as it builds on the foundation of work by Duncan & Humphreys (1989) who provide an alternative view to theories incorporating feature maps (A. Treisman & Gelade, 1980; Wolfe, 1994a, 2007). Their theory focuses on stimulus similarity as a determinant of efficiency in visual search paradigms. Here, search efficiency increases with increased similarity between non-targets, and decreases with increased similarity between targets and non-targets. Duncan & Humphreys (1989) propose that all items in the display are processed in a parallel first stage which provides a structured representation of the input. In this parallel stage, items are thought to be organized in a part-whole structure whereby items which share properties (such as same color, same shape, same motion, or even simply proximity) are linked together via gestalt grouping. This is followed by a selection process whereby the input from the parallel stage is matched against a template of the information needed to complete the search. Finally, information enters into visual short-term memory (VSTM) and thus reaches conscious awareness and allows for a response to be initiated.

As access to VSTM is limited, items must compete for entry. Increasing attentional weight to one structural unit, or gestalt group, must therefore naturally lead to a decrease in attentional weight to another. It is further assumed that items gain and lose attentional weighting together via a process called weight linkage. This means that items which are strongly grouped will tend to be either selected together or efficiently rejected together if they are non-targets which do not match the target template. Thus, search efficiency is determined by two factors: the degree to which targets and non-targets match the target template and the degree of similarity between items allowing for spreading suppression or activation of groups. For example, increasing non-target and target similarity will reduce search efficiency as the weight of each non-target depends on its match to the target template. Decreasing the similarity between

non-targets will further reduce search efficiency as the opportunity for spreading suppression is decreased. Indeed, this effect of target-distractor similarity was shown to modulate processing architecture in Fifić, Townsend, & Eidels (2008). When also considering the current work in the context of Duncan & Humphreys (1989), the notion that items of the same color would be grouped together as a whole is consistent with our finding of coactivity, however, as their theory does not include a temporal component, or specify an underlying architecture, it could also be seen to be consistent with other accounts.

The question of decision-making, independent from the search process itself, has been investigated by Wolfe et al. (1990) who had participants complete a "search" with a set size of one (i.e., a simple identification task). In contrast to a standard visual search where they found a processing advantage for between-dimension features, they found that there was no difference in RT for identifying whether a color × color conjunction was a target versus a color × orientation conjunction was a target. This suggests that there is no cost to identifying whether or not a conjunction is a target for within-dimension conjunctions versus conjunctions which comprise different features. However, this result alone does not address whether the individual colors in a color × color target are processed independently or not. Our results show that for many individuals, these color × color targets are treated as a single source of information. The current methods could therefore be usefully extended to fully characterizing decision-making in complex visual search task which necessitate target categorization, for example when asking "is this item a gun or a hair-dyer?" in baggage screening (Wolfe et al., 2005), or "is this tumor malignant or benign?"in cancer screening (Drew et al., 2013).

The current findings, and SFT more generally, can therefore be used to inform the underlying architecture of decision-making in complex visual search. This applies to a variety of settings including security baggage-screening and visual search of medical images, as well as other complex searches such as foraging studies (Wolfe, 2013), and visual search of natural scenes (Wolfe, 1994b), familiar

scenes (Hout & Goldinger, 2010), and in everyday life (where are my wallet and car-keys?; Wolfe et al., 2011)[9].

Encoding, Selection, and VSTM

Within-dimension feature stimuli have also been investigated in studies of visual selection, visual encoding, and VSTM (see e.g., Huang et al., 2007; Mance et al., 2012; Sewell et al., 2014). However, most models of visual attention do not explicitly speak to how decisions are actually made, and thus the current findings provide an interesting insight for these models. For example, this finding is consistent with P. L. Smith & Sewell (2013)'s conceptualization of visual information, perhaps represented in VSTM, feeding into a single diffusion process. In P. L. Smith & Sewell (2013)'s multi-stage model of visual attention, different stages of processing may be subject to different capacity limitations and, consequently, governed by different architectures. In their model, although the information retained by VSTM is determined by the selection and encoding stages which occur in parallel, with competitive interaction, the decision-making stage accumulates noisy samples from a single VSTM trace until a criterion is reached. Our coactive model embodies a similar set of assumptions. P. L. Smith & Sewell (2013)'s implementation of selection as a competitive parallel process is also consistent with theories of visual search (Wolfe, 1994a, 2007) which propose that attentional selection of color $\times$ color conjunction targets cannot occur efficiently as they belong to the same feature map and attention, therefore, cannot be effectively guided to the target location. In P. L. Smith & Sewell (2013)'s model, when performing tasks such as visual search, items which contain task-relevant attributes or features excite the "where" pathway of attentional selection and in turn, mutually inhibit each other via competitive interaction. These relevant attributes or features further self-excite, modulated by attention. In a color $\times$ color conjunction search task, distractor items also contain task relevant attributes and therefore compete with the target for selection, leading

[9]We thank A/Prof Michael Hout for raising these points

to inefficient search that can be completed only with the application of attention.

P. L. Smith & Sewell (2013) state that little is currently known about capacity limitations and the processing architecture of decision-making. The logical rules approach is useful in this regard as it provides a way to investigate decision-making independently of sensory and memory representations. In our task, the stimuli are presented supra-threshold and are available to participants until a response is made. Given this manipulation, sensory and memory load are held constant while only the decisional difficulty is varied factorially. Although our findings do not speak explicitly to items held in VSTM, they are nonetheless consistent with P. L. Smith & Sewell (2013)'s analysis of briefly presented, multi-element displays.

While our design is useful in that it holds constant memory load, it does this by not requiring the retention of any information at all, removing the need to encode information and hold it in VSTM. It is consequently difficult to know whether decision-making which is driven by information held in VSTM is equivalent to decision-making using stimuli which remain present in the visual scene. For example, in the current design, participants are able to accumulate information from the available scene until a decision is reached. In VSTM experiments, however, the information needed to make a decision must be accumulated from a potentially imperfect memory representation which could alter the decision strategy used.

We set as a goal for future research to investigate decisions in a VSTM task, again, holding constant sensory and memory load and factorially varying the difficulty of decision-making. Some progress has been made towards this using a SFT change detection paradigm (C.-T. Yang, 2011; C.-T. Yang et al., 2011, 2013). However, these investigations have only investigated single item displays and have systematically varied sensory information such as relative saliency of changes between dimensions and the probability of changes. In order to accurately isolate the decision-making stage for multi-element displays, environmental cues and stimulus properties that guide attention such as relative saliency of the changes

and probability of changes should be held constant. Further, changes in two or more items should be investigated. This is the current focus of our research group.

3.4.3 Relationship to the Race Model Inequality

Several studies (Mordkoff & Yantis, 1993; Mordkoff & Danek, 2011; Poom, 2009) have used an alternative measure, violations of the Race Model Inequality (J. Miller, 1982), to infer coactive processing of certain visual features and their location in space. Given our inference of coactivity using SFT in the present work, some further discussion of the connection to the RMI is warranted. The RMI is typically measured using a redundant target task (see Egeth & Mordkoff, 1991; Snodgrass & Townsend, 1980). In this task, two locations (e.g., left and right) are monitored for the onset of stimuli or targets. These targets can appear in one of the two locations, as well as both locations, or neither location. In one condition (an OR task), participants are instructed to respond in the affirmative whenever any target is present in either or both locations. RTs are compared between the double target displays (i.e., displays containing a redundant target) and single target displays. An RT advantage for detecting the redundant target could arise solely due to statistical facilitation in an independent parallel race model (i.e., the minimum time expected for the detection of two possible targets is smaller than the detection of any single target alone; Raab, 1962). On the other hand, faster redundant target detection times could also be due to what J. Miller (1982) refers to as "coactivity", where both items contribute to a target present response. To distinguish these two accounts J. Miller (1982) formulated the RMI, which provides an upper limit on the speeding of responses which can be accounted for by statistical facilitation. According to J. Miller (1982) violation of the RMI therefore provides evidence for coactivity.

Mordkoff & Yantis (1993) used a modification of the redundant targets detection task to investigate single objects comprising two between-dimension features (i.e., color and shape), as well as these between-dimension features separated

in space. This task is different to detection in that it requires participants to provide a stimulus identification in the presence of distracting information using a go/no-go response. For example in the first experiment, stimuli comprised of green and purple X's and O's and participants were required to respond to the presence of the color green, or the letter X (a green X therefore constituted the double target, whereas a purple O required no response). In Experiments 2 and 3, the features were separated in space. For example, in Experiment 3 shape was represented by a white letter presented at fixation, and color was represented as a colored border surrounding this shape. This was further compared in Experiments 4 and 5 to a within-dimension version of the experiment in which two letters or colors separated in space were presented either side of fixation (drawn from a pool of six, two of which were given target designations). Overall, they found a violation of the RMI for all versions of the experiment using between-dimension stimuli (regardless of location), but not for the within-dimension stimuli separated in space. They therefore concluded that decisions requiring the integration of between-dimension occurred coactively, whereas decisions requiring the integration of within-dimension interactions can be explained by statistical facilitation.

However, Feintuch & Cohen (2002), suggested that Mordkoff and Yantis' (1993) findings could be accounted for by perceptual grouping (Duncan, 1984). That is, the between-dimension items which were thought to coactivate were actually being perceived as part of the same object, whereas items which did not coactivate (i.e., which happened to be within-dimension) were perceived as two separate objects. Feintuch & Cohen (2002) therefore suggested that grouping, rather than feature type was responsible for the coactivation of information in Mordkoff and Yantis' work. In Feintuch and Cohen's experiments, between and within-dimension items were presented as both separated in space (e.g., in the between-dimension condition, a color and shape were presented either side of fixation and in the within-dimension condition two colors were presented either side of fixation), and as parts of the same object (achieved simply by drawing an

ellipse around the two items). They found that features which were considered as part of a single object (and therefore were contained in a single locus of attention) were processed coactively, regardless of feature type, whereas features which did not comprise a single object were not. This interpretation was later confirmed by Mordkoff & Danek (2011) in a follow up study.

While a cursory consideration of the findings of Feintuch & Cohen (2002) and Mordkoff & Danek (2011) seem counter to the current work, it is important to take a number of factors into account. A violation of the RMI is not isomorphic with the concept of coactivity used in SFT. In SFT, coactivity refers to pooling of information and can be conceptualized as an interactive model with a complete facilitation between decision-making channels. In J. Miller (1982), Mordkoff & Danek (2011), and Feintuch and Cohen's (2002) work, coactivity is defined as any process violating the RMI (although pooling of information to form a decision is often also therefore inferred). However, there is now a large body of work showing that the RMI can be violated by models which decidedly do not show complete pooling: serial exhaustive models (Townsend & Nozawa, 1997), models violating the assumptions of the RMI (Cheng et al., 2017; Otto & Mamassian, 2016; C.-T. Yang et al., 2017), and for exhaustive models in the presence of distractors (Little et al., 2015, 2018). Consequently, the link between the RMI and architecture is not one-to-one and these concepts should be treated independently.

In SFT, violation of the RMI is instead isomorphic to the concept of super-capacity. In SFT, capacity refers to the efficiency of information processing when workload varies (i.e., the amount of information that needs to be processed increases or decreases; Townsend & Ashby, 1983; Wenger & Townsend, 2000). Capacity is also measured using a redundant target detection task. By using the expected minimum processing time of the parallel independent race model as a baseline, performance can be evaluated as either less efficient than expected or more efficient than expected (Townsend & Ashby, 1983; Townsend et al., 2007; Wenger & Townsend, 2000). A violation of the RMI therefore indicates

that workload efficiency is more efficient than expected under an independent parallel race model. Super capacity and violations of the RMI have recently been shown to be formally related: the RMI provides a bound on the super capacity expected from an unlimited capacity parallel system (Townsend & Eidels, 2011). The experiments of Mordkoff & Danek (2011) and Feintuch & Cohen (2002) can therefore be better interpreted in the context of the current paper as measures of workload capacity. These papers therefore provide evidence for super capacity processing for between- and within-dimension features comprising a single object, and unlimited-capacity processing for between- and within-dimension features separated in space. However, coactivity in the single object condition is only one explanation for the difference between these conditions.

Caution should also be exercised when interpreting these results as both non-target and single target displays in these experiments necessarily include irrelevant/distracting information and therefore do not constitute a pure single-target as per redundant target detection tasks. Cheng et al. (2017) and Little et al. (2015, 2018) have shown that utilizing capacity (and hence the RMI) in the presence of distractors is not straightforward due to distracting information potentially slowing responses for single target items leading to an overestimation of capacity and which can also lead to violations of the RMI. Nonetheless, Feintuch and Cohen's (2002) finding that within-dimension stimuli do not violate the RMI warrants further investigation and characterizing the capacity of within-dimension stimuli independent from distracting information would be a worthwhile pursuit for future work.

Measuring the capacity of the current within-dimension stimuli would also be a worthwhile target for future work. The current work assumes channel independence, however, there is a possibility that the luminance level of one disc may affect the processing speed of the other disc. Eidels et al. (2011) describe a series of parallel models which allowed for facilitatory and inhibitory interaction between the two channels and showed that a combination of SIC and capacity coefficients could allow for the identification of particular interactive systems.

Measuring capacity would therefore allow for an even more comprehensive characterization of processing and would assist further investigations into the possibility of interactions between processing channels in the current work.

3.5 Conclusion

In summary, the current study adds to the body of literature investigating perceptual categorization and provides insight into the relationship between visual attention and decision-making when stimuli are comprised of within-dimension features. In particular, we have made the novel finding that within-dimension stimuli in separate locations are processed coactively by most participants, rather than in serial as seen with between-dimension stimuli in separate locations (Little et al., 2011; Fifić et al., 2010). Whether or not this generalizes to non-integral stimuli would be a worthy avenue for future research. The current work further emphasizes areas where theories of visual attention, categorization decision-making, and the logical rule models converge, and highlights avenues for synthesis between these theories. This same focus on featural information as a determinant of information processing is also a cornerstone of Anne Treisman's work, demonstrating that the influence of her legacy spans well beyond the scope of visual search, and continues to inform other areas of cognitive science.

3.6 Appendix A

Figure 3.15 shows the Survivor functions for Experiment 1. Table 3.11 shows KS-tests for Experiment 1.

Table 3.11: P-Values from KS-tests of stochastic dominance.

Participant	Dominance Test for Selective Influence							
	HH > HL	HH > LH	HL > LL	LH > LL	HH < HL	HH < LH	HL < LL	LH < LL
B1	<.001	<.001	<.001	<.001	1.000	1.000	0.999	1.000
B2	0.003	<.001	<.001	<.001	0.992	1.000	0.982	0.992
B3	<.001	0.002	<.001	<.001	0.982	0.998	0.927	0.981
B4	<.001	<.001	<.001	<.001	1.000	1.000	0.982	0.993
B5	<.001	<.001	<.001	<.001	1.000	0.998	0.950	1.000
B6	<.001	<.001	<.001	0.002	0.997	0.997	1.000	0.997
B7	<.001	<.001	<.001	<.001	1.000	1.000	1.000	1.000

3.7 Appendix B

Figure 3.16 shows the Survivor functions for Experiment 2.

3.8 Appendix C

For each stimulus, we assumed that each dimension was represented by a normal distribution with mean locations, μ_{Left} and μ_{Right}, and standard deviations, σ_{Left} and σ_{Right}. For the coactive model, we assumed that the representation was the joint distribution over both dimensions; hence, a bivariate normal distribution with covariance set to 0. For simplicity, we assumed that the locations of each of the items were simply the logical values of those items in the category space. That is, $\mu_{Left} = 1, 2$, or 3 and $\mu_{Right} = 1, 2$, or 3. Each participant is assumed to implement a decision boundary along each dimension in order to separate each category. This gives two further parameters: D_{Left} and D_{Right}. The integral of these distributions within each category provides the mean drift rate for the LBA for that stimulus.

In the LBA, there is a parameter for capturing the variability in the starting point of accumulation which varies as a uniform distribution between 0 and A from trial to trial. There were response threshold parameters for both the

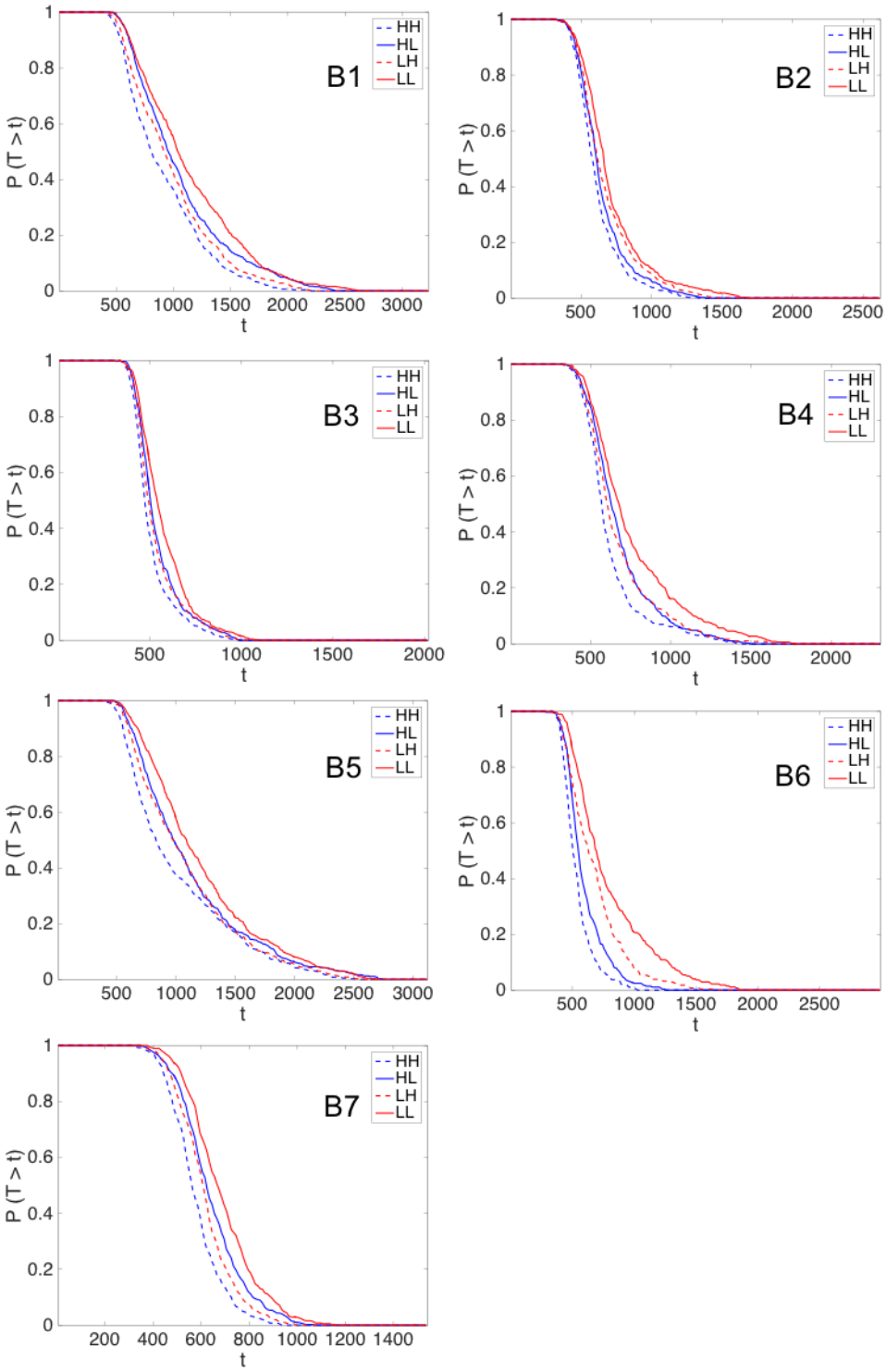

Figure 3.15: Survivor functions for individual participants in Experiment 1

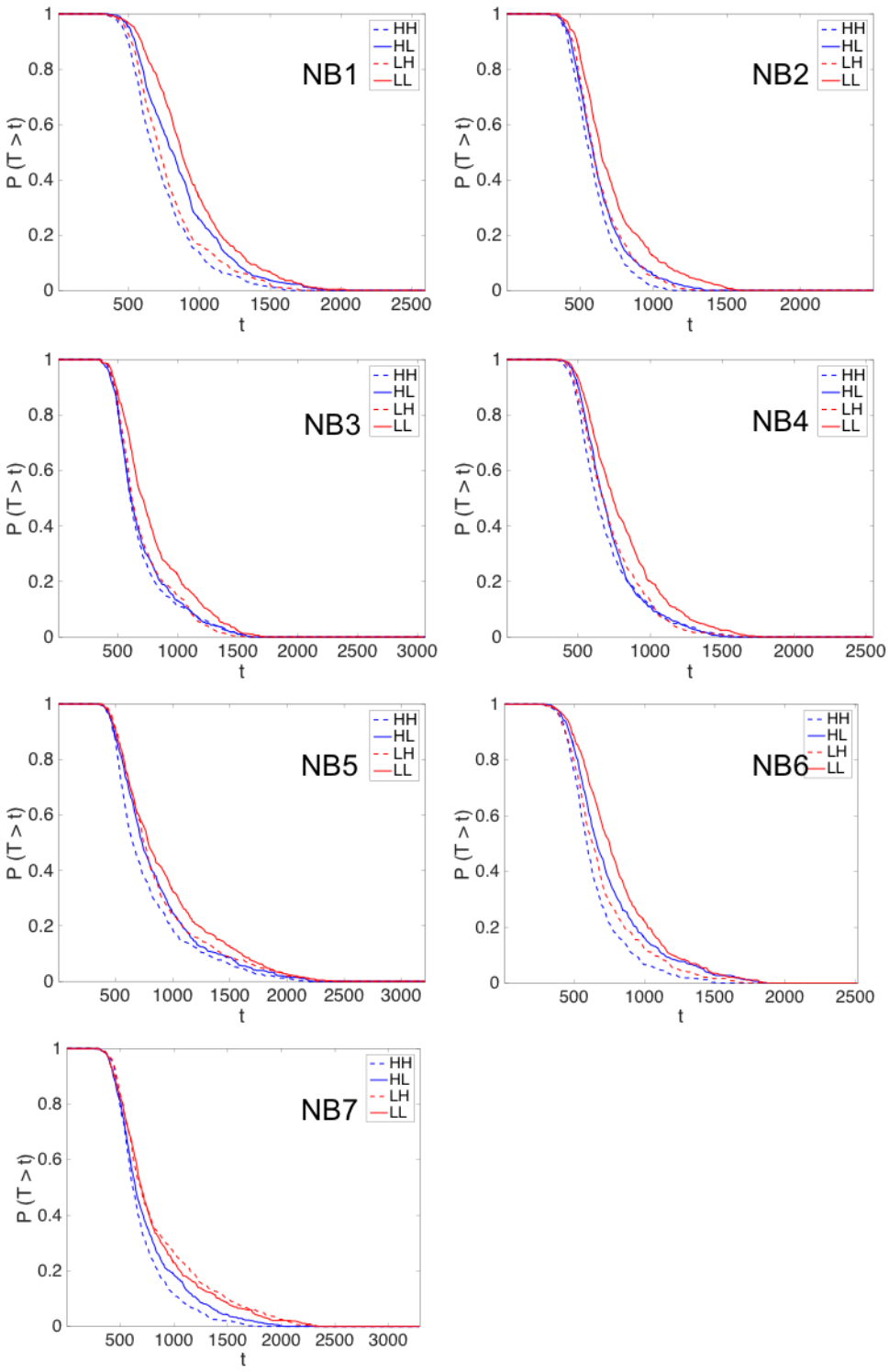

Figure 3.16: Survivor functions for individual participants in Experiment 2

target, b_{Target}, and contrast category, $b_{Contrast}$ accumulators, which capture the distance between the starting point, A, and the decision threshold. For the serial and parallel model, there are separate target and contrast accumulators for each dimension. We assume that the response thresholds and starting points are the same for each of these dimensional accumulators. Drift rate is assumed to vary from trial to trial normally with standard deviation, s. Both accumulators also incorporate non-decision time (t_0) in order to account for time taken to complete actions not associated with the decision process (e.g., encoding responses or the time taken to initiate a motor response). The final RT is therefore comprised of the sum of the non-decision time and the decision time predicted by the LBA.

In summary, there are nine free parameters in the parallel self-terminating and coactive models. The serial self-terminating model has one further parameter (p_x) representing the probability that one dimension is processed before the other. For the free drift rate model, which estimates the drift rate for each stimulus independently rather than using the GRT framework, there are 14 parameters: the five LBA parameters plus the nine drift rates.

3.8.1 DE-MCMC Details

In order to fit the logical rules models, we used a Bayesian framework. We approximated the likelihood for each model using probability density approximation (PDA) as described in Turner & Sederberg (2014). The likelihood of each trial was computed by simulating 50,000 data points from the model and then using Holmes' (2015) method. The log likelihood for each trial was then summed over trials and items. Each participant was fit separately. For each parameter, we first transformed the parameter to lie on the whole real line and then adopted reasonably informative priors based on our prior work with these models. The transformations and priors are shown in Table 5.10.

We used DEMCMC (Turner et al., 2013) to generate proposals from the posterior distributions of each parameter in an efficient manner. However, the variability in the likelihood approximation can cause the chains to become stuck

Table 3.12: Prior parameter distributions and transformations for each parameter.

Parameter	Transformation	Distribution	Prior Parameter	Values
D_{Left}	$\hat{D}_{left} = logit\left[D_{left} - x_0\right]$	Normal	$\mu = 0$	$\sigma = .5$
D_{Right}	$\hat{D}_{right} = logit\left[D_{right} - y_0\right]$	Normal	$\mu = 0$	$\sigma = .5$
σ_{Left}	$\hat{\sigma}_{left} = \log(\sigma_{left})$	Normal	$\mu = -1.5$	$\sigma = .2$
σ_{Right}	$\hat{\sigma}_{right} = \log(\sigma_{right})$	Normal	$\mu = -1.5$	$\sigma = .2$
A	$\hat{A} = log(A)$	Normal	$\mu = -1.05$	$\sigma = .2$
B_{Target}	$\hat{b}_{Target} = log(b_{Target} - A)$	Normal	$\mu = -1.05$	$\sigma = 1$
$B_{Contrast}$	$\hat{b}_{Contrast} = log(b_{Contrast} - A)$	Normal	$\mu = -1.05$	$\sigma = 1$
s	$\hat{s} = log(s)$	Normal	$\mu = -1.39$	$\sigma = .5$
T_0	$\hat{t}_0 = log(t_0)$	Normal	$\mu = -1.51$	$\sigma = .2$
p_x	$\hat{p}_x = logit(p_x)$	Normal	$\mu = 0$	$\sigma = 2$

if an accepted parameter set results in an usually high likelihood. To prevent the chains getting stuck in this manner, we re-sampled the likelihood of any existing chains each time the current proposal was rejected (Holmes, 2015). Following this method led to a good mixing of the chains and strong convergence after the burn-in period.

We used a burn-in period of 2150 iterations with a deterministic migration step Turner et al. (2013) every 20 iterations between iterations 501-700. The remaining sampling used a probabilistic migration step instead of a cross-over step with a probability of .05. The number of chains was determined by taking three times the amount of parameters for each model. A minimum of 20000 posterior samples were taken per parameter, with 750 iterations estimated for each chain.

For model comparison we used the Deviance Information Criterion (DIC; Gelman et al., 2014). The deviance of a posterior sample of parameters, denoted θ, is calculated as:

$$D(\theta) = -2lnL(y \mid \theta) \tag{3.3}$$

The DIC is calculated as:

$$DIC = \overline{D}(\theta) + 2p_D \tag{3.4}$$

Here, $\overline{D}(\theta)$ is the mean of the distribution of posterior deviances and $p_D = 2\ var[lnL(y|\theta)]$. The DIC punishes for model complexity. This is achieved by penalizing the average negative log likelihood by a term which accounts for the functional form complexity of the model.

Chapter 4

Characterizing the processing capacity of within-dimensions features

A pervasive issue in the current literature is that workload capacity is measured in varying ways, and often researchers mistakenly use measures of workload capacity in an attempt to diagnose processing architecture (see e.g., Mance et al., 2012; Mordkoff & Danek, 2011; Huang & Pashler, 2007). Although these measures are related, importantly, they are not equivalent (Townsend, 1990a), meaning conclusions about architecture which are based on measures of workload capacity may be misleading. Nonetheless, when coupled with direct measures of architecture, measures of workload capacity can provide a more precise method of diagnosing serial, parallel, and coactive processing, particularly in cases where channel interactions are a possibility (Eidels et al., 2011). This chapter will therefore focus on characterizing the workload capacity of decision-making using the within-dimension stimuli presented in Chapter 3. Although the evidence presented in Chapter 3 suggests that within-dimension features could be processed coactively in a single decision-making channel, coupling

decision-making architecture with a measure of workload capacity would provide stronger insight into the decision-making architecture. I first turn to a discussion of the current literature, using the example of the Boolean Map Theory of visual attention (Huang & Pashler, 2007) as a salient example which highlights issues in utilizing measures of capacity to infer architecture. This issue has been highlighted by Fitousi (2019), whose work I discuss in some detail, including a refit of his data using the computational modeling described in Chapter 3.[1] I then outline the current experiment.

The Boolean Map Theory of Visual Attention (Huang & Pashler, 2007) attempts to address the question of what visual information can be apprehended at any one given instant? A Boolean Map is a representation of the visual scene which is divided into a binary delineation: the selected region and the non-selected region. For example, in a visual scene containing only one red object and one green object, either the red object is selected, whereas the green object is not, or, both objects are selected but are indistinguishable in terms of their individual features such as color, size, shape etc. A central claim of Boolean Map theory is that each dimension only contains one feature label; however, the same feature (here the value of red is considered the feature; color would be a higher level dimension) at multiple locations is comprised within the one Boolean Map. Boolean Map theory is therefore very similar to Feature Integration Theory (FIT; A. Treisman & Gelade, 1980) and the later Guided Search theory (Wolfe, 1994a, 2007) in that these theories suggest visual representations are best described using feature maps. In the case of FIT and Guided Search, these are simply termed feature maps, where basic features each have their own representative map, which can subsequently be used to guide further processing (usually visual search). In Boolean Map theory, these featural maps are represented in binary form and inform processing via the combination of Boolean maps using Boolean operations.

Huang & Pashler (2007) propose that there are two important attentional

[1] We thank Dr Daniel Fitousi for kindly providing his data.

limitations which Boolean Maps describe. The first is selection (i.e., the way the Boolean Map is created) and the second is access (i.e., the information that can be obtained from a Boolean Map)[2]. In Boolean Map theory, an initial map can only be created from a single feature value, and only a single feature value can be accessed at any given instant. Importantly, this puts a limitation on the architecture of selection and access. Boolean maps must be created in sequence, meaning that selection must be described by a serial process. In turn, if more than one feature needed to be accessed, this too must occur one at a time, in serial.

Of particular interest to the current work were Huang and Pashler's (2007) Experiments 3-4. In these experiments, participants were shown two colors, presented either simultaneously or sequentially. This was followed by a probe stimulus for which participants indicated whether there was a match or no match to the previously seen items. Huang and Pashler found that performance for the simultaneous condition was significantly worse than the sequential condition. From this they concluded that participants were not able to access the colors at the same time. However, as Mance et al. (2012) suggest, there are a number of limitations to this conclusion. Firstly, change detection tasks may overestimate the capacity limitations of simple visual encoding (or perhaps, access; Hollingworth, 2003; Mitroff et al., 2004; Simons & Rensink, 2005)[3]. This is because the task requires visual encoding of both a memory display and probe display. Further, the stimuli used by Huang & Pashler (2007) were small and positioned close together (in Experiment 4 they were less than 0.8cm $\times$ 0.8cm and were placed less than 2cm apart). As Mance et al. (2012) note, items within a given cortical area which are small and close together may compete for representational access in the simultaneous but not the sequential condition (Desimone & Duncan,

[2]In FIT (A. Treisman & Gelade, 1980) and Guided Search (Wolfe, 1994a, 2007) feature maps are created at an early pre-attentive parallel stage. Subsequent access to this information is only limited by the need to integrate information. This may be represented in terms of a top-down salience map which combines information from each saliency map and subsequently guides attention (Wolfe, 1994a, 2007).

[3]Mance et al. (2012) argue that the concepts of encoding and access are synonymous (see also Huang, 2007 for a discussion of this issue).

1995; Duncan, 2006; Kastner et al., 2001).

In their task, Mance and colleagues (2012) presented participants with 1-4 colored squares (measuring $1^\circ \times 1^\circ$ and appearing 6° apart) either simultaneously or sequentially but found no performance advantage for the sequential condition over the simultaneous condition for up to two items, concluding that participants were able to do the task in parallel. Using an analogous design, no sequential advantage has also been shown using orientation (Sewell et al., 2014).

Mance et al.'s (2012) finding that there is no sequential processing advantage for up to two colors is counter to the findings of Huang & Pashler (2007). However, inferences about architecture made using this paradigm are flawed. That is, the results of Huang & Pashler (2007) and Mance et al. (2012) do not describe whether access is occurring in serial, one after the next, or simultaneously, in parallel. Rather, the simultaneous-sequential paradigm provides an indication of *workload capacity* (Townsend & Ashby, 1983; Wenger & Townsend, 2000). To refresh, workload capacity refers to the efficiency of processing given varying workloads. For example, a limited capacity system becomes less efficient as the number of to-be-processed items increases. An unlimited capacity system on the other hand, is unaffected as the number of to-be-processed items increased. Finally, a super capacity system becomes more efficient as the number of to-be-processed items increases. Importantly, while measures of architecture and capacity are related, they are not equivalent. A sequential advantage in the simultaneous-sequential paradigm could indicate serial processing, or it could indicate certain types of limited capacity parallel processing, which is nonetheless simultaneous (Townsend, 1990a).

In his recent paper, Fitousi (2019) acknowledged this issue and sought to directly test the processing architecture of within-dimension color stimuli using the Logical-Rules paradigm (Fifić et al., 2010) and complementary non-parametric SFT analyses (Townsend & Nozawa, 1995; Little, Altieri, et al., 2017). His stimuli were red and green colored squares ($1.83^\circ \times 1.83^\circ$) arranged vertically ($5^\circ$) apart, such that the red square was always on the top whereas the green

square was always on the bottom. The red and green squares varied in three levels of "typicality" (low, medium, and high) which combined orthogonally to create the nine item stimulus space. As per the standard Logical-Rules paradigm, participants were asked to categorize items as belonging to either the target category (Category A) or the contrast category (Category B). If participants were unable to process the items simultaneously, Fitousi suggested that the processing architecture would be best described as serial. This would be taken as evidence for Boolean Map theory. If however, participants were able to process the stimuli in parallel or coactively, this would be taken as evidence against Boolean Map theory.

From the non-parametric results, Fitousi (2019) concluded that the target category was best described as a serial exhaustive process, whereas the contrast category results were consistent with either coactive or parallel self-terminating processing. To gain further confidence in this finding, Fitousi (2019) fit the computational models similar to those described by Fifić et al. (2010), excepting that a Linear Ballistic Accumulator (LBA; Brown & Heathcote, 2008) was used to replace the random-walk component, and only the contrast category was fitted. The parallel self-terminating model best fit three participants for their contrast category RTs, whereas the coactive model best fit the remaining participant.

While it may seem like the results of Fitousi (2019) provide evidence against Boolean Map theory (at least for the contrast category), there are some important issues with the experimental design and data analysis to consider. Firstly, the salience manipulation which was used created an experiment in which highly saturated items were allocated to the target category, and items with low saturation were allocated to the contrast category. This meant that participants may have been able to make an accurate categorization decision simply by judging the overall saturation, rather than requiring specific processing of the color value of each individual item. Of course, the key prediction of Boolean Map theory which Fitousi (2019) was seeking to test was whether or not colors can be consciously accessed simultaneously. However, given the confound in

experimental design, any evidence of coactivity can be interpreted as participants making an overall saturation judgement, rather than their ability to access the colors simultaneously or not.

On the other hand, Fitousi (2019) also suggested that the non-significant MIC from his target category constituted strong evidence for serial exhaustive processing, which is in line with the predictions of Boolean Map theory. However, the MIC is a test of the *point prediction* of the serial model, which predicts that the interaction term should equal zero (cf. Sternberg, 1969). This presents a different goal to the typical null hypothesis significance testing case, where the goal of the significance cut-off is to place some criteria on the false positive rate. In the present case, an alpha criterion of .05 is biased toward the serial model (Fox & Houpt, 2016). Consequently, caution must be taken when interpreting a non-significant result in this context. This could mean that the underlying architecture could actually still be better described by a coactive model, which would be in line with Fitousi's (2019) contrast category predictions.

Typically, in the Logical-Rules paradigm, questions arising from the non-parametric analyses can be clarified with parametric modeling (Blunden et al., 2020; Cheng et al., 2018; Moneer et al., 2016). Importantly, parametric modeling can resolve some important limitations of the non-parametric analyses. For example, the non-parametric analyses only take into account a subset of the data. They do not, for instance, account for error rates or error RT distributions, nor do they consider data from both the target category and the contrast category simultaneously. Conversely, computational modeling can consider the data set as a whole, offering a more complete picture of participants' categorization strategy. Nonetheless, Fitousi (2019) only fit the parametric models for the contrast category, rather than considering all of the available items. Without further analysis, the alternative interpretation presented above cannot be ruled out.

The findings of Chapter 3 are therefore interesting to consider in the context of Huang and Pashler's (2007) and Fitousi's (2019) work. The previous chapter

also tested the decision-making architecture of spatially separated luminance discs using the Logical-Rules paradigm. Importantly, unlike the work of Fitousi (2019), catch trial stimuli were introduced to prevent participants being able to do the task using an overall judgment of the brightness. Instead, participants were required to process the values of each item. Despite this, in Chapter 3 we find evidence of coactive processing among the majority of participants.

It is interesting to note that, like Fitousi (2019), we also found positive, albeit non-significant MICs, which could be construed as evidence for a serial exhaustive process. However, given the significance test of the MIC is biased towards a serial model, we also conducted one-sided Kolmogorov-Smirnov (KS) tests (Houpt et al., 2013) of the survivor interaction contrast (SIC) which provided evidence for coactive processing for the majority of participants. We further fit the computational models for all available target, contrast, and error data and found that most participants were best fit by a coactive model, or a free-drift model which could capture coactive patterns which are accompanied by violations of perceptual separability. This suggests that within-dimension feature stimuli, regardless or decision rule, are processed simultaneously in a single coactive decision-making channel.

To test this theory further we refit the data presented by Fitousi (2019), including both target and contrast RTs simultaneously. This was done using the same methods as in Chapter 3 although for simplicity we only fit the three main candidate models: serial self-terminating, parallel self-terminating, and coactive. The DICs for each individual participant and model are shown in Table 4.1. The preferred model is the model with the lowest DIC.

