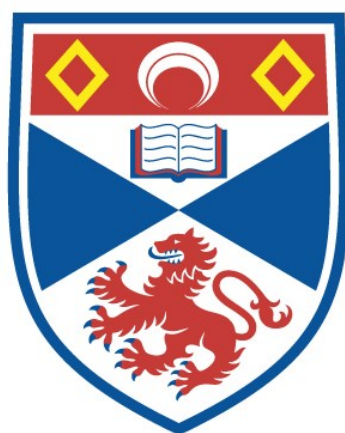


TARGETING BENEFITS AND INTERACTIONS IN  
MULTISENSORY PROCESSING : A NOVEL MODELLING  
FRAMEWORK FOR EXPLORING THE REDUNDANT SIGNAL  
EFFECT

Bobby Richard Innes

A Thesis Submitted for the Degree of PhD  
at the  
University of St Andrews



2019

Full metadata for this thesis is available in  
St Andrews Research Repository  
at:

<http://research-repository.st-andrews.ac.uk/>

Identifiers to use to cite or link to this thesis:

DOI: <https://doi.org/10.17630/10023-17968>

<http://hdl.handle.net/10023/17968>

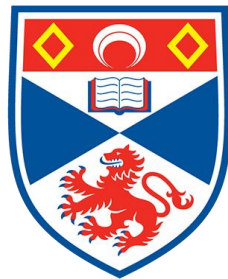
This item is protected by original copyright

This item is licensed under a  
Creative Commons License

<https://creativecommons.org/licenses/by-nc-nd/4.0>

Targeting Benefits and Interactions in Multisensory  
Processing: A Novel Modelling Framework for Exploring  
the Redundant Signal Effect

Bobby Richard Innes



University of  
St Andrews

This thesis is submitted in partial fulfilment for the degree of  
Doctor of Philosophy (PhD)  
at the University of St Andrews

March 2019

## Abstract

An overarching goal of multisensory research is to understand the rules by which multiple sensory signals are combined to produce behavioural benefits. One example of a multisensory benefit is that response times (RTs) to multisensory signals are faster than RTs to either component unisensory signal (the *redundant signal effect*). In contrast to other areas of multisensory research, a common explanatory model framework (and thus the basic *combination rule*) has yet to emerge for multisensory RT benefits. One key reason for this lack of progress would seem to be that additional *processing interactions* are infrequently quantified, and are rarely incorporated into a formal model of RTs. In this thesis, I assess the ability of a previously-neglected model class – so called *race models* – to account for both benefits and interactions. To do so, I develop and apply a *comparative approach* to RT analysis. This approach tests different experimental factors which attempt target sources of benefits and interactions. The effect of these manipulations is compared across 3 key analytical steps. First, the size of the multisensory benefit (i.e. the redundant signal effect) is computed, and compared to the basic prediction of the race model combination rule. Second, indicators of processing interactions beyond the basic combination rule are quantified. Third, a formal race model is applied to explain benefits and processing interactions. By consistently applying this comparative approach, I demonstrate that the race model framework offers a simple yet powerful account of multisensory benefits across participants. This suggests a basic combination rule for multisensory RT benefits. In addition, a new interpretation of interactions emerges by applying a race model framework, which highlights an important role of context. This approach suggests a common foundation for future studies of multisensory RTs, which can implement the same analytical steps within a wide range of experimental contexts.

# Contents

1. General Introduction.....	11
1.1. Thesis Topic.....	11
1.1.1. Chapter Overview .....	13
1.2. Unisensory Response Times (Background) .....	14
1.2.1. Measuring RTs: The Simple Detection Task.....	14
1.2.2. Evaluating RTs: The Cumulative Distribution Function.....	15
1.2.3. Explaining RTs: Models of Unisensory Decision-Making .....	17
1.3. Multisensory Response Times .....	21
1.3.1. The Redundant Signal Paradigm (RSP).....	21
1.3.2. The Redundant Signal Effect (RSE).....	23
1.3.3. Models of Multisensory Decision-Making .....	24
1.3.4. Deciding Between Multisensory Models: An Evaluation of Current Practice.....	27
1.4. Formulating the Basic Race Model .....	29
1.4.1. Assumption 1: Statistical Independence.....	30
1.4.2. Assumption 2: Context Invariance .....	31
1.5. Developing a Comparative Approach .....	31
1.6. Comparative Approach Step 1: Understand Benefits .....	32
1.6.1. Race Model Principles for Multisensory RT Benefit.....	32
1.6.2. Key Analysis: Compare Predicted and Empirical Benefits.....	36
1.7. Comparative Approach Step 2: Quantify Interactions.....	36
1.7.1. Interaction 1: Statistical Dependence (Trial History Effects) .....	36
1.7.2. Interaction 2: Context Variance (Violations of Miller’s Bound) .....	37
1.7.3. Key Analysis: Quantify History Effects and Violations .....	39
1.8. Comparative Approach Step 3: Apply and Evaluate a Model .....	40
1.8.1. The Context Variant Race Model .....	41
1.8.2. Key Analysis: Observe Changes in Model Parameters .....	42