Across all participants the coactive model provides the best overall fit for both target and contrast categories simultaneously. Along with model comparison estimates, it is essential to provide an estimate of how well a model fits the data to ensure that the model is capturing the data well (Heathcote et al., 2015). To show that the coactive model provides a good fit to the data we present the posterior predictions for each participant in Figure 4.1. These fits suggest that

Table 4.1: DIC values for each individual participant and candidate model from Fitousi (2019). Lower values indicate better fit.

	Model		
Subject	Serial ST	Parallel ST	Coactive
P1	-191.26	-463.31	**-628.86**
P2	4596.5	4489.7	**4469.1**
P3	3806.6	3912.1	**3489.8**
P4	762.32	648.08	**558.32**

the coactive model is a viable alternative explanation for the data presented in Fitousi (2019).

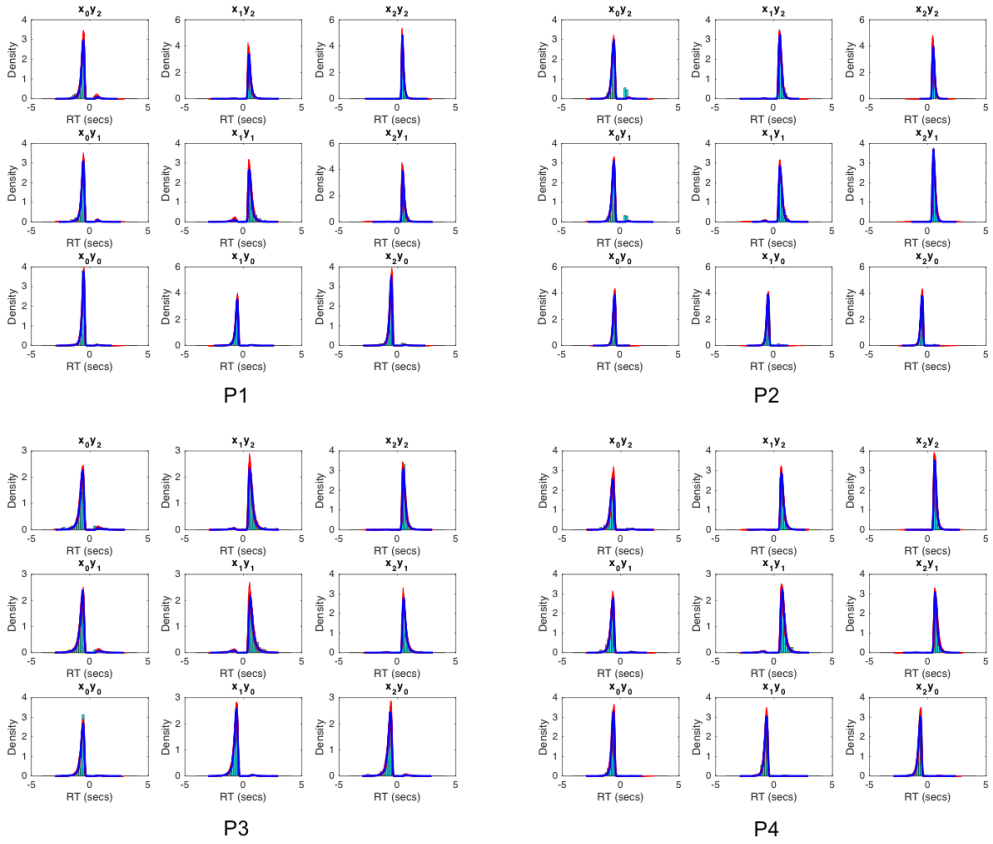

Figure 4.1: Posterior predictions from the coactive model for each participant in Fitousi (2019). The data are shown as bars. The solid blue line is the posterior mean prediction, and the red lines are draws from the posterior predictive distribution.

Nonetheless, there are still some further issues to consider. For example, in the Logical-Rules paradigm and therefore in the model fits above, we assume that the means for the two items are aligned to a grid (as shown, e.g., in Figure 1 of Chapter 3). That is, we assume that the technical property of perceptual separability holds (Ashby & Townsend, 1986). Given that a number of participants in Chapter 3 were best fit by the free-drift model, it is a possibility that this standard assumption of SFT was not met, and instead the salience of each individual item depends on the combination of items presented on any given trial. Further, the multidimensional scaling (MDS) space presented by Fitousi (2019) also does not adhere to the standard nine item grid assumed by the Logical-Rules framework. This suggests that the within-dimension stimuli in his study may not be perceived independently.[4] Returning to the findings of Chapter 3, it is important to consider a situation whereby the salience of one disc inhibits or facilitates the salience of the other disc thus violating perceptual separability, particularly as luminance and color perception is best characterized as relative in nature (Gilchrist et al., 1999).

Eidels et al. (2011) investigated a series of parallel models which allowed for facilitatory and inhibitory interaction between two decision-making channels. Of particular interest are models which incorporate excitation or inhibition at the input stage, known as pre-accumulator models. In these models the inputs are dependent, such that in an excitatory model, higher levels of input into one channel would lead to more activation feeding into the accumulator in the other channel. Conversely, in an inhibitory model, high levels of input into one channel lead to lower input in the other channel. These types of models could therefore be used to usefully describe a situation of perceptual dependence, as facilitation or inhibition is happening before the decision-making process has occurred, at

[4]Fitousi (2019) suggests that the target category items maintain their structure, but notes violations in the contrast category. However, it is clear when considering the data as a whole that there are marked overlaps between target and contrast category stimuli. In particular the LL stimulus is grouped with the redundant and interior stimuli, whereas the exterior stimuli are more closely grouped with the remaining target category stimuli. It is likely that this is due to the saturation of the items, as the exterior items contain one highly saturated item (and is therefore more like the target category stimuli), whereas the LL item is of comparatively lower saturation (and is therefore similar to the contrast category stimuli).

the level of more basic perceptual encoding.

Through a series of simulations Eidels et al. (2011) showed that these inhibitory and excitatory parallel models can be diagnosed when coupled with a measure of workload capacity. For example, a parallel inhibitory model in an AND design was shown to have an SIC which is less than zero for fast RTs, and greater than zero for slow RTs (this pattern describes a classically coactive SIC), but with limited capacity. If, however, capacity was super, the model would be considered an excitatory model.[5].

Given the additional explanatory power of workload capacity coupled with a diagnosis of processing architecture, we therefore seek to provide a measure of workload capacity to complement the findings presented in Chapter 3. While we expect to find super capacity (in line with a coactive or parallel facilitatory model) we do not rule out the possibility of finding limited capacity processing (in line with a inhibitory parallel model).

4.1 Current Work

To measure capacity, we implemented the redundant target detection task using the same stimuli as presented in Chapter 3. As in Chapter 3, we added additional catch trial stimuli to ensure the experiments were analogous. This resulted in one double target (white on the left and black on the right) which participants were required to respond "YES" to, and one double target (black on the left and white on the right) which participants were required to respond "NO" to. The experiment also included four single target items and one blank trial for which "NO" was also the correct response (see Figure 4.2). We again ran two versions of this experiment: a version with a dividing boundary comprised of discs with randomly sampled luminance levels and a version without such a boundary.

[5]A coactive model is equivalent to an excitatory parallel model where both channels interact completely.

4.2 Method

4.2.1 Participants

Ten participants from the University of Melbourne community (six males and four females, aged between 19 and 54 years) with normal or corrected-to-normal vision participated in the boundary condition. Participants in this condition were naïve to the experiment. Participants were recruited via advertising placed on notice boards within the Melbourne School of Psychological Sciences and through the school's online recruiting system.

The seven participants from the non-boundary condition had previously participated in the non-boundary experiment presented in Chapter 3 and so were familiar with the general task requirements but were otherwise naïve to the purpose of the experiment.

Participants gave informed consent and were reimbursed $10 per session, plus an extra $3 bonus for accuracy within a session greater than 90%. The participants from the boundary condition are referred to as CB1 - CB10[6]. The participants from the non-boundary condition retain their labeling from Chapter 3 and are referred to as NB1 - NB7. Testing was approved by the Melbourne Human Research Ethics Committee (Approval Number 1034866).

4.2.2 Apparatus and Stimuli

An example of the stimuli shown in the two experiments is shown in Figure 4.2. Stimuli were presented at a monitor resolution of 1280 $\times$ 1024 and participants viewed the screen at a distance of approximately 60 cm. There were seven stimuli with each stimulus containing either two, one, or zero discs presented on a gray background (RGB color space values [128 128 128]). The discs subtended a visual angle of 1.91° with centers 11.34° of visual angle to the left and/or right of fixation (the center of the screen).

[6]Here the B denotes "boundary" and the C denotes "capacity". This is to distinguish these participants from the boundary participants in Chapter 3.

There were two luminance levels; one lighter than the background (RGB color space values [200 200 200]) and one darker than the background (RGB color space values [64 64 64]). For single target trials, lighter and darker discs could appear in either location. All single targets were associated with a "NO" response. The redundant target trial (blank stimulus) was also associated with a "NO" response. There were two double target stimuli. The stimulus which was lighter on the left and darker on the right was associated with a "YES" response, whereas the stimulus which was darker on the left and lighter on the right was associated with a "NO" response (see Figure 4.2).

As in Chapter 3, for the boundary condition, the screen was divided by a boundary of 29 discs (also subtended at a visual angle of 1.91° at 60cm viewing distance) presented as a central vertical column. The luminance values of the boundary discs were randomized from trial to trial using six possible RGB color space values (drawn from the values used to implement the salience manipulation). This boundary was removed for the no boundary condition. All other aspects were the same. RTs for categorization were collected using a calibrated response time box (Li et al., 2010).

4.2.3 Procedure

Participants completed two one-hour sessions on consecutive or near-consecutive days. At the beginning of the task, participants were shown experimental instructions, as well as an example of the stimuli. Each session consisted of 714 trials (14 practice trials and 700 experimental trials). Each unique stimulus was presented 100 times per block. All stimuli were presented in randomized order within blocks. In between each block, participants were shown their percent correct on the current block and given the option to take a short break. During each trial a fixation cross was presented for 1500 ms. A stimulus was then presented and participants were asked to decide whether the stimulus belonged to "Group A" or "Group B" (see Figure 4.2). Stimuli were presented for 5000 ms or until a response was made. Feedback was presented for incorrect responses.

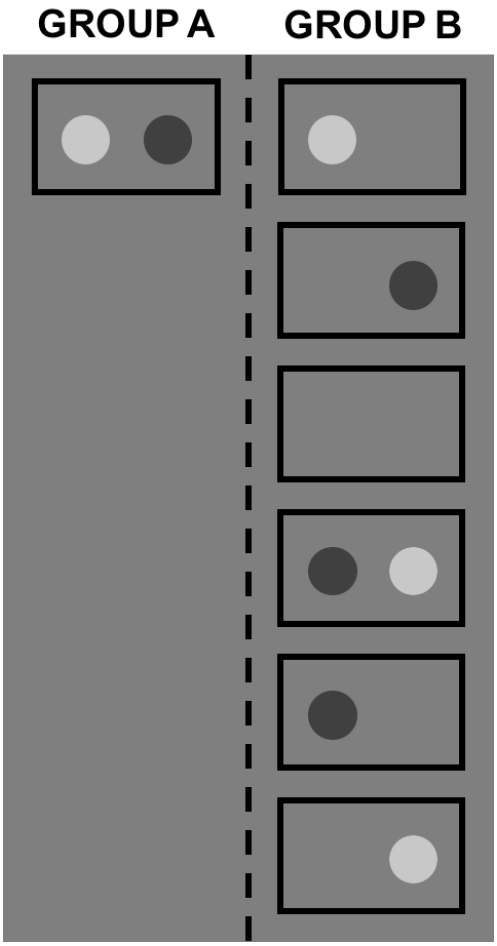

Figure 4.2: The capacity decision space. Items belonging to Group A require a "YES" response, whereas items belonging to group B require a "NO" response.

For responses greater than 5000 ms, the feedback "Too Slow" was presented and the trial was removed from the analysis.

4.3 Results

RTs less than 200 ms or greater than two standard deviations above the mean were excluded from further analysis. Less than 1% of trails were excluded using this method. The mean correct RT and error proportions for both boundary and no boundary conditions are shown in Tables 4.2 and 4.3, respectively.

Table 4.2: Observed Mean Correct and Error RTs (ms), and Error Rates for Individual Stimuli for each Participant in the Boundary Condition

		Items						
Participant	Variable	HW-HB	HB-HW	X-HW	X-HB	HW-X	HB-X	X-X
CB1	RT correct	599.96	577.10	436.85	424.54	520.66	455.99	467.28
	p(error)	0.06	*	*	*	0.01	*	*
CB2	RT correct	744.59	654.42	544.38	541.26	621.45	608.60	588.54
	p(error)	0.05	*	*	*	0.01	*	*
CB3	RT correct	409.04	433.97	306.00	307.76	399.08	351.03	319.36
	p(error)	0.08	0.05	*	*	0.03	*	*
CB4	RT correct	545.29	535.45	347.11	339.85	411.51	375.55	340.69
	p(error)	0.02	*	*	*	*	*	*
CB5	RT correct	538.08	492.41	378.42	363.90	427.47	381.95	366.57
	p(error)	0.05	0.01	*	*	0.01	*	*
CB6	RT correct	1003.40	1176.90	859.02	815.83	874.19	817.69	772.01
	p(error)	0.03	*	*	*	*	*	*
CB7	RT correct	540.03	458.30	373.35	382.98	398.95	386.91	394.74
	p(error)	0.02	*	*	*	0.01	0.01	0.01
CB8	RT correct	448.36	404.97	307.36	323.29	415.11	332.64	295.76
	p(error)	0.15	0.04	0.01	*	0.04	0.01	*
CB9	RT correct	628.21	606.45	536.07	522.00	575.66	531.58	543.23
	p(error)	0.01	*	*	*	*	0.01	*
CB10	RT correct	579.24	538.30	400.14	394.35	525.01	429.68	419.97
	p(error)	0.03	*	*	*	0.01	0.01	*

Note: * indicates error free performance; CB1 = Capacity boundary participant 1; HW-HB = stimulus which is high salience white on the left and high salience black on the right.

Capacity coefficients for the boundary condition and no-boundary condition are displayed in Figures 4.3 and 4.4, respectively. In this figure, the red line indicates the capacity coefficient, the blue lines represent 95% bootstrapped confidence intervals.

In the boundary condition, $C(t) < 1$ for the majority of the participants, providing evidence of limited capacity processing. While some participants show

Table 4.3: Observed Mean Correct and Error RTs (ms), and Error Rates for Individual Stimuli for each Participant in the No Boundary Condition

		Items						
Participant	Variable	HW-HB	HB-HW	X-HW	X-HB	HW-X	HB-X	X-X
NB1	RT correct	841.10	890.28	704.16	722.56	778.13	747.06	903.11
	p(error)	0.03	0.01	*	*	*	*	*
NB2	RT correct	688.03	470.82	387.89	370.19	467.29	396.66	487.44
	p(error)	0.09	*	*	*	*	0.01	*
NB3	RT correct	729.46	649.81	531.04	506.09	592.97	547.33	675.97
	p(error)	0.01	*	*	*	*	*	*
NB4	RT correct	756.63	738.18	652.08	602.64	672.60	632.91	741.58
	p(error)	0.03	*	*	*	0.01	*	*
NB5	RT correct	740.32	632.09	515.70	471.14	611.75	530.33	556.32
	p(error)	0.12	*	*	*	*	*	*
NB6	RT correct	670.38	553.81	387.87	378.17	476.13	435.87	416.02
	p(error)	0.05	*	*	*	*	*	*
NB7	RT correct	931.22	938.92	570.31	631.63	606.98	594.15	563.02
	p(error)	0.07	0.01	*	0.01	0.01	0.02	0.01

Note: * indicates error free performance; NB1 = No-boundary participant 1; HW-HB = stimulus which is high salience white on the left and high salience black on the right.

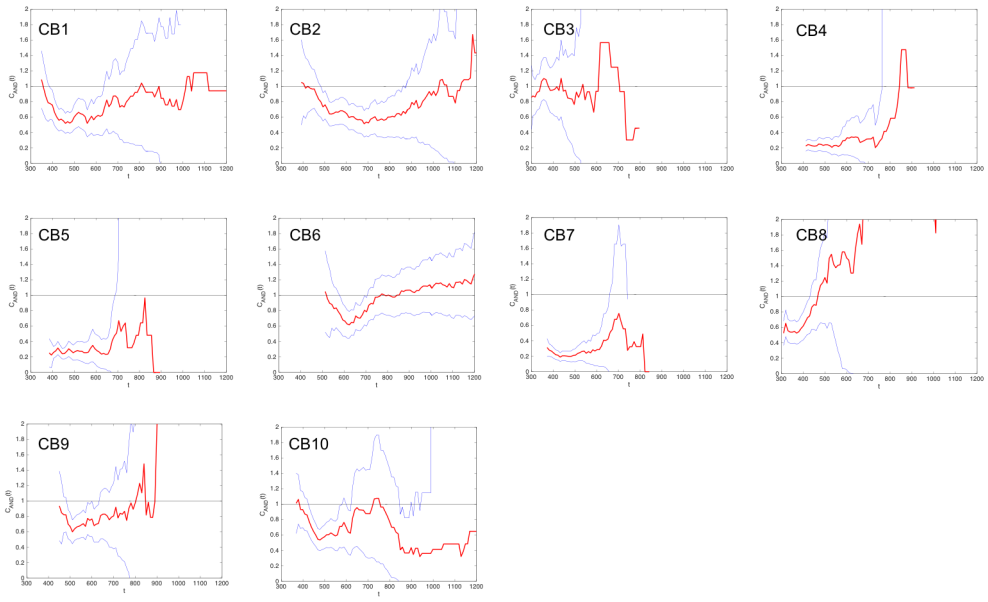

Figure 4.3: Capacity coefficients for each participant in the boundary condition. Red line indicates capacity coefficient across time, t. $C(t) = 1$ indicates the unlimited capacity parallel model and is the standard of comparison. $C(t) > 1$ indicates super capacity, and $C(t) < 1$ indicates limited capacity. Blue lines are 95% confidence intervals.

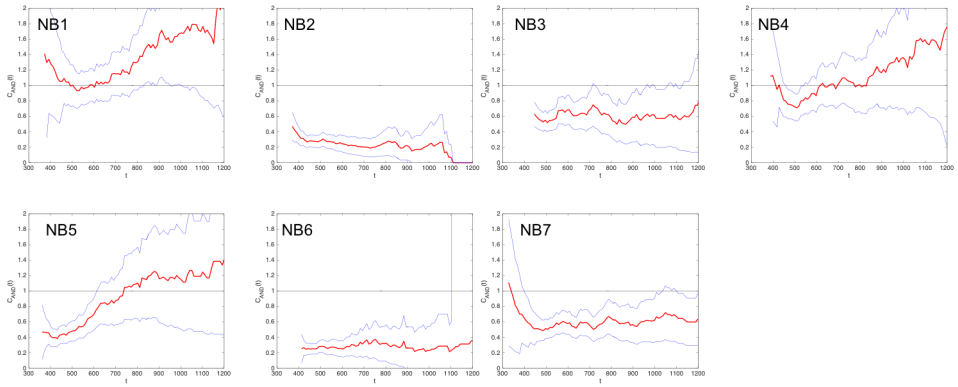

Figure 4.4: Capacity coefficients for each participant in the no-boundary condition. Red line indicates capacity coefficient across time, t. $C(t) = 1$ indicates the unlimited capacity parallel model and is the standard of comparison. $C(t) > 1$ indicates super capacity, and $C(t) < 1$ indicates limited capacity. Blue lines are 95% confidence intervals.

$C(t) > 1$ for longer RTs, this result must be interpreted with caution as there are less trials for the slower RTs. The faster RTs therefore provide a better estimation of c(t).

In the non-boundary condition, four out of seven participants clearly show $C(t) < 1$. For participant NB5, capacity is limited for faster RTs but for longer RTs demonstrates super capacity. Again, this result must be interpreted with caution as there are less trials for the slower RTs. NB1 and NB4 demonstrate a $C(t) > 1$ indicating super capacity.

4.4 Discussion

In this chapter, we have shown that the workload capacity for spatially separated luminance discs are best described as limited in nature for most individuals (15 out of 17 participants). The combination of a limited workload capacity and a coactive architecture (as shown in Chapter 3) requires further thinking regarding possible explanations for this result.

As discussed earlier, one explanation may be that the results can be explained by a pre-accumulator inhibitory parallel model. In particular, this model predicts

the combination of limited capacity with an SIC which is < 0 for small RTs and > 0 for large RTs (Eidels et al., 2011). In this type of model high levels of input into one channel lead to lower input in the other channel. As the inhibition between discs increases, the positive portion of the SIC for higher RT also increases which could explain our finding of a positive, yet non-significant MIC in Chapter 3.

This could also potentially explain why the MDS space presented in Fitousi's (2019) paper did not adhere to the standard grid assumed in the Logical Rules paradigm. In his experiment, instead of perceiving the color patches independent and as equidistant from one another, the particular combination of individual colors may have affected input to the decision-making process leading to an atypical MDS space. Although this would not be consistent with an explanation of coactivity, an inhibitory parallel model is also still inconsistent with Huang and Pashler's (2007) Boolean Map Theory, as although we have found limited capacity, the items are still processed simultaneously in parallel. This again highlights the need to decouple measures of capacity from measures of processing architecture.

A pre-accumulator inhibitory parallel account is in line with theories of lightness perception such as Gilchrist et al.'s (1999) *Anchoring Theory of Lightness Perception*. Gilchrist et al. emphasise the problem of lightness constancy, whereby the luminance values projected on to the retinal image vary only as a factor of 30-1 as a function of surface reflectance; however, under varying levels of illumination, the variation can be up to a billion to one. Despite this, individuals are able to perceive lightness with relative accuracy, meaning that despite varying levels of illumination, participants can perceive the luminance of a surface correctly. Gilchrist et al. propose that in order to accurately perceive the luminance values of a visual scene, at least one surface must act as an anchor (in this theory the highest luminance value) and the rest are perceived relative to this anchor. This would suggest, that by virtue of our biological system, it is not surprising that luminance and color would influence each other at an earlier

perceptual encoding (pre-accumulator) stage as lightness perception is inherently relative in nature.

Corbett & Smith (2017) provide a more sophisticated characterization of capacity limitations. In particular they investigated the *double-target deficit* which is the fundamental finding that when multiple locations must be monitored for the presence of a target, there are some small to moderate costs in identifying a single target compared to when monitoring only a single location. However, identification of two simultaneous targets is substantially worse than the cost predicted from divided attention alone (Duncan, 1980; Corbett & Smith, 2017). Corbett & Smith (2017) suggested that the double-target deficit is caused by a capacity limitation in access to VSTM. This prediction relies on a two-stage system which can pre-attentively select targets while rejecting non-targets. The double-target deficit therefore arises because two target representations need to be formed in the double-target case, but only one needs to be formed in the single-target case. If memory resources are represented as noisy samples shared equally among target-like stimuli, then representing more targets will naturally lead to a capacity limitation. Using a visual search task where participants searched for either one or two targets amongst two or three distractors, they found that this capacity limitation was well described by the sample-size relationship $1/\sqrt{2}$ (Palmer, 1990; Shaw, 1980) for the double-target deficit[7]. In the context of the current study, the findings of Corbett & Smith (2017) suggest that the capacity limitation is due to competition between target-like items at the point of encoding into VSTM, rather than at the decision stage. Limited capacity may therefore arise during encoding where memory resources are shared between the double target item, but remain undivided for the single target item.

Such an account would indicate a systematic violation of context invariance that is caused by limitations in the amount of perceptual information that can be sampled at this early encoding stage predicted by a sample-size account (Palmer, 1990; Shaw, 1980). Interestingly, P. L. Smith & Sewell (2013) utilize

[7]This can be generalized to $1/\sqrt{m}$ for an m-item display.

both a sample-size relationship and inhibitory interaction in their model of visual selection for briefly presented, multi-element visual displays. They propose a parallel selection process whereby relevant stimuli are selected for encoding using competitive interaction. When two stimuli are equally task relevant they mutually inhibit each other in accordance with the sample size account. However, this account could be commensurate both with an early inhibitory parallel encoding stage which has a coactive decision-making stage (indeed, this is the model used in P. L. Smith & Sewell (2013)) and also with a pre-accumultor inhibitory model with a parallel decision-making stage (Eidels et al., 2011).

However, a potential limitation to these interpretations is that the experimental design used in the current experiment may have artificially induced limited capacity processing. In the standard AND capacity task participants are presented with a single double target item which is assigned a "YES" response and two single targets and a blank target which are assigned a "NO" response. In the current design, however, there are two double target items, one of which is assigned a "YES" response and one of which is assigned a "NO" response. This is accompanied by four single target items and blank target item which are also assigned a "NO" response. This experimental design was adopted to be commensurate with the design of experiments presented in Chapter 3 whereby participants were required to process the exact luminance values of each disc. However, this created a situation whereby single targets could be potentially categorised quite quickly as belonging to Group B (as no judgement on the specific luminance value needs to take place) whereas the double target items may take longer to distinguish from one another. This would naturally cause the double target item to be slower compared to the single target items, leading to limited capacity processing. That being said, in simple detection experiments using luminance values, the common finding is also limited capacity processing (Townsend & Nozawa, 1995). The question therefore remains whether a measure of capacity which requires some further categorization or identification decision, rather than simple detection is still desirable, or whether capacity limitations are

necessarily measuring limitations which occur before the decision stage (Corbett & Smith, 2017).

One potential avenue for investigation could be an adaptation of the full identification task described in Howard et al. (2019). The main purpose of this work was to show that the absence of information (previously thought to have no effect on processing time) could affect processing capacity. In the context of the current work, if participants were utilizing the AND decision rule outlined above, the location without any stimulus must be checked and would therefore contribute to the overall processing time of the system. Alternatively, participants may not have utilized this decision rule strictly and instead may have used the absence of information to make more efficient decisions (i.e., to self-terminate where possible). To avoid this, an experiment such as that used in Howard et al. (2019) where each stimulus instead requires a unique identification (e.g., using numbers 1-8) could be utilized. This would prevent self-termination on the single target trials as both locations need to be checked in order to make an accurate identification.

4.4.1 Conclusion

The finding of limited capacity processing in the current experiments suggests that the processing architecture for categorization decisions using spatially separate luminance discs may not be best described as coactive (as was the conclusion in Chapter 3). Instead, processing can perhaps be better described as a pre-accumulator inhibitory parallel model whereby stimuli interact with one another at an early visual encoding stage. This interpretation is in line with biological theories of lightness perception (Gilchrist et al., 1999) and sample-size model accounts of visual perception (Corbett & Smith, 2017).

Chapter 5

Decision-making using visual short term memory representations

At the time of submission this chapter is in revision following peer review by Psychological Review.

In the 15 years since Wilken & Ma (2004) introduced the continuous report paradigm, the major foci in the study of change detection have been (a) the development of models that differentiate "slot-based" accounts from the use of continuous resources (see e.g., Wilken & Ma, 2004; Bays & Husain, 2008), (b) the development of mixture models which encompass elements of guessing and information-based responding (Zhang & Luck, 2008; Cowan & Rouder, 2009), and (c) the characterization of capacity limits in the visual short-term memory processes that underlie change detection (Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974). While this work has provided insight into the structure and formation of the underlying memory representations, the other primary question posed in Wilken and Ma's paper "how is information from different possible change locations integrated?" has been largely overlooked. Answering this question provides fundamental insight into the way change detection decisions are made; namely, is change detection better explained by a *first-order, differencing* information integration model (Sorkin, 1962), in which the evidence from each location is summed or pooled into a single source of evidence or as a *second-order, independent decisions* integration model, in which each location is considered in parallel and compared to individual criteria (Noreen, 1981; Shaw, 1982)?

In more detail, as applied to related tasks like detection or visual search, first-order integration models assume that each potential location or feature value is represented by a noisy strength variable of differing magnitude. Higher magnitudes correspond to stronger evidence for the presence of a stimulus. In order to make a target present or target absent decision, evidence from each location is pooled together and compared to a single decision criterion. Conversely, second-order integration models assume that information from each location is processed separately and independently. The strength of each location is compared to its own criterion. A target present response is made when any criterion is exceeded. Gardner's (1973) *independent channels* model and Smith's (1998) *separable decisions* model are both notable examples of second-order

models. If the same criterion is used at all locations, the model is equivalent to assuming that only the location or feature with the largest strength is evaluated (Koopman, 1956; P. L. Smith, 2010). That is, the probability that *any* of the locations exceeds the criterion is the same as the probability that the location with maximum strength exceeds the criterion. Graham, Kramer, & Yager's (1987) *maximum of outputs* model or *MAX* model is one such model which assumes that the location with maximum output is selected via pre-attentive filtering and then compared to a decision criterion.

These models can be instantiated as signal detection models and then compared using, for instance, receiver operating characteristics (ROCs; Wilken & Ma, 2004). A limitation of this approach is that the models only predict or account for choice probabilities or accuracy, without taking into account processing over time. Further, the predictions of the first- and second-order models make qualitatively similar ROC predictions and are therefore difficult to distinguish on the basis of detection accuracy alone. In this paper, we develop a framework for generalizing first- and second-order models of change detection across two locations to account for response time. We present a set of novel parametric models that combine first- and second-order signal detection models with alternative decision architectures (e.g., serial, parallel, or pooled) coupled with different decision stopping rules. We also introduce a novel methodology which makes use of the strong inference techniques available via Systems Factorial Technology (SFT; Townsend & Nozawa, 1995; Townsend & Wenger, 2004; Little, Altieri, et al., 2017) to provide converging nonparametric analyses to distinguish the models.

In generalizing first- and second-order models to timing, it is necessary to consider (a) how the representation of the strength of a change develops at each potential location, (b) how that information is integrated (or not), and (c) how the resulting evidence is accumulated from that representation over time to drive a decision. We draw on recent developments in perceptual categorization (Fifić, Nosofsky, & Townsend, 2008; Fifić et al., 2010; Little et al., 2011, 2013; Blunden

et al., 2015; Moneer et al., 2016; see Griffiths et al., 2017 for a review) and models that have synthesized multidimensional signal detection (e.g., Generalized Recognition Theory, GRT; Ashby & Townsend, 1986; Townsend et al., 2012) with decision-bound theories (Ashby & Gott, 1988), sequential sampling models (Busemeyer, 1985; Link, 1992; Luce, 1986; Ratcliff, 1978; Ratcliff & Rouder, 1998; Townsend & Ashby, 1983), and mental architectures (Kantowitz, 1974; Schweickert, 1992; Schweickert et al., 2012; Sternberg, 1969; Townsend, 1984).

To foreshadow our approach and the major contribution of our present work, (1) we developed a novel representation of the change signal by taking the difference between a Gaussian distribution of perceptual effects representing the memory array and a Gaussian distribution of perceptual effects representing the probe array. (2) We further allow for additional variability in accordance with the strength of the change present in the visual scene. (3) We develop a novel change detection task which allows us to test the qualitative predictions of a large number of different mental architectures and stopping rules.

Like other information integration models based on signal detection theory, our model treats the change signal as a distribution over the strength of the change at different possible locations. In a first-order model, we represent the pooled strength as a joint distribution over all possible locations. In a second-order model, we represent the change at each location independently. From these distributions we derive a change signal which is then used as the sampling rate in an evidence accumulation model.

While few models have considered the time course of change detection, some exceptions should be noted. Donkin et al. (2013) accounted jointly for choice probabilities as well as RTs using a Linear Ballistic Accumulator (LBA; Brown & Heathcote, 2008) to represent evidence accumulation for change or no-change responses. In this work, Donkin et al. (2013) developed different "front-end" models to compare slot-based and resource-based accounts of change detection (these two accounts are discussed in further detail below).[1] Lilburn (2016)

[1]We use the term "front-end" model to refer to the set of processes, mechanisms, or algorithms that are used to derive the drift rate of the evidence accumulation process. These

used a diffusion model to characterize decision processes while comparing near-threshold orientation discrimination and change detection and P. L. Smith (2016) has modeled RTs in a closely-related continuous report task using a circular diffusion process. These approaches each either assume that the information which drives decision-making is accumulated in a single decision-making channel or circumvent the need to explicitly model decision-making architecture by only requiring the retrieval of a single item. By contrast, our framework considers not only single-channel models but also parallel and serial processing architectures. A further development of our approach is that it allows us to differentiate whether the second-order integration of information considers only the maximum signal strength, terminating after a decision on that channel alone (i.e., using a self-terminating stopping rule) or whether more than one channel is considered (i.e., perhaps exhaustively checking the outcome of all channels). To assess this distinction, we consider multiple versions of the change detection task: one in which only a single change needs to be detected (an OR task) and one in which changes at all locations must be detected (an AND task). In both types of decision task, we further contrast redundant changes with changes at a single location to characterize the workload capacity of the system (Townsend & Nozawa, 1995; Townsend & Wenger, 2004; Townsend et al., 2012) along with the processing architecture. In the following, we review the treatment of first- and second-order models in the literature and introduce our modeling framework.

5.0.1 First- versus second-order integration

Wilken & Ma (2004) contrasted both first- and second-order models with a high threshold model (cf., Swets, 1961; Swets et al., 1961; Tanner & Swets, 1954a,b). High threshold models are akin to slot-based models as they assume that items are either encoded into a limited number of available "slots", or not encoded at all (Cowan, 2001), in which case that participant resorts to guessing. Across a

drift rates are then used to drive a "back-end" model which predicts the response time and choice for that trial.

number of conditions, Wilken & Ma (2004) found that hit rate decreased and false alarms increased as a function of increasing set size. The ROC data from their experiments were curvilinear, ruling out high threshold models, but were equivocal in support for a first- or second-order model. In particular, although their curvilinear ROC data ruled out a high threshold model, the ROC results were unable to differentiate the independent and pooled differencing models. Further, although the second-order model quantitatively fit better in some instances (e.g., using χ^2), the pooled model was quantitatively better in other conditions. In their first-order model, termed the *Sum of Absolute Differences model* (SAD) model, Wilken & Ma (2004) assumed that participants monitor noisy information from a number of different locations or features (representing dimensions along which a possible change may occur). In these models only the absolute difference of the change is important, not its sign (e.g., a change from a red to green is treated the same as a change from green to red). If an item in the display changes, the signal in that channel increases. In order to make a decision, the participant sums the change strength signal across all locations. If this summed signal exceeds a decision threshold, a change response is emitted. The pooled signal was modeled as a sum of normal random variables across all locations. When a change is present (in at least one location), the average value of the pooled signal is increased relative to when a change is absent. The effect of increasing the number of distractors causes an additional increase in the overall mean and variance of both the change present and change absent distributions. False alarms can occur in this context due to the variability in the change absent strength distribution, that is the pooled signal can exceed the change threshold on the basis of noise alone.

In their second-order model, Wilken & Ma (2004) assume that a participant monitors the noisy change signal from a number of different locations independently. If any of these signals exceeds its respective threshold then a change response is emitted. This model is equivalent to Shaw's (Shaw, 1980) independent decisions model. Wilken and Ma term this model the *maximum*

absolute difference model (MAD); however, it is equivalent to the *minimum evidence* integration rule in signal detection theory (Macmillan & Creelman, 2004; see Figure 5.1, bottom right panel).[2]

To contrast these integration rules, it is useful to consider how a variable representation gives rise to hit and false alarm rates. The top panel of Figure 5.1 shows a two dimensional signal detection representation. The circles centered at the origin are the iso-probability contours of a bivariate normal density with a mean of $(0, 0)$. This distribution represents the variable strength when there are no changes in the display. By contrast, the second distribution (in the top right) represents the variable strength when there are changes in both locations; the mean of this distribution is (d'_x, d'_y) where d' is the difference between the means at each location. It is evident that if separate freely-varying thresholds are established on each dimension, the bivariate representation is equivalent to Wilken and Ma's (2004) SAD model. The SAD model, in effect, samples a vector of change strengths from the model shown in Figure 5.1, sums the values on each dimension, and compares this to a single threshold. The integration within each region of the space in Figure 5.1 can be made equivalent to comparing the sum to a single threshold. That is, without further empirical constraint, you cannot tell these models apart using accuracy or choice probability.

The univariate marginal distributions shown in Figure 5.1 (top panel) represent the strength of change in each location independently. In Wilken and Ma's (2004) MAD model, separate decision criteria are established along each location. Change strengths are sampled from each independent distribution and compared to the individual thresholds. This is equivalent to integration using a minimum evidence rule (see Figure 5.1, bottom right).

When applied to change detection (as for example in Wilken & Ma, 2004), the evidence signal is modeled as a difference between the study and test arrays. The differencing approach was initiated by Hefner (1958) as a method of scaling

[2]The difference in terminology likely arises because Wilken & Ma (2004) emphasize the strength of the signal; whereas Macmillan & Creelman (2004) emphasize the amount of evidence required to make a decision. Hence, a maximum strength rule is equivalent to a minimum evidence rule. In this paper, we use the terminology which emphasizes evidence.

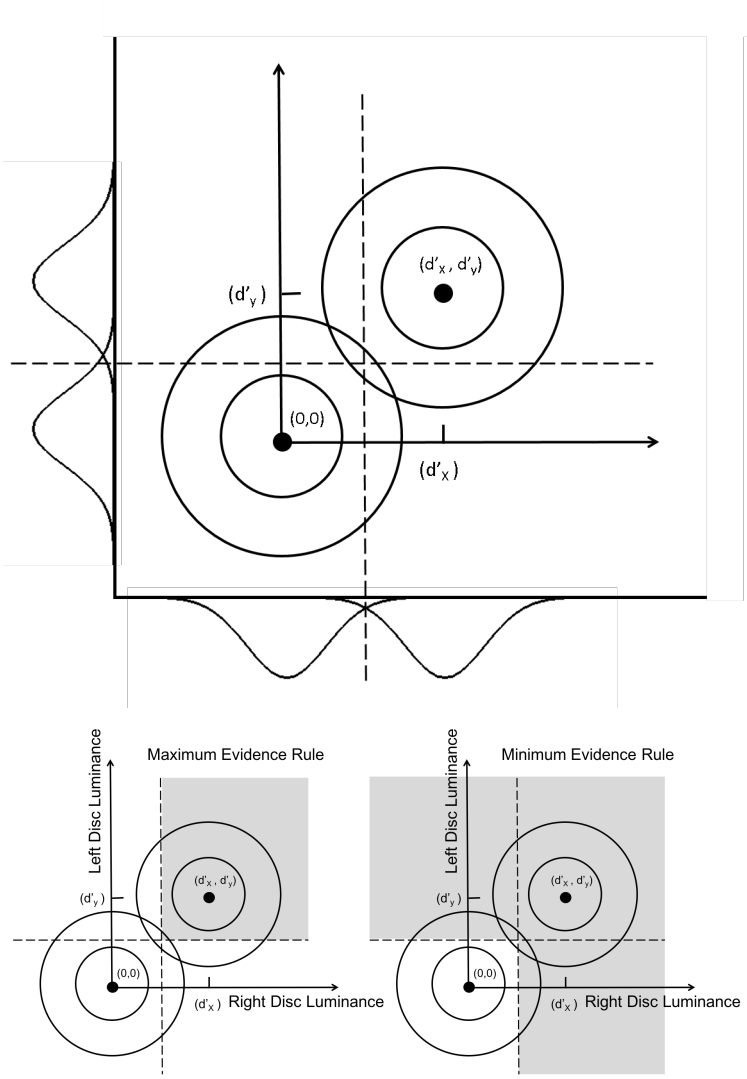

Figure 5.1: Top panel: A two dimensional signal detection representation for a task requiring detection of a change in one of two locations. The point (0,0) represents the variable strength at the point of origin and the point (d'_x, d'_y) represents the variable strength when there are changes in both locations. Circles surrounding these points represent isoprobability contours of a bivariate normal density. The dotted line represent decision thresholds. Bottom left panel: The maximum evidence integration rule (Macmillan & Creelman, 2004). Bottom right panel: The minimum evidence integration rule (Macmillan & Creelman, 2004). The shaded regions represent the area which must be integrated to determine the evidence for a change decision. Note: The minimum evidence integration rule is equivalent to Wilken and Ma's (2004) terminology of a maximum strength rule.

differences between dimensions. Hefner assumed that each study and test value was drawn from a normal distribution and that the difference between these values was compared to a threshold. Sorkin (1962) extended this approach to same-different auditory judgments. In Sorkin's model, differences were represented both above and below zero such that the direction of the change was also available as part of the representation. That is, two criteria were applied depending on whether the test tone was incremented or decremented from the study tone. A less general but potentially more parsimonious account of the change detection signal is to represent the absolute change between the study and the test arrays. Such a representation can be modeled using a folded-normal distribution (Leone et al., 1961). This is the approach taken in the current paper.

Let the signal at each presentation location, L, be represented by a normally distributed random variable capturing the idea that the perceptual signal from each location is noisy:

$$L_i \sim N\left(\mu_{L_i}, \sigma_{L_i}\right) \tag{5.1}$$

where μ_{L_i} and σ_{L_i} are the mean and standard deviation of the perceptual distributions. The subscript i denotes whether the presentation of the stimulus at that location was in the memory array or the test array (e.g., time 1 or time 2). For each location, we assume that the representation of the change signal is provided by the distribution of the absolute differences at time 1 and time 2, $|Z| = |L_1 - L_2|$. We model this as a folded-Normal distribution:

$$|Z| \sim \text{folded-}N\left(\mu_{L_1} - \mu_{L_2}, \sqrt{\sigma_{L_1}^2 + \sigma_{L_2}^2}\right). \tag{5.2}$$

In the following, we implement both the minimum evidence model and the pooled model using either independent folded-normal distributions at each location or using a joint multivariate folded-normal distribution. In order to differentiate these models, we take a different approach to Wilken & Ma (2004) by generalizing the models to the time domain and designing experiments that contrast the response time predictions of the models. While Dai et al.

(1996) showed that a model assuming independence was optimal when there was a fixed set of possible change detection stimuli and when the test display was not correlated with the study display, when a test change varied as a increment or decrement of the study display, a pooled differencing model is optimal. Despite these differences, individual performance often falls short of these optimal strategies (Noreen, 1981) and as such it is not clear whether a first- or second-order model will be best describe participant performance in the current paradigm. We now turn to a more detailed consideration of how reaction time can be used to effectively differentiate first- and second-order integration models.

5.0.2 Using reaction time to differentiate integration models

When we begin to consider how change detection decisions play out over time, the question of first- and second-order models necessarily becomes more complicated. First, let us consider second-order models. In this class of models the signal strength representing a change for each location is compared to an individual criterion, and if the strength of the variable exceeds the criterion, then a target present response is made. Importantly, decisions are made based on each location separately and independently.

Let us consider a task where a participant must detect a change in luminance in either, or both, of two discs present in the display as this is the task used in the current experiments (see Figure 5.2). If we consider second-order models in their general form, we can see that independence over time can be thought of in two ways. One may process the discs simultaneously but in independent parallel channels, or one may process the discs one at a time (i.e., in serial) in independent channels. Further, returning to Figure 5.1, evidence from each disc can be integrated in two ways within each of these architectures. It may be that evidence from both discs must be incorporated in order to make a decision (i.e., a *maximum* integration rule is used). Alternatively, evidence from only one

disc may need to be incorporated in order to make a decision (i.e., a *minimum* integration rule is used). Both the minimum and maximum integration rules can be implemented in both parallel and serial cases. The minimum and maximum integration rules also map on to self-terminating and exhaustive decision rules, respectively. For example, a parallel model with a minimum integration rule would be equivalent to a parallel self-terminating model. This is because RT is determined by the fastest of the two channels and only information from a single disc is therefore being used in order to make a decision. A parallel self-terminating model is logically consistent with a second-order model in its general form (although importantly, not equivalent to the MAX strength model, a distinction which we will turn to momentarily). Conversely, a parallel model with a maximum integration rule would be equivalent to a parallel exhaustive model. In this model, RT is determined by the slowest of the decision channels, since both discs must be processed before a decision can be made. This represents a more conservative, but nonetheless independent, decision-making process.

A similar mapping applies in the case of the serial model. A serial model with a minimum integration rule would be equivalent to a serial self-terminating model. If a decision can be made after processing the first location, then the process terminates. If a decision cannot be made after processing the first location, then the process continues to the second location. Again, this is logically consistent with second-order models in their general form. Conversely, a serial model with a maximum integration rule would be equivalent to a serial exhaustive model. That is, a participant would complete processing of the first location which is then necessarily followed by the processing of the second location, again representing an independent, albeit conservative decision-making process.

For first-order models, evidence from each disc is summed or pooled into a single source. If we consider decision-making over time, a first-order model would be equivalent to a coactive model where information from both locations is pooled into a single decision-making channel. Although this pooling results in a single decision-making channel, this class of models can also incorporate

both integration rules. The information which drives a decision can be either integrated using the maximum integration rule (left panel of Figure 5.1) or using the minimum integration rule (right panel of Figure 5.1).

A MAX model predicts that decisions are determined by a process in which only the strongest signal drives decision-making, that is, a single channel decision model where only the signal with the maximum strength contributes to the decision. For accuracy or detection probabilities, a parallel self-terminating model is equivalent to a MAX model (Koopman, 1956); however, at the level of RTs, only the strongest location is actually processed. Hence, the predicted RTs from a MAX model would be different from a parallel model, in which the weaker channel would influence the response time, and also different from a coactive model, in which the weaker channel would be integrated into the change signal. However, the extent of the qualitative difference is unclear; we consequently include a MAX model in our comparison set.

Logical rules applied to the detection of changes

In order to model each combination of integration rule with each architecture, we extended a framework initially developed to model response times in perceptual categorization, namely the Logical-Rules models (Fifić et al., 2010). In categorization, the goal of these models is to characterise decisions that require the integration of multiple sources of information. The framework assumes that individuals make categorization decisions based on individual dimensions, which are then combined via logical rules to form a final categorization decision. As explored earlier, these assumptions are similar to those of first- and second-order integration models of change detection decision-making. However, in the current work, rather than making a categorization judgment, participants are instead required to make a "change" or "no-change" response.

To generalize the representation of perceptual variability to the representation of change variability, we assume that there are two distributions of perceptual effects: one representing the memory array and one representing the probe

array. In order for a comparison to be made, we take the absolute difference of these two representations, leaving us with a folded bivariate normal distribution representing the degree to which a stimulus has changed which we term the *change distribution*.[3].

The evidence rates (i.e., drift rates for the "change" and "no change" accumulators) are generated by integrating the folded normal change distribution with respect to the relevant decision rule. For the maximum rule (Macmillan & Creelman, 2004), this would involve integrating over the region of space where both dimensions have changed (as shown previously in Figure 5.1). For the minimum rule, this would involve integrating over the region of space where only one dimension has changed (see Figure 5.1). In serial and parallel models, it is assumed that individuals make separate decisions along each dimension. In the current framework, this gives four accumulators; one representing "change" and one representing "no change" for each of the two dimensions. The decision time in a serial exhaustive process is the sum of the time it takes each accumulator pair to reach a decision. For a serial self-terminating model, the decision time is the time taken for the first accumulator pair to complete, providing a decision can be logically made. If a decision cannot be logically made, then the second dimension is processed, and the decision time is the sum of time taken for each pair to reach a decision. For a parallel exhaustive model, decisions take place simultaneously. Hence, the decision time is the slowest time it takes for both accumulator pairs to make a decision. Conversely, in a parallel self-terminating model the decision time is determined by the faster of the two dimensions, providing a decision can be logically made. For the coactive model, the perceptual information from each dimension is pooled and decision time is determined by a single pair of accumulators. Although there is only one pair of accumulators for this model, the drift rate can still be determined by using either a maximum or minimum integration rule giving two alternative coactive models which we term *coactive max* and *coactive min*, respectively.

[3]This idea of a change distribution has also been theoretically (although not empirically) explored by e.g. Liesefeld et al. (2017)

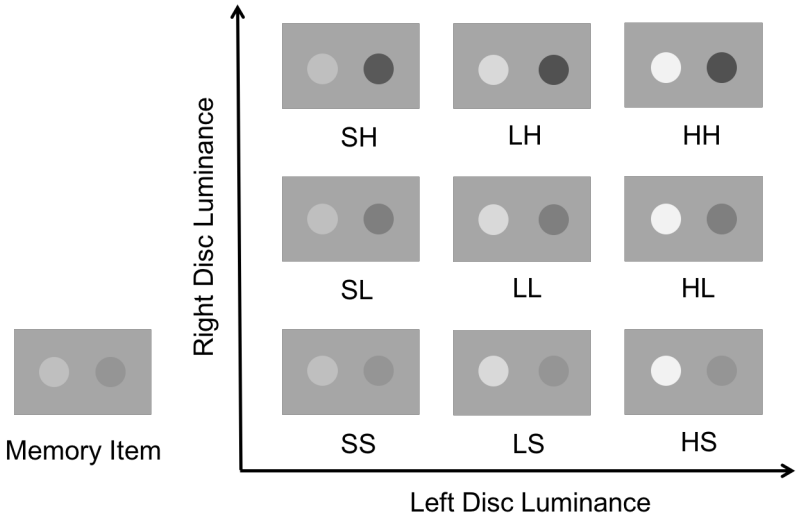

Figure 5.2: Schematic of stimulus space with an example memory array and the corresponding trial types. S denotes same or no change; L denotes low magnitude change; H denotes high magnitude change. These are combined to form nine possible change types. For example, LH indicates that a low magnitude change in the left channel and a high magnitude change in the right. In practice, any item from the nine item stimulus space could appear as either the memory item or the probe item.

An example of our change detection variant of the double factorial design (cf. Fifić et al., 2010) is shown in Figure 5.2. While the double factorial paradigm has previously been implemented using single items which comprise both dimensions (e.g., halves of a face, Cheng et al., 2017; or two parts of a lamp, Fifić et al., 2010), recent work has used within-dimension feature stimuli which comprise different values of the same feature (discs of varying luminance) at different respective locations in space (Blunden et al., 2020).

Following the work of Blunden et al. (2020) in categorization, in our one-shot change detection task, participants are presented with a memory and probe array, both containing two luminance discs separated in space and are asked to respond "YES" if any of the discs have changed (for the OR task), or alternatively to respond "YES" only if both items have changed (for the AND task). The double factorial design combines two experimental manipulations: one affecting workload - in which two changes are presented instead of one - and one affecting

within channel processing speed - changes can be easy or hard to detect[4].

To manipulate workload we systematically varied the number of changes that could occur on any given trial. That is, on each trial there may be no changes, one change, or two changes. Further, to manipulate the processing speed, we varied the magnitude of each change. Changes could either be high salience (H) or low salience (L). The orthogonal combination of these two manipulations gives a nine item stimulus space comprising of two dimensions (left and right disc) each which vary over three levels (no change, a low salience change, and a high salience change; see Figure 5.2). We assume that a high magnitude change provides a stronger signal than does a low magnitude change. A stimulus which comprises two high salience changes (e.g., stimulus HH in Figure 5.2) will therefore involve a faster comparison process than a stimulus with one high salience and one low salience change (e.g., stimulus HL). Hence, stimulus dimensional values of either high discriminability (H) or low discriminability (L) combine to form four stimuli which are defined by their relative detection difficulty: HH, HL, LH, and LL.

Similar work by Thiele et al. (2017) has used SFT in the context of change detection to test whether the architecture mediating perception was the same as that mediating working memory tasks for object features. They used simple screw head stimuli which could differ in size and orientation of the slot, and participants were required to state whether both features had changed, or if at least one had remained the same. In the perception condition, participants were presented with two screwheads simultaneously, side-by-side, whereas in the working memory condition the screwheads were presented sequentially. Thiele et al. (2017) found serial processing across both tasks, suggesting that the architecture mediating both perception and working memory tasks were the same. Whether this finding applies to how information from multiple items is

[4]Although our double factorial change detection design manipulates whether there are zero, one or two changes in the display, it is the case that for a single change, the other location remains the same. Hence, it is a misnomer to call this a manipulation of workload since the location which has not changed may provide some evidence which affects processing. This type of design has been explored in Little et al. (2015, 2018). For present purposes, the important comparison is the relative speed of processing of the redundant change targets and the single change trials, which is computed via the capacity measure.

integrated in change detection remains to be seen.

Further, work by C.-T. Yang (2011) as well as C.-T. Yang et al. (2011) and C.-T. Yang et al. (2013) have also used SFT in the context of change detection. Again, using a single-item display, they investigated how various task manipulations affected attentional deployment and, as a consequence, processing architecture. For example, C.-T. Yang et al. (2011) asked participants to detect contrast and orientation changes in a single Gabor patch. The experiment consisted of two conditions: equal and unequal salience changes. In the equal salience condition, psychometric testing was done prior to the task to equate the discriminability of orientation changes with contrast changes. However, in the unequal salience condition, a single set of change magnitudes was used for all participants, resulting in the relative salience of the two changes differing across participants. In the equal salience condition, performance was uniformly parallel, while in the unequal salience condition performance varied between serial and coactive.

Further, C.-T. Yang et al. (2013) found that change probability also affected the response architecture. In cases where spatial frequency changed more often than did orientation, the two changes were processed serially. However, when a change in either was equally likely, parallel processing was more likely to be adopted. This shift from serial to parallel occurred despite participants reporting that they were unaware of using information regarding the relative probability of a change to complete the task (see also Chang et al., 2016). Additionally, when there was a lack of cognitive resources, participants reverted to parallel processing even when relative change probability was not equated. While these studies provide further insight into how varying dimensional change probability and salience within a single object affects processing architecture, our main question is even more fundamental. In our study we hold constant salience and change probability, and instead look simply at the architecture of combining information from multiple locations in change detection decisions.

Although the main focus of this paper is quantitative model fitting, we also

consider the qualitative non-parametric predictions afforded by SFT (Townsend & Nozawa, 1995). These predictions provide a powerful diagnostic test for mental architecture as they are parameter and distribution free. In short, these analyses utilize the empirical RT distributions in different contrasts to provide strong, qualitative distinctions between the predictions of different model classes. We describe each of these analyses along with the experiments that are designed to test them. In Experiment 1, we examine the architecture underlying change detection in OR and AND tasks, respectively. In Experiment 2, we examine measures of workload capacity.

5.0.3 Current Experiment

In Experiment 1, we sought to diagnose the processing architecture of change detection decisions using two luminance discs of different polarity separated in space (see Figure 5.2). We limited the display size to two to avoid exceeding a potential item capacity limit (see e.g., Sewell et al., 2014). We investigated both an OR and an AND task. In the OR task, participants needed to detect any change in the probe array, whereas in the AND task, participants needed to verify that both items had changed. It is possible that participants may engage in different decision-making/integration strategies across these two tasks. To diagnose these strategies, we tested a wide variety of models consistent with both first- and second-order accounts which predict both choice probability and RT across the entire distribution.

By characterizing the processing architecture of change detection decisions we will answer Wilken and Ma's (2004) question of whether information integration is best described by a first- or second-order integration model. Further, we will provide a detailed characterization of the decision stage independent from other processes such as encoding and storage capacity.

5.1 Experiment 1

5.1.1 Method

Participants

Eleven individuals with normal or corrected-to-normal vision (visual acuity of at least 20/25 as determined by a near-field Snellen eye chart; normal color vision as determined by Ishihara plates) from the University of Melbourne community participated in the study. Participants from the OR condition are labeled O1 - O4 (O denoting "OR"), and participants from the AND condition are labeled A1 - A5 (A denoting "AND"). Participants were recruited via advertising placed on notice boards within the Melbourne School of Psychological Sciences and through the school's online recruiting system. All participants were naïve to the purpose of the experiment. Six participants were assigned to the OR condition and five participants were allocated to the AND condition. Two participants (one from the OR condition and one from the AND condition) did not complete all sessions and so were excluded from the study. Inspection of the data showed that one participant from the OR condition was not following the task instructions - this participant only considered changes at one location and performed poorly overall - and was consequently also removed from the analysis. All participants were reimbursed $10 per session. Testing was approved by the Melbourne Human Research Ethics Committee (Approval Number 1034866).

Apparatus and Stimuli

Stimuli were two discs subtending a visual angle of 1.91° at 60 cm viewing distance presented on a gray background (RGB color space value [128 128 128]) with centers 11.34° of visual angle to the left and right of fixation (center of the screen). The disc on the left was always an increment relative to the background while the disc on the right was always a decrement. Each disc could take on one of three different luminance levels. In RGB color space these were [228 228 228; 192 192 192; 156 156 156] for the left disc and [40 40 40; 76 76 76;

112 112 112] for the right disc. These values were combined orthogonally to create a 3 x 3 matrix of possible stimuli presented in Figure 5.2. Both memory items and probe items were drawn from this display giving nine possible change types (see Figure 5.2), although the individual memory item and probe item varied. On a given trial, there could be no change, one change, or two changes between the memory and probe array. For change trials, luminance increased or decreased by one step (a low discriminability change) or two steps (a high discriminability change). For example, when considering only the left disc, a one-step change may involve a decrement from the highest level of luminance [228 228 228] to the middle luminance level [192 192 192]. Alternatively, it may involve an increment from the lowest level of luminance [156 156 156] to the middle level of luminance [192 192 192]. If the memory item was the middle level of luminance [192 192 192] this could then be either incremented to the highest level of luminance [229 228 228] or decremented to the lowest level of luminance [156 156 156] for the probe item. A two-step change, however, could only involve a decrement from the highest level [228 228 228] to the lowest level [156 156 156] and vice versa. In total, this gave 16 unique stimulus combinations for LL trials, eight unique stimulus combinations for HL and LH trials respectively, and four unique stimulus combinations for the HH item. To ensure equal numbers or trials for each stimulus type we used repetition of certain trial types. Change and no-change trials were equated in number. Frequencies for all unique trial types and their repetitions are listed in Table 5.1. RTs were collected using a calibrated response time box (Li et al., 2010).

5.1.2 Procedure

The task was a one-shot change detection task with either an OR or an AND response rule. For the OR condition, participants were able to respond "YES" if either or both of the discs changed. For the AND condition, participants responded "YES" only if both discs changed.

Participants in the OR condition completed five sessions each. Participants

Table 5.1: Trial frequencies per block and total presentation numbers for each trial type for Experiment 1.

	Trial Types								
	Double Change Trails				Single Change Trials				
	LL	LH	HL	HH	SL	LS	SH	HS	SS
Number of Unique Items per Block	16	8	8	4	12	12	6	6	9
Number of Repetitions per Block	1	2	2	4	1	1	2	2	12
Total per Experiment	160	160	160	160	120	120	120	120	1080

in the AND condition completed six sessions[5]. Sessions began with 15 practice trials, followed by two blocks of 220 trials each. Trials were randomized within each block. At the start of each session participants were shown a diagram of the stimulus space displaying all nine possible stimulus displays. Participants were allowed a short break every 74 trials. Each trial began with a fixation cross for 500 ms, followed by the memory array for 1500 ms. The memory array was then masked for 1000 ms, followed by a retention interval of 1000 ms. Finally, the probe array was presented for 5000 ms or until a response was made.

For the OR condition, upon presentation of the probe array, participants were asked whether the probe was the same as or different to the memory array. They responded "DIFFERENT" by pressing the right button on the RT Box and "SAME" by pressing the left button on the RT Box. For the AND condition, upon presentation of the probe array, participants were asked whether *both* objects were different. Participants responded "YES" by pressing the right button and "NO" by pressing the left button. Participants were then told if their response was correct, incorrect, or too slow (in the event of taking longer than 5000 ms to respond). Feedback was presented for 1000 ms. A schematic illustration of a single trial is shown in Figure 5.3. Participants were given feedback regarding their overall accuracy at the end of each session.

[5]This was done as a matter of convenience

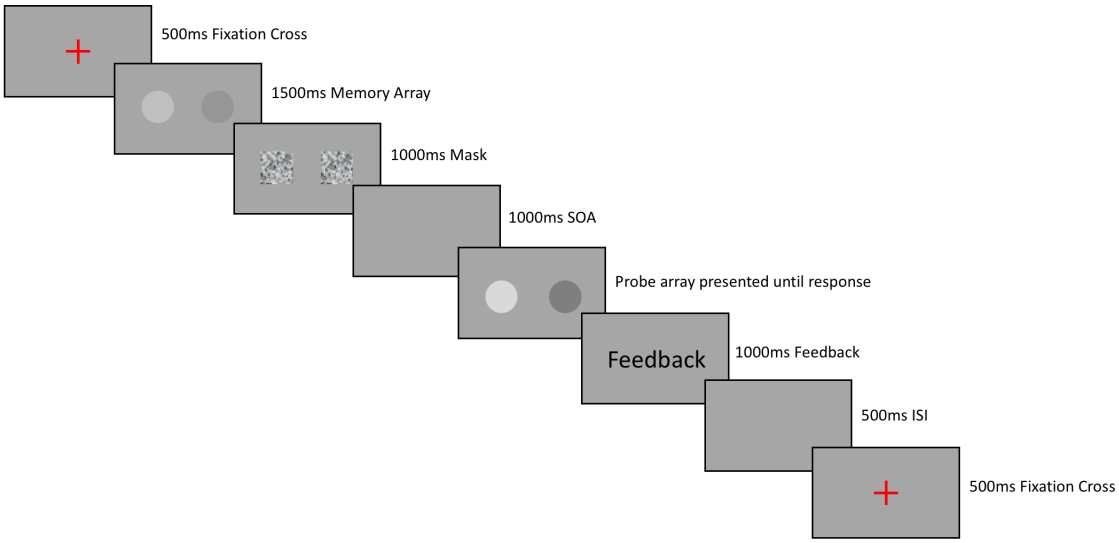

Figure 5.3: Schematic Representation of a Trial

Data Analysis

In addition to quantitative model fitting we also used the qualitative non-parametric analyses afforded by SFT (Townsend & Nozawa, 1995). The mean interaction contrast (MIC) and the survivor interaction contrast (SIC) are two useful measures used in SFT for diagnosing processing architecture. The MIC is calculated as follows:

$$MIC = (\bar{RT}_{LL} - \bar{RT}_{LH}) - (\bar{RT}_{HL} - \bar{RT}_{HH}) \tag{5.3}$$

where $\bar{RT}_L L$ is the mean RT from the item condition which comprised two low salience changes (LL). LH comprised a low salience change on the left and a high salience change on the right, HL comprised a high salience change on the left and a low salience change on the right, and HH comprised two high salience changes.

An additive pattern ($MIC = 0$) is evidence for a serial process; this is true for both self-terminating and exhaustive models. An under-additive MIC

($MIC < 0$) is indicative of a parallel exhaustive model and an over-additive MIC ($MIC > 0$) is predicted by both a parallel self-terminating model and a coactive model. While on its own the MIC provides a simple measure of processing architecture, when combined with the full RT distribution (the SIC which is outlined below), the MIC further provides a useful means of statistically assessing the interactions seen in the SIC function.

The SIC provides a more sensitive contrast which utilizes the entire RT distribution. The SIC is defined as follows:

$$SIC(t) = [S_{LL}(t) - S_{LH}(t)] - [S_{HL}(t) - S_{HH}(t)] \tag{5.4}$$

where $S(t)$ is the survivor function for a specific item condition. The SICs for serial, parallel, and coactive models as well as the relevant stopping rules are shown in Figure 5.4. While both serial self-terminating and serial exhaustive models predict an $MIC = 0$ the SIC functions have very different forms. The serial self-terminating models give an entirely flat curve (i.e., every time point is equal to zero) whereas serial exhaustive models predict an initially negative function which becomes positive, with the entire function integrating to zero (i.e., since the integral of the survivor function for an item equals the mean RT for that item; Townsend, 1990b, the integral of the SIC equals the MIC).

For serial self-terminating models, only one item need be processed before a decision can be made. This means that only one dimension contributes to the overall RT. Hence, the functions cancel out leaving a flat SIC curve. For serial exhaustive models on the other hand, an additive pattern arises, as both LH and HL items will show some slowing relative the to HH item due to their lower discriminability on one of the dimensions. The increase of RT for the LL item compared to the HH item is therefore simply the sum of the individual sources of slowing.

Parallel exhaustive models predict an entirely negative function, whereas parallel self-terminating models predict an entirely positive function. For parallel

exhaustive models, RTs for the double change items are determined by the slower of the two decisions (i.e., the maximum processing time). The LH and HL item RTs will be therefore much longer than the HH item RTs. The LL item RTs will be only slightly longer than either the LH or HL item RTs, leading to an entirely negative SIC function. The converse is true for parallel self-terminating models as RT is determined by the fastest of the two decisions. This means that the HH, HL, and LH item RTs are all much shorter than the LL item RTs.

Coactive models predict an initial negative deflection in the SIC function, with the majority of the SIC being positive for longer RTs. Coactive models therefore do not integrate to zero but rather integrate to a positive value. This over-additive pattern of results has been shown by Townsend & Nozawa (1995) in their mathematical proof, and this finding has been corroborated by simulations done by Fifić, Nosofsky, & Townsend (2008).

The predictions outlined above are non-parametric in that they do not depend on the particular form of the RT distributions and can be computed directly from the data; hence, the qualitative contrasts apply to the entire class of serial, parallel, and coactive models.

The SIC and MIC analyses rely on the assumption of *selective influence* (Dzhafarov, 2003; Schweickert et al., 2012; Townsend & Schweickert, 1989). Selective influence implies that the experimental manipulation of salience in a single location only affects a single channel within a mental architecture. When testing architecture, the mean RTs for the high level channel should be lower than the mean RTs in the low level channel. In order to interpret the MIC and the SIC functions, it is therefore necessary that the survivor functions are ordered such that $S_{HH}(t) \leq S_{HL}(t) \approx S_{LH}(t) \leq S_{LL}(t)$ with the strict inequality holding for at least one time point (Townsend & Nozawa, 1995). If this ordering is maintained, *stochastic dominance* holds and the SIC and MIC are interpretable.

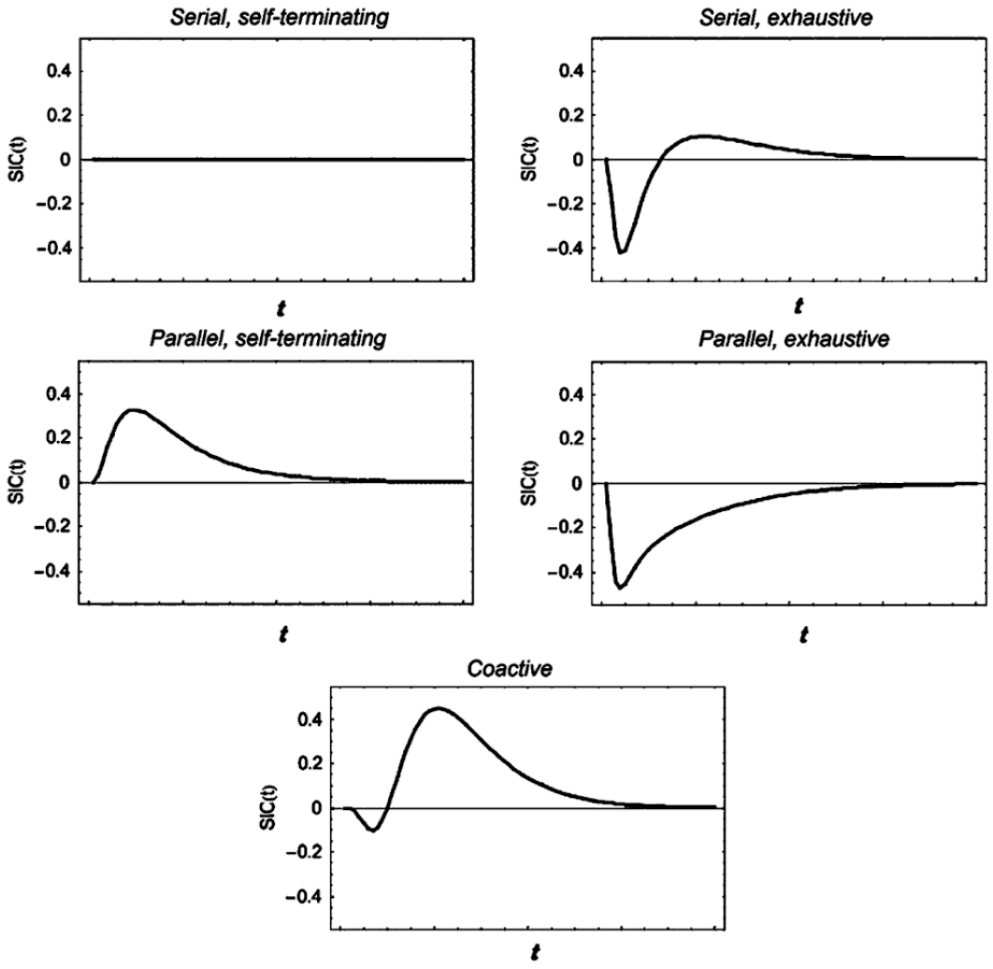

Figure 5.4: SIC predictions for the serial, parallel, and coactive models. The top panel displays SIC predictions for the serial self-terminating and serial exhaustive models, the middle panel displays SIC predictions from the parallel self-terminating and parallel exhaustive models, and the bottom panel displays SIC predictions from the coactive model.

5.1.3 Results

For all participants, the first session was excluded as practice. Additionally, RTs less than 200 ms or greater than 3000 ms were excluded. In total, less than 3% of trials were removed using this method. Mean correct RTs, mean error RTs, and error rates for the OR Task are presented in Table 5.2 and for the AND Task in Table 5.3.

Although some items showed a higher error rate, Townsend & Wenger (2004) support the robustness of the SIC functions up to an error rate of 30%. This is further supported by the simulations of Fifić, Nosofsky, & Townsend (2008) who demonstrate that estimates of the SIC are robust to violations of the assumption of perfect accuracy. More recent investigation have shown that the so long as stochastic dominance is preserved, the SIC conditioned on correct responding is robust to any error level (H. Yang et al., 2019).

Table 5.2: Observed Mean Correct and Error RTs (ms), and Error Rates for Individual Stimuli for each Participant in Experiment 1: OR Task

		Items								
Participant	Variable	HH	HL	LH	LL	SH	SL	HS	LS	SS (R)
O1	RT correct	539.23	616.46	577.8	685.6	634.67	732.63	619.51	739.82	619.71
	RT error	1051.4	681.12	720.4	772.79	767.85	664.42	787.59	684.72	790.15
	p(error)	0.039	0.057	0.065	0.296	0.158	0.723	0.077	0.372	0.084
O2	RT correct	560.69	579.71	645.07	774.51	578.91	805.71	681.48	939.53	755.32
	RT error	489.68	-	-	1304.4	2022.6	952.88	943.1	930.37	1170.8
	p(error)	0.008	-	-	0.073	0.011	0.379	0.021	0.170	0.026
O3	RT correct	479.38	522	504.13	581.6	511.99	645.35	495.38	590.79	519.07
	RT error	764.44	627.79	494.49	563.31	671.47	562.68	645.83	544.22	657.02
	p(error)	0.008	0.074	0.024	0.304	0.095	0.547	0.056	0.394	0.148
O4	RT correct	465.96	515.33	493.83	598.43	522.58	719.74	504.74	710.67	623.1
	RT error	-	-	-	658.18	-	646.48	-	600.41	923.91
	p(error)	-	-	-	0.008	-	0.372	-	0.087	0.018

Note: - indicates error free performance; O1 = participant 1.

A series of Kolmogorov-Smirnov tests (KS tests; Houpt et al., 2013) were used to check that each participant's survivor functions followed the necessary ordering to meet the assumption of stochastic dominance. If the assumption of stochastic dominance holds, the first four columns of Table 5.18 should be significant, whereas the last four should not. Table 5.18 and Figures 5.20 and 5.21, depicting the survivor functions for OR and AND tasks respectively are shown in the Appendix.

Table 5.3: Observed Mean Correct and Error RTs (ms), and Error Rates for Individual Stimuli for each Participant in Experiment 1: AND Task

		Items								
Participant	Variable	HH	HL	LH	LL	SH	SL	HS	LS	SS (R)
A1	RT correct	696.82	703.91	688.1	779.27	766.84	732.86	761.25	738.87	687.47
	RT error	779.76	755.7	786.28	753.59	810.91	807.93	711.97	782.6	844.44
	p(error)	0.07	0.283	0.077	0.507	0.574	0.221	0.943	0.424	0.131
A2	RT correct	623.16	793.2	749.42	931.63	712.66	814.24	758.14	807.11	678.6
	RT error	896.21	995.75	837.75	948.01	994.78	1070.1	896.49	942.48	1164.8
	p(error)	0.013	0.052	0.2	0.241	0.023	0.086	0.121	0.127	0.015
A3	RT correct	924.56	1102	1126.2	1325.1	1000.4	1035	902.64	907.97	799.61
	RT error	733.24	1245.00	1231.90	1379.60	1199.70	1335.60	1406.30	1241.20	914.73
	p(error)	0.004	0.051	0.096	0.181	0.079	0.034	0.125	0.092	0.008
A4	RT correct	967.46	1295.6	1228.8	1447.1	1143	1123.8	1021.6	952.53	917.39
	RT error	1004.90	1410.60	1442.80	1399.50	1407.60	1772.40	1421.2	1395.60	1032.60
	p(error)	0.021	0.085	0.140	0.160	0.086	0.062	0.067	0.068	0.008
A5	RT correct	1164.5	1462.5	1492	1710.8	1433.9	1479.7	1888.5	1794.5	1498.3
	RT error	1977.8	1821.3	1871.9	1841.5	1633.7	1944.7	1768.3	1852	2090.9
	p(error)	0.009	0.067	0.219	0.307	0.144	0.159	0.271	0.295	0.041

Note: - indicates error free performance; A1 = participant 1.

For the OR Task, both O2 and O3 show some violations of stochastic dominance. For O3, the HH > HL comparison was marginally significant and inspection of the survivors functions in Figure 5.20 does not reveal any serious violations and so this participant was retained for further analysis. For O2, this violation was due to some crossover between the HL and HH survivor functions. Again, this violation was not considered serious and this participant was retained for further analysis, although caution should be taken when interpreting the SFT results in this case.

For the AND task, only A1 showed a severe violation of stochastic dominance. For this participant, the violation related to cross over between the HL and HH survivors. As this participant also had a very high error rate for many items, they were removed from the non-parametric analysis.

5.1.4 Double Change Items

Figure 5.5 and Figure 5.6 show the mean RTs and MICs for the OR and AND Tasks respectively. For the OR Task the MICs are uniformly positive. This result rules out serial processing and parallel exhaustive processing, leaving either coactive or parallel self-terminating processing as viable possibilities. For the AND Task, two participants had MICs close to zero; this result rules out parallel and coactive processing and could indicate serial exhaustive processing.

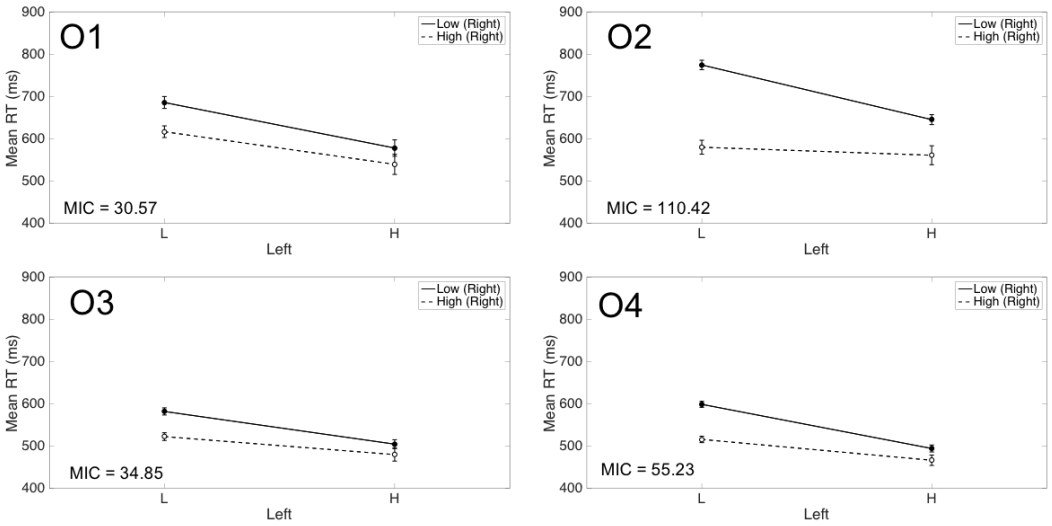

Figure 5.5: Observed double change item mean RTs and MICs for individual participants in Experiment 1: OR Task. The two left hand points represent low discriminability on the left disk and the two right hand points represent high discriminability on the left disk. The solid line represents low discriminability on the right disk, and the dotted line represents high discriminability on the right disk. Error bars represent one standard error.

The remaining two participants had negative MICs which could indicate parallel exhaustive processing.

In order to analyze the RTs for the double change items for the OR task, we conducted a series of 4 (sessions 2:5) $\times$ 2 (left disc change: L or H) $\times$ 2 (right disc change: L or H) ANOVAs on the double change item RTs for each individual participant (see Table 5.4). As participants in the AND Task completed an additional session we conducted a series of 5 (sessions 2:6) $\times$ 2 (left disc change: L or H) $\times$ 2 (right disc change: L or H) ANOVAs for these participants (see Table 5.5).

For most (or all) participants, we find the following:

1. There was a main effect of session for all participants except for A2, indicating that RTs become shorter over the course of the experiment.