1.9.	Thesis Chapter Outline.....	43
2.	General Methods .....	45
2.1.	Experimental Procedures.....	45
2.1.1.	Apparatus.....	45
2.1.2.	Stimulus Generation .....	45
2.1.3.	Volume Setting & Calibration .....	48
2.1.4.	Calibrating Stimulus Timing .....	49
2.2.	Analytic Procedures .....	51
2.2.1.	Assessing General Performance.....	51
2.2.2.	Outlier Correction .....	51
2.3.	Creating RT Distributions .....	52
2.3.1.	Linear Interpolation of RTs.....	53
2.3.2.	Measures of Benefit (The RSE).....	54
2.3.3.	Measures of Interactions .....	57
2.4.	Model Fitting Procedures.....	60
2.4.1.	Parameter Search Methods .....	60
2.4.2.	Error Measures.....	63
2.4.3.	Model Selection .....	64
2.4.4.	Models of Response Time Distributions .....	66
2.5.	Simulation Procedures .....	67
2.5.1.	Monte Carlo Sampling.....	67
2.5.2.	Bootstrapping.....	68
2.5.3.	Simulating Response Times Using Models.....	69
3.	Experiment 1: Stimulus Construction vs. Signal Features.....	72
3.1.	Introduction .....	72
3.1.1.	Factor 1: Stimulus Construction.....	72
3.1.2.	Factor 2: Signal Features.....	74
3.1.3.	Experimental Hypotheses .....	75

3.2.	Methods .....	75
3.2.1.	Participants .....	75
3.2.2.	Stimulus Design .....	75
3.2.3.	Procedure .....	77
3.3.	Results .....	78
3.3.1.	General Performance .....	78
3.3.2.	Unisensory RTs .....	79
3.3.3.	Comparative Approach Step 1: Multisensory Benefits .....	81
3.3.4.	Comparative Approach Step 2: Quantifying Interactions .....	83
3.3.5.	Comparative Approach Step 3: Applying the Context Variant Race Model .....	84
3.4.	Discussion .....	86
4.	Experiment 2: Signal Duration vs. Task-Irrelevant Stimulation .....	90
4.1.	Introduction .....	90
4.1.1.	Factor 1: Signal Duration .....	90
4.1.2.	Factor 2: Task-Irrelevant Stimulation .....	92
4.1.3.	Experimental Hypotheses .....	93
4.2.	Methods (Part 1) .....	94
4.2.1.	Participants .....	94
4.2.2.	Stimulus Design .....	94
4.2.3.	Procedure .....	95
4.3.	Results (Part 1) .....	97
4.3.1.	General Performance .....	97
4.3.2.	Unisensory RTs .....	98
4.3.3.	Comparative Approach Step 1: Multisensory Benefits .....	100
4.3.4.	Comparative Approach Step 2: Quantifying Interactions .....	102
4.3.5.	Comparative Approach Step 3: Applying the Context Variant Race Model .....	104
4.4.	Interim Summary (Part 1) .....	106
4.5.	Methods (Part 2) .....	107

4.5.1.	Participants .....	107
4.6.	Results (Part 2) .....	108
4.6.1.	General Performance .....	108
4.6.2.	Unisensory RTs .....	110
4.6.3.	Comparative Approach Step 1: Multisensory Benefits .....	111
4.6.4.	Comparative Approach Step 2: Quantifying Interactions .....	113
4.6.5.	Comparative Approach Step 3: Applying the Context Variant Race Model .....	115
4.7.	Interim Summary (Part 2) .....	117
4.8.	Can Increased False Alarms Account For Reductions in Violation? .....	117
4.9.	Discussion .....	119
5.	Experiment 3: Signal Strength vs. Response Effector .....	122
5.1.	Introduction .....	122
5.1.1.	Factor 1: Signal Strength .....	123
5.1.2.	Factor 2: Response Effector .....	124
5.1.3.	Experimental Hypotheses .....	125
5.2.	Methods .....	126
5.2.1.	Participants .....	126
5.2.2.	Stimulus Design .....	126
5.2.3.	Procedure .....	127
5.3.	Results .....	128
5.3.1.	General Performance .....	128
5.3.2.	Unisensory RTs .....	129
5.3.3.	Comparative Approach Step 1: Multisensory Benefits .....	131
5.3.4.	Comparative Approach Step 2: Quantifying Interactions .....	133
5.3.5.	Comparative Approach Step 3: Applying the Context Variant Race Model .....	134
5.4.	Discussion .....	136
6.	Examining Trends in Experimental Data .....	140
6.1.	Race Model Principles and Empirical Benefits .....	140