2. There was a significant main effect of change discriminability for both discs indicating that the change discriminability manipulation was effective.

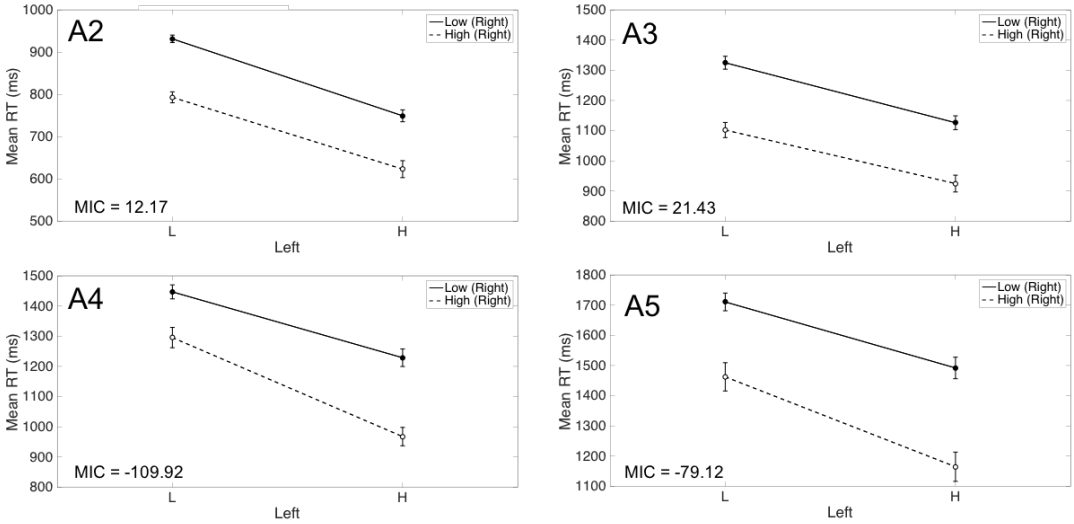

Figure 5.6: Observed double change item mean RTs and MICs for individual participants in Experiment 1: AND Task. The two left hand points represent low discriminability on the left disk and the two right hand points represent high discriminability on the left disk. The solid line represents low discriminability on the right disk, and the dotted line represents high discriminability on the right disk. Error bars represent one standard error.

3. Session did not interact with either dimension, indicating the left and right channel changes tended to be processed at the same speed across sessions. The sole exception was participant A5 who exhibited a Session × Left Change interaction.

4. The three-way interaction was not significant, indicating a stable relationship between double change items across sessions. This indicates that participants were not changing processing strategy from session to session.

In the OR Task, the Left Change × Right Change interaction (see Figure 5.5) was significant for O2 and O4 which provides support for either a parallel self-terminating or coactive account of processing strategy. (Note that the interaction effect is a direct test of the signficance of the MIC; Sternberg, 1969). However, for O1 and O3 this interaction was not significant. Although the MIC was positive for these participants, the non-significant interaction is consistent with serial processing. Nonetheless, caution must be taken when interpreting a

null result in this context since the MIC with a p-value criterion of .05 is biased toward the serial model. We clarify this ambiguity by considering comparison of computations model fit to the data in order to clarify the most likely processing strategy for all participants.

In the AND Task, the Left Change × Right Change interaction (see Figure 5.6) was significant for A4 which provides support for a parallel exhaustive processing account for this participant. For the remaining participants, the interaction was not significant. As for the OR task, we rely on the computational model comparisons presented below to clarify processing strategy for these participants.

Table 5.4: Double Change Item Statistical Results for Individual Participants in Experiment 1: OR Task

Variable	*df*	*F*	p	*df*	*F*	p
		O1			O2	
Session	3	20.97	<.001	3	46.89	<.001
Left	1	11.97	<.001	1	106.62	<.001
Right	1	31.68	<.001	1	30.2	<.001
Session x L	3	0.91	0.438	3	0.93	0.427
Session x R	3	0.3	0.829	3	0.18	0.910
Left x Right	1	1.36	0.244	1	16.69	<.001
Sess x L x R	3	0.57	0.632	3	0.69	0.558
Error	425			467		
		O3			O4	
Session	3	16.32	<.001	3	9.04	<.001
Left	1	17.58	<.001	1	39.12	<.001
Right	1	34.49	<.001	1	75.76	<.001
Session x L	3	1.55	0.120	3	1.39	0.256
Session x R	3	1.64	0.180	3	0.5	0.683
Left x Right	1	2.88	0.090	1	9.58	0.002
Sess x L x R	3	0.51	0.677	3	0.15	0.930
Error	428			478		

Figures 5.7 and 5.8 show the SICs along with the 95% bootstrapped confidence intervals for the OR Task and AND Task, respectively. For the OR Task, it can be seen that all SICs have a positive portion, which is greater than the negative portion. This is most consistent with a coactive or parallel self-terminating

Table 5.5: Double Change Item Statistical Results for Individual Participants in Experiment 1: AND Task

Variable	*df*	*F*	p	*df*	*F*	p
		A2			A3	
Session	4	2.34	0.054	4	36.54	<.001
Left	1	85.81	<.001	1	95.29	<.001
Right	1	156.73	<.001	1	73.56	<.001
Session x L	4	0.47	0.786	4	0.51	0.731
Session x R	4	1.29	0.274	4	0.62	0.646
Left x Right	1	0.11	0.736	1	0.47	0.495
Sess x L x R	4	0.06	0.992	4	0.78	0.535
Error	796			831		
		A4			A5	
Session	4	17.30	<.001	4	13.95	<.001
Left	1	56.94	<.001	1	62.04	<.001
Right	1	90.61	<.001	1	46.28	<.001
Session x L	4	0.27	0.895	4	2.84	0.024
Session x R	4	1.79	0.128	4	0.68	0.608
Left x Right	1	3.97	0.047	1	1.20	0.273
Sess x L x R	4	0.54	0.707	4	1.18	0.318
Error	818			726		

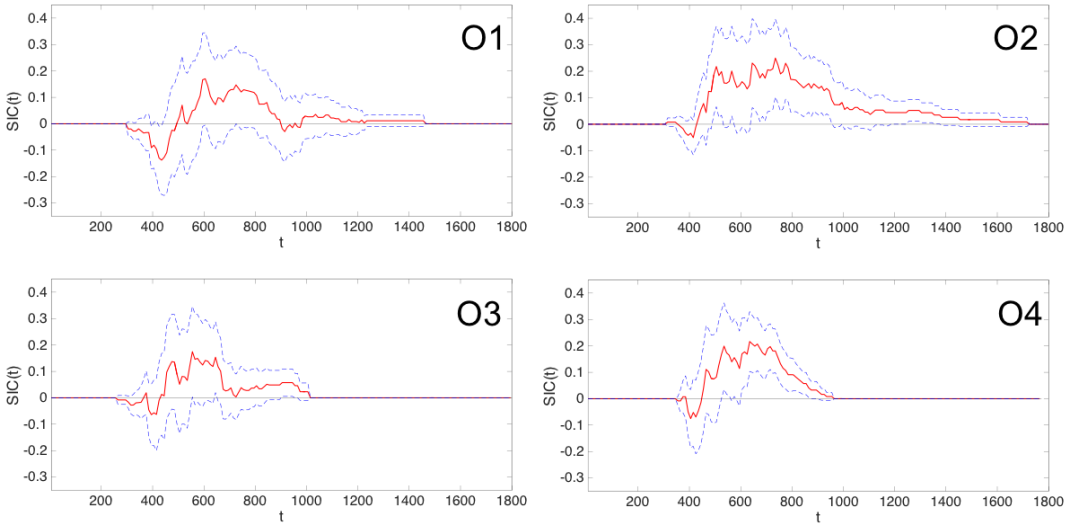

Figure 5.7: Observed double change item SICs (red line) for individual participants in Experiment 1: OR Task. Blue lines represent 95% bootstrapped CIs.

architecture. Although each SIC appears to have an initial negative deflection, this part of the curve always contains 0 within its confidence bounds. This suggests that the evidence for coactivity is not strong.

For the AND Task, the SICs for A2 and A3 appear to have positive and negative portions which are approximately equal, which is most indicative of serial exhaustive processing. For A4 and A5 the SIC is primarily negative, which is consistent with parallel exhaustive processing.

To determine whether the positive and negative portions of the SICs were significantly different to zero, we used two one-sided KS Tests from Houpt et al. (2013)'s SFT analysis package. Two null-hypothesis tests were performed: one which determines whether the largest value of the SIC is significantly greater than zero (D+) and one which determines whether the lowest value of the SIC is significantly lower than zero (D-; see Houpt & Townsend, 2010). Like the MIC, the null hypothesis for the Houpt-Townsend statistic is $SIC(t) = 0$ for all times t. A conservative significance level therefore biases the test toward retaining the null hypothesis (and therefore inferring serial processing). We therefore adopted

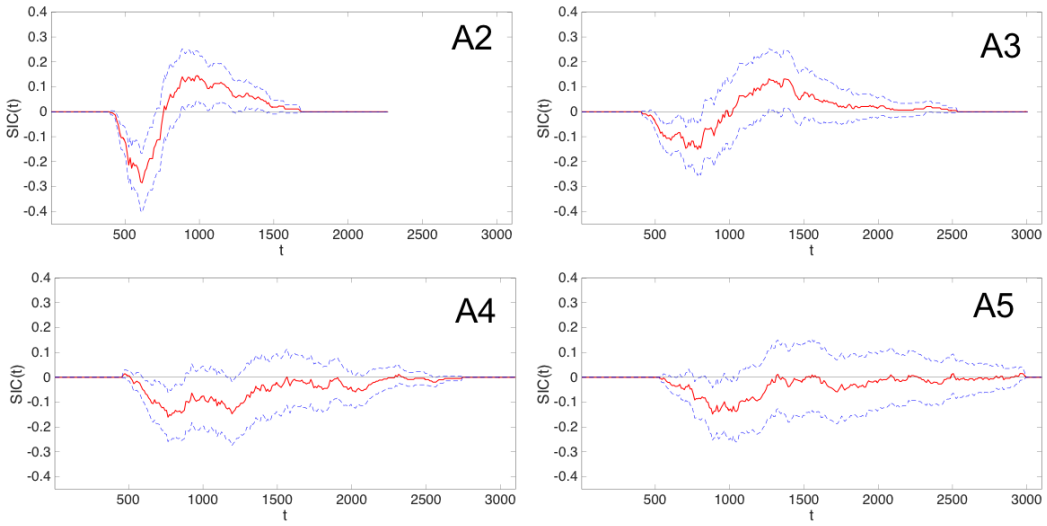

Figure 5.8: Observed double change item SICs (red line) for individual participants in Experiment 1: AND Task. Blue lines represent 95% bootstrapped CIs.

a less conservative cut-off of $\alpha = 0.33$. This value has been used in previous research (Blunden et al., 2020) and has shown to work well in model recovery tests using this statistic (Fox & Houpt, 2016). Both positive and negative D-tests are displayed in Table 5.6 for the OR Task and Table 5.7 for the AND Task.

For the majority of participants in the OR Task, the positive deflection was significantly different to zero, whereas the negative deflection was not. This is consistent with either parallel self-terminating processing. For O3, however, neither positive nor negative deflections were significantly different from zero.

For the AND Task, A2 and A3 both had positive and negative portions which were significantly different to zero, consistent with serial processing. A4 and A5, however, only had negative portions which were significantly different from zero, indicative of parallel processing.

5.1.5 OR Task Single Change Trials

Single change items, like double change items, can be referenced in relation to the decision boundary (see Figure 5.2). The SS item can be considered a

Table 5.6: Directional KS tests for individual participants in Experiment 1: OR Task

Participant	*D+*	*p*	*D-*	*p*
O1	0.208	0.067*	0.112	0.460
O2	0.280	0.008*	0.049	0.862
O3	0.129	0.355	0.100	0.539
O4	0.222	0.047*	0.116	0.437

Note: * indicates a significant difference with an alpha level of .33. D+ tests whether the SIC is significantly greater than zero. D- tests whether the SIC is significantly lower than zero.

Table 5.7: Directional KS tests for individual participants in Experiment 1: AND Task

Participant	*D+*	*p*	*D-*	*p*
A2	0.11	0.243*	0.297	<.001*
A3	0.116	0.212*	0.182	0.021*
A4	0.009	0.990	0.194	0.013*
A5	0.02	0.959	0.181	.029*

Note: * indicates a significant difference with an alpha level of .33. D+ tests whether the SIC is significantly greater than zero. D- tests whether the SIC is significantly lower than zero.

redundant stimulus as it satisfies both decision boundaries (i.e., neither item changes) and is therefore referred to as R. The low-magnitude (L) items lay closest to the redundant stimulus and involve one low salience change, while the high-magnitude (H) items lay on the outer edges and involve one high salience change. This gives five single change items in total: R, LS and SL, HS and SH.

Although the single change trials are not diagnostic of processing architecture for the OR task using SFT, there are still aspects of the data worth considering. A summary of the mean RTs for the single change trials, alongside the double change trials, for each participant is presented in Figure 5.9.

We expect that trials with two changes should be easier to detect than trials with one change (i.e., a redundant target effect). This expectation generally held across participants for the low salience single change (SL and LS) stimuli. However, there were some exceptions. For O1 the mean RTs for the LL, LS, and SL trials were approximately equivalent, with RTs for the SL item actually being slightly shorter.

For O3, the LL and LS items were approximately equivalent, with RTs for the SL item being shorter. In addition, RTs for high salience changes always tended to be shorter than low salience changes, although there was generally an advantage for the double change HH, HL, and LH items over the single change HS and SH items.

When specifically comparing the low salience single changes to the high salience single changes, we would expect an RT advantage for the high salience change. This generally seemed to be the case for all participants. However, when any change occurred on the left, participants were generally quicker to respond, meaning that the HS and LS items had more similar RTs than SH and SL items. It is unclear whether this was due to a preference for location in the left (e.g., due to a left-right English language reading bias) or due to a strategic preference or unconscious processing advantage for discs of lighter luminance polarity (as lighter luminance levels always appeared on the left). For the OR task, RTs for the SS item, were shorter than the LS and SL items for all participants, but

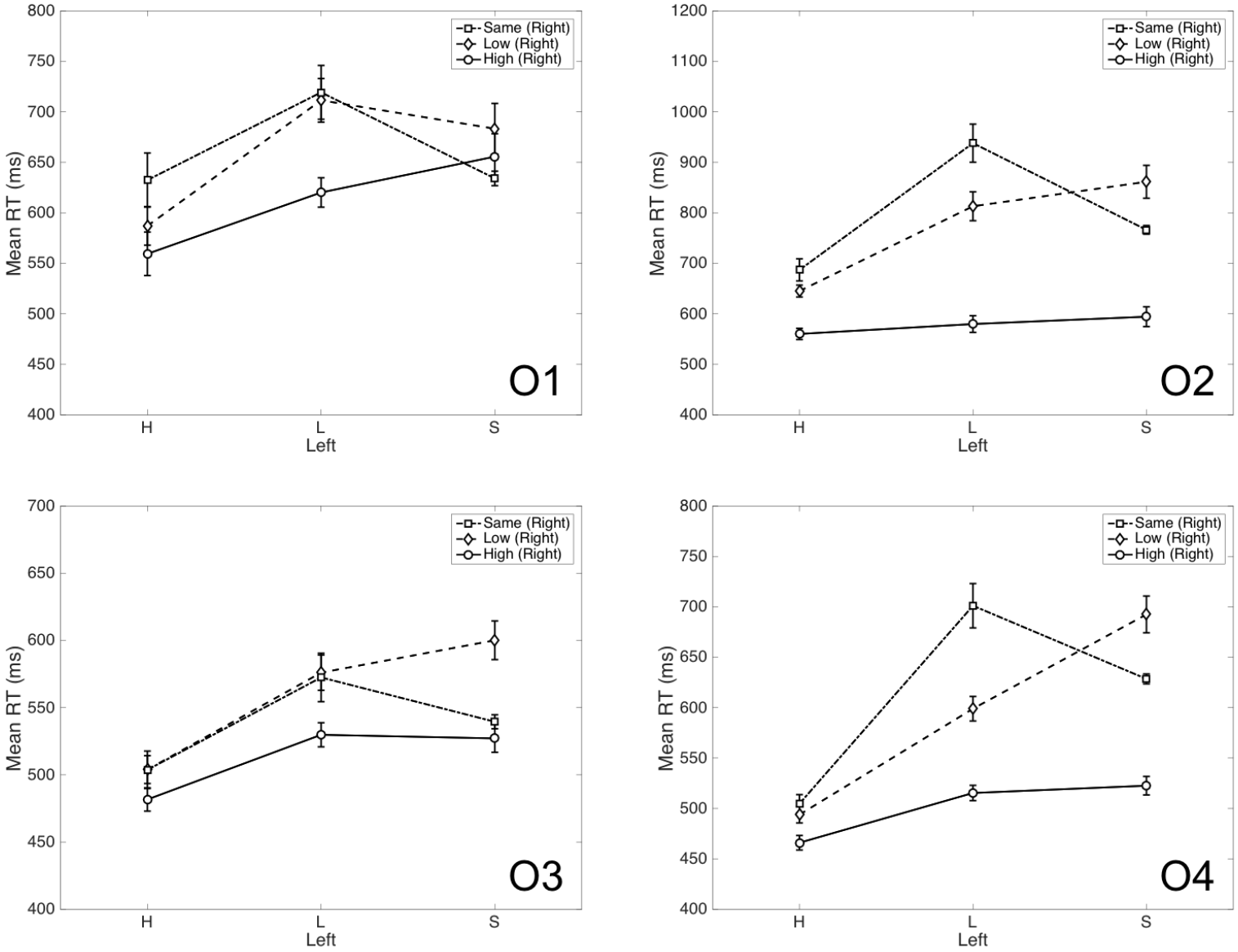

Figure 5.9: Mean RTs for no change, low salience change, and high salience change items for the left and right discs, respectively.

were equivalent with or longer than HS and HS items.

5.1.6 AND Task Single Change Trials

Unlike the OR task, the single change trials in the AND task also offer us additional nonparametric diagnostic information (Fifić et al., 2010; see also Little et al., 2015, 2018). Diagnostic predictions for each architecture are shown in Figure 5.10.

In a fixed-order serial self-terminating model, one dimension (either left disc or right disc) is always processed first, followed, if necessary, by the other dimension. If the first dimension that is processed does not involve a change, then a decision can be made and the process will self-terminate. If, however, the first dimension that is processed involves a change, then processing must continue to the second dimension. This leads to the general RT prediction that the first-processed dimension will be approximately the same, whereas the second-processed dimension will be comparatively longer (as it must wait until the first process is finished before it can begin). Further, RTs for the high-magnitude item for the second-processed dimension are comparatively shorter than the low-magnitude item. This is because identifying a large magnitude change is easier than identifying a low magnitude change and so the time taken to switch from the first-processed dimension to the second-processed dimension should be shorter for the HS/SH items.

The single-change items for the AND task also allow for distinction between fixed-order and mixed-order serial processing. While fixed-order processing assumes that the same dimension is always processed first, a mixed-order model assumes that there is variation in which dimension is processed first from trial to trial. Like the fixed-order self-terminating model, the mixed-order model predicts that the redundant stimulus (SS) will have the shortest processing time (as both dimensions lead to a self-terminating stopping rule regardless of which is processed first). The first-processed dimension is still shorter that the second-processed dimension (again, as the second-processed dimension must wait

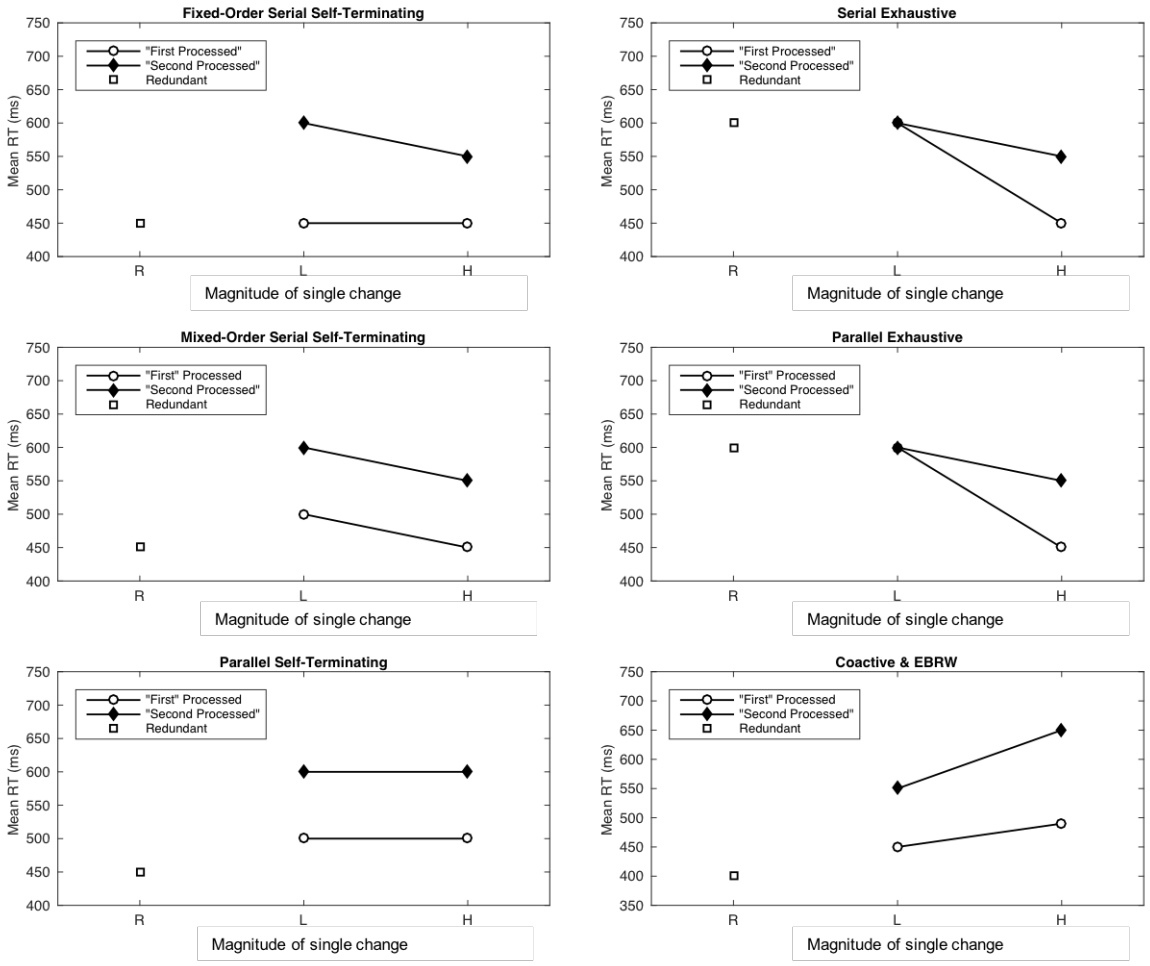

Figure 5.10: RT predictions for the single and no-change items. R = Redundant Stimulus, L = Low-magnitude single change items, H = High magnitude single change items.

for the first-processed dimension to complete). However, in a mixed-order model, the processing time for the high-magnitude stimulus is shorter for both first- and second-processed dimensions. This is because, on average, the time taken to switch from one dimension to the other dimension will be shorter for high magnitude changes compared to low magnitude changes.

For the parallel self-terminating model, processing time is equivalent to the minimum processing time to detect "no-change" across both dimensions. That is, the time it takes to determine whether SL or SH contains an item which has not changed is equivalent for both stimuli. This is also the case with LS or HS. The RTs for low-magnitude and high-magnitude items are therefore equivalent, however, the redundant stimulus should have a faster processing time due to statistical facilitation (Raab, 1962; as there are more chances to self-terminate). Note that parallel and coactive models, the distinction between first-processed and second-processed is less relevant; however, it may be the case that one location is processed faster than the other. This is not a strong or central prediction of the models.

For the serial and parallel exhaustive models, both dimensions must be processed regardless of whether one, or both, dimensions do not involve a change. For the serial exhaustive model, this means that the total processing time comprises the RTs for each individual dimension. The RT for the redundant stimulus should therefore be longer than for the self-terminating models. The parallel exhaustive model predicts that the processing time is equivalent to the maximum processing time across both dimensions. Again, redundant and interior item RTs are longer than for the self-terminating models.

Finally, a general prediction of the coactive model is that the low-magnitude item RTs will be shorter than the high-magnitude items. In the coactive model, each stimulus is represented by a bivariate normal distribution. As one moves from the high-magnitude items towards the left-most corner of the decision space, a greater proportion of the distribution will lie in the single/no-change region. Hence, the chance that a sample will come from the no-change distribution

increases for the redundant and low-magnitude items compared to the high-magnitude items, also resulting in a shorter RT for the redundant stimulus and low-magnitude items.

For the AND Task, the mean RTs for the single change items are displayed in Figure 5.11. For most participants, the RTs for the redundant stimulus was shorter than the other stimuli, suggesting a self-terminating stopping rule. This difference was predominantly significant, excepting the the HS-R comparison for A2 and the SL-R comparison for A4. For A5 the RT for the redundant stimulus was longer than the HS and LS items. For A2 and A3 the RTs for the low-magnitude items were longer than the high-magnitude items, which is consistent with serial processing. However, for A4 the RTs for the low-magnitude items were shorter than the high-magnitude items which could suggest coactive processing. For A5 RTs for the low-magnitude item were shorter on one dimension and longer on the other.

For AND Task single change items, we conducted a series of planned t tests comparing low-magnitude and high-magnitude items on both dimensions and comparing the redundant stimulus to the other single change items (see Table 5.8). For most items the RTs for the redundant stimulus was significantly shorter than the other stimuli, indicating a self-terminating stopping rule[6].

For A2 the low-magnitude item RTs were significantly longer than the high-magnitude item RTs, indicating serial processing. For the remaining participants there were no significant differences between low-magnitude and high-magnitude items which is consistent with parallel processing, however, caution must be exercised when extrapolating from a null result.

5.1.7 Discussion

Taken together, the SFT results suggest independent processing for the AND task (whether serial or parallel, with some minor individual differences in the single change items). For the OR task, however, the SIC functions are primarily

[6]The HS item for A2, SL item for A4 and LS and HS item for A5 were non-significant.

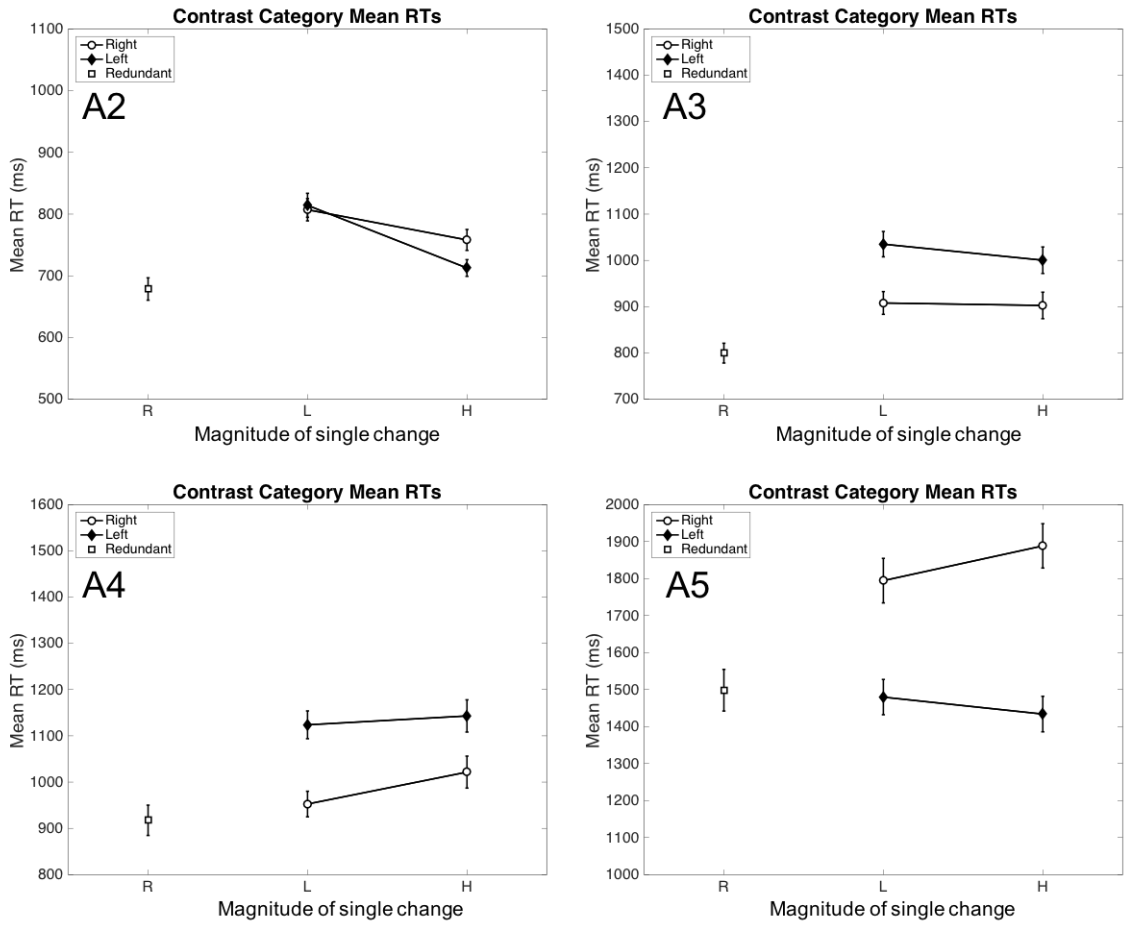

Figure 5.11: Observed single change item mean RTs for individual participants in Experiment 1: AND Task. Error bars represent one standard error. R = redundant stimulus, I = interior stimulus, E = exterior stimulus.

Table 5.8: Single Change Item Statistical Results for Individual Participants in Experiment 1: AND task

Stimulus Pair	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*
	A2				A3			
HS - LS	-101.58	-4.34	331	<.001	-34.68	-0.88	332	0.382
SH - SL	-48.97	-1.98	302	0.049	-5.34	-0.14	310	0.887
HS - R	34.06	1.54	300	0.125	200.79	5.37	292	<.001
LS - R	135.64	5.01	287	<.001	235.44	6.45	298	<.001
SH - R	79.54	3.20	280	0.002	103.03	2.8	282	0.006
SL - R	128.51	4.99	278	<.001	108.37	3.25	286	0.001
Stimulus Pair	*Mdiff*	*t*	*df*	*p*	*Mdiff*	*t*	*df*	*p*
	A4				A5			
HS - LS	19.19	0.42	325	0.676	-45.84	-0.68	279	0.499
SH - SL	69.09	1.56	328	0.120	93.99	1.11	201	0.270
HS - R	225.63	4.64	288	<.001	-64.41	-0.88	258	0.382
LS - R	206.44	4.62	295	<.001	-18.57	-0.25	253	0.800
SH - R	104.23	2.14	294	0.033	390.17	4.75	220	<.001
SL - R	35.15	0.83	292	0.410	296.18	3.59	213	<.001

positive, and although there is a slight negative blip for shorter RTs which could indicate coactive processing, the 95% CIs incorporate zero and so parallel self-terminating processing cannot be ruled out. Further, the D- test did not indicate a significant difference which would be expected in a coactive model. This suggests that processing is likely independent for both OR and AND tasks. To provide further support for this interpretation, we now turn to quantitative model fitting.

5.1.8 Computational Model Fitting

Although the non-parametric analysis of architecture provides strong diagnostic information for both processing architecture and capacity, there are a number of additional benefits to supplementing these with parametric computational modeling. First, in an OR design, the single change items are not used in the SFT analysis. Further, even in an AND task, in the SFT analysis the single change and double change items are considered separately.

Computational modeling allows us to consider both single and double change

items in both OR and AND designs simultaneously. Second, the non-parametric analyses only utilize correct responses, whereas with computational modeling we are able to fit both correct and error RT distributions, affording a more comprehensive understanding of the data. Third, it is reasonable to assume there may be differences in the variability of the distribution of perceptual effects depending on the strength of the change. One of the benefits of computational modeling is that it allows us to explore a more flexible account of change detection decision-making, where the variance of perceptual representations may vary when there is low versus high salience changes, or no-change in the display.

A summary of each candidate model is presented in Table 5.9. In short, the evidence accumulation process is modelled using the LBA (Brown & Heathcote, 2008). The LBA represents a choice between N alternatives using N accumulators. In the LBA, each trial begins with a starting amount of evidence (k) that increases linearly at an amount given by the drift rate (d). Evidence accumulates until a response threshold (b) is reached. The first accumulator to reach a decision provides the response time, plus non-decision time (t_0) which is a constant time for non-decisions processes.

In the current study, each channel (left disc and right disc) is represented by a pair of racing accumulators (one for a change response and one for a no-change response). The final RT and response choice for that channel is determined by the fastest of these accumulators. The RTs for each channel are then combined as per the rules of each candidate architecture (as explained further below) to yield a final RT/ response choice. To generate a drift rate for each of the LBA channels we assume that the magnitude of the changes can be represented by a bivariate folded normal distribution. We then integrated over the region of space relevant to each evidence integration rule. That is, we determine the volume of the distribution that falls into the each response region.

For example, a minimum evidence integration rule would involve integrating over the region of space accounting for either of one or two changes to represent a change response, whereas a maximum evidence integration rule would involve

only integrating over the region of space accounting for two changes to represent a change response.

Table 5.9: Summary of each candidate model, including architecture and decision-rules, and corresponding integration rule.

Model	Architecture	Integration Rule
MAD	Coactive	Minimum Evidence
SAD	Coactive	Maximum Evidence
MAX	Parallel/Single Channel	Minimum Evidence
Decision Rule	**Architecture**	**Integration Rule**
Self-terminating	Parallel	Minimum Evidence
Self-terminating	Serial	Minimum Evidence
Exhaustive	Parallel	Maximum Evidence
Exhaustive	Serial	Maximum Evidence

Note: The MAD model is Wilken and Ma's 2004 *Maximum Absolute Differences* model. This model assumes that a participant monitors the noisy change signal from a number of different locations independently. If any of these signals exceeds its respective threshold then a change response is emitted. When extended to RT this is equivalent to a minimum evidence integration rule and could be represented by either coactive or parallel self-terminating architectures. The SAD model is Wilken and Ma's 2004 *Sum of Absolute Differences* model. This model assumes that the participant sums the change strength signal across all locations. If this summed signal exceeds a decision threshold, a change response is emitted. This when extended to RT this is equivalent to a coactive model with a maximum integration rule. The MAX model considers only the strongest signal in its decision and is therefore represented by only a single decision-making channel, however, a parallel self-terminating architecture is a useful point of comparison and could be equivocal in terms of RT.

Both serial and parallel models assume that decisions about changes in each of the discs occur independently and can be modeled using independent evidence accumulation processes for each dimension. Therefore, the minimum and maximum integration rules map on to the self-terminating and exhaustive stopping rules, respectively.

This means that, for the maximum integration rule, the final decision times for each channel (resulting from the race between each accumulator pair) are

either summed, for the serial model, or used to find the maximum time prediction, for the parallel model. When applying the minimum integration rule to the serial model, the decision time is represented by a single accumulator pair providing a decision can be logically made without considering the additional disc. For the parallel model, the minimum decision time for both accumulator pairs is used to determine decision-time. For the coactive model, the change distributions are not modeled independently. Instead, a single LBA is employed with the drift estimated using either a maximum or minimum integration rule. Similarly, the single channel model assumes a single pair of racing accumulators, however, in this model the drift is only determined by the greatest change strength in the display.

We further fitted two versions of each of these models: one in which the variance of the normal perceptual distribution was assumed to be equal for each level of change strength (*fixed variance* models), and another in which the variance of the no-change, low discriminability change, and high discriminability change distributions were allowed to vary (*free variance* models). This second version of the models was to account for the possibility that the variability in the overall change distribution could vary based on the strength of the change in each individual disc. For example, it is possible that a low salience change in the left, lighter disc may be more or less variable, compared to the right, darker disc etc.

Formal model fitting procedure

We assumed that a change in each individual stimulus was represented by a folded normal distribution with mean locations μ_{Left} and μ_{Right}, and standard deviations, σ_{Left} and σ_{Right}. For the fixed variance models, σ_{Left} and σ_{Right} were equivalent for each of the three salience levels (S, L, and H), however, these were allowed to vary in the free variance model.

For the coactive model a bivariate folded-normal distribution was used to represent the joint distribution for changes in either location. The covariance of

this distribution was set to zero. For simplicity, we assumed that the locations of each of the changes was $\mu_{Left} = 1, 2$, or 3 and $\mu_{Right} = 1, 2$, or 3, respectively. The evidence integration rule used to instantiate a decision boundary gives two further parameters D_{Left} and D_{Right}. The integral of these distributions within each region provides the mean drift rate for the LBA for that stimulus.

Starting point variability was represented using a uniform distribution which varied between 0 and $A0$ from trial to trial. There were parameters for decision thresholds for both "change", b_{Change}, and "no change", $b_{NoChange}$ accumulators. These parameters capture the distance between the starting point and the decision threshold. For the serial and parallel models there are separate change and no-change accumulators for each disc. However, for simplicity, we assume that the starting point and response thresholds are the same for each of the dimensional accumulators. For the MAX model, there was only a single set of change and no-change accumulators. Drift rate is assumed to vary from trial to trial according to a normal distribution with a standard deviation, s. To account for time taken to complete processes and actions not associated with the decision process, both accumulators also incorporate non-decision time (t_0). The final RT therefore comprises the sum of the non-decision time and the decision time predicted by the LBA. The serial self-terminating model has one further parameter (p_x) representing the probability that one dimension is processed before the other.

The likelihood for each model was approximated using probability density approximating (PDA; Turner & Sederberg, 2014) by simulating 50,000 data points from the model and then using Holmes' method (2015). The log likelihood for each trial was then summed over items. Each participant was fit separately. For each parameter, we first transformed the parameter to lie on the whole real line and then adopted reasonably informative priors based on our prior work with these models. The transformations and priors are shown in Table 5.10.

We used Differential Evolution Markov Chain Monte Carlo (DEMCMC; Turner et al., 2013) to efficiently generate proposals from the posterior distributions of each parameter. However, the variability in the likelihood approximation

Table 5.10: Prior parameter distributions and transformations for each parameter.

Parameter	Transformation	Distribution	Prior Parameter	Values
D_{Left}	$\hat{D}_{Left} = logit\,[D_{Left} - x_0]$	Normal	$\mu = 0$	$\sigma = .5$
D_{Right}	$\hat{D}_{Right} = logit\,[D_{Right} - y_0]$	Normal	$\mu = 0$	$\sigma = .5$
σ_{Left}	$\hat{\sigma}_{Left} = \log(\sigma_{Left})$	Normal	$\mu = -1.5$	$\sigma = .2$
σ_{Right}	$\hat{\sigma}_{Right} = \log(\sigma_{Right})$	Normal	$\mu = -1.5$	$\sigma = .2$
A	$\hat{A} = log(A)$	Normal	$\mu = -1.05$	$\sigma = .2$
B_{Change}	$\hat{b}_{Change} = log(b_{Change} - A)$	Normal	$\mu = -1.05$	$\sigma = 1$
$B_{No-Change}$	$\hat{b}_{No-Change} = log(b_{No-Change} - A)$	Normal	$\mu = -1.05$	$\sigma = 1$
s	$\hat{s} = log(s)$	Normal	$\mu = -1.39$	$\sigma = .5$
T_0	$\hat{t}_0 = log(t_0)$	Normal	$\mu = -1.51$	$\sigma = .2$
p_x	$\hat{p}_x = logit(p_x)$	Normal	$\mu = 0$	$\sigma = 2$

can cause the chains to become stuck if an accepted parameter set results in an usually high likelihood. To prevent the chains getting stuck in this manner, we re-sampled the likelihood of any existing chains each time the current proposal was rejected (Holmes, 2015). This led to a good mixing of the chains and strong convergence after the burn-in period.

For the fixed variance models we used a burn-in period of 2150 iterations and for the free variance models we used a burn-in period of 3150 iterations. Between iterations 501-700 there was a deterministic migration step (Turner et al., 2013) every 20 iterations. The remaining sampling used a probabilistic migration step instead of a cross-over step with a probability of .05. The number of chains was determined by taking three times the amount of parameters for each model. A minimum of 20000 posterior samples were taken per parameter, with 750 iterations estimated for each chain.

For model comparison we used the Deviance Information Criterion (DIC; Gelman et al., 2014). The deviance of a posterior sample of parameters, denoted θ, is calculated as:

$$D(\theta) = -2lnL(y \mid \theta) \tag{5.5}$$

The DIC is calculated as:

$$DIC = \overline{D}(\theta) + 2p_D \tag{5.6}$$

Here, $\overline{D}(\theta)$ is the mean of the distribution of posterior deviances and $p_D = 2var[lnL(y|\theta)]$. The DIC punishes for model complexity. This is achieved by penalizing the average negative log likelihood by a term which accounts for the functional form complexity of the model. The DIC can be thought of as a Bayesian version of a maximum likelihood-based fit statistic such as the Akaike Information Criteria (AIC; Akaike, 1974). To summarise, we fit serial, parallel, and coactive models using minimum and maximum integration rules (see Table 5.9) as well as the single channel MAX model. We investigated both first-order (i.e., pooled) and second-order (i.e., independent) integration rules via coactive (pooled) architectures and serial and parallel (independent) architectures. We further mapped integration rule to architectures as self-terminating and exhaustive stopping rules for the parallel and serial models, and a minimum and maximum integration rule for the coactive model.

In this section, we report the results of competitive model comparison. Because the computational models do not rely on the assumption of stochastic dominance, nor on low error rates (these are explicitly modeled), we also fit each of the models to participant A1, who was omitted in our previous analyses. We also further included the participant who was excluded for not following task instructions in the OR task. They are labeled O5 in the following analyses.

The DICs for each individual participant in the OR fixed variance, OR Free variance, AND fixed variance, and AND free variance are displayed in Tables 5.11 and 5.12, respectively. The preferred model is the model with the lowest DIC.

In the OR task, the majority of participants were best fit by a parallel self-terminating model (with one participant being best fit by a minimum integration coactive model). For two participants (O2 and O5) the more flexible free variance model was preferred. For the remaining three participants (O1, O3, and O4) the DIC favoured the simpler fixed variance account.

For the AND task, the majority of participants were best fit by a serial self-terminating model, although two of the five participants were best fit by the

Table 5.11: DIC values for each individual participant and candidate model for the OR task with fixed variance (top rows) and free variance (bottom rows). Lower values indicate better fit; the best model for each fixed and free variance set is bolded. The overall best model is underlined.

Fixed Variance Model							
Subject	Coactive Max	Coactive Min	Parallel ST	Parallel Ex	Serial ST	Serial Ex	MAX model
O1	639.58	**309.47**	466.79	863.06	551.53	766.65	362.02
O2	425.83	-110	**-129.93**	382.09	-98.131	336.63	-103.65
O3	-850.27	-881.72	**-986.38**	-852.65	-971.68	-839.9	-959.01
O4	-1613.4	-2056	**-2104.9**	-1806	-2002.3	-1796.6	-2056.3
O5	363.81	188.99	**119.02**	393.26	263.3	345.23	177.01
Free Variance Model							
Subject	Coactive Max	Coactive Min	Parallel ST	Parallel Ex	Serial ST	Serial Ex	MAX Model
O1	623.15	**324.51**	432.93	638.49	538.78	634.94	365.81
O2	416.54	-84.68	**-140.49**	336.24	-101.79	305.83	160.77
O3	-868.98	-864.05	**-976.32**	-803	-967.56	-793.57	-917.7
O4	-1603.8	-2061.3	**-2101.1**	-1796.2	-2005.6	-1804.1	-1681.9
O5	337.13	129.86	**63.77**	401.85	222.12	386.13	97.67

Table 5.12: DIC values for each individual participant and candidate model for the AND task with fixed variance (top rows) and free variance (bottom rows). Lower values indicate better fit; the best model for each fixed and free variance set is bolded. The overall best model is underlined.

Fixed Variance Model							
Subject	Coactive Max	Coactive Min	Parallel ST	Parallel Ex	Serial ST	Serial Ex	MAX Model
A1	**588.68**	912.98	662.21	632.7	677	639.12	767.32
A2	354.3	2221.1	319.68	387.05	**315.48**	409.57	2120.5
A3	2091.9	3832.8	2097.7	2345.3	**2047.5**	2333.7	3845.3
A4	2796.5	4531.7	2828.4	3050.1	**2786**	3011.4	4626.7
A5	**3897.8**	5280.6	3958.7	4051.7	3928.9	4027.3	5329.7
Free Variance Model							
Subject	Coactive Max	Coactive Min	Parallel ST	Parallel Ex	Serial ST	Serial Ex	MAX Model
A1	**611.39**	850.47	669.12	662.88	658.08	652.37	708
A2	349.82	1667.4	315.36	383.51	**304.24**	406.64	2020.5
A3	2091.6	3287.2	2097	2364.3	**2049.1**	2324.9	3692
A4	2797.2	4058.9	2840.9	3055	**2795.4**	3011.9	4397.7
A5	**3923.6**	4987.7	3952.8	4046.3	3926.6	4023.1	5161.9

maximum integration rule coactive model. For the majority of participants (A1, A3, A4, and A5) the simpler, fixed variance models were preferred. For A2, the free variance maximum integration rule coactive model provided the best fit to the data.

Along with model comparison estimates, it is essential to provide an estimate of how well a model fits the data to ensure that the model is capturing the data well (Heathcote et al., 2015). In the interest of space, we present the posterior predictions of the best fitting models for each item along with posterior parameter estimates of the parameters for the fixed variance models. The posterior parameter estimates for the free variance models and the best fitting parameters and 95% highest density intervals (HDIs) are shown in the supplementary material.

In Figure 5.23, we show the model predictions for O3 for the OR rule and in Figure 5.28, we show the model predictions for A3 for the AND rule to show that the model does provide a good fit to the data. The supplementary material shows that the fits to the other participants are comparable. The posterior parameter estimates for each participant for the overall best fitting models for the AND and OR tasks are shown in Figures 5.14 and 5.15.

5.1.9 Discussion

For the OR task, the best fitting model for most participants was the parallel self-terminating model. Although there were differences in whether a fixed or free variance version was preferred, the parallel self-terminating model was the best fitting model for all participants except O5 for whom the coactive model with a minimum integration rule was preferred (although recall that O5 was excluded from the SFT analyses for not following task instructions). These results clarify the results of the SFT analyses supporting an inference of a parallel self-terminating architecture. For the AND task, the best fitting model overall for three of the five participants (A2, A3, and A4) was the serial self-terminating model and there were again differences in whether a fixed or free variance version

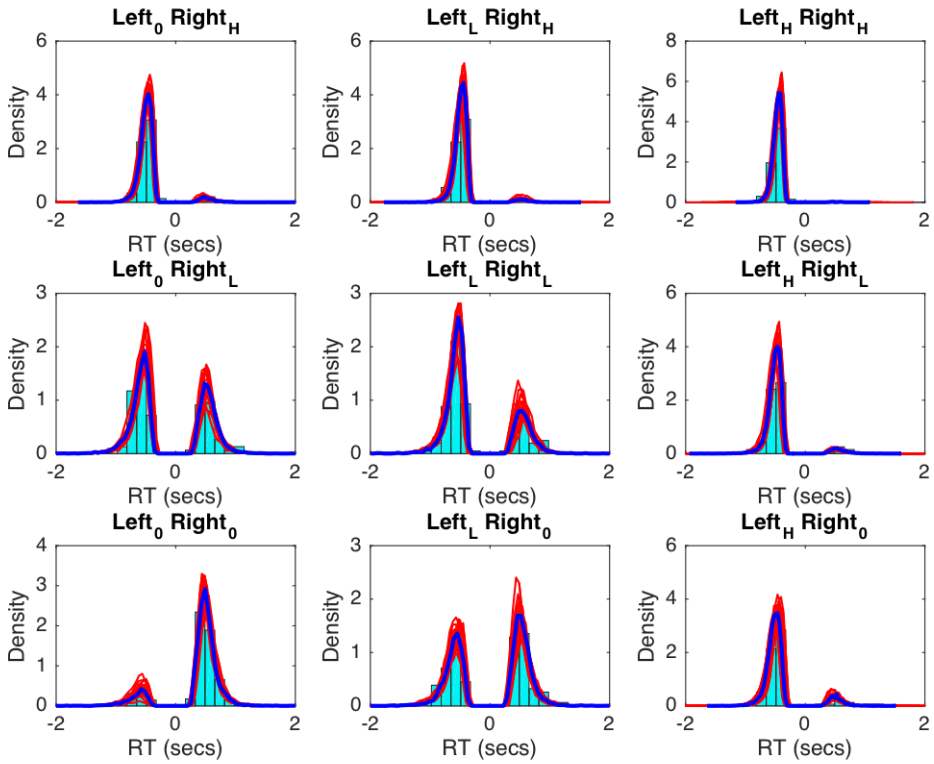

Figure 5.12: Posterior predictions from the Parallel Self-Terminating model (fixed-variance version) for observer O3. Each subplot shows the predictions for one item. Positive RTs indicate change responses. Negative RTs indicate no change responses. Data are plotted in the histogram. Each red line indicates a posterior sample. The solid blue line indicates the predictions using the average posterior parameter values.

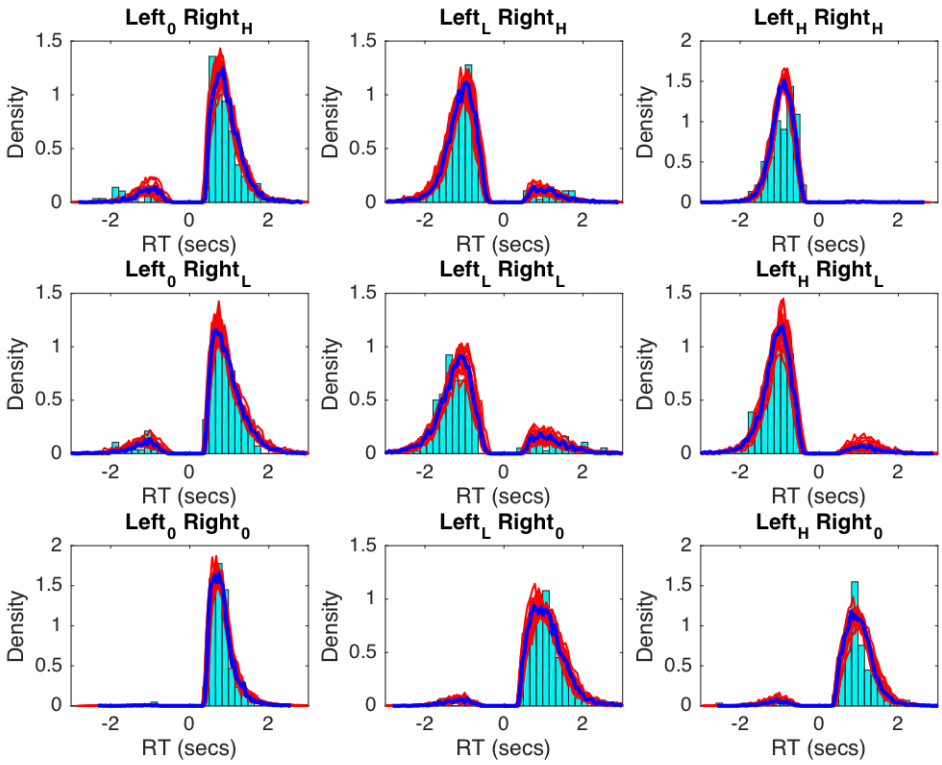

Figure 5.13: Posterior predictions from the Serial Self-Terminating model (fixed-variance version) for observer A3. Each subplot shows the predictions for one item. Positive RTs indicate a "YES" response (i.e., both items have changed). Negative RTs indicate "NO" responses (i.e., both items did not change). Data are plotted in the histogram. Each red line indicates a posterior sample. The solid blue line indicates the predictions using the average posterior parameter values.

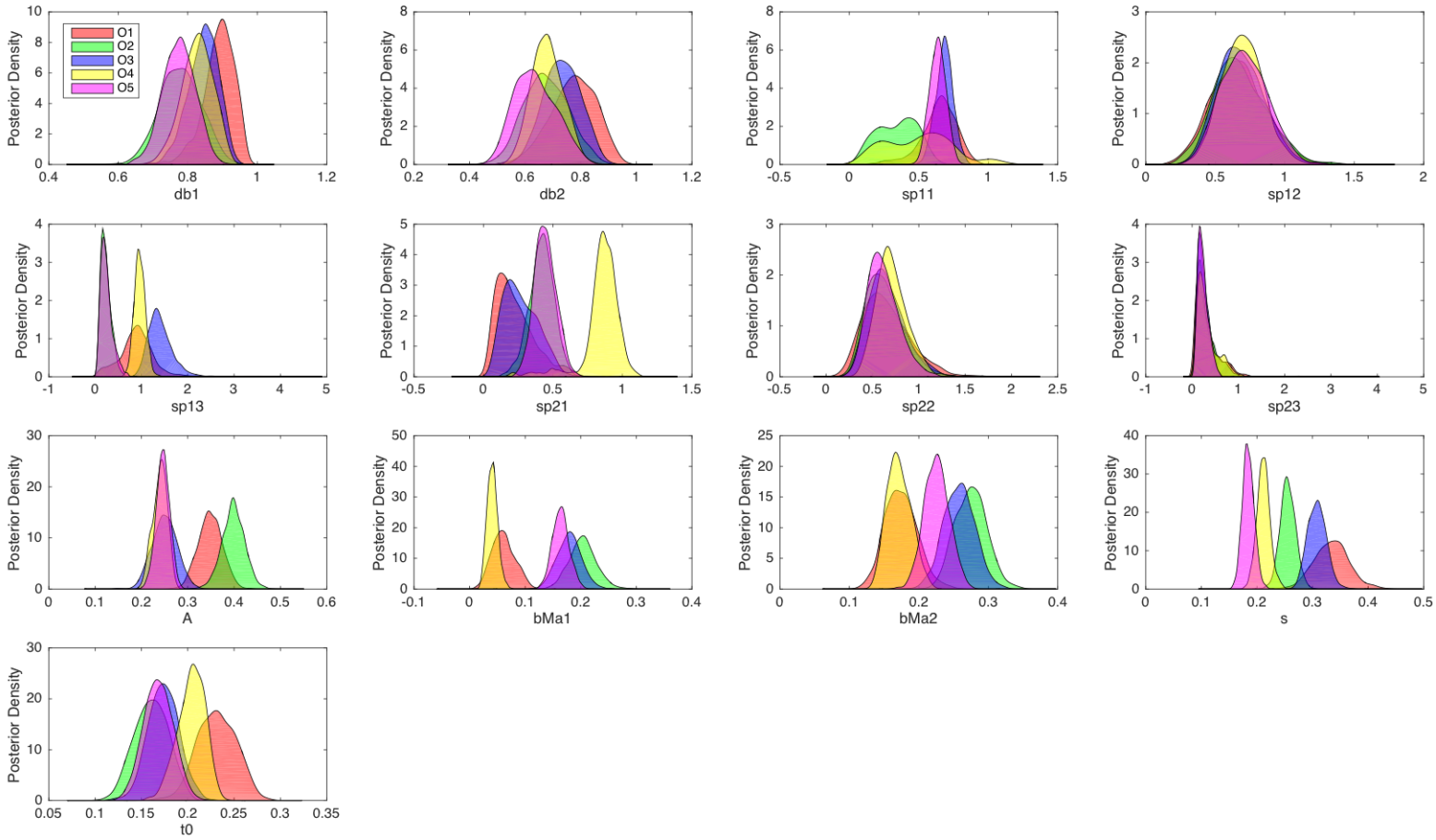

Figure 5.14: Posterior parameter estimates for each participant in the fixed variance OR condition, parallel self-terminating model. db1 and db2 = decision boundary for the left and right items, respectively. sp11, sp12, and sp13 = the standard deviation of the change distribution for the left item for S, L, and H salience changes, respectively. sp21, sp22, and sp23 = the standard deviation of the change distribution for the right item for S, L, and H salience changes respectively. A = starting point. bMa1 and bMa2 represent change and no-change decision thresholds, respectively. s = drift rate standard deviation. t_0 = non-decision time.

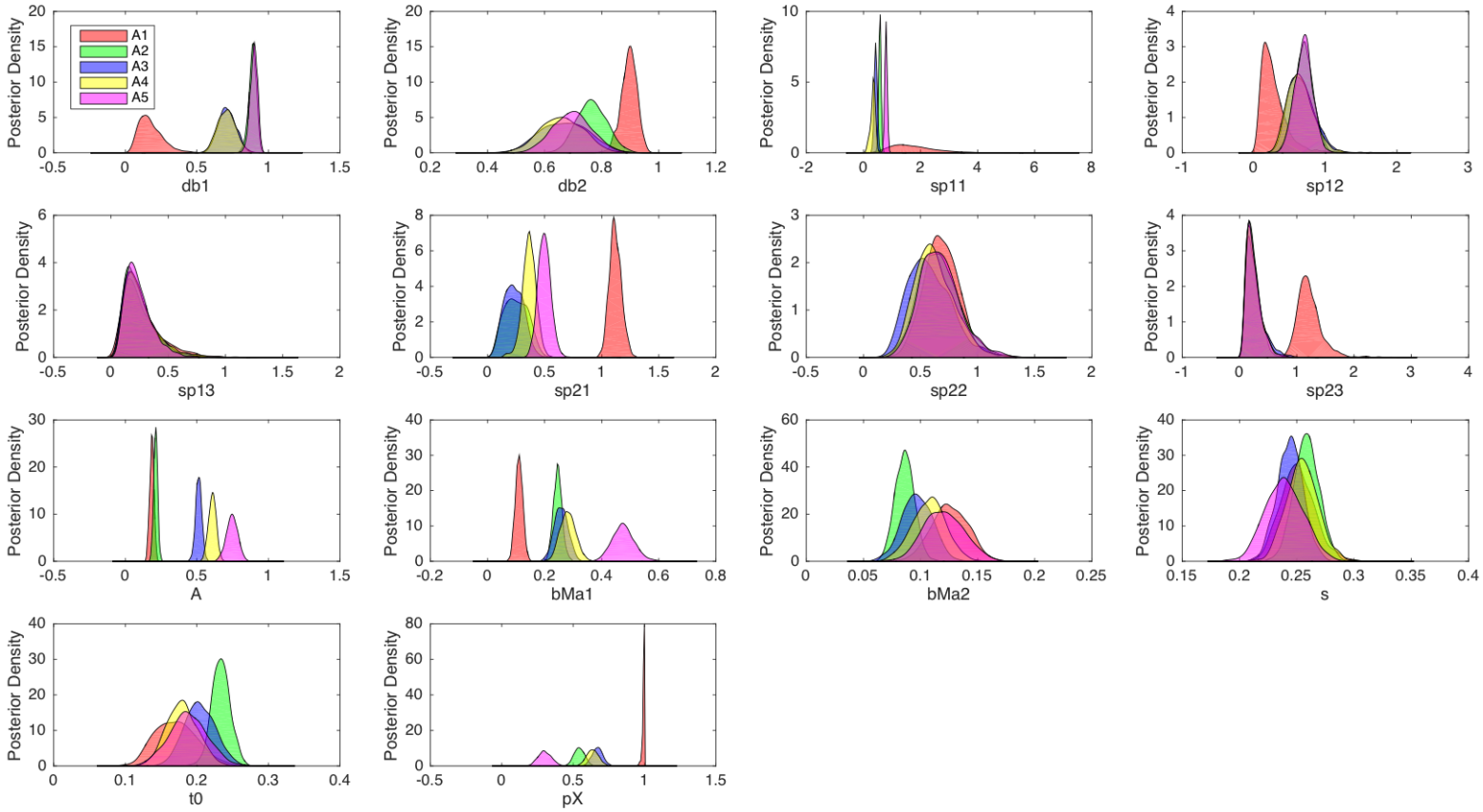

Figure 5.15: Posterior parameter estimates for each participant in the fixed variance AND condition, serial self-terminating model. db1 and db2 = decision boundary for the left and right items, respectively. sp11, sp12, and sp13 = the standard deviation of the change distribution for the left item for S, L, and H salience changes, respectively. sp21, sp22, and sp23 = the standard deviation of the change distribution for the right item for S, L, and H salience changes respectively. A = starting point. bMa1 and bMa2 represent change and no-change decision thresholds, respectively. s = drift rate standard deviation. t_0 = non-decision time. p_x = probability one dimension is processed before the other.

was preferred. The coactive model with a maximum integration rule was the best fit for A1 and A5.

Our strong inference methods allow us to effectively rule out several models of change detection of performance. Although there were some differences between individuals, the evidence for a coactive model was weaker than the evidence for models based on the independent accumulation of evidence at each location across both tasks. Further, the different task demands of the OR task and the AND task appear to result in different attentional strategies. The fact that some participants were better fit by the free variance version of the models also suggests there is some evidence that the change signal varied differentially as a function of the strength of the change.

Given the difference in architecture between OR and AND decision rules, a naturally arising question is whether the difference in strategy is also reflected by differences in the efficiency with which detections are made. In particular, whether the finding of serial processing in the AND task will be associated with decreased efficiency compared with the finding of parallel processing in the OR task. In Experiment 2, we measured the efficiency of change detection by examining how the capacity of detection changes at the number of changes in the display increases.

5.2 Experiment 2

In Experiment 2, we utilized a redundant target change detection task in order to obtain a measure of workload capacity. The stimuli were spatially separated luminance discs as in Experiment 1; however, Experiment 2 also incorporated trials where only a single item was presented in both probe and memory arrays. The design of this task allow us to compare performance on the redundant change trials to the predictions of a baseline unlimited capacity parallel model. This comparison provides a measure of the efficiency with which change detection is altered by the number of possible changes (Townsend & Ashby, 1983; Wenger &

Townsend, 2000).

5.2.1 Workload Capacity

Increasing the number of features or objects that need to be processed before a decision can be made might increase, decrease, or leave unaffected the rate of processing changes (Townsend & Ashby, 1983; Townsend et al., 2007; Wenger & Townsend, 2000). Importantly, this differs from the common idea of item capacity often investigated in the change detection literature (see e.g. Luck & Vogel, 1997) as it refers to limitations in processing rate, rather than a limit to the number of items held in memory. Workload capacity is also a continuous measure, measured as a function of time rather than a discrete item bottleneck.

In the redundant target detection task, two locations are monitored for the onset of stimuli or targets. Targets can appear in either of the locations, as well as both locations or neither location. In an OR design, participants respond affirmatively if any target is present; otherwise, participants respond negatively. In an AND design (see Figure 5.17), participants respond affirmatively only if both targets are present; otherwise, participants respond negatively. In the present change detection task, we implement this design using possible changes from one display to the next rather than target onsets (cf., C.-T. Yang et al., 2014).

For the OR task, the baseline comparison model predicts that the RT for the double change trials should equal the minimum processing time derived from the single change trials. Capacity is inferred by comparing performance to this baseline prediction.

The capacity coefficient is computed using the integrated hazard functions, $H(t)$, of the single and double targets. $H(t)$ gives the cumulative processing completed by the system until time, t. The hazard function is defined by the probability density function, $f(t)$, which gives the probability of observing a RT at time, t, divided by the survivor function, $S(t)$, which gives the probability that a process will finish after time, t:

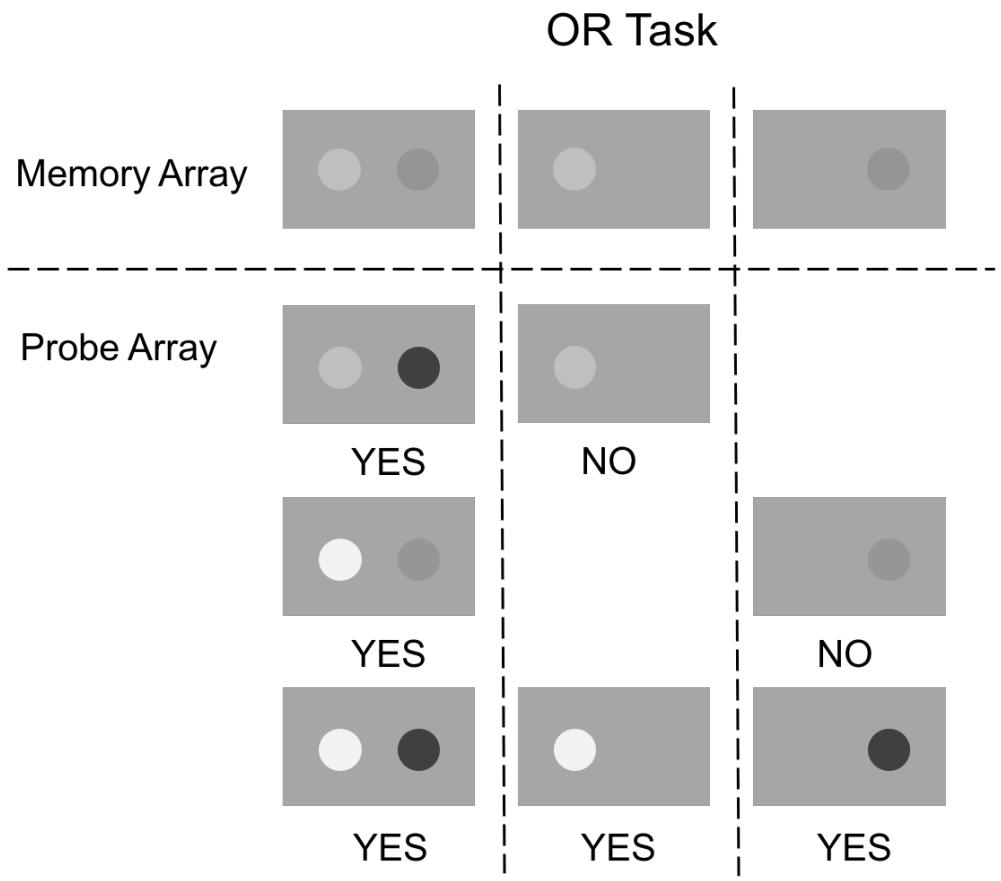

Figure 5.16: Example of redundant targets paradigm extended to a change detection task with an OR decision rule. Note: as in Experiment 1 the memory and probe arrays varied and double change items could comprise no changes, one change, or two changes.

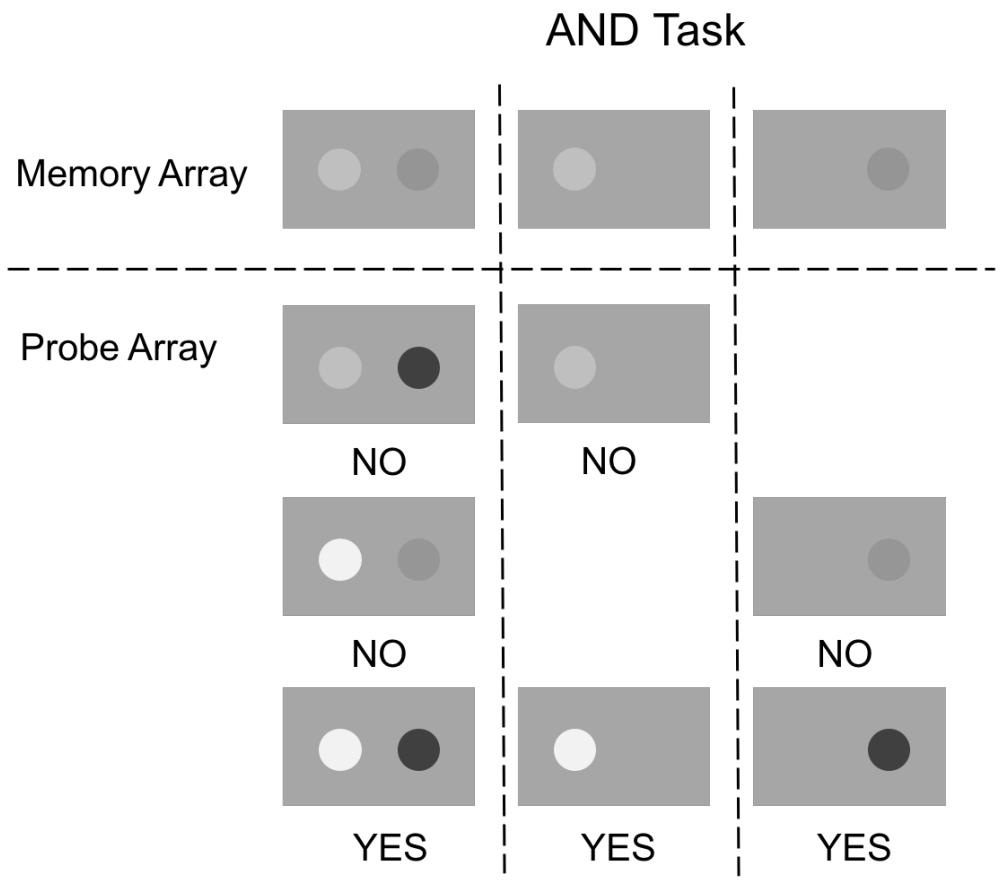

Figure 5.17: Example of redundant targets paradigm extended to a change detection task with an AND decision rule. Note: as in Experiment 1 the memory and probe arrays varied and double change items could comprise no changes, one change, or two changes.

$$H(t) = \int \frac{f(t)}{S(t)} dt = -\log(S(t)) \tag{5.7}$$

If we represent a change in the left position as A, a change in the right position as B, and a change in both positions as AB, then for the OR task, the capacity coefficient is defined as the ratio of the integrated hazard functions for the double target, H_{AB}(t), over the sum of the two single target integrated hazard functions, H_A(t) and H_B(t):

$$C_{or}(t) = \frac{H_{AB}(t)}{H_A(t) + H_B(t)} \tag{5.8}$$

The sum of the single target integrated hazard functions is the log of the derived minimum time of the single targets. The use of this value in the denominator of Equation 5.8 reflects the fact that the capacity coefficient uses the unlimited capacity parallel model as the standard of comparison. The standard parallel processing model predicts that the rate of processing will be unaffected by the workload and therefore will satisfy the equality:

$$H_{AB}(t) = H_A(t) + H_B(t) \tag{5.9}$$

A system with unlimited capacity will therefore produce a capacity coefficient equal to one across the entire time course of processing; $C_{OR}(t) = 1$. A system with limited capacity, for example a serial system, will produce a capacity coefficient less than one; $C_{OR}(t) < 1$. Additionally, a coactive system with super capacity will produce a capacity coefficient greater than one; $C_{OR}(t) > 1$.

Because the AND task requires exhaustive processing, a different measure is required in order to maintain the same interpretation of performance relative to the standard parallel model. That is, in the AND task, the baseline model is an exhaustive parallel model. Consequently, the $C_{AND}(t)$ requires use of the reverse integrated hazard function; $K(t)$ (Townsend & Wenger, 2004). $K(t)$ can be defined as the log of $F(t)$:

$$K(t) = \log(F(t)) \tag{5.10}$$

Capacity for the AND case is defined as the ratio of the sum of the reverse integrated hazard functions for the single targets, K_A(t) and K_B(t), over the double target, K_{AB}(t):

$$C(t)_{AND} = \frac{K_A(t) + K_B(t)}{K_{AB}(t)} \tag{5.11}$$

The use of the reverse integrated hazard function allows for $C_{AND}(t)$ to have the same interpretation as $C_{OR}(t)$. That is, $C_{AND}(t) = 1$ indicates unlimited capacity, $C_{AND}(t) < 1$ indicates limited capacity, and $C_{AND}(t) > 1$ indicates super capacity.

5.2.2 Predictions

In the OR task, we might expect capacity to be unlimited in line with the finding that the architecture is consistent with parallel self-termination. By contrast, in the AND task, we might expect limited capacity since the architecture was consistent with serial exhaustive processing. On the other hand, there are theoretical indications that capacity might be limited. Lilburn (2016) showed that change detection decisions were consistent with the idea that a fixed number of samples were divided between the elements of the display. The implication is that the processing rate would be less efficient in the two item condition compared to the one item condition resulting in limited capacity. Finally, we note that in Experiment 1, some participants had architecture results consistent with coactivity. We might therefore expect to see supercapacity in our redundant change task.

5.2.3 Method

Participants

Six individuals with normal or corrected to normal vision (visual acuity of at least 20/25 as determined by a near-field Snellen eye chart; normal color vision as determined by Ishihara plates) from the University of Melbourne community participated in the study. Participants were reimbursed $10 per session. The majority of the participants were new to the experimental design and had not participated in previous experiments. One participant (C5), however, had participated in Experiment 1 (O2).

Apparatus and Stimuli

The apparatus and stimuli were the same as in Experiment 1 with the addition of six single item stimuli comprising each left and right channel luminance level presented in isolation.

Procedure

The experiment was a one-shot change detection task in the same manner as Experiment 1 but with three conditions: single-item, double-item AND, and double item OR. In the AND condition participants were asked to respond with "different" only if both items had changed. In the OR condition participants were asked to respond "different" if either one or the other target changed. In all conditions change trials were equated with no change trials. The exact number of unique items and repetitions per block are printed in Table 5.13.

Participants completed five experimental sessions (one single-item session, two AND sessions, and two OR sessions). Each session began with instructions explaining the response corresponding to the current condition followed by eight practice trials. Sessions were comprised of six blocks of 48 trials each. The order of trials was randomized within each block. The order conditions were completed was also randomized.

Table 5.13: Trial frequencies per block and total presentation numbers for each trial type for Experiment 2.

		Trial Types						
		LL	HH	SL	LS	SH	HS	SS
Experiment 2 AND	Number of unique items per block	4	4	4	4	4	4	8
	Number of repetitions per block	3	3	1	1	1	1	1
	Total per session	72	72	24	24	24	24	48
Experiment 2 OR	Number of unique items per block	4	4	4	4	4	4	8
	Number of repetitions per block	1	1	1	1	1	1	3
	Total per session	24	24	24	24	24	24	144
		XL	XH	LX	HX	XS	SX	
Experiment 2 Single Items	Number of unique items per block	2	2	2	2	2	2	
	Number of repetitions per block	4	4	4	4	4	4	
	Total per session	48	48	48	48	48	48	

5.2.4 Results

RTs less than 200 ms or greater than three standard deviations above the mean were excluded from further analysis. Less than 1% of trials in total were removed using this method. Mean correct RTs, mean error RTs, and error rates for Experiment 2 are presented in Tables 5.14 to 5.16.

Capacity coefficients are displayed in Figure 5.18 for the OR task and in Figure 5.19 for the AND task. We also computed upper and lower bounds on the baseline model prediction (i.e., the so-called, Miller inequality and Grice inequality in the OR task and the upper and lower Colonius-Vorberg bounds; Colonius & Vorberg, 1994, in the AND task; see Townsend & Eidels, 2011). For each participant, capacity was computed twice (once for the high salience changes and once for the low salience changes); however, the results were similar across both so we pooled the data across salience level.

For the OR condition, capacity was below the lower bound for participants C1-C4 indicating limited capacity, although C4 shows some evidence of unlimited or even super capacity at shorter RTs.

Participants C5 and C6 showed $C(t) < 1$, indicating limited capacity; however, the function was above the lower bound consistent with what one might expect from the baseline model. For the AND condition, capacity was limited for participants C1-C3; however, there was some evidence of super capacity for

Table 5.14: Observed Mean Correct and Error RTs (ms), and Error Rates for Individual Stimuli for each Participant in Experiment 2, OR Condition.

Participant	Variable	OR Condition				
		HH	LL	HS/SH	LS/SL	SS
C1	RT correct	807.06	820.1	857.71	934.97	1022.8
	RT error	-	-	1056	1156.2	1599.3
	p(error)	-	-	0.250	0.430	0.110
C2	RT correct	905.12	1148.3	1042.5	1189.2	1073.6
	RT error	-	1605.8	1270.5	1300.9	1389.7
	p(error)	-	0.13	0.094	0.48	0.06
C3	RT correct	485.79	560.74	536.25	583.17	552.6
	RT error	-	601.96	618.58	568.51	614.16
	p(error)	-	0.21	0.14	0.43	0.27
C4	RT correct	669.09	882.27	756.1	995.18	774.51
	RT error	465.28	985.92	651.72	1132.5	1766.8
	p(error)	0.02	0.02	0.03	0.26	0.03
C5	RT correct	757.46	930.55	779.19	918.63	833.13
	RT error	775.86	927.25	786.62	785.54	1008.5
	p(error)	0.1	0.13	0.05	0.28	0.05
C6	RT correct	1015.9	916.89	940.21	1321.1	1115.9
	RT error	1429.8	872.61	1086	1187	1089.6
	p(error)	0.08	0.04	0.08	0.11	0.06

Note: - indicates error free performance; C1 = participant 1.