6.2.	Modelling Empirical Benefits and Interactions .....	144
6.3.	Evaluating the Context Variant Race Model .....	147
6.3.1.	Comparing Fitting Procedures .....	147
6.3.2.	Nested Model Comparison .....	149
6.3.3.	Limitations and Future Directions.....	154
7.	General Discussion.....	158
7.1.	Race Models Offer a Common RSP Framework.....	158
7.1.1.	Variability is the Source of Multisensory Benefits .....	159
7.1.2.	Race Model Principles as a Guiding Framework for Multisensory Research.....	160
7.2.	Understanding Sources of Processing Interactions .....	162
7.2.1.	Higher-Level Processing is Important for Trial History Effects .....	162
7.2.2.	Context is Crucial for Violations of Miller’s Bound .....	164
7.3.	Application and Development of Models in the RSP .....	165
7.4.	Experimental Applications of the Comparative Approach .....	167
7.4.1.	Beyond Bisensory Pairings .....	168
7.4.2.	Between-Subjects Comparisons .....	169
7.4.3.	Involving Cognitive Neuroscience Techniques.....	171
7.5.	Beyond the RSE: Implications for Multisensory Research .....	172
7.5.1.	Is The Race Model Part of a Wider Set of Combination Rules? .....	172
7.5.2.	Linking Frameworks for Response Times and Accuracy .....	173
7.5.3.	Towards Conceptual Clarity in Multisensory Research.....	174
8.	References .....	176
9.	Appendices.....	186
9.1.	Illustrating the Basic Race Model Mechanism .....	186
9.2.	Literature Review of RSE Studies .....	187
9.3.	Simulation Parameters for Validation of Down-Sampled Benefit Procedure.....	188
9.4.	Equations for the Context Variant Race Model .....	188
9.5.	Analysis of Misses for Chapter 4 (Part 2) .....	189



9.6.	History Effects for Each Modality.....	190
9.7.	Simulating the Effect of Non-Decisional Components on Model Fitting .....	191
9.8.	Average Model Parameters for Experimental Chapters .....	193
9.8.1.	Two-Step Fitting Procedure .....	193
9.8.2.	One-Step Fitting Procedure .....	194
10.	Abbreviations .....	195

# Figure Contents

Figure 1.1 Unisensory and multisensory signals in everyday life .....	11
Figure 1.2 Combining multiple senses for a single response .....	13
Figure 1.3 The simple detection task .....	15
Figure 1.4 Evaluating RTs .....	16
Figure 1.5 The unisensory decision-making process .....	18
Figure 1.6 Explaining empirical observations in response times .....	19
Figure 1.7 The LATER model .....	20
Figure 1.8 The redundant signal paradigm (RSP).....	22
Figure 1.9 The basic race model architecture.....	25
Figure 1.10 The basic pooling model architecture.....	25
Figure 1.11 Race and pooling model explanations of the RSE.....	26
Figure 1.12 A Venn diagram illustration of the race model equation (Otto & Mamassian, 2017).....	30
Figure 1.13 Calculating benefit .....	33
Figure 1.14 Predicted benefits according to the simple race model (Otto et al., 2013) .....	35
Figure 1.15 Trial history effects .....	37
Figure 1.16 Limits of the race model (assuming context invariance) .....	38
Figure 1.17 Violation of Miller’s bound .....	40
Figure 1.18 Fitting the context variant race model .....	42
Figure 1.19 Overview of the comparative approach .....	44
Figure 2.1 Setup of experimental equipment.....	46
Figure 2.2 The RTbox.....	46
Figure 2.3 Auditory signal examples .....	47
Figure 2.4 Visual signal examples .....	47
Figure 2.5 Calibrating volume of auditory stimuli .....	48
Figure 2.6 Procedure for calibrating auditory stimulus volume .....	49
Figure 2.7 Measuring visual onset delay.....	50
Figure 2.8 The effect of fast outliers on modelling.....	53
Figure 2.9 Linear interpolation of RTs.....	54
Figure 2.10 Measuring multisensory benefit.....	56
Figure 2.11 Predicting multisensory benefit.....	56
Figure 2.12 Validating the down-sampling procedure for benefit calculation .....	58
Figure 2.13 Quantifying violations of Miller's bound.....	59
Figure 2.14 The grid search method .....	61