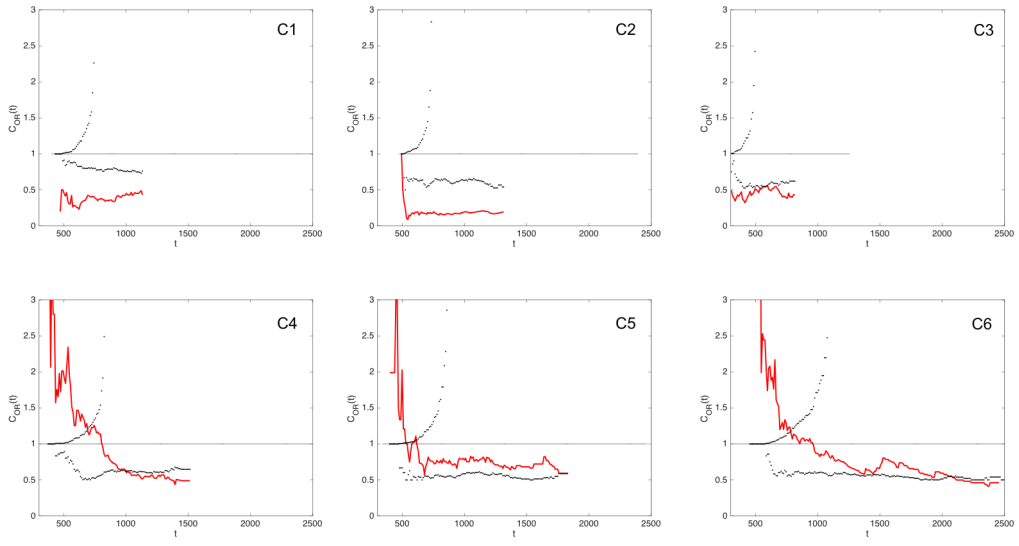

Figure 5.18: Capacity estimates for each individual participant in the OR condition. Red line represents the capacity coefficient. Black dotted lines represent the Miller (upper) and Grice (lower) bounds, respectively.

Table 5.15: Observed Mean Correct and Error RTs (ms), and Error Rates for Individual Stimuli for each Participant in Experiment 2, AND Condition.

		AND Condition				
		HH	LL	HS/SH	LS/SL	SS
C1	RT correct	1179	1561	1473.6	1443.3	1423.9
	RT error	1823.3	1544.3	1405.3	1962.2	1586.3
	p(error)	0.08	0.48	0.35	0.23	0.01
C2	RT correct	1140.1	1469.3	1398.2	1409.7	1199.9
	RT error	1691.6	1609.9	1265.9	1445.1	1974.1
	p(error)	0.06	0.44	0.19	0.15	0.01
C3	RT correct	662.86	791.19	724.81	732.56	697.47
	RT error	656.89	741.46	743.2	771.74	817.79
	p(error)	0.08	0.51	0.61	0.35	0.26
C4	RT correct	943.09	1321.1	1256.7	1316.5	967.02
	RT error	-	1614.6	1151.9	1149.2	1345.5
	p(error)	-	0.17	0.14	0.13	0.03
C5	RT correct	866.45	1008.2	896.53	845.36	761.08
	RT error	979.84	961.64	920.46	1241.3	1135
	p(error)	0.11	0.25	0.18	0.19	0.09
C6	RT correct	1089.5	1824.1	1451.2	1630.1	1301
	RT error	1419.5	1846.7	1657.9	1890.9	2472.5
	p(error)	0.05	0.26	0.29	0.22	0.03

Note: - indicates error free performance; C1 = participant 1.

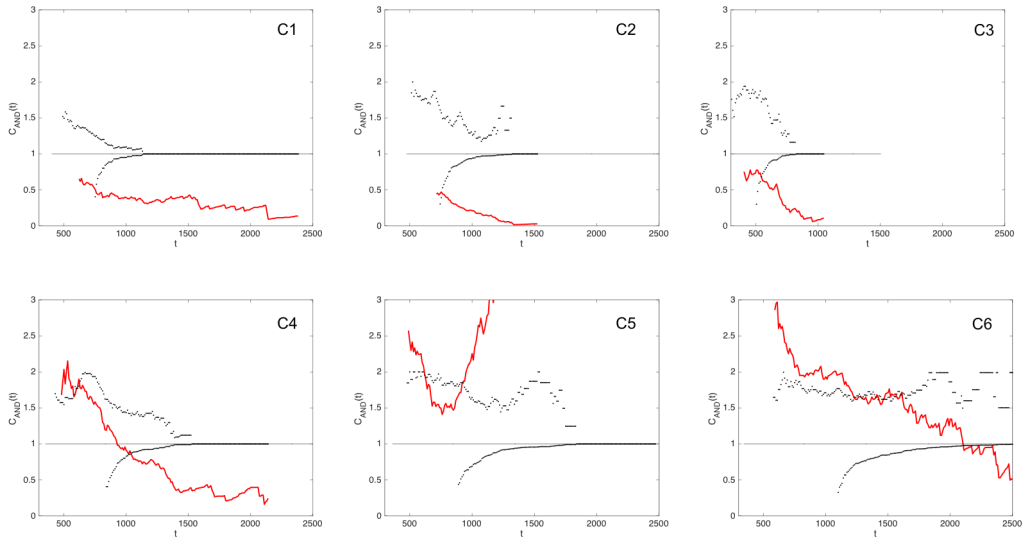

Figure 5.19: Capacity estimates for each individual participant in the AND condition. Red line represents the capacity coefficient. Black dotted lines represent the upper and lower Colonius-Vorberg bounds.

Table 5.16: Observed Mean Correct and Error RTs (ms), and Error Rates for Individual Stimuli for each Participant in Experiment 2, Single Item Condition.

		Single Item				
		HX	LX	XH	XL	SX, XS
C1	RT correct	782.49	1088.2	679.25	669.77	1025.5
	RT error	1150.6	1343	-	-	1344.9
	p(error)	0.02	0.44	-	-	0.13
C2	RT correct	714.63	1022.1	698.05	792.02	698.05
	RT error	966.19	844.66	738.93	907.76	738.93
	p(error)	0.04	0.44	0.02	0.06	0.02
C3	RT correct	493.88	528.67	469.05	523.53	606.1
	RT error	558.5	634.09	-	611.95	637.06
	p(error)	0.13	0.54	-	0.17	0.29
C4	RT correct	886.3	1030.2	830.4	864.4	744.68
	RT error	-	1020.8	-	855.53	1425.5
	p(error)	-	0.21	-	0.15	0.07
C5	RT correct	919.9	1113.5	792.06	991.05	947.2
	RT error	828.83	1048.5	653.48	1325	1024.4
	p(error)	0.11	0.34	0.04	0.17	0.09
C6	RT correct	1099.4	1278.8	941.05	1316.4	1365.4
	RT error	-	1778.9	966.89	1320.9	1977.3
	p(error)	-	0.17	0.02	0.06	0.13

Note: - indicates error free performance; C1 = participant 1.

C4-C5. The finding of super capacity is similar to findings of Eidels et al. (2015) who investigated evidence integration in a simple detection task. Eidels et al. (2015) suggested that an "AND" decision rule may more strongly engage attention across both items.

To confirm this interpretation we used Houpt and Townsend's (2012) null-hypothesis-significance test for the workload capacity analysis where conclusions about the capacity coefficient can be drawn using a z-test (see Table 5.17). If the related z-statistic is negative then processing is limited in capacity, while if it is positive then related processing is super capacity. In the OR condition, capacity was uniformly limited when compared with the baseline model while in the AND condition performance was a mixture of limited and super capacity. Interestingly, those individuals who demonstrated super capacity in the AND condition also showed super capacity for early RTs in the OR condition. However, again, since there are relatively few data points in the leading edge of the distributions used to calculate the capacity coefficient, this area of the curve should be interpreted with caution.

Table 5.17: Houpt-Townsend UCIP Test of Capacity for both AND and OR conditions

	$C_{AND}(t)$			$C_{OR}(t)$		
Participant	Capacity	z	p	Capacity	z	p
C1	Limited	-7.15	<.001	Limited	-6.2	<.001
C2	Limited	-8.35	<.001	Limited	-9.16	<.001
C3	Limited	-5.38	<.000	Limited	-5.92	<.001
C4	Super	3.53	<.001	Limited	-3.02	.002
C5	Super	4.87	<.001	Limited	-3	.002
C6	Super	6.69	<.001	Limited	-5.41	<.001

5.2.5 Discussion

Taken together, the capacity results suggest uniformly limited capacity processing for the OR task, with a mixture of limited and super capacity processing for the AND task. On the whole, the general pattern of results shows that adding

additional items slowed decision-making. Such a result is consistent with the idea of a *double-target deficit* in visual attention literature that when attention must be divided between multiple sources of information, there are small to moderate costs in detecting a target compared to focusing attention on a single source. However, when two targets that must be detected independently are presented simultaneously, there are large additional costs (Duncan, 1980; Corbett & Smith, 2017). Ultimately, a model of change detection needs to accord both with the finding that change detection decisions are made independently in each location (i.e., for many participants, parallel in the OR task and serial in the AND task) but with limited capacity for both.

5.3 General Discussion

In this paper, we sought to identify whether change-detection is best described by a first-order differencing integration model (Sorkin, 1962) or a second-order independent decisions integration model (Shaw, 1982; Noreen, 1981). We extended a set of change detection models, using different integration rules, to form a synthesized account of the time course of change detection. This account draws on prior work on mental architectures in categorization (Fifić et al., 2010) to provide a framework that allows for powerful non-parametric anlayses coupled with computational models of detection and response times. Our novel experimental method allows us to draw inferences about the processing architecture and workload capacity of detection decisions. We further investigated how task requirements affect both architecture and capacity by utilizing disjunctive (OR) and conjunctive (AND) decision rules. We characterize the strategies of individual participants in line with the small-N approach (P. L. Smith & Little, 2018).

Overall, our main empirical results were as follows: First, we found that change-detection decisions for the luminance level of two spatially separated discs generally occurred independently, consistent with second-order integration

models for most participants. Second, we found that the task requirements affected processing strategy. For participants in the OR task, change detection decision-making was best described as parallel self-terminating in nature. This pattern of parallel self-terminating processing was consistent across both non-parametric, MIC and SIC results for the double change items, and parametric, computational modeling analyses. For the AND task the SFT results were inconclusive, however, for three out of five participants computational modeling showed that decision-making proceeded in serial, self-terminating fashion.

We further found some evidence that the change signal varied differentially as a function of the strength of the change. Finally, the third major finding of the current work was that workload capacity tends to be limited, in both OR and AND tasks.

5.3.1 Implications for Theories of Change Detection

Our first major finding was that that change-detection decisions for the luminance level of two spatially separated discs occurred independently, consistent with second-order integration models for most participants. While Wilken & Ma (2004) ruled out a high threshold account of change detection decision-making, considering only choice probability meant it was difficult to differentiate between first- and second-order accounts. This question has now been addressed in the current work.

Our second major finding that the attentional strategy (e.g., parallel in the OR task but serial in the AND task) changes as a function of the task is consistent with other findings in change-detection. For example, C.-T. Yang (2011) and C.-T. Yang et al. (2011, 2013) found that varying the probability and salience of a changes in dimensions of a single object affected the processing strategy used. Further, Donkin et al. (2016) found that when the environment was predictable (i.e., set size remained constant during each block), participants were able to flexibly allocate their attention across items. However, when the environment was unpredictable (i.e., set size varied from trial to trial) participants focused

their attention on a smaller subset of items.

Within this context, studies of change detection often utilize set sizes of four item locations or more (Alvarez & Cavanagh, 2004; Cowan, 2001; Donkin et al., 2016; Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974). In order to make further inferences regarding change detection decision-making in a one-shot change detection task, a worthwhile pursuit for future research would be investigating potential differences in the decision architecture for varying set-sizes in both OR and AND tasks. While it is well known that accuracy decreases with increasing set size (Lilburn, 2016; Luck & Vogel, 1997; Pashler, 1988; P. L. Smith, 2016; Vogel et al., 2006; Wilken & Ma, 2004), especially above a limit of about four items (Cowan, 2001; Luck & Vogel, 1997), it is unclear whether the processing architecture will change, reflecting a change in strategy for integrating information. One important question is whether the tools we use here are applicable to examining increasing set size. This is not immediately evident since SFT was developed using the assumption of error-free performance in a simple detection task. Although SFT has been successfully evolved into many other domains (see Little, Altieri, et al., 2017 for review), including change detection in the current work, increasing set size necessarily will increase errors in responding. However, in practice, SFT is robust to errors and recent results by H. Yang et al. (2019) show that providing the ordering of the LL, LH, HL, and HH distributions is preserved, the SIC predictions hold. Consequently, it should be possible to examine changes in processing architecture with set size. Further, given change detection strategy has been shown to be sensitive to task demands (Donkin et al., 2016), a worthwhile pursuit for future research would be to test change detection performance in blocked and un-blocked experimental designs. These two questions form the current focus of work within our research group.

The third major finding of the current work was that workload capacity tends to be limited, regardless of whether the task requires an OR rule or an AND rule. We consider three possible explanations for this results:

(1) The capacity analyses depends on an assumption of context invariance in order to draw conclusions about how efficiency varies compared to a baseline parallel model. The assumption of context invariance implies that the processing of a channel does not change based on the contents of the other channel. For example, a left hand luminance disc is assumed to be processed at the same rate regardless of the strength of the right hand luminance disc - including the case where nothing is presented on the right at all. If context invariance holds, then capacity can be limited simply because the processing architecture is not parallel as assumed by the baseline model. For instance, in the OR task, if processing is serial or exhaustive, then capacity will be limited. In the AND task, if processing is serial, then capacity will be limited. In the present case, our finding of serial processing in the AND task of Experiment 1 is consistent with the finding of limited capacity in Experiment 2. However, the limited capacity in the OR task of Experiment 2 is not commensurate with the parallel self-terminating architecture inferred in Experiment 1.

(2) A second possibility is that there is a systematic violation of context invariance such that the processing rates for detecting changes in the redundant target are slower than processing rates for detecting changes in the corresponding disc in the single targets. One explanation which predicts a systematic violation of context invariance is based on the idea that only a fixed amount of evidence or a fixed number of *samples* can be accumulated from a display in a given amount of time. When set size is increased from one item to two items, the amount of samples is divided between two items instead of one leading to less efficient processing of each location with the larger set. In other words, the processing rate is slower when the sample size is higher. This would naturally violate the assumption of context invariance, consequently predicting limited capacity (see Altieri et al. (2017) for a discussion of how violation of context invariance affects capacity). We discuss this sample size account in detail below, after discussing several empirical effects which can be elucidated by our findings.

(3) A third possibility for the OR task is that processing is parallel but that

the parallel channels interact with each other. Eidels et al. (2011) investigated a series of parallel models which allowed for facilitatory and inhibitory interaction between the two channels. Inhibitory parallel models, in which one channel influences the other such that it slows evidence accumulation can lead to limited capacity. Consequently, an inhibitory parallel model is a plausible model of performance in the OR decision task. An example of this in practise is work by P. L. Smith & Sewell (2013) who proposed a model which uses parallel inhibitory interactions to capture the sample size model's effect in a dynamic fashion. We describe this model further below.

Our three major findings that change detection is best described as a second-order model, that different tasks lead to different architectures, and that workload capacity is primarily limited, can inform several results in the change detection and related literatures.

Scene statistics in change-detection and change-blindness

Change blindness is a failure to notice otherwise perceptible changes in a visual environment when those changes are masked in some way (e.g., with an intervening blank screen or occlusions in the scene). This failure is thought to be associated with a lack of transient signals that would otherwise accompany a change (Rensink et al., 1997). One view suggests that change-blindness reflects a limit of visual attention (Rensink, 2000). Consequently, if changes occur in an attended region of space, then a comparison between the memory and visual scene can take place and changes are readily detected. The object or feature that has changed can then be identified. Such a finding accords well with the idea that processing occurs independently at different locations.

On the other hand, some studies have suggested that participants are able to use more general knowledge of the statistics of a scene to detect changes (Brady & Tenenbaum, 2013; Howe & Webb, 2014). For example, using a one-shot change detection paradigm which varied the proportion of green discs versus red discs, Howe & Webb (2014) showed that participants were accurately able to

detect a change in the scene but were not always able to identify the location of the change. In a follow up experiment, Howe & Webb (2014) used the same number of individual item changes as the previous experiment but ensured that the overall proportion of colors remained the same between the displays. In this condition, participants were no longer able to accurately detect a change independently of identifying an item which had changed. This suggests that the underlying summary statistics (e.g., overall redness) were driving the detection of changes. In the latter experiment, however, the change had to be directly attended.

The finding that change detection performance can be affected by changes in scene statistics also suggests some implications for our understanding of processing architecture. In an OR design, where participants need to identify any change in the scene, it is possible that participants are simply monitoring the overall scene statistics in parallel. In the AND task, however, participants must instead verify that changes have occurred in both discs, resulting in a serial comparison. That serial analysis implies that attention must be applied to each location in the AND task.

A limitation of our current analysis is that more detailed differences in the overall composition of the scenes are not taken into account. By taking the absolute value of the perceptual distributions from the memory and probe arrays, we average across a number of potential different types of changes, assuming that the change magnitude for each disc (regardless of location or direction) have identical properties. For example, we treated HH trials where the left disc became darker and the right disc became lighter as identical to trials where both discs became darker. In the first example, the change in the two discs increases the overall range of luminance values present in the scene (i.e., the contrast) while the scene's mean luminance is left unaltered. In the second example, however, the contrast is unaltered while the mean luminance of the visual scene is decreased. A worthwhile direction for future research would be identifying whether or not changes which alter the overall scene statistics in these

two ways alter the decision-making architecture. This could be accomplished by controlling the proportion of trials that changed in each manner.

Any-sameness and any-difference tasks

In a comparative visual search study, Taylor (1976) presented participants with two four-letter arrays side-by-side. In one condition, participants were asked was there any *difference* between the two arrays (i.e., an OR task). In the other condition, participants were asked if any of the items had remained the *same* (this is effectively an AND task since it is equivalent to asking have *all* items changed). Taylor (1976) found that in the any-difference task, RTs became shorter as the number of differences between the arrays (varying from one to four differences) increased. For their any-sameness condition, however, while the general pattern of RTs remained the same (RTs increased as the number of critical features decreased), the effect was much larger than in the any-difference task. Further, in the any-difference condition, there was an advantage for identical arrays, compared to arrays where only one or two items changed. However, when the critical number of features in the any-sameness condition was zero, there was no RT advantage. Taken together, these results suggest that any-sameness decisions are more difficult than any-difference decisions. These results would further be expected from a change in the decision strategy from parallel in the OR task to serial in the AND task.

Hyun et al. (2009) further investigated the effects of any-sameness and any-difference decision rules in a change detection paradigm, hypothesizing that a change detection task would have a number of similarities to effects seen in visual search. In particular, they hypothesized that there would be a fundamental difference between the two decision-rules, as there is a fundamental difference in identifying presence versus absence of features (M. Treisman, 1985; Taylor, 1976). As in Taylor's (1976) comparative visual search task, Hyun et al. (2009) found that set size slopes for the any-difference condition were significantly shallower than set size slopes for the any-sameness condition, suggesting that

detecting the absence of changes is more difficult than detecting the presence of changes. Additionally, both set size slopes were steeper than what would typically be expected for an unlimited capacity system. Again, these results would be expected if processing were parallel in the any-difference task but serial in the any-sameness task. Hyun and colleagues' finding that mean RT increased with set size even in the any-difference task is also commensurate with our finding of limited capacity in our OR task.

Hyun et al. (2009) looked in further detail at the locus of capacity limitations using the N2pc component which reflects the focusing of attention onto an object (Woodman & Luck, 1999). They found that the latency of the N2pc component remained the same with varying set sizes, suggesting firstly, that changes attract spatial attention, and secondly, that identifying changes is unlimited in capacity. However, the P3 component (which represents a limited-capacity process that follows stimulus categorization; Kok, 2001) indicated a later limited-capacity comparison process. Hyun et al. (2009) suggested that this could be consistent with an effortful post-change-detection verification stage of all of the items. Alternatively, they suggested that the comparison process could be limited to the attended changes, but that this process becomes less efficient when more items are contained within VSTM.

5.3.2 The double-target deficit

When attention must be divided between multiple sources of information, there are some small to moderate costs in detecting a target compared to focusing attention on a single source. However, when two targets which must be detected independently are presented simultaneously, there are large additional costs. This is commonly referred to as the *double-target deficit* and has been studied in the auditory literature (Moray, 1970a,b; Sorkin & Pohlmann, 1973; Sorkin et al., 1973, 1976) and later in visual perception (Duncan, 1980; Corbett & Smith, 2017). Corbett & Smith (2017) suggested that the double-target deficit is caused by a capacity limitation in VSTM. This prediction relies on a two-stage

system which can pre-attentively select targets while rejecting non-targets. The double-target deficit therefore arises because two target representations need to be formed in the double-target case, but only one needs to be formed in the single-target case. Using a visual search task where participants searched for either one or two targets amongst three or two distractors, they found that this capacity limitation was well described by the sample-size relationship $1/\sqrt{2}$ for the double-target deficit.

Burmester & Wallis (2012), compared a low threshold model with a high threshold (i.e., "slots-based") account. Specifically, they were interested in testing a sample-size account (Palmer, 1990; Shaw, 1980), which assumes, like signal-detection theories, that stimuli are represented by a normally distributed noisy strength signal, that observers set a threshold for responding that can be exceeded based on noise alone, and that noise increases with the number of items which must be encoded in order to make a decision. The sample-size account has the additional constraint that a finite number of samples can be taken from the display in a given time period. If samples are taken uniformly from each location, then increasing the amount of relevant information in a visual scene necessarily decreases the quality of the samples. This process can be described by the equation $1/\sqrt{m}$ where m is the number of stimuli relevant to a decision. Burmester & Wallis (2012) found that their sample-size model better accounted for participant choice probability in a change detection task varying changes in the color, speed, size, and orientation of Gabor patches, as well as in set size (1, 2, 4, or 6 items) when compared to the high threshold, slot-based model. Sewell et al. (2014) also found that workload capacity limitations were well described by a sample-size relationship.

P. L. Smith & Sewell (2013) incorporated a sample-size relationship in their model of visual selection for briefly presented, multi-element visual displays. Using this relationship they characterized capacity limitations as occurring at the level of the visual representation in VSTM. In their model, the information retained by VSTM is determined by the selection and encoding stages which occur

in parallel with competitive interaction. In P. L. Smith & Sewell (2013)'s model, when performing tasks such as visual search, items which contain task-relevant attributes or features excite the "where" pathway of attentional selection and in turn, mutually inhibit each other via competitive interaction. The excitation rate is modulated by attention. If the attention weights for items in VSTM are equivalent, the information capacity is described using a sample-size account ($1/\sqrt{m}$ where m is the number of items selected into VSTM). Their model therefore suggests that capacity limitations occur at the level of the visual representation, rather than, for example, at the decision-making stage.

Our account is largely in agreement with their model, with the following exception: In their model, the decision-making stage is represented by a diffusion model accumulating noisy samples from this single, imperfect VSTM trace. In our account, the decision stage varies depending on the task demands. A diffusion model would be most consistent with coactivity, whereas we show here that processing is best accounted for by a parallel processing architecture in the OR task and a serial architecture in the AND task. This difference can be explained due to differences in experimental paradigm. For example, in Sewell et al. (2014) participants are asked to identify the value of a probed item presented after a to-be-remembered display is presented. In this paradigm, the probe display must be encoded into VSTM, but the decision only requires information from a single location unlike in our paradigm, where all locations are needed for the decision. The difference in decision architecture therefore maps closely to the task demands, but the fundamental sampling limitation of VSTM predict our corresponding limited capacity results.

5.4 Conclusion

In summary, the current work provides novel insight into the integration of information in change detection decision-making by accounting for the time course of information processing. This was achieved through the synthesis of

the mental architectures approaches (Fifić et al., 2010) with signal detection theories of change detection (Wilken & Ma, 2004). In particular, we made the novel finding that for most participants change detection decision-making is best described by a second-order integration model whereby decisions are made about each stimulus independently. Furthermore, we found that the task requirements affected processing strategy. When change detection decisions involved a disjunctive rule (OR task), decision-making was best described as a parallel self-terminating process. When change detection decisions involved a conjunctive rule (AND task), decision-making was generally best described as a serial self-terminating process, although there were some individual differences. Taken together with our findings of limited capacity in both tasks, our strong inference method allows us to rule out a large number of different hypotheses and to provide a basis which can inform future modeling of change detection decisions using RT. Lastly, we find our results in accord with recent theorizing from related domains, and we hope that our results, are in conjunction, mutually informative. These findings provide a way-point for moving forward in the study of change detection. In particular we have identified a number of important avenues for future study, the most important of which are a) examining whether and how architecture changes with set size in both the OR and AND tasks b) examining whether changing the predictability of the number of changes affects processing architecture and c) examining whether changes affecting the overall scene statistics changes processing architecture.

5.5 Appendix

Table 5.18 shows KS-tests for Experiment 1.

Figures 5.20 and 5.21 show survivor functions for the OR and AND task, respectively, for Experiment 1.

Table 5.18: P-Values from KS-tests of stochastic dominance Experiment 1: OR Task Participants and AND Task Participants.

Participant	Dominance Test for Selective Influence							
	HH > HL	HH > LH	HL > LL	LH > LL	HH < HL	HH < LH	HL < LL	LH < LL
O1	0.015	0.013	<.001	<.001	0.992	0.969	0.992	0.992
O2	**0.344**	<.001	<.001	<.001	0.658	1.000	1.000	0.992
O3	0.008	**0.061**	0.004	<.001	1.000	0.969	0.931	0.969
O4	<.001	0.039	<.001	<.001	0.992	0.968	1.000	1.000
A1	0.035	**0.834**	0.046	<.001	0.746	0.589	0.982	0.961
A2	<.001	<.001	<.001	<.001	1.000	1.000	0.943	0.996
A3	<.001	<.001	<.001	<.001	1.000	1.000	0.983	0.983
A4	<.001	<.001	0.002	<.001	1.000	0.963	0.983	0.935
A5	<.001	<.001	<.001	<.001	1.000	1.000	1.000	1.000

Note: O1 = participant 1 for the OR condition. A1 = participant 1 for the AND condition.

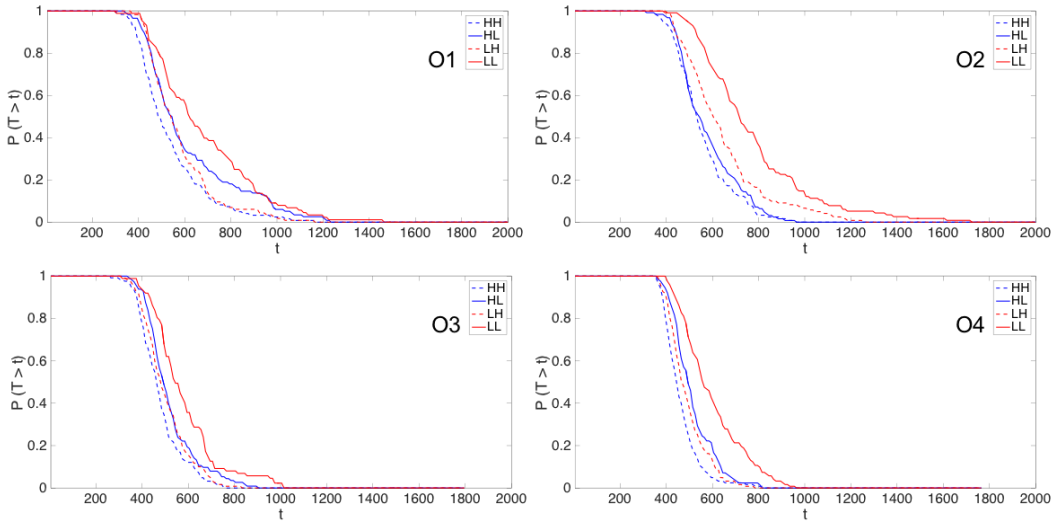

Figure 5.20: Survivor functions for individual participants in Experiment 1: OR Task

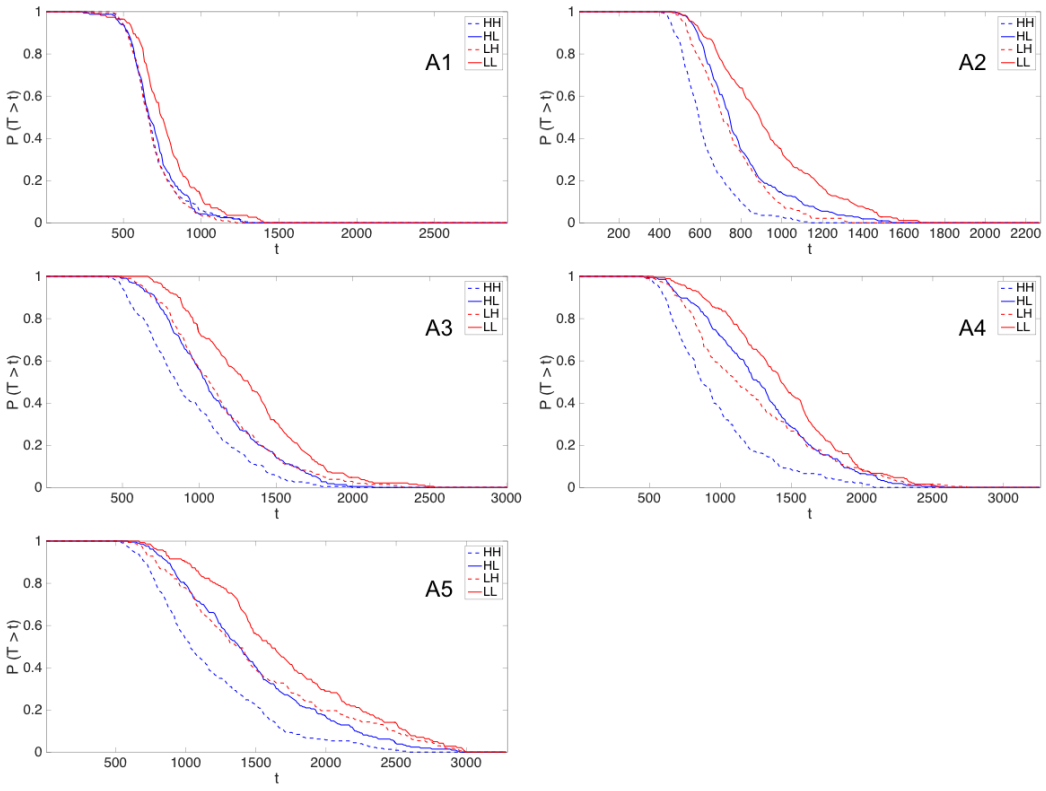

Figure 5.21: Survivor functions for individual participants in Experiment 1: AND Task

5.6 Supplementary Material

In the following posterior predictive distribution plots, each panel captures the data for a single item. The RTs are plotted as a histogram; positive values indicate target category response, and negative values indicate contrast category response. The four items in the top right of each figure belonged to the target category; consequently, most of the RTs should be positive, and any negative RTs indicate an error response for these items. The remaining items belonged to the contrast category and should therefore have mostly negative RTs. Any positive RTs for these items are error RTs.

For each observer, we took 40 samples of parameters from the posterior and generated predictions. Each sample prediction is plotted as a red line. The solid blue line is the prediction based on the average posterior parameters. The likelihood of the average posterior samples is used in the computation of the DIC.

The data were binned into 50 msec bins. The posterior predictive densities found using a kernel density estimate with a bandwidth of 10 msec.

5.6.1 OR TASK

Posterior Predictive Distributions

The following plots show the model fit to each observer for each item. First, we plot the parallel self-terminating fits for each observer. We then plot the best fitting model for the observer who was not best fit by the parallel self-terminating model.

Parallel Self-Terminating Model Figures 5.22 to 5.25 show the posterior predictions for the parallel self-terminating model for the observers in the OR condition.

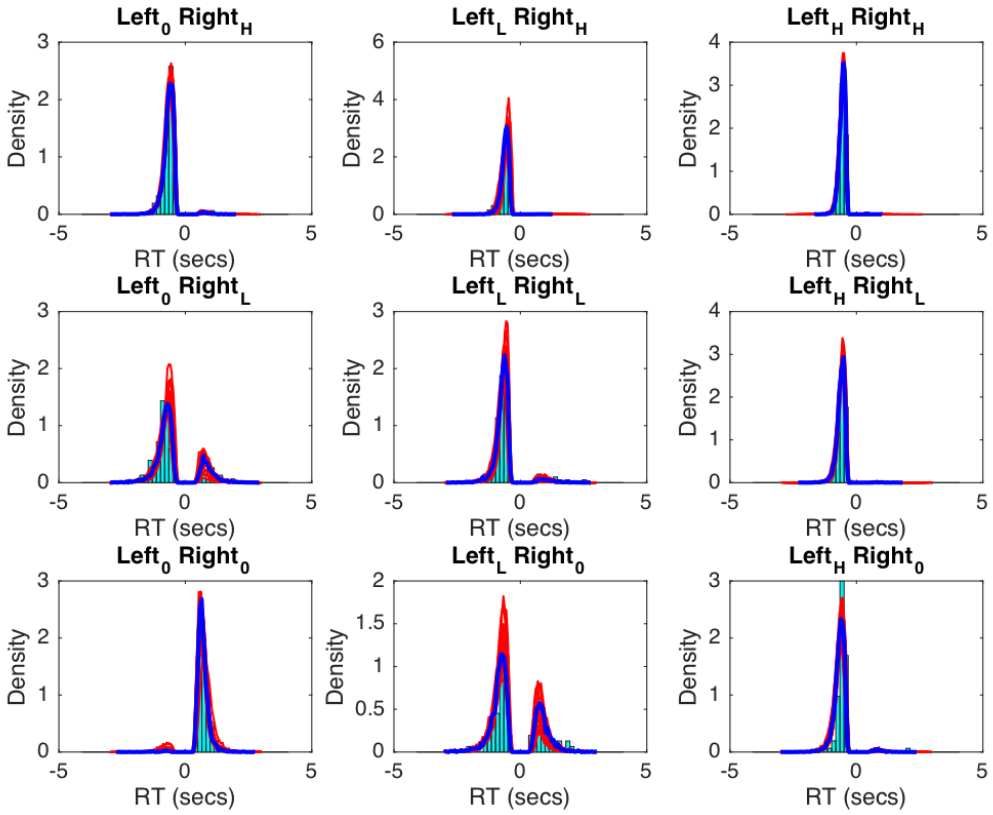

Figure 5.22: Posterior predictions from the Parallel Self-Terminating model (free-variance version) for Observer O2. Each subplot shows the predictions for one item. Positive RTs indicate target category response. Negative RTs indicate contrast category responses. Data are plotted in the histogram. Each red line indicates a posterior sample. The solid blue line indicates the predictions using the average posterior parameter values.

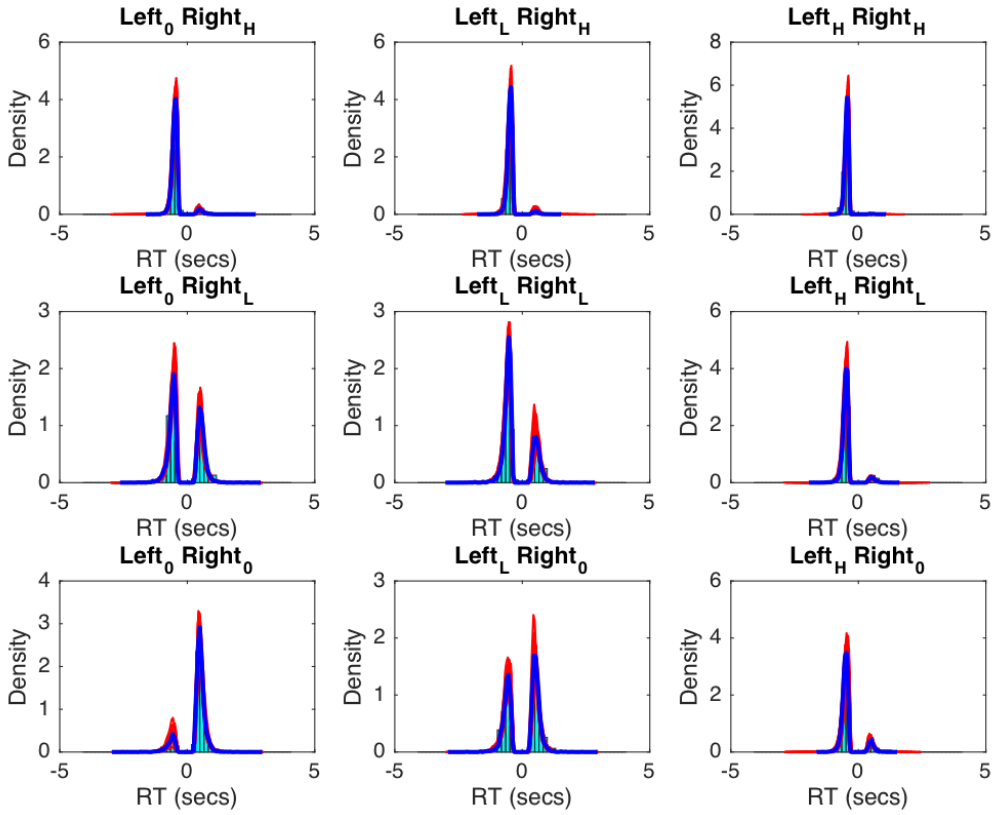

Figure 5.23: Posterior predictions from the Parallel Self-Terminating model (fixed-variance version) for Observer O3.

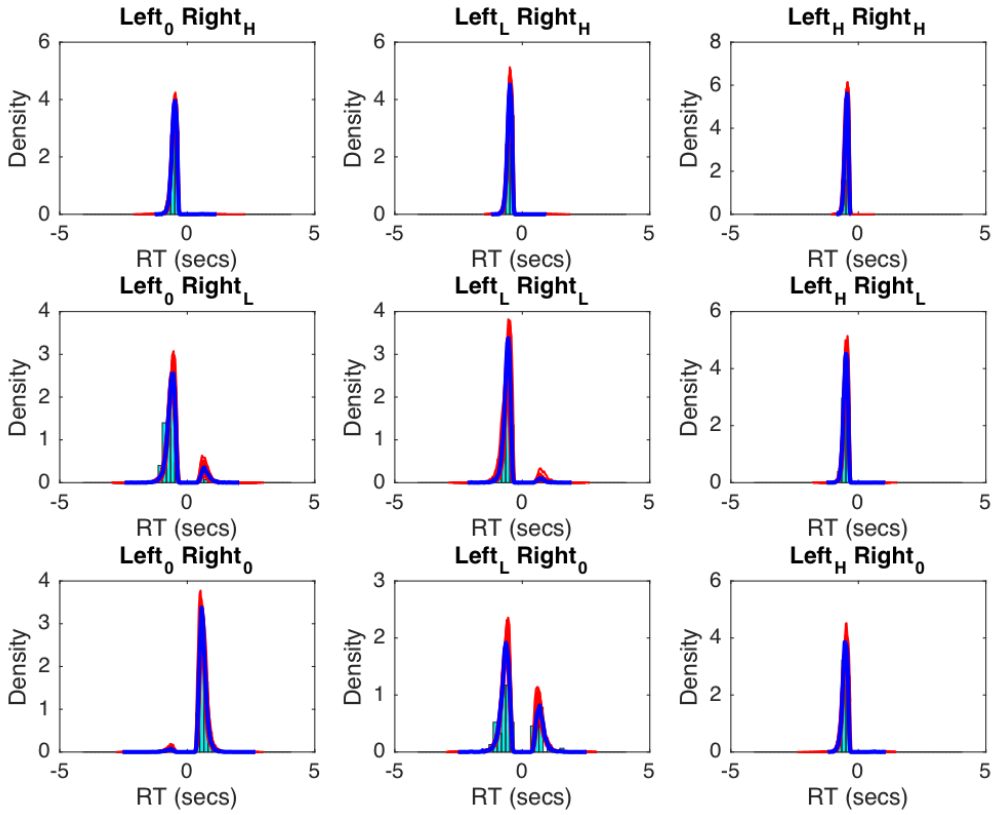

Figure 5.24: Posterior predictions from the Parallel Self-Terminating model (fixed-variance version) for Observer O4.