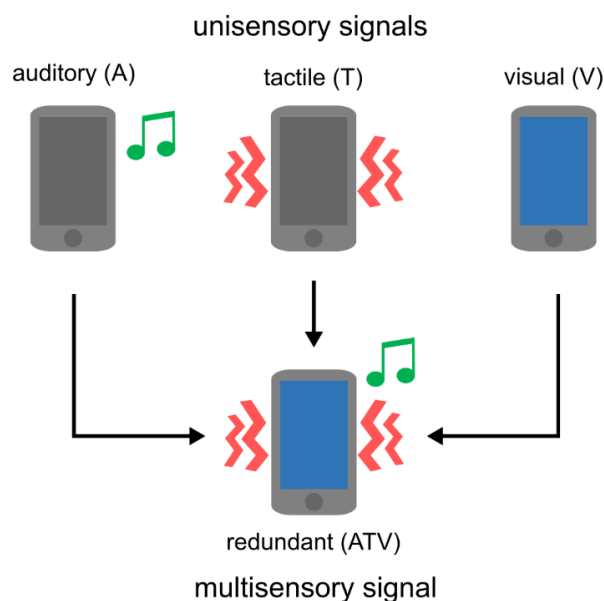
Figure 2.15 The optimisation algorithm method .....	62
Figure 2.16 Monte Carlo simulation .....	68
Figure 2.17 Bootstrap sampling .....	69
Figure 2.18 Bootstrapping model fits .....	69
Figure 2.19 Simulating unisensory RTs using a LATER model .....	70
Figure 3.1 Stimulus construction .....	73
Figure 3.2 Signal features.....	75
Figure 3.3 Stimulus design for Chapter 3.....	76
Figure 3.4 Unisensory RT analysis.....	80
Figure 3.5 Predicting and measuring multisensory benefits .....	82
Figure 3.6 Quantifying processing interactions .....	84
Figure 3.7 Fitting the context variant race model .....	85
Figure 4.1 Signal duration .....	91
Figure 4.2 Task-irrelevant stimulation .....	93
Figure 4.3 Stimulus design for Chapter 4 (Part 1).....	95
Figure 4.4 Placement of the factor .....	96
Figure 4.5 Unisensory RT analysis (Part 1).....	100
Figure 4.6 Predicting and measuring multisensory benefits (Part 1) .....	101
Figure 4.7 Quantifying interactions (Part 1) .....	104
Figure 4.8 Fitting the context variant race model (Part 1) .....	105
Figure 4.9 Stimulus design for Chapter 4 (Part 2).....	107
Figure 4.10 Unisensory RT analysis (Part 2).....	111
Figure 4.11 Predicting and measuring multisensory benefits (Part 2) .....	112
Figure 4.12 Quantifying interactions (Part 2) .....	114
Figure 4.13 Fitting the context variant race model (Part 2) .....	116
Figure 4.14 Simulating the effect of false alarms on violation area .....	118
Figure 5.1 Pre-decisional and post-decisional stages in the race model architecture .....	122
Figure 5.2 Signal strength .....	124
Figure 5.3 Response effector .....	125
Figure 5.4 Stimulus design for Chapter 5.....	126
Figure 5.5 Unisensory RT analysis.....	130
Figure 5.6 Predicting and measuring multisensory benefits .....	132
Figure 5.7 Quantifying interactions .....	134
Figure 5.8 Fitting the context variant race model .....	135

Figure 5.9 Problems modelling changes in non-decisional RT components with LATER models.....	139
Figure 6.1 Evaluating the quantitative predictions of the simple race model.....	141
Figure 6.2 Evaluating the race model principles.....	143
Figure 6.3 Evaluating the context variant race model account of benefits.....	145
Figure 6.4 Relating interactions to model parameters.....	146
Figure 6.5 Comparing fitting procedures for the context variant race model.....	148
Figure 6.6 Nested model comparison using AIC.....	151
Figure 6.7 Nested model comparison using BIC.....	152
Figure 6.8 Nested model comparison using AIC difference score.....	155
Figure 6.9 Evaluating model residuals.....	156
Figure 7.1 The race model as the combination rule for multisensory response time benefits.....	159
Figure 7.2 Adding third modality intermediate signals to remove history effects.....	164
Figure 7.3 Interaction between behavioural and formal modelling approaches.....	167
Figure 9.1 Statistical facilitation with runners' lap times.....	186

# 1. General Introduction

## 1.1. Thesis Topic

In everyday life, even some of the most common tasks involve the processing of signals from multiple senses at once. One simple example is receiving a call on a smartphone. Smartphones are capable of signalling their users to make a response (i.e. take the call) in multiple sensory modalities (**Figure 1.1**). For instance, it might play a ringtone (auditory signal), vibrate (tactile signal) or flash something onscreen (visual signal). In isolation, each of these is a *unisensory* signal. Alternatively, the device may present two or more of these signals simultaneously; this would be a *multisensory* signal. In many real-world cases, from phone calls (as in **Figure 1.1**) to alerts regarding potential dangers (e.g. sirens and flashing lights on alarm systems), multisensory signals are used. This is often despite the fact that a unisensory signal alone would suffice to elicit the necessary response (i.e. the multiple signals are *redundant*). One important reason for this is that multisensory signals are often advantageous, eliciting behavioural *benefits* above unisensory signals alone: individuals respond to them both more *quickly* (e.g. Miller, 1982; Raab, 1962; Todd, 1912) and more *accurately* (e.g. Frassinetti, Bolognini, & Ladavas, 2002; Lovelace, Stein, & Wallace, 2003; Spence, Baddeley, Zampini, James, & Shore, 2003).



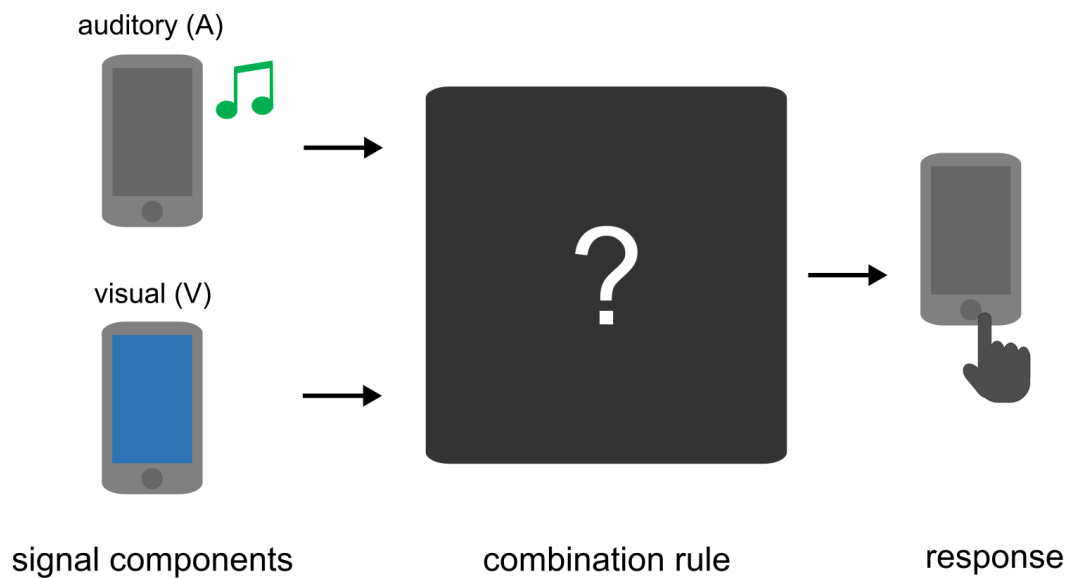
**Figure 1.1 Unisensory and multisensory signals in everyday life**

Devices such as smartphones provide signals to different senses. Each of these signals often requires the same response (e.g. answer a call). On the top row, examples of unisensory signals are shown (where only one sensory modality is activated). In the bottom row, these unisensory component signals are combined into a multisensory signal. This can also be considered a *redundant* signal, as response to any one component is sufficient to successfully answer the call. Such multisensory signals are reliably associated with benefits (e.g. faster response times than the individual unisensory components).

A proper understanding of the processes which lead to these multisensory benefits has direct implications for these everyday tasks. To return to the simple examples above, it would allow us to more effectively design and interact with technological devices (see Pomper, Brincker, Harwood, Prikhodko, & Senkowski, 2014 for an implementation of the smartphone example). Further, multisensory research is already targeted toward the development of effective warning systems (e.g. van Erp, Toet, & Janssen, 2015) and driver safety systems (e.g. Biondi, Strayer, Rossi, Gastaldi, & Mulatti, 2017; Ho, Reed, & Spence, 2007; Keyes, Whitmore, Naneva, & McDermott, 2018; Reinmueller, Koehler, & Steinhauser, 2018; Spence & Ho, 2008). In short, gaining a better understanding of the underlying brain mechanisms of multisensory benefits allows us to exploit them in important scenarios.

Perhaps to this end, multisensory research has seen a surge in recent years, whereas understanding of perception had previously focused on senses in isolation (Alais, Newell, & Mamassian, 2010). One advantage of this, however, is that our understanding of unisensory perceptual decision-making is now advanced on multiple levels, describing empirical effects on both behaviour and the neural level (Gold & Shadlen, 2007; Smith & Ratcliff, 2004). Further, links between these levels of understanding exist in the form of clear *model frameworks* i.e. formalised descriptions of the underlying processes which give rise to empirical effects (e.g. Mulder, van Maanen, & Forstmann, 2014 for a review). Multisensory research, similarly, now has a wealth of data demonstrating empirical effects on behaviour (e.g. Gondan & Minakata, 2016) and individual neurons (Stein & Stanford, 2008; Stein, Stanford, & Rowland, 2014). What must also be developed, then, are clear multisensory model frameworks, which can build upon the well-established unisensory framework. To effectively extend the unisensory framework to multisensory responses, however, we need to understand how multiple unisensory inputs are combined (i.e. the fundamental *combination rule* for multiple sensory inputs). Beyond this fundamental step, we also need to understand how implementation of the combination rule differs depending on the situation (i.e. understand processing *interactions* between different sensory inputs, which may influence how effectively signals can be combined).

In some fields of multisensory research, the modelling of combination rules has already been enormously successful. For example, model explanations of how multiple senses combine to allow *more accurate* estimates, e.g. using vision and touch in combination to better estimate the size of an object (Ernst & Banks, 2002; Ernst & Bulthoff, 2004), provide clear accounts of behaviour and offer links to neuronal implementations. However, for understanding *response time* (RT), progress in understanding the underlying combination rules has been lacking. Indeed, a recent review of the literature on a classic multisensory RT paradigm (with similar demands to that shown in **Figure 1.2**)



**Figure 1.2 Combining multiple senses for a single response**

In the above example, auditory and visual signal components are presented simultaneously. Following signal input from two senses, the observer must somehow process both unisensory components to produce a single response (e.g. answer the call). Though this *combination rule* cannot be observed directly (illustrated by the black box), we can develop predictions and explanations for responses by constructing a model of it.

has suggested that there simply may be no common model which can explain behavioural results across different experiments (Gondan & Minakata, 2016, p. 731). The lack of a common model framework for multisensory RTs is the problem which this thesis attempts to address.