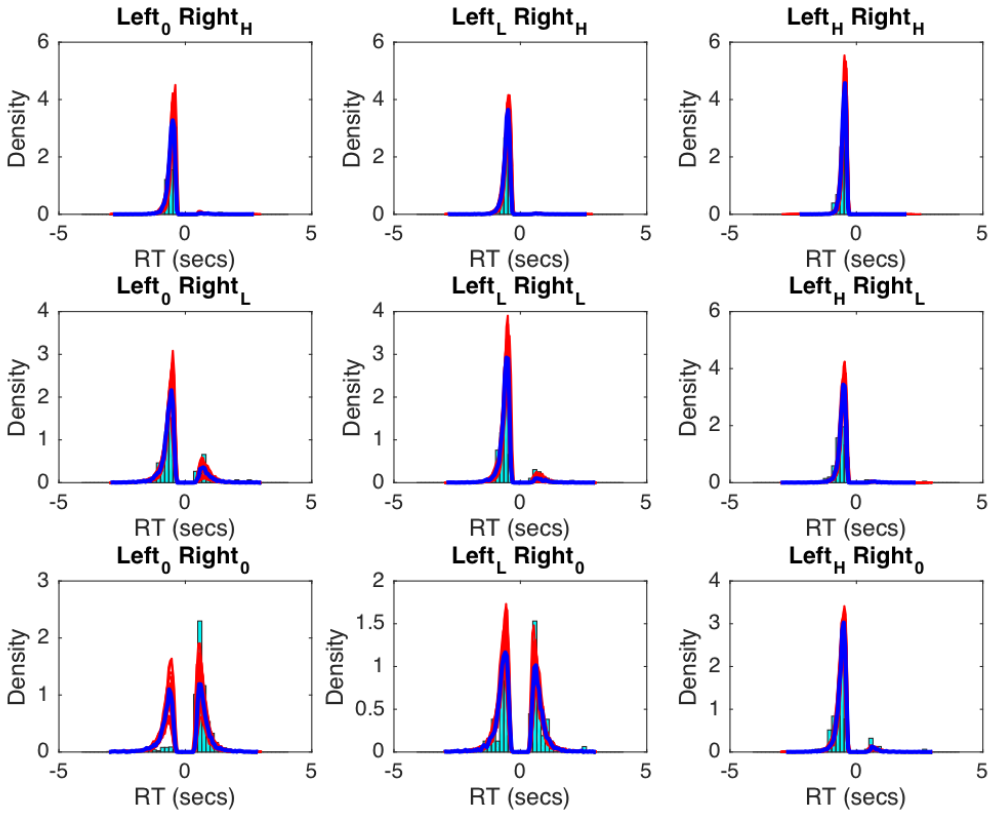

Figure 5.25: Posterior predictions from the Parallel Self-Terminating model (free-variance version) for Observer O5.

Observers O1 For one of the participants, the best fitting model was not the parallel self-terminating model, but the Coactive Minimum model.

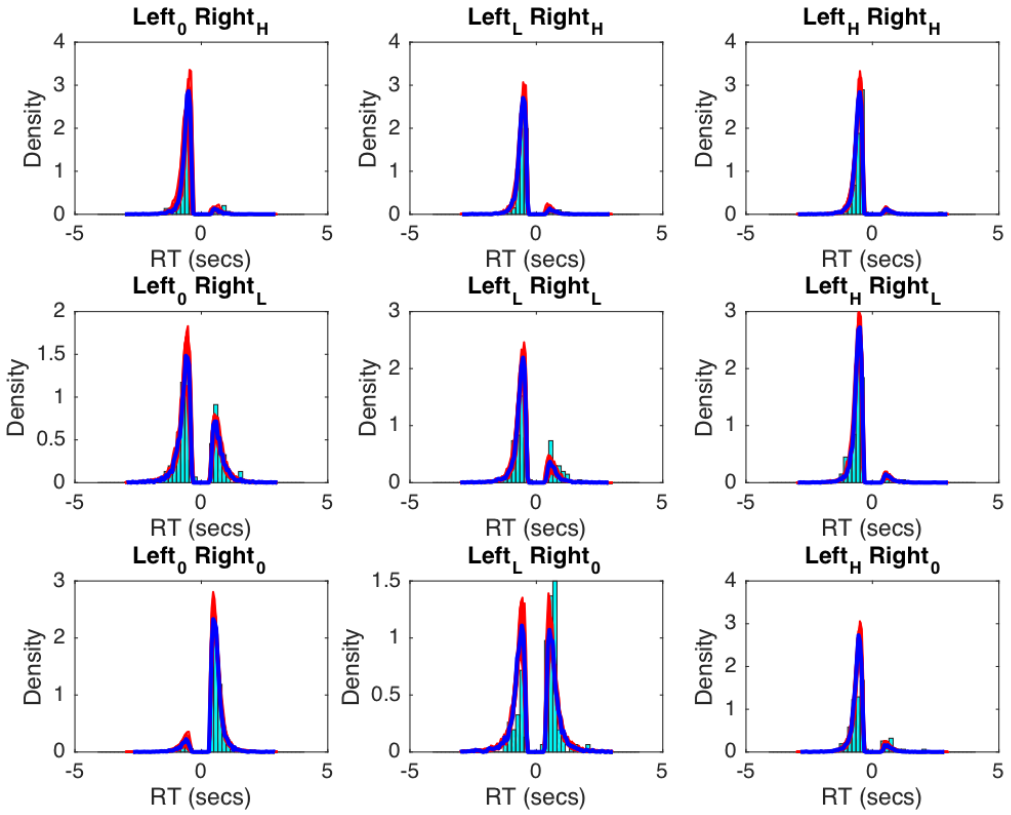

Figure 5.26: Posterior predictions from the Coactive Min model for Observer O1.

5.6.2 AND TASK

Posterior Predictive Distributions

The following plots show the model fit to each observer for each item. First, we plot the serial self-terminating fits for A2, A3 and A4 who were best fit by this model. We then plot the best fitting model for observers who were not best fit by the serial self-terminating model.

Serial Self-Terminating Model Figures 5.27 to 5.29 show the posterior predictions for the Coactive model for observers A2, A3, and A4.

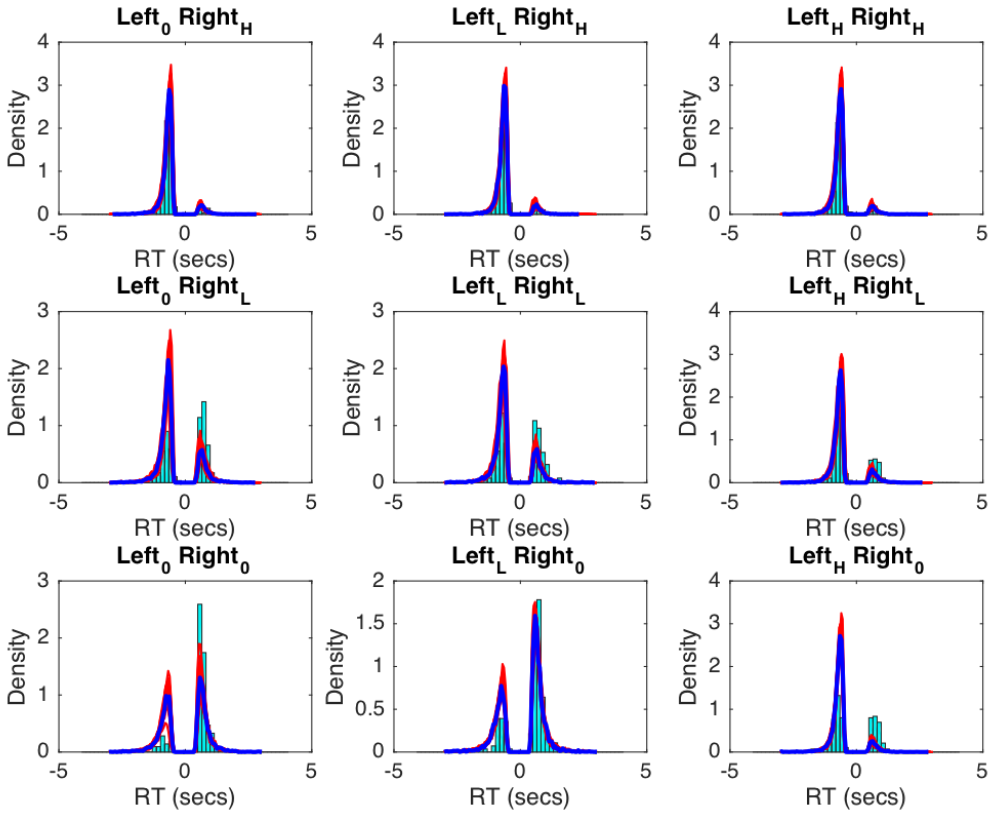

Figure 5.27: Posterior predictions from the Serial Self-Terminating model for Observer A2 (free-variance version). Each subplot shows the predictions for one item. Positive RTs indicate target category response. Negative RTs indicate contrast category responses. Data are plotted in the histogram. Each red line indicates a posterior sample. The solid blue line indicates the predictions using the average posterior parameter values.

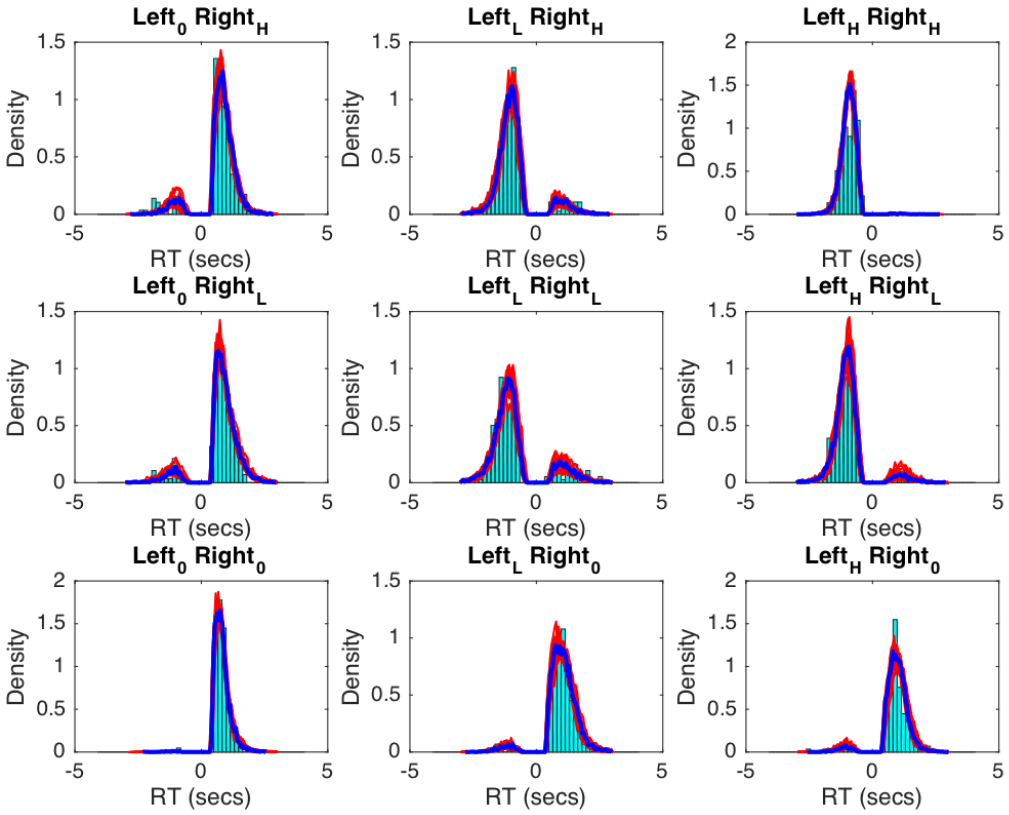

Figure 5.28: Posterior predictions from the Serial Self-Terminating model (fixed-variance version) for Observer A3.

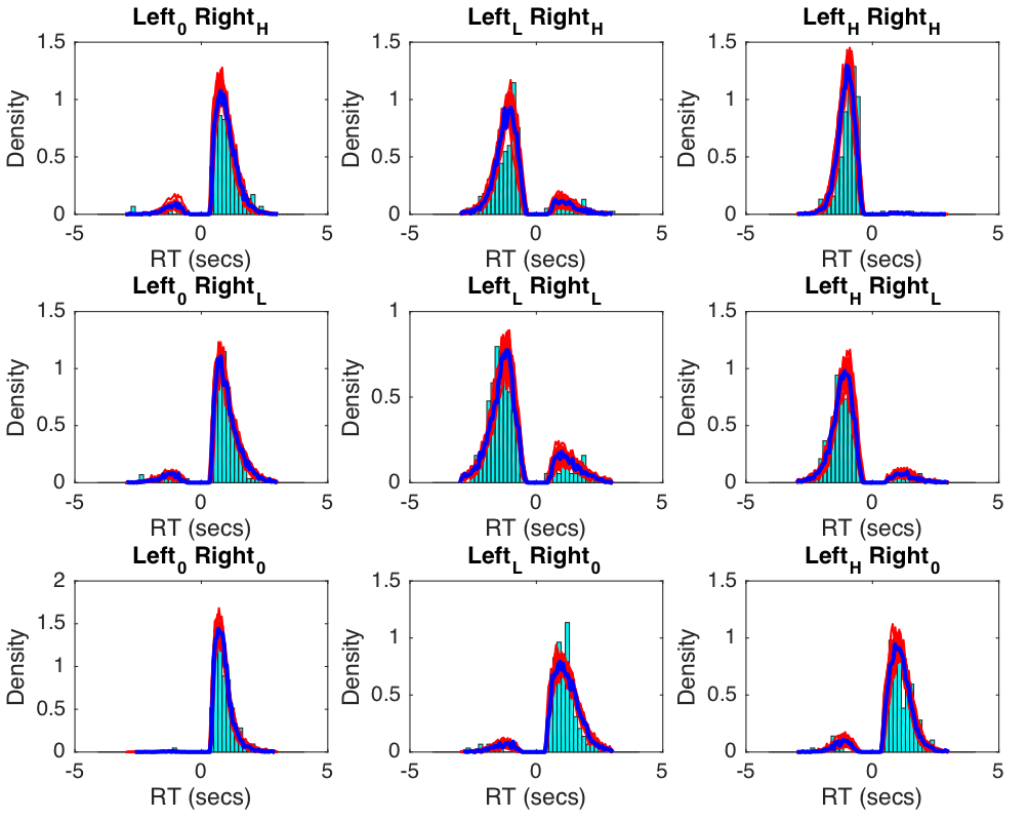

Figure 5.29: Posterior predictions from the Serial Self-Terminating model (fixed-variance version) for Observer A4.

Observers A1 and A5 For two of the participants, the best fitting model was not the serial self-terminating model, but the coactive model with a maximum integration rule.

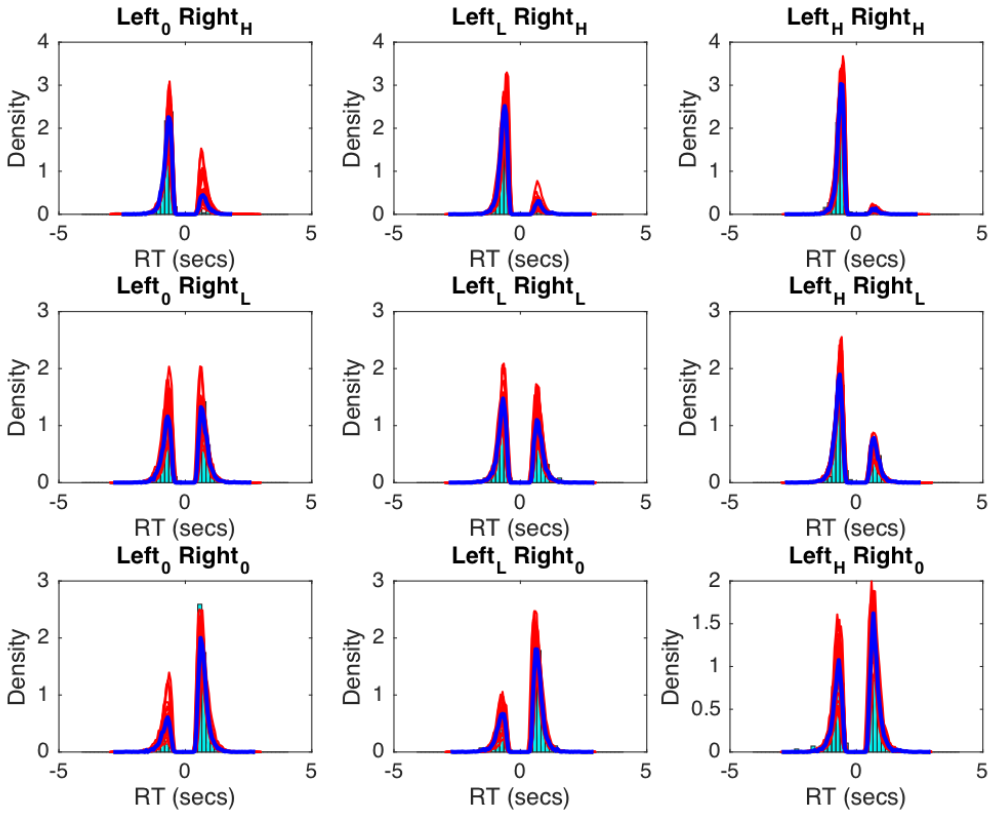

Figure 5.30: Posterior predictions from the Coactive Max model (fixed variance version) for Observer A1.

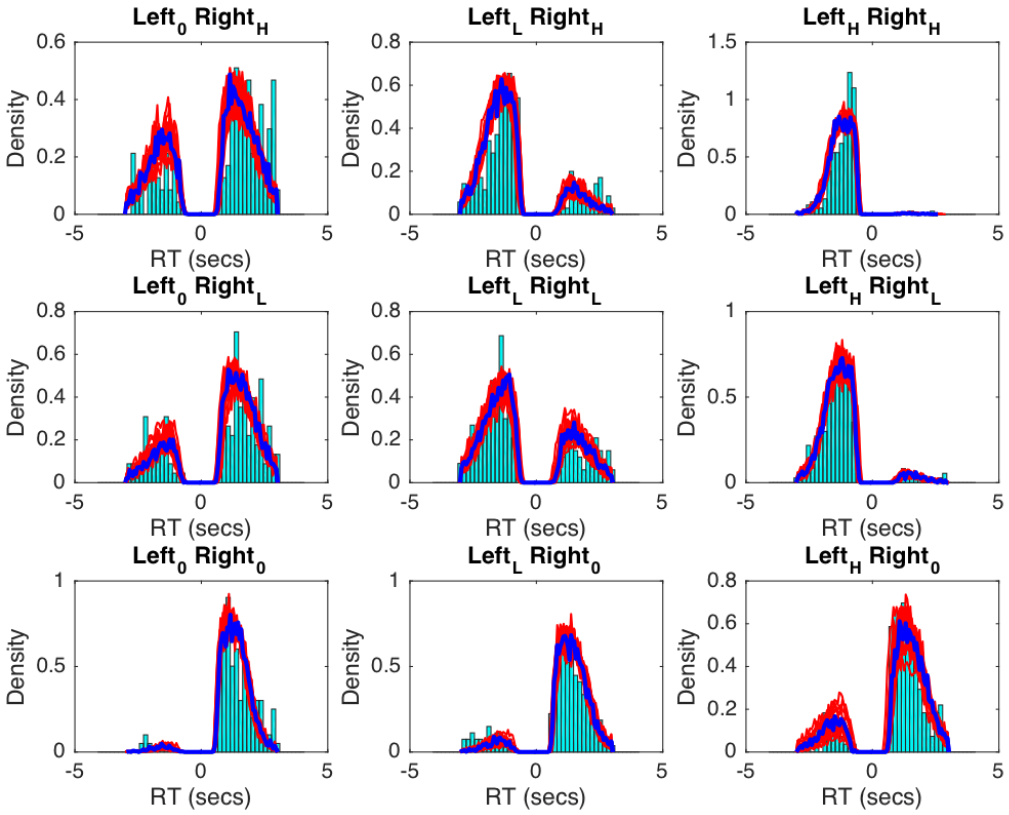

Figure 5.31: Posterior predictions from the Coactive Max model for Observer A5.

Best fitting parameter values and 95% HDIs for best fitting models

Posterior parameter estimates for the free variance models

Table 5.19: Best fitting parameter values and 95% HDIs for the best fitting model for each participant: Free variance models

PARTICIPANT/ MODEL	D_{LEFT}	D_{RIGHT}	σ_{LEFT}	σ_{RIGHT}	A	B_{CHANGE}	$B_{NOCHANGE}$	S	T_0	P_X
O1 COACTIVE MIN	0.93 [0.88-0.96]	0.91 [0.85-0.96]	0.67 [0.56-0.78]	0.36 [0.24-0.48]	0.32 [0.26-0.37]	0.13 [0.09-0.17]	0.17 [0.12-0.23]	0.34 [0.30-0.38]	0.24 [0.20-0.28]	-
O2 PARALLEL ST	0.82 [0.73-0.90]	0.74 [0.66-0.82]	0.47 [0.31-0.59]	0.48 [0.35-0.59]	0.39 [0.33-0.44]	0.21 [0.16-0.26]	0.29 [0.24-0.34]	0.35 [0.22-0.28]	0.15 [0.11-0.19]	-
O3 PARALLEL ST	0.79 [0.70-0.87]	0.71 [0.61-0.81]	1.01 [0.86-1.13]	0.72 [0.53-0.90]	0.21 [0.19-0.24]	0.05 [0.03-0.07]	0.20 [0.17-0.23]	0.20 [0.18-0.23]	0.17 [0.14-0.21]	-
O4 PARALLEL ST	0.78 [0.72-0.84]	0.69 [0.62-0.76]	0.66 [0.57-0.74]	0.51 [0.41-0.58]	0.24 [0.21-0.27]	0.17 [0.14-0.20]	0.23 [0.20-0.27]	0.18 [0.16-0.20]	0.16 [0.12-0.19]	-
O5 PARALLEL ST	0.88 [0.79-0.94]	0.81 [0.79-0.94]	0.74 [0.60-0.87]	0.38 [0.24-0.52]	0.26 [0.20-0.32]	0.16 [0.11-0.21]	0.26 [0.21-0.31]	0.31 [0.28-0.35]	0.18 [0.15-0.22]	-
A1 COACTIVE MAX	0.47 [0.27-0.68]	0.68 [0.54-0.82]	1.58 [1.14-2.04]	1.16 [1.00-1.35]	0.19 [0.15-0.24]	0.23 [0.19-0.27]	0.27 [0.23-0.32]	0.19 [0.17-0.22]	0.17 [0.12-0.22]	-
A2 SERIAL ST	0.91 [0.86-0.94]	0.84 [0.79-0.90]	0.61 [0.54-0.69]	0.42 [0.31-0.50]	0.21 [0.18-0.24]	0.25 [0.21-0.28]	0.09 [0.07-0.11]	0.25 [0.23-0.27]	0.23 [0.20-0.25]	0.53 [0.45-0.61]
A3 SERIAL ST	0.78 [0.72-0.83]	0.77 [0.71-0.85]	0.49 [0.42-0.56]	0.37 [0.25-0.46]	0.51 [0.47-0.56]	0.25 [0.21-0.30]	0.10 [0.07-0.13]	0.24 [0.21-0.26]	0.20 [0.16-0.25]	0.66 [0.58-0.74]
A4 SERIAL ST	0.79 [0.73-0.85]	0.75 [0.69-0.81]	0.44 [0.35-0.52]	0.44 [0.36-0.51]	0.61 [0.55-0.66]	0.28 [0.23-0.33]	0.11 [0.08-0.14]	0.25 [0.22-0.28]	0.18 [0.13-0.22]	0.64 [0.56-0.73]
A5 COACTIVE MAX	0.67 [0.59-0.73]	0.48 [0.43-0.53]	0.86 [0.76-0.97]	0.52 [0.46-0.57]	1.16 [1.07-1.26]	0.40 [0.33-0.47]	0.39 [0.32-0.46]	0.18 [0.16-0.21]	0.20 [0.14-0.27]	-

Table 5.20: Best fitting parameter values and 95% HDIs for the best fitting model for each participant: Fixed variance models

PARTICIPANT/ MODEL	D_{LEFT}	D_{RIGHT}	$\sigma_{NOCHANGELEFT}$	$\sigma_{LOWCHANGELEFT}$	$\sigma_{HIGHCHANGELEFT}$	$\sigma_{NOCHANGERIGHT}$	$\sigma_{LOWCHANGERIGHT}$	$\sigma_{HIGHCHANGERIGHT}$	A	B_{CHANGE}	$B_{NOCHANGE}$	S	T_0	P_X
O1 COACTIVE MIN	0.92 [0.86-0.96]	0.86 [0.76-0.93]	0.64 [0.52-0.76]	0.68 [0.41-0.94]	1.00 [0.46-1.64]	0.20 [0.06-0.38]	0.69 [0.39-1.04]	0.27 [0.07-0.73]	0.32 [0.28-0.38]	0.12 [0.08-0.16]	0.16 [0.11-0.21]	0.34 [0.30-0.38]	0.25 [0.21-0.28]	-
O2 PARALLEL ST	0.78 [0.66-0.89]	0.68 [0.53-0.83]	0.34 [0.09-0.58]	0.68 [0.33-1.11]	0.23 [0.07-0.48]	0.42 [0.23-0.58]	0.64 [0.31-1.10]	0.30 [0.06-0.83]	0.40 [0.35-0.45]	0.20 [0.16-0.26]	0.28 [0.23-0.32]	0.26 [0.23-0.28]	0.16 [0.12-0.20]	-
O3 PARALLEL ST	0.83 [0.74-0.91]	0.68 [0.58-0.80]	0.51 [0.11-1.04]	0.70 [0.41-1.01]	0.97 [0.74-1.22]	0.85 [0.38-1.03]	0.71 [0.42-1.09]	0.33 [0.08-0.86]	0.24 [0.21-0.27]	0.04 [0.02-0.06]	0.17 [0.14-0.21]	0.21 [0.19-0.24]	0.20 [0.17-0.23]	-
O4 PARALLEL ST	0.77 [0.68-0.87]	0.64 [0.51-0.79]	0.64 [0.52-0.75]	0.70 [0.41-1.05]	0.25 [0.07-0.56]	0.45 [0.30-0.61]	0.61 [0.33-1.00]	0.24 [0.07-0.50]	0.25 [0.22-0.27]	0.16 [0.14-0.20]	0.23 [0.19-0.26]	0.19 [0.17-0.21]	0.17 [0.14-0.20]	-
O5 PARALLEL ST	0.85 [0.75-0.92]	0.73 [0.60-0.86]	0.68 [0.56-0.80]	0.68 [0.39-1.02]	1.42 [1.00-2.07]	0.27 [0.08-0.51]	0.66 [0.33-1.09]	0.23 [0.07-0.49]	0.25 [0.20-0.30]	0.18 [0.14-0.22]	0.26 [0.22-0.30]	0.31 [0.28-0.34]	0.17 [0.14-0.21]	-
A1 COACTIVE MAX	0.72 [0.53-0.87]	0.60 [0.47-0.77]	1.93 [1.51-2.34]	0.44 [0.22-0.72]	1.64 [1.17-2.32]	1.78 [0.93-1.53]	0.69 [0.36-1.13]	0.30 [0.06-0.82]	0.21 [0.17-0.25]	0.21 [0.17-0.25]	0.26 [0.21-0.31]	0.21 [0.18-0.23]	0.19 [0.15-0.25]	-
A2 SERIAL ST	0.89 [0.83-0.94]	0.77 [0.66-0.87]	0.58 [0.50-0.65]	0.70 [0.45-0.96]	0.25 [0.07-0.57]	0.25 [0.08-0.43]	0.67 [0.36-1.07]	0.23 [0.08-0.50]	0.21 [0.18-0.24]	0.25 [0.22-0.28]	0.09 [0.07-0.10]	0.26 [0.24-0.28]	0.23 [0.21-0.26]	0.54 [0.47-0.61]
A3 SERIAL ST	0.71 [0.60-0.83]	0.67 [0.51-0.83]	0.43 [0.34-0.53]	0.67 [0.37-1.08]	0.25 [0.07-0.59]	0.23 [0.08-0.40]	0.58 [0.28-1.02]	0.26 [0.07-0.66]	0.51 [0.47-0.56]	0.26 [0.21-0.31]	0.10 [0.07-0.12]	0.24 [0.22-0.27]	0.20 [0.16-0.25]	0.67 [0.60-0.74]
A4 SERIAL ST	0.71[0.59-0.83]	0.66 [0.53-0.81]	0.34 [0.14-0.48]	0.65 [0.35-1.04]	0.26 [0.08-0.66]	0.36 [0.24-0.48]	0.63 [0.33-1.04]	0.25 [0.08-0.56]	0.61 [0.56-0.66]	0.28 [0.23-0.34]	0.12 [0.08-0.14]	0.26 [0.23-0.28]	0.18 [0.14-0.22]	0.64 [0.55-0.72]
A5 COACTIVE MAX	0.67 [0.55-0.81]	0.52 [0.34-0.73]	0.86 [0.71-1.06]	0.67 [0.37-1.06]	0.32 [0.08-0.86]	0.53 [0.34-0.75]	0.54 [0.28-0.94]	0.33 [0.08-1.05]	1.18 [1.08-1.27]	0.39 [0.32-0.46]	0.39 [0.32-0.46]	0.20 [0.17-0.22]	0.22 [0.16-0.29]	-

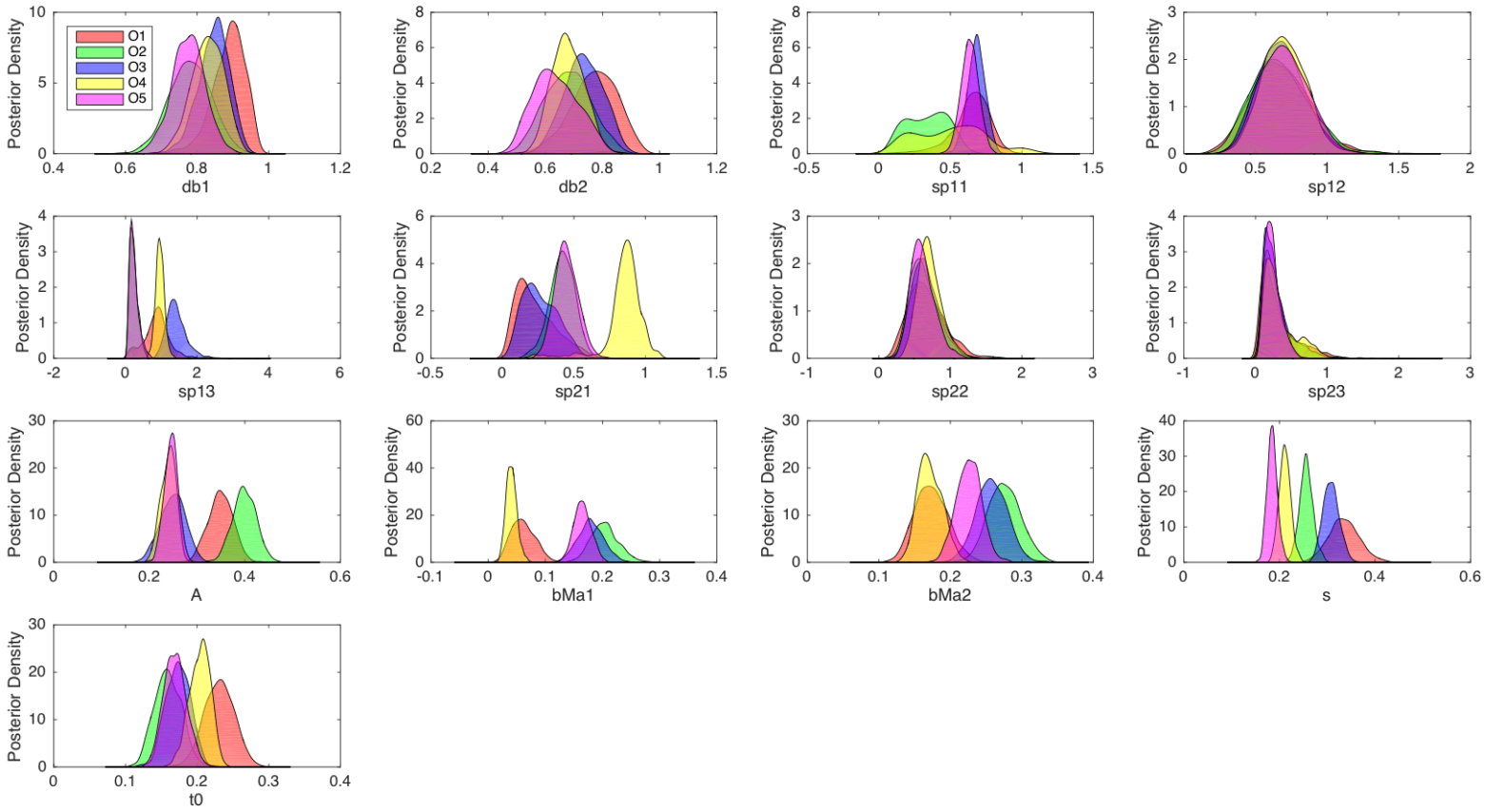

Figure 5.32: Posterior parameter estimates for each participant in the free variance OR condition, parallel self-terminating model. db1 and db2 = decision boundary for the left and right items, respectively. sp11, sp12, and sp13 = the standard deviation of the change distribution for the left item for S, L, and H salience changes, respectively. sp21, sp22, and sp23 = the standard deviation of the change distribution for the right item for S, L, and H salience changes respectively. A = starting point. bMa1 and bMa2 represent change and no-change decision thresholds, respectively. s = drift rate standard deviation. t_0 = non-decision time.

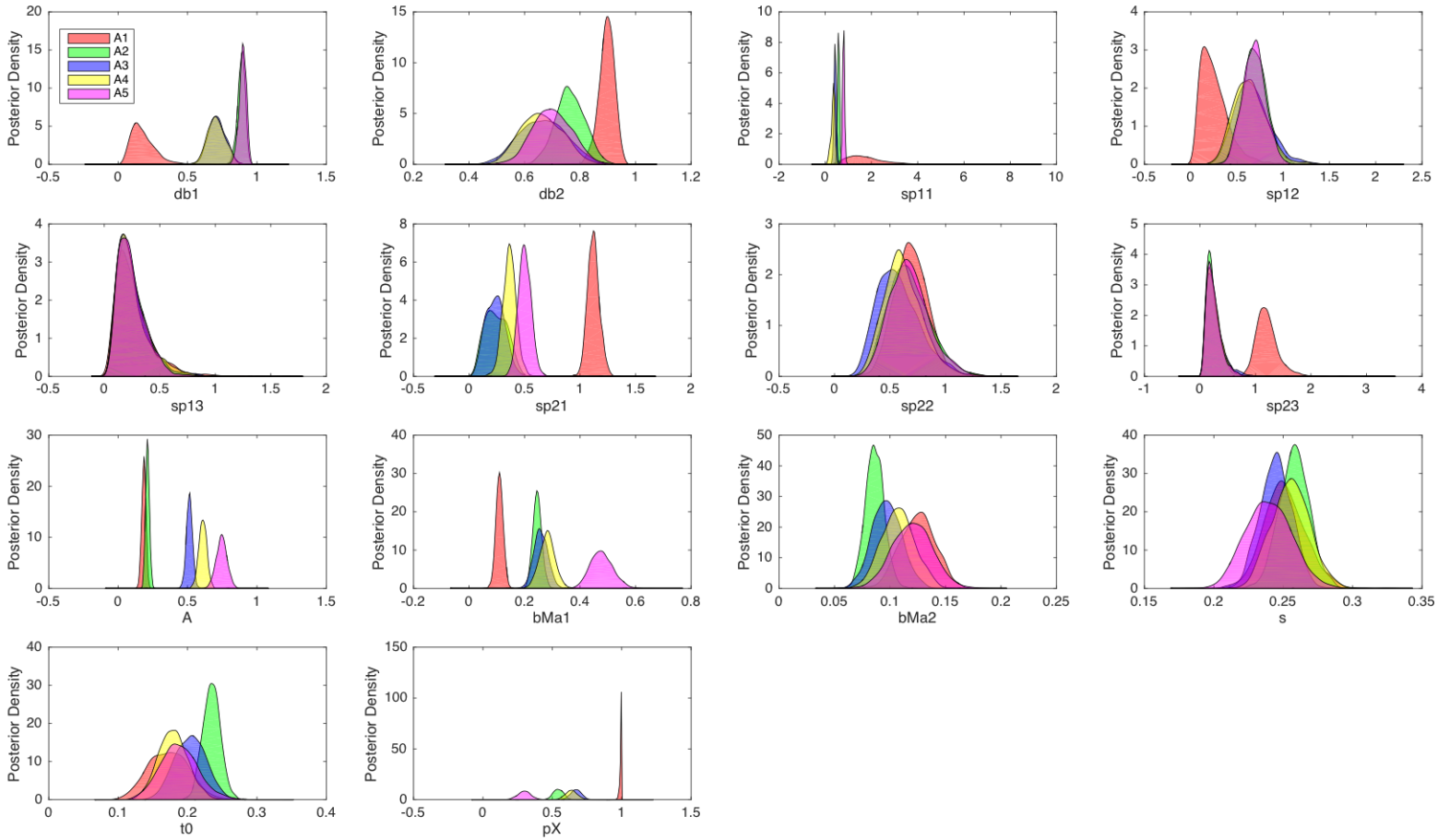

Figure 5.33: Posterior parameter estimates for each participant in the free variance AND condition, serial self-terminating model. db1 and db2 = decision boundary for the left and right items, respectively. sp11, sp12, and sp13 = the standard deviation of the change distribution for the left item for S, L, and H salience changes, respectively. sp21, sp22, and sp23 = the standard deviation of the change distribution for the right item for S, L, and H salience changes respectively. A = starting point. bMa1 and bMa2 represent change and no-change decision thresholds, respectively. s = drift rate standard deviation. t_0 = non-decision time. p_x = probability one dimension is processed before the other.