1.1.1. Chapter Overview

In the General Introduction, I show how we might develop a common model framework for understanding simple multisensory RTs. To do this, I first cover the necessary background on the unisensory RT framework. Second, I extend this to the multisensory RT task of interest (known as the *redundant signal paradigm*). Third, I review candidate models which explain multisensory RT benefits observed in this task, and describe why neither is yet considered to be a common framework. After this overview, I detail what I consider the essential analytical steps for understanding multisensory RT benefits. I then describe how these can be applied within a *comparative approach* (i.e. an experimental approach which will allow understanding of multisensory RT benefits across factors). Ultimately, what I suggest after applying the comparative approach is that an excellent candidate for a common model framework has long been known, but simply neglected.

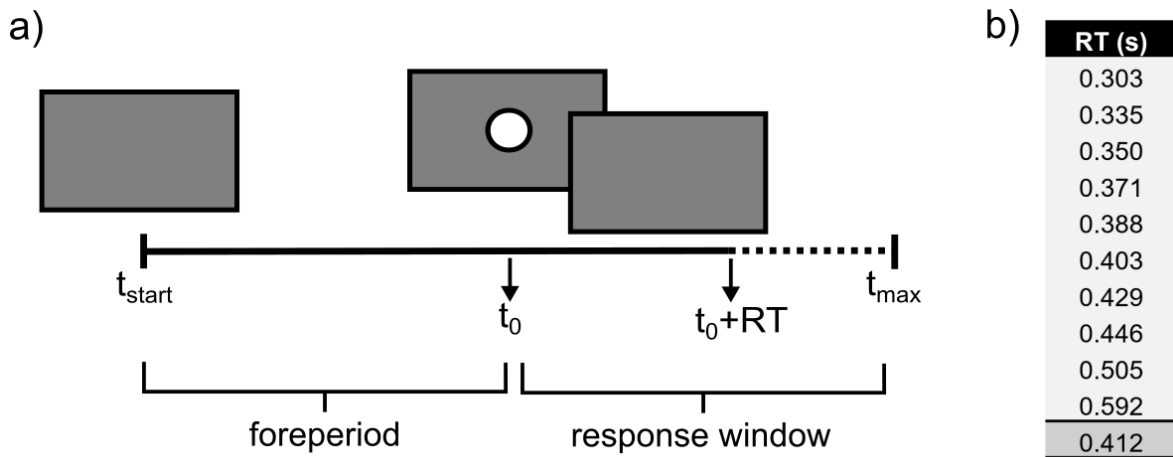
## 1.2. Unisensory Response Times (Background)

The development of a multisensory RT framework begins by understanding the simpler case of unisensory RTs. By studying unisensory RTs and comparing them between different conditions, researchers have been able to understand the timing of the processing stages which make up behaviour, as part of a field known as *mental chronometry* (Posner, 2005). In the field of perceptual decision-making, studying RTs is also useful because they mostly reflect decision-related processing (e.g. Oswal, Ogden, & Carpenter, 2007). As such, the underlying ‘rules’ of decision-making can be understood by deconstructing elements of RT distributions according to a formal model. For unisensory RTs, a widely agreed-upon model framework already exists, which multisensory researchers have attempted to extend in various ways. In this section, I detail this unisensory framework, and the approaches to measuring and evaluating RTs used in this thesis. For demonstration, I also detail some empirical effects in unisensory RTs, and show how the unisensory model framework allows us to understand them.

### 1.2.1. Measuring RTs: The Simple Detection Task

One of the most straightforward paradigms for RT research is the *simple detection task* (Luce, 1986; Woods, Wyma, Yund, Herron, & Reed, 2015). Here, participants respond (e.g. by button-press) as quickly as possible following the onset of a signal, and their RT is measured across trials (see **Figure 1.3** for a more detailed overview). To ensure RTs genuinely reflect decision-making (and not, for example, a response in anticipation of the signal) the signal onset times are typically randomised, and on some trials the signal is not shown at all (catch trials). The experimenter can also record any responses made when a signal was not present (*false alarms*), or record when any signal was not responded to (*misses*); however, as signals presented are usually easy to detect (and there is only a single response option) the number of these errors should be close to zero (i.e. participants demonstrate *ceiling performance*). It should be noted that there are also more complex forms of RT task. For example, some paradigms involve discrimination or choice between multiple stimuli and response options (e.g. Luce, 1986), which allow researchers to consider performance as well as RT. For present purposes, I have detailed only the simple detection task as it forms the basis of all of the experimental work in this thesis. Regardless, as a window onto perceptual decision-making, the simple detection task is incredibly effective as evaluating simple RTs alone (as seen in the following sections) allows significant insight into the architecture of perceptual decisions.





**Figure 1.3 The simple detection task**

**a)** Trial example. During the foreperiod ( $t_{\text{start}}-t_0$ ), the participant waits for the onset of a signal. The foreperiod is randomly varied from trial-to-trial to make signal onset unpredictable. When the signal is presented (e.g. white circle), an onset timestamp ( $t_0$ ) is recorded. A response window then follows when the participant should respond to the signal ( $t_0-t_{\text{max}}$ ). The time of the response ( $t_0+RT$ ) is also recorded. RTs (in s) can be calculated by subtracting  $t_0$  from the response timestamps.

**b)** Simple RT data. 10 simulated RTs, similar to those recorded in detection tasks, are shown ranked fastest to slowest. The central tendency (mean) is shown below (shaded grey).

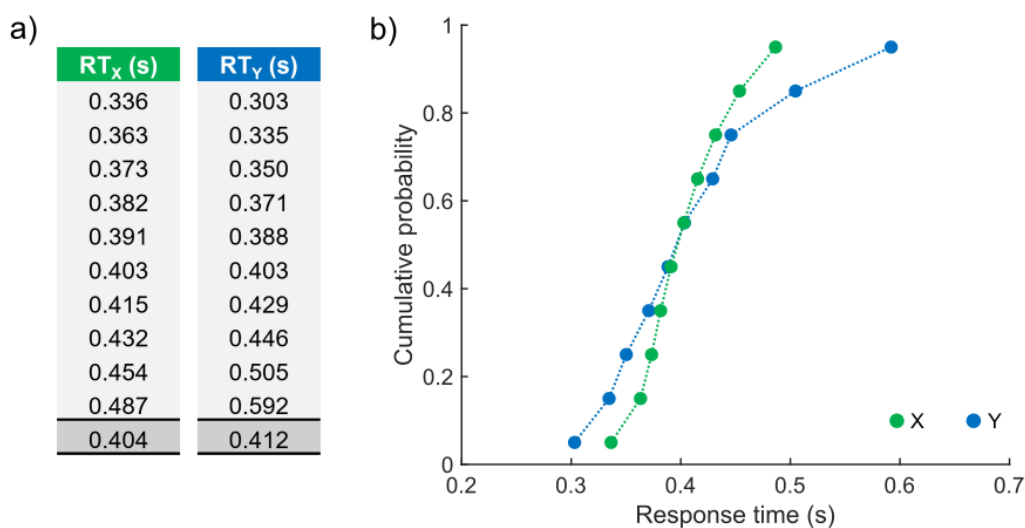
### 1.2.2. Evaluating RTs: The Cumulative Distribution Function

Once collected, RTs must be meaningfully evaluated (**Figure 1.4**). This involves analysing summary statistics (**Figure 1.4a**), such as the central tendency (e.g. mean, median) and variability (e.g. standard deviation, median absolute deviation). Additionally, the *distribution* of RTs (**Figure 1.4b**) in its entirety should be considered. A straightforward representation of this is the *cumulative distribution function* (CDF), which has long been used in RT research (Houghton & Grange, 2011; Ratcliff, 1979). As well as displaying the whole range of RTs measured for a given signal, these plots link RT to probability, which forms a useful basis for further modelling work.

To plot the CDF, all RTs collected for a given signal are first ranked from fastest to slowest. Next, the corresponding *cumulative probability* (CP) value is assigned. CP, for any given value of RT, essentially describes the probability that a response has been recorded up to and including that timepoint. For example, assume an RT value of 0.4 s corresponds to a CP of 0.3. This would mean 30% of RTs were recorded 0.4 s after signal onset. As for any probability, CP values range from 0 to 1. To compute CP, the following equation is applied

$$CP_i = \frac{i - 0.5}{N}, \quad (1)$$

where  $N$  indicates the total number of RTs, and  $i$  is a rank index from 1 to  $N$ . According to **Equation 1**, the median of RT directly corresponds to a CP value of 0.5.



**Figure 1.4 Evaluating RTs**

**a)** Summary statistics (central tendency). 10 simulated RTs, ranked fastest to slowest, are shown for two different kinds of signal (X and Y). Central tendency (mean) is shaded grey. As the means are very similar, it might be concluded that the signals are processed in a similar way.

**b)** Cumulative Distribution Functions (CDFs). The same RTs are now shown plotted as distributions, with fastest RTs on the left of the plot and slowest RTs on the right. This shows that while RTs may be similar in terms of central tendency, the shape of the distributions is very different. For this reason, it is better to work on the CDF to understand how RTs change depending on different signals.

The CDF is a useful graphical tool to represent RTs, particularly if one is interested in variability, as it allows for a detailed comparison of RTs from different experimental conditions. For instance, as shown in **Figure 1.4b**, a steeper curve immediately indicates a *less variable* RT distribution (X), whereas a flatter curve indicates a *more variable* RT distribution (Y). It is also highlighted here because it forms a main level of RT analysis. Working on a distributional level is important as using summary statistics alone to describe RTs is insufficient (e.g. Houghton & Grange, 2011; Whelan, 2008). For instance, two signals may elicit a very similar mean RT, but completely different *distributions*. Any statistical test based only on central tendency, or sometimes even based on a simple approximation of the distribution, may fail to detect these differences (Noorani & Carpenter, 2011).

There are several additional reasons to work with RT distributions, which are important for later chapters. First, certain empirical effects are known to occur at the extremes of the CDF. Benefits between conditions (e.g. congruent vs. incongruent stimulus-response compatibility in Simon Tasks) sometimes occur primarily on the *fastest* RTs, whereas interactions (e.g. task-switching effects) have different impacts across parts of the whole distribution (see Houghton & Grange, 2011). The CDF is therefore important for any approach interested in benefits or interactions between signals. Second,

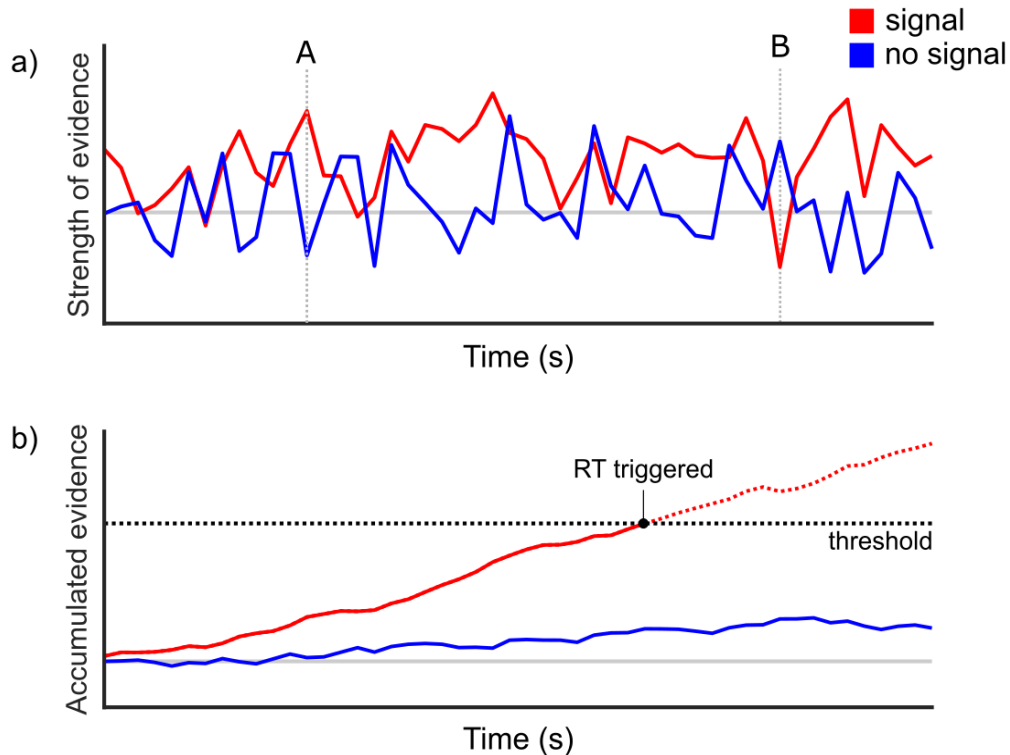
RT distributions act as an additional test of model fit (Lacouture & Cousineau, 2008; Wolfe, Palmer, & Horowitz, 2010). Many models may be able to capture the mean RT, however it is much more difficult to accurately capture entire distributions, which can help distinguish between model explanations.

### 1.2.3. Explaining RTs: Models of Unisensory Decision-Making

Consistent empirical effects can be observed even in simple RTs (e.g. **Figure 1.4**). One example is that RTs are typically much slower and more variable than can be accounted for by elementary sensory processing and motor response (e.g. Genest, Hammond, & Carpenter, 2016; Noorani & Carpenter, 2011, 2016). Another example is that distributions of RTs are not normal, but are instead skewed. In particular, RTs are *right-skewed*, with larger variability at the slow tail than at the fast tail of the distribution. It is difficult to make sense of these effects without an idea of the underlying decision-making process; a useful approach, therefore, is to construct a decision-making model which can account for them.

The most widely agreed-upon modelling explanation of unisensory RTs is *sequential sampling* (for reviews, see Forstmann, Ratcliff, & Wagenmakers, 2016; Gold & Shadlen, 2007; Smith & Ratcliff, 2004). This broad class of models assumes the following common processes in unisensory decisions, such as the simple detection of a signal. To begin, sensory *evidence* is available from the environment relating to the presence or absence of the signal. The observer is able to acquire samples of evidence from the environment at various time points (**Figure 1.5a**). However, the amount of evidence that can be sampled at each time point is not constant. Rather, it