Chapter 6

Summary and conclusion

The aim of this thesis was to characterize the underlying organization of processing over time for within-dimension features across both categorization and change detection tasks. Specifically, we sought to characterize whether information processing proceeds in serial, parallel, or is pooled into a single decision-making channel (Kantowitz, 1974; Sternberg, 1969; Schweickert, 1993; Townsend, 1984), the stopping rules (i.e., whether the process is exhaustive or self-terminating), and the efficiency of processing (i.e. workload capacity; Townsend & Ashby, 1983; Wenger & Townsend, 2000). The major contributions of this thesis are twofold. Firstly, we extended previous work in categorization to a novel stimulus type: within-dimension features. Here, we demonstrated that within-dimension luminance features are pooled together into a single decision channel in categorization tasks. This finding is novel because it is not commensurate with the idea that spatially separated dimensions need to be resolved serially due to limits on the spatial aspects of visual attention as one might conclude from previous work (see e.g., Little et al., 2011). Secondly, building from this work in categorization, we presented a novel change detection task which enables testing of the qualitative predictions of a large number of different mental architectures and stopping rules. Accompanying this, we proposed a novel extension of the Logical-Rules modeling framework to a change detection paradigm. Here, we developed a novel

representation of the change signal by taking the difference between a Gaussian distribution of perceptual effects representing the memory array and a Gaussian distribution of perceptual effects representing the probe array. We also further allowed for additional variability in accordance with the strength of the change present in the visual scene. This work is a departure from investigations in the broader VSTM literature which tend to focus on characterizing the memory representation and storage limits (Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974). Instead, this thesis provides a unique contribution via its theoretical focus on the way in which information is integrated in order to make change detection decisions.

The current thesis was divided into two main parts. The first half comprised three experiments and examined decision-making using within-dimension stimuli in a categorization task. The second half comprised four experiments focusing on developing a novel framework to characterize processing architecture, stopping rule, and capacity in a change detection task. In the following I summarise the major conclusions from these experiments before turning to a summary of the major implications and avenues for future research.

6.1 The architecture and capacity of categorization decision-making

The aim of the first section of the thesis was to characterize the architecture, stopping rule, and capacity of within-dimension stimuli. In the first two experiments architecture and stopping rule were assessed through the analysis of RTs using Systems Factorial Technology (SFT; Townsend & Nozawa, 1995; Little, Altieri, et al., 2017), the Logical-Rules paradigm (Little, Altieri, et al., 2017; Fifić et al., 2010) and relevant computational modeling (Fifić et al., 2010). In the third experiment, processing capacity was assessed using the redundant targets task and the capacity coefficient (Townsend & Ashby, 1983; Wenger & Townsend, 2000). These experiments, as well as the experiments in the second

part of the thesis, focused specifically on spatially separated luminance discs. These within-dimension stimuli are unique to investigations in the categorization literature although the distinction between within- and between-dimension features forms a key theoretical difference in other areas of the visual attention literature.

In Experiments 1 and 2 of Chapter 3 participants were asked to categorize pairs of luminance discs as belonging to either Category A or Category B in accordance with the Logical-Rules paradigm. In Experiment 1 the screen was divided by a boundary of luminance discs to reduce the probability that the discs would be treated as a single object. This boundary was removed in Experiment 2. For ten of the fourteen participants across both experiments both non-parametric analysis and computational modeling provide evidence for coactive processing. For the target category, the non-parametric MIC and SIC results across both experiments showed a pattern of coactivity which was statistically supported by directional KS-Test results. There was individual variation in the contrast category, however, these results also tended towards a pattern of coactivity. Finally, the computational modeling provided further evidence that the majority of participants favored a strategy in which information was pooled into a single channel.

In Chapter 4 we implemented the redundant targets detection paradigm (Egeth & Mordkoff, 1991; Snodgrass & Townsend, 1980) using the same within-dimension stimuli as in Chapter 3. This paradigm differed from the standard redundant targets paradigm in that we introduced a series of catch trial stimuli. This was done to ensure the experiments were analogous. This resulted in one double target (white on the left and black on the right) which participants were required to respond "YES" to, and one double target (black on the left and white on the right) which participants were required to respond "NO" to. The experiment also included four single target items and one blank trial for which "NO" was also the correct response. Again, to decrease the likelihood that objects would be treated as a single object, we introduced a boundary of

luminance discs in the centre of the screen. This boundary was again removed for the second, *no-boundary*, condition. In the boundary condition, the majority of participants showed $c(t) < 1$ indicating limited capacity processing. In the non-boundary condition, four out of seven participants clearly showed $c(t) < 1$ also indicating limited capacity processing. For participant NB5, capacity was limited for faster RTs but for longer RTs showed super capacity. NB1 and NB4 showed a $c(t) > 1$ indicating super capacity.

At face value, the results from Chapters 3 and 4, seem to indicate a coactive model with limited capacity. However, the capacity analysis depends on the assumption of context invariance in order to draw conclusions about how efficiency varies in comparison to the baseline parallel unlimited capacity model. That is, a given luminance disc located on the left hand side will always be processed at the same speed regardless of whether the disc on the right is high salience or low salience, black or white etc. If context invariance holds, then limited capacity would indicate that the architecture was either serial or limited capacity parallel. However, the finding of a coactive architecture with limited capacity processing is not commensurate with the assumption of context invariance. Instead, it could be that there is a systematic violation of context invariance occurring such that there is a cost to processing the double target item compared to the corresponding discs in single target trials.

One explanation which predicts a systematic violation of context invariance is the sample-size account (Palmer, 1990; Shaw, 1980). This account assumes that only a fixed number of *samples* can be accumulated from a display in a given amount of time. When set size is increased from one item to two items, the amount of available samples is divided between the two items leading to less efficient processing of each location. This leads to slower processing for higher set sizes.

Another alternative is that the architecture of categorization decision-making utilising spatially separated luminance discs is actually parallel but where the channels are not independent of one another. For example, Eidels et al. (2011)

describe a pre-accumulator inhibitory parallel model whereby higher levels of input into one channel leads to lower input in the other channel. As the inhibition between the discs increases, the positive portion of the SIC for higher RT also increases leading to a pattern which looks like coactivity, however, capacity is limited. This finding would be in line with theories of vision such as Gilchrist et al.'s (1999) *Anchoring Theory of Lightness Perception* which propose that luminance perception is relative.

P. L. Smith & Sewell (2013) utilize both a sample-size relationship and inhibitory interaction in their model of visual selection for briefly presented, multi-element visual displays. They propose a parallel selection process whereby relevant stimuli are selected for encoding using competitive interaction. When two stimuli are equally task relevant they mutually inhibit each other in accordance with the sample size account. This finding is therefore important as it is commensurate with other major theories of visual attention.

Nonetheless, a limitation to this interpretation is that the experimental design may have artificially induced limited capacity processing. Additional catch trials were introduced to make the task commensurate with the categorization experiment introduced in Chapter 3; however, this may have created a situation whereby single target items were easy to dismiss, whereas the two double targets took longer to categorize as they each were associated with different responses, leading to to limited capacity processing. Although in simple detection experiments using luminance values the common finding is also limited capacity processing (Townsend & Nozawa, 1995), the current literature would nonetheless benefit from further investigation into a measure of capacity which requires a further categorization or identification decision. For example, to overcome this issue an adaptation of a full identification task such as that described in Howard et al. (2019) could perhaps be used.

6.2 The architecture, stopping rule, and capacity of change detection decision-making

The aim of the second section of the thesis was to test whether information from different locations in a change detection task are best described by a *first-order* evidence integration model, where information from each change location is pooled, or by a *second-order integration model*, where information from each change location is considered independently (see e.g., Noreen, 1981; Sorkin, 1962; Shaw, 1982). Following the research of Wilken & Ma (2004) using accuracy, we developed a novel modeling framework for characterizing the time course of change detection based on information held in visual short-term memory. We combined non-parametric SFT analysis with Bayesian model comparison using posterior density estimation to approximate the likelihoods of each model and differential evolution MCMC to estimate posteriors. This framework allowed us to answer whether change detection is better captured by a pooled first-order integration model (i.e., a coactive model) or by an independent second-order integration model (i.e., a serial or parallel model). We further investigate stopping rule (i.e., detecting any change vs. detecting all changes), and how the efficiency of detection is affected by the number of changes in the display.

We created a novel change detection task which varied change magnitude in order to implement the non-parametric SFT analyses and introduced the novel representation of the change signal by taking the difference between a Gaussian distribution of perceptual effects representing the memory array and a Gaussian distribution of perceptual effects representing the probe array to allow for computational modeling. In Experiment 1 of Chapter 4 participants were either asked to respond "YES" if either one or two discs changed (OR task) or alternatively, to respond "YES" only if both discs changed (AND task).

We found that change detection decisions occurred independently, consistent with second-order integration models for most participants. Further, we found differences in decision-making strategy depending on task requirements. Par-

ticipants in the OR task were generally best fit by a parallel self-terminating model, whereas participants in the AND task were generally best fit by a serial self-terminating model. For the OR task, this pattern was consistent across both non-parametric, MIC and SIC results for the double change items, and parametric, computational modeling analyses. For the AND task the SFT results were mostly inconclusive, however, for three out of five participants computational modeling showed that decision-making proceeded in serial, self-terminating fashion.

We further fit two versions of each of the candidate models. In the *fixed variance* models, the perceptual distribution representing change strength was assumed to be equal for each level of change magnitude within a dimension. In the *free variance* models, the variance of the no-change, low discriminability change, and high discriminability change distributions were allowed to vary within a dimension. This second version of the models was to account for the possibility that the variability in the overall change distribution could vary based on the strength of the change in each individual disc. There was some evidence that the change signal varied differentially as a function of the strength of the change as the best fitting model for some participants was the free-variance rather than fixed-variance version.

Finally, we found that across both OR and AND tasks, workload capacity tended to be limited. Although there was some minor individual variation between limited and super capacity in the AND task, overall capacity tended to be limited for the majority of participants. This limited capacity processing is again consistent with a violation of context invariance which can again be explained using a sample size account (Palmer, 1990; Shaw, 1980).

6.3 Methodological implications

One of the major contributions of this thesis was the introduction of a novel change detection task which allowed for the implementation of non-parametric

SFT analyses and complementary computational modeling which were able to account for change detection decisions over time. Using this task we were able to effectively diagnose whether change detection decisions were best described by a first- or second-order evidence integration model. More specifically, we were able to diagnose decision-making as either serial, parallel, or coactive, as well as investigate the relevant stopping rules and workload capacity. Our finding of independent, second-order processing (whether serial or parallel) with limited capacity provides a direction for future models of change detection. Ultimately, any future model of change detection needs to be able to account for independent processing of each change, with limited capacity across both items.

To allow for computational modeling we introduced a novel representation of the change signal by taking the difference between a Gaussian distribution of perceptual effects representing the memory array and a Gaussian distribution of perceptual effects representing the probe array. We further fit two versions of the models, one whereby the variability of the change signal was fixed within a dimension, and another where it was free to vary between no-change, low magnitude, and high magnitude changes within a dimension. These models follow the same approach as the Logical Rule-Based models in categorization where each of the two dimensions comprises three levels of a given feature (e.g., saturation and orientation). From this perspective, fixing variation along a dimension is sensible as the dimensions are generally different features. In the current work the dimensions comprise three change magnitudes in a lighter disc located in the left of the screen and three change magnitudes in a darker disc located in the right of the screen. In this sense, fixing the level of variation in the change signal within each dimension also formed a valid starting point, as was allowing the distribution to vary for each change magnitude, however, these are by no means the only ways to consider variation in the change signal. For example, an alternative approach could be increasing the variability of the memory array compared to the probe array to account for the fact that the memory array is likely represented with less fidelity than the probe array which

remains present in the visual display. Exploring the best way to represent the change signal would be a worthy pursuit for future research.

6.4 Theoretical implications and directions for future research

The experiments presented in Chapter 3 showed that categorization decisions using within-dimension stimuli can occur coactively. These findings broadly cohere with other theories of visual attention such as in visual search where the difference between within-dimension and between-dimension features is central to search efficiency. These findings further add to the literature examining categorization decisions using visual stimuli by demonstrating that within-dimension luminance features are pooled together into a single decision channel. This provides a novel contribution to the categorization literature as it is not commensurate with the idea that spatially separated dimensions need to be resolved serially due to limits on the spatial aspects of visual attention as one might conclude from the results of Little et al. (2011). The experiments presented in Chapter 4 further showed that categorization decisions using within-dimension stimuli are limited in capacity. This finding is consistent with a violation of context invariance in the sample-size account (Palmer, 1990; Shaw, 1980).

The findings presented in Chapter 5 showed that change-detection decisions for the luminance level of two spatially separated discs generally occurred independently. Further, we found that the task requirements affected processing strategy. For participants in the OR task, change detection decision-making was best described as parallel self-terminating in nature, whereas for the AND task, decision-making proceeded in serial, self-terminating fashion. These findings cohere with many major theories and findings of the change detection literature.

Our finding of differences in processing strategy depending on variations in task requirements coheres with the findings of C.-T. Yang (2011) and C.-T. Yang et al. (2011, 2013) who showed that varying the probability and salience

of a changes in dimensions of a single object affected the processing strategy used. They are also commensurate with the findings of Donkin et al. (2016) who showed that when the environment was predictable (i.e., set size remained constant during each block), participants were able to flexibly allocate their attention across items. However, when the environment was unpredictable (i.e., set size varied from trial to trial) participants focused their attention on a smaller subset of items. Importantly, studies of change detection often utilize set sizes of four item locations or more (Alvarez & Cavanagh, 2004; Cowan, 2001; Donkin et al., 2016; Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974). In order to make further inferences regarding change detection decision-making in a one-shot change detection task, a worthwhile pursuit for future research would be investigating potential differences in the decision architecture for varying set-sizes in both OR and AND tasks. While it is well known that accuracy decreases with increasing set size (Lilburn, 2016; Luck & Vogel, 1997; Pashler, 1988; P. L. Smith, 2016; Vogel et al., 2006; Wilken & Ma, 2004), especially above a limit of about four items (Cowan, 2001; Luck & Vogel, 1997), it is unclear whether the processing architecture will change, reflecting a change in strategy for integrating information. In addition, a further exploration of processing strategy within blocked and unblocked experiments could be implemented using the current experimental design and accompanying SFT diagnoses.

Investigations into scene statistics provide another avenue for future research. For example, using a one-shot change detection paradigm which varied the proportion of green discs versus red discs, Howe & Webb (2014) showed that participants were accurately able to detect a change in the scene but were not always able to identify the location of the change. In a follow up experiment, Howe & Webb (2014) used the same number of individual item changes as the previous experiment but ensured that the overall proportion of colors remained the same between the displays. In this condition, participants were no longer able to accurately detect a change independently of identifying the location of a changed item. This suggests that the underlying scene statistics could be used

to identify a change. In the latter experiment, however, the change had to be directly attended in order to be identified.

It is possible that scene statistics could inform the difference in processing strategy between OR and AND tasks. In an OR design, where participants simple need to identify any change in the scene, participants could potentially monitor the overall scene statistics in parallel, rather than attend to specific changes in individual items. In the AND task, however, participants must verify that changes have occurred in both discs, resulting in a serial comparison. However, a limitation of the current design is that differences in the overall compositions of the memory and probe arrays were not taken into account. Instead we use the absolute value of the perceptual distributions from the memory and probe arrays thus averaging across a number of different changes. For example, both discs could become either lighter or darker, altering the overall mean luminance of the display. Alternatively, one disc could become lighter while the other becomes darker, maintaining the mean luminance, but altering the range of luminance values present in the scene. A worthwhile direction for future research would be identifying whether or not changes which alter the overall scene statistics in these two ways also alter the decision-making architecture.

As a whole, this thesis found both differences and similarities in decision-making using within-dimension stimuli across categorization and change detection tasks. In the categorization experiments, decision-making was able to occur coactively, despite the spatial separation between the dimensions. In the change detection tasks, decision-making instead occurred in either parallel or serial, depending on further task instructions. However, across all experiments a common pattern of limited capacity processing was found. These findings highlight that different perceptual operations can yield a variety of experimental results. It may be expected that other tasks such as visual search, identification, and detection might also diverge. However, the common finding of limited capacity suggests these within-dimension stimuli share a common bottle neck in processing efficiency regardless of task type. This coheres well with a sample

size account. In the interests of developing a unified view of visual perception and cognition, investigating different task types utilizing the same experimental stimuli and, ideally, the same participant pool is a worthwhile avenue for future research.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716-723.

Altieri, N., Fifić, M., Little, D. R., & Yang, C.-T. (2017). A tutorial introduction and historical background to systems factorial technology. In D. R. Little, N. Aliteri, M. Fifić, & C.-T. Yang (Eds.), *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms.* Elsevier.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106-111.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409-429.

Ashby, F. G. (1992). Multidimensional models of categorization.

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, *44*, 210-329.

Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorization response times with multidimensional stimuli. *Perception & Psychophysics*, *55*, 11-27.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 33-53.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, *120*(2), 150.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154-179-.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851-854.

Becker, M. W., Miller, J. R., & Liu, T. (2013). A severe capacity limit in the consolidation of orientation information into visual short-term memory. *Attention, Perception, & Psychophysics*, *75*(3), 415-425.

Biederman, I., & Checkosky, S. F. (1970). Processing redundant information. *Journal of Experimental Psychology*, *83*, 486-490.

Blunden, A. G., Howe, P. D. L., & Little, D. R. (2020). Evidence that within-dimension features are generally processed coactively. *Attention, Perpection, & Psychophysics*, *82*(1), 193-227.

Blunden, A. G., Wang, T., Griffiths, D., & Little, D. (2015). Logical-rules and the classification of integral dimensions: Individual differences in the processing of arbitrary dimensions. *Frontiers in Psychology*, *5*.

Bourne, L. E. (1970). Knowing and using concepts. *Psychological Review*, *77*, 546-556.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*, 85-109.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Burmester, A., & Wallis, G. (2012). Contrasting predictions of low- and high-threshold models for the detection of changing visual features. *Perception*, *41*, 505-516.

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *11*, 538-564.

Chang, T.-Y., Little, D. R., & Yang, C.-T. (2016). Selective attention modulates the effect of target location probability on redundant signal processing. *Attention, Perception, & Psychophysics*, *78*, 1603–1624.

Cheng, X. J., McCarthy, C. J., Wang, T. S., Palmeri, T. J., & Little, D. R. (2018). Composite faces are not (necessarily) processed coactively: A test using systems factorial technology and logical-rule models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *44*(6), 833.

Cheng, X. J., Moneer, S., Christie, N., & Little, D. R. (2017). Capacity, categorization, conflict, and resilience. In D. R. Little, N. Aliteri, M. Fifić, & C.-T. Yang (Eds.), *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms* (p. 158-174). Academic Press.

Colonius, H., & Vorberg, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology*, *38*, 35-58.

Corbett, E. A., & Smith, P. L. (2017). The magical number one-on-square-root-two: The double-target deficit in brief visual displays. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(7), 1376-1369.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-114.

Cowan, N., & Rouder, J. N. (2009). Comment on "dynamic shifts of limited working memory resources in human vision". *Science*, *323*(5916), 877.

Dai, H., Versfeld, N. J., & Green, D. M. (1996). The optimum decision rules in the same-different paradigm. *Perception & Psychophysics*, *58*(1), 1-9.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193-222.

Donkin, C., Brown, S., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology*, *55*(2), 140-151.

Donkin, C., Kary, A., Tahir, F., & Taylor, R. (2016). Resources masquerading as slots: Flexible allocation of visual working memory. *Cognitive Psychology*, *85*, 30-42.

Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review*, *120*, 873-902.

Donkin, C., Tran, S., & Nosofsky, R. M. (2014). Landscaping analyses of the roc predictions of discrete-slots and signal-detection models of working memory. *Attention, Perception & Psychophysics*, *76*(7), 2103-2116.

Drew, T., Evans, K., Võ, M. L. H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics*, *33*(1), 263-274.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*, 272-300.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501-517.

Duncan, J. (2006). Brain mechanisms of attention. *The Quarterly Journal of Experimental Psychology*, *59*(1), 2-27.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433-458.

Dzhafarov, E. N. (2003). Selective influence through conditional independence. *Psychometrika*, *68*, 7-26.

Egeth, H. E., & Mordkoff, T. J. (1991). Redundancy gain revisited: Evidence for parallel processing of separable dimensions. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure* (p. 131-143). Washington, DC: American Psychological Association.

Eidels, A., Houpt, J. W., Altieri, N., Pei, L., & Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology*, *55*, 176-190.

Eidels, A., Townsend, J. T., Hughes, H. C., & Perry, L. A. (2015). Evaluating perceptual integration: uniting response-time and accuracy-based methodologies. *Attention, Perception, & Psychophysics*, *77*(2), 659-680.

Estes, W. (1986). Array models for category learning. *Cognitive Psychology*, *18*, 500-549.

Feintuch, U., & Cohen, A. (2002). Visual attention and coactivation of response decisions for features from different dimensions. *Psychological Science*, *13*(4), 361-369.

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, *407*, 630-632.

Fifić, M., Little, D. R., & Nosofsky, R. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review*, *117*, 309-348.

Fifić, M., Nosofsky, R. M., & Townsend, J. (2008). Information-processing architectures in multidimensional classification: A validation of test of the

systems factorial technology. *Journal of Experimental Psychology: Human Perception & Performance*, *34*, 356-375.

Fifić, M., Townsend, J., & Eidels, A. (2008). Studying visual search using systems factorial methodology with target-distractor similarity as the factor. *Attention, Perception, & Psychophysics*, *70*, 583-603.

Fitousi, D. (2019). Can we perceive two colors at the same time? a direct test of huang and pashler's (2007) boolean map theory of visual attention. *Attention, Perception, & Psychophysics*, *81*(5), 1532-1550.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(3), 1-12.

Fox, E. L., & Houpt, J. W. (2016). The perceptual processing of fused multispectral imagery. *Cognitive Research: Principles and Implications*, *1*(31). Retrieved from `10.1186/s41235-016-0030-7`

Gardner, G. T. (1973). Evidence for independent parallel channels in tachistoscopic perception. *Cognitive Psychology*, *4*(1), 130-155.

Garner, W. R. (1974). *The processing of information and structure*. New York: Psychology Press.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*(3), 225-241.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, *24*(6), 997-1016.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., ... Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106*(4), 795-834.

Graham, N., Kramer, P., & Yager, D. (1987). Signal-detection models for multidimensional stimuli: Probability distributions and combination rules. *Journal of Mathematical Psychology*, *31*(4), 366-409.

Green, D., & Swets, J. (1966). Signal detection theory and psychophysics. *New York*.

Grice, G. R., Canham, L., & Gwynne, J. W. (1984). Absence of a redundant-signals effect in a reaction time task with divided attention. *Perception & Psychophysics*, *36*, 565-570.

Griffiths, D. W., Blunden, A. G., & Little, D. R. (2017). Logical-rule based models of categorization: Using systems factorial technology to understand feature and dimensional processing. In D. R. Little, N. Aliteri, M. Fifić, & C.-T. Yang (Eds.), *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms* (p. 245-269). Academic Press.

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An introduction to good practices in cognitive modeling. In *An introduction to model-based cognitive neuroscience* (pp. 25–48). Springer.

Hefner, R. A. (1958). *Extensions of the law of comparative judgment to discriminable and multidimensional stimuli* (Unpublished doctoral dissertation). University of Michigan.

Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 388-403.

Holmes, W. R. (2015). A practical guide to the probability density approximation (pda) with improved implementation and error characterization. *Journal of Mathematical Psychology*, *68-69*, 13-24.

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., & Townsend, J. T. (2013). Systems factorial technology with r. *Behavior Research Methods*, *46*, 307-330.

Houpt, J. W., & Townsend, J. T. (2010). The statistical properties of the survivor interaction contrast. *Journal of Mathematical Psychology*, *54*, 446-453.

Houpt, J. W., & Townsend, J. T. (2012). Statistical measures for workload capacity analysis. *Journal of Mathematical Psychology*, *56*, 341-355.

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics*, *72*(5), 1267-1282.

Howard, Z. L., Garrett, P., Little, D. R., Townsend, J. T., & Eidels, A. (2019). Nice guys check twice. Retrieved from `https://psyarxiv.com/b98wn/download?format=pdf`

Howe, P. D., & Webb, M. E. (2014). Detecting unidentified changes. *PLoS-ONE*, *9*(1), e84490.

Huang, L. (2007). Characterizing the limits of human visual awareness. *Science*, *317*(5839), 823-825.

Huang, L., & Pashler, H. (2007). A boolean map theory of visual attention. *Psychological Review*, *114*, 599-631.

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, *317*(5839), 823-825.

Hyun, J. S., Woodman, G. F., Vogel, E. K., Hollingworth, A., & Luck, S. J. (2009). The comparison of visual working memory representations with perceptual inputs. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1140-1160.

Jolicoeur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology*, *36*(2), 138-202.

Kantowitz, B. H. (1974). *Human information processing: Tutorials in performance and cognition.* Oxford, England: Lawrence Erlbaum.

Kastner, S., DeWeerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology*, *86*, 1398-1411.

Keshvari, S., Van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology*, *9*(2), e1002927.

Kok. (2001). On the utility of the p3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557-577.

Koopman, B. O. (1956). The theory of search. *Operations Research*, *4*, 5.

Kruschke, J. K. (1992). Alcove: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22-44.

Laming, D. R. J. (1968). *Information theory of choice reaction times.* New York: Wiley.

Leone, F. C., Nelson, L. S., & Nottingham, R. B. (1961). The folded normal distribution. *Technometrics*, *3*(4), 543-550.

Levine, M. (1975). *A cognitive theory of learning: Research on hypothesis testing.* Hillsdale, NJ: Erlbaum.

Li, X., Liang, Z., Kleiner, M., & Lu, Z.-L. (2010). Rtbox: A device for highly accurate response time measurements. *Behavior Research Methods*, *42*(1), 212-225.

Liesefeld, H. R., M., L. A., Müller, H. J., & Rangelov, D. (2017). Saliency maps for finding changes in visual scenes? *Attention, Perception & Psychophysics*, *79*, 2190-2201.

Lilburn, S. D. (2016). *Information limits within visual short term memory* (Unpublished doctoral dissertation). The University of Melbourne.

Link, S. W. (1992). *The wave theory of difference and similarity.* Psychology Press.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Pscyhometrika*, *40*, 77-105.

Little, D. R., Altieri, N., Fifić, M., & Yang, C.-T. (2017). *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms.* New York: Academic Press.

Little, D. R., Eidels, A., Fific, M., & Wang, T. (2015). Understanding the influence of distractors on workload capacity. *Journal of Mathematical Psychology*, *68*, 25-36.

Little, D. R., Eidels, A., Fific, M., & Wang, T. (2018). How do information processing systems deal with conflicting information? differential predictions for serial, parallel and coactive processing models. *Computational Brain & Behavior*, *1*, 1-21.

Little, D. R., Eidels, A., Houpt, J. W., & Yang, C. T. (2017). Set size slope still does not distinguish parallel from serial search. *Behavioral and Brain Sciences*, *40*, 32-33.

Little, D. R., Nosofsky, R., & Denton, S. E. (2011). Response time tests of logical rule-based models of categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *37*, 1-27.

Little, D. R., Nosofsky, R. M., Donkin, C., & Denton, S. E. (2013). Logical-rules and the classification of integral dimensioned stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *39*, 801-820.

Little, D. R., & Smith, P. L. (2018). Replication is already mainstream: Lessons from small-N designs. *Behavioral and Brain Sciences*, *41*, E141.

Lockhead, G. R., & King, M. C. (1977). Classifying integral stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *72*, 95-104.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527.

Luce, R. D. (1963). Detection and recognition.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Luck, S., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279-81.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide.* Psychology Press.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (p. 147-180). Erlbaum.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human perception and performance*, *22*(4), 795.

Mance, I., Becker, M. W., & Liu, T. (2012). Parallel consolidation of simple features into visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 429-438.

Medin, D., & Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247-279.

Miller, J. R., Becker, M. W., & Liu, T. (2014). The bandwidth of consolidation into visual short-term memory depends on the visual feature. *Visual Cognition*, *22*(7), 920-947.

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Attention, Perception, & Psychophysics*, *66*(8), 1268–1281.

Moneer, S., Wang, T., & Little, D. R. (2016). The processing architectures of whole-object features: A logical rules approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1443-1465.

Moray, N. (1970a). Introductory experiments in auditory time sharing: Detection of intensity and frequency increments. *Journal of the Acoustical Society of America*, *47*, 1071-1073.

Moray, N. (1970b). Time sharing in auditory detection: Effects of stimulus duration. *Journal of the Acoustical Society of America*, *47*, 660-661.

Mordkoff, J. T., & Danek, R. H. (2011). Dividing attention between color and shape revisited: Redundant targets coactivate only when parts of the same perceptual object. *Attention, Perception and Psychophysics*, *73*(1), 103-112.

Mordkoff, J. T., & Yantis, S. (1993). Dividing attention between color and shape: Evidence for coactivation. *Attention, Perception and Psychophysics*, *54*(4), 357-366.

Noreen, D. L. (1981). Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology* (p. 237-280). Providence, RI: Psychology Press.

Normand, M. P. (2016). Less is more: Psychologists can learn more by studying fewer people. *Frontiers in Psychology*, *7*. Retrieved from `https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00934/full`

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-61.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *14*, 700-708.

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics*, *45*, 4.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, *43*, 25-53.

Nosofsky, R. M., & Palmeri, T. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266-300.

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science*, *21*(3), 164-169.

Otto, T. U., & Mamassian, P. (2016). Multisensory decisions: the test of a race model, its logic, and power. *Multisensory Research*, *30*, 1-24.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 1227-1268.

Pashler, H. (1988). Familiarity and visual change detection. *Current Directions in Psychological Science*, *4*(4), 369-378.

Peterson, W. W., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 171-212.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Attention, Perception, & Psychophysics*, *16*(2), 283–290.

Phillips, W. A., & Baddeley, A. D. (1971). Reaction time and short-term visual memory. *Psychonomic Science*, *22*(2), 73-74.

Poom, L. (2009). Integration of colour, motion, orientation, and spatial frequency in visual search. *Perception*, *38*(5), 708-718.

Posner, M. I., & Keele, S. W. (1967). Decay of visual information from a single letter. *Science*, *158*(3797), 137-139.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas [Journal Article]. *Journal of Experimental Psychology*, *77*, 353-363.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160-174.

Raab, D. (1962). Statistical facilitation of simple reaction time. *Transaction of the New York Academy of Science*, *43*, 574-590.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *5*, 347-356.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*, 382-407.

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition*, *7*(1-3), 345-376.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368-373.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975-5979.

Rouder, J. N., Speckman, P. L., D., S., Morey, R. D., & Iverson, G. (n.d.). A bayes factor calculation for accepting the null hypothesis. *Retrieved from:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.138.1717rep=rep1type=pdf*.

Schweickert, R. (1992). Information, time, and the structure of mental events: A twenty-five year review. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance: Vol. 14. synergies in experimental psychology, artificial intelligence, and cognitive neuroscience – a silver jubilee* (p. 535-566). Cambridge, MA: MIT Press.

Schweickert, R. (1993). Information, time, and the structure of mental events. In D. E. Meyer & S. Kornblum (Eds.), *Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience - a silver jubilee, vol. 14* (p. 535-566). MIT Press.

Schweickert, R., Fisher, D. L., & Sung, K. (2012). *Discovering cognitive architecture by selectively influencing mental processes* (Vol. 4). World Scientific.

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(6), 2214-2242.

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2016). Object selection costs in visual working memory: A diffusion model analysis of the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1673-1693.

Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. In R. S. Nickerson (Ed.), *Attention and performance viii* (p. 277-295). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shaw, M. L. (1982). Attending to multiple sources of information. *Cognitive Psychology*, *14*, 353-409.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54-87.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Science*, *9*, 16-20.

Smith, L. B., & Kilroy, M. C. (1979). A continuum of dimensional separability. *Perception & Psychophysics*, *25*, 285-291.

Smith, P. L. (1998). Attention and luminance detection: A quantitative analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 105-133.

Smith, P. L. (2010). Spatial attention and the detection of weak visual signals. In V. Coltheart (Ed.), *Tutorials in visual cognition* (p. 211-259). Psychology Press.

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, *123*(4), 425-451.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defence of the small-N design. *Psychonomic Bulletin & Review*, *25*(6), 2083-2101.

Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychological Review*, *120*(3), 589-627.

Snodgrass, J. G., & Townsend, J. T. (1980). Comparing parallel and serial models: Theory and implementation. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 330-354.

Sorkin, R. D. (1962). Extension of the theory of signal detectability to matching procedures in psychoacoustics. *Journal of the Acoustical Society of America*, *34*, 1945-1751.

Sorkin, R. D., & Pohlmann, L. D. (1973). Some models of observer behavior in two-channel auditory signal detection. *Perception & Psychophysics*, *53*, 1045-1050.

Sorkin, R. D., Pohlmann, L. D., & Gilliom, J. (1973). Simultaneous two-channel signal detection: Iii. 630- and 1400-hz signals. *Journal of the Acoustical Society of America*, *53*, 1045-1050.

Sorkin, R. D., Pohlmann, L. D., & Woods, D. D. (1976). Decision interaction between auditory channels. *Perception & Psychophysics*, *19*, 290-295.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1-29.

Sternberg, S. (1969). Memory scanning: Memory processes revealed by reaction-time experiments. *American Scientist*, *4*, 421-457.

Swets, J. A. (1961). Is there a sensory threshold? *Science*, *134*(3473), 168-177.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*(5), 301-340.

Tanner, W. P., & Swets, J. A. (1954a). A decision-making theory of visual detection. *Psychological Review*, *61*(6), 401-409.

Tanner, W. P., & Swets, J. A. (1954b). The human use of information-i: Signal detection for the case of the signal known exactly. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 213-221.

Taylor, D. A. (1976). Effect of identity in the multiletter matching task. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 199-806.

Thiele, J. E., Haaf, J. M., & Rouder, J. N. (2017). Is there variation across individuals in processing? Bayesian analysis for systems factorial technology. *Journal of Mathematical Psychology*, *81*, 40-54. Retrieved from `psyarxiv.com/auf9n`

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review*, *114*, 71-103.

Townsend, J. T. (1971). A note on the indentifiability of parallel and serial processes. *Perception & Psychophysics*, *10*, 161-163.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology*, *28*, 363-400.

Townsend, J. T. (1990a). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*, 46-54.

Townsend, J. T. (1990b). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, *108*, 551.

Townsend, J. T. (2016). A note on drawing conclusions in the study of visual search and the use of slopes in particular. *i-Perception*, *7*(6), 2041669516674220.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.

Townsend, J. T., & Eidels, A. (2011). Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied. *Psychonomic Bulletin & Review*, *18*, 659-681.

Townsend, J. T., Fific, M., & Neufeld, R. W. (2007). Assessment of mental architecture in clinical/cognitive research. In T. A. Treat, R. R. Bootzin, & T. B. Baker (Eds.), *Psychological clinical science: Papers in honour of richard m. mcfall* (p. 223-258). Mahwah, NJ: Lawrence Erlbaum Associates.

Townsend, J. T., Houpt, J. W., & Silbert, N. H. (2012). General recognition theory extended to include response times: Predictions for a class of parallel systems. *Journal of Mathematical Psychology*, *56*(6), 476-494.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, *39*, 321-340.

Townsend, J. T., & Nozawa, G. (1997). Serial exhaustive models can violate the race inequality: Implications for architecture and capacity. *Psychological Review*, *104*, 595-602.

Townsend, J. T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology*, *33*(3), 309–327.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003-1035.

Trabasso, T., & Bower, G. H. (1968). *Attention in learning: Theory and research.* New York, NY: Wiley.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97-136.

Treisman, M. (1985). The magical number seven and some other features of category scaling: Properties of a model for absolute judgment. *Journal of Mathematical Psychology*, *29*(2), 175–230.

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review*, *21*(2), 227-250.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*, 368.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 92-114.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436-1451.

Wenger, M. J., & Townsend, J. T. (2000). Basic response time tools for studying general processing capacity in attention, perception, and cognition. *Journal of General Psychology*, *127*(1), 67-99.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48-64.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*, 1120-1135.

Wolfe, J. M. (1994a). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202-238.

Wolfe, J. M. (1994b). Visual search in continuous, naturalistic stimuli. *Vision Research*, *34*(9), 1187-1195.

Wolfe, J. M. (1998). What do 1,000,000 trials tell us about visual search? *Psychological Science*, *9*(1), 33-39.

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (p. 99-119). Oxford: Oxford University Press.

Wolfe, J. M. (2013). When is it time to move to the next raspberry bush? foraging rules in human visual search. *Journal of Vision*, *13*(3), 10.

Wolfe, J. M. (2016). Visual search revived: the slopes are not that slippery: a reply to kristjansson (2015). *i-Perception*, *7*(3), 2041669516643244.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception & Psychophysics*, *73*(6), 1650-1671.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare targets are often missed in visual search. *Nature*, *435*(7041), 439-440.

Wolfe, J. M., Yu, K. P., Stewart, M. I., Shorter, A. D., Friedman-Hill, S. R., & Cave, K. R. (1990). Limitations on the parallel guidance of visual search: Color x color and orientation x orientation conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 879-892.

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, *400*, 867-869.

Yang, C.-T. (2011). Relative saliency in change signals affects perceptual comparison and decision processes. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1708-1728.

Yang, C.-T., Altieri, N., & Little, D. R. (2017). An examination of parallel versus coactive processing accounts of redundant-target audiovisual signal processing. *Journal of Mathematical Psychology*, *21*, 138-158.

Yang, C.-T., Chang, T.-Y., & Wu, C.-J. (2013). Relative change probability affects the decision process of detecting multiple feature changes. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1365-1385.

Yang, C.-T., Hsu, H.-Y., & Yeh, Y.-Y. (2011). Relative salience affects the process of detecting changes in orientation and luminance. *Acta Psychologica*, *138*, 377-389.

Yang, C.-T., Little, D. R., & Hsu, C.-C. (2014). The influence of cuing on attentional focus in perceptual decision making. *Attention, Perception & Psychophysics*, *76*(8), 2256-2275.

Yang, H., Little, D. R., Eidels, A., & Townsend, J. T. (2019). Survivor interaction contrasts for error response times: Non-parametric contrasts for serial and parallel systems. *Manuscript in preparation. Indiana University.*.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233-235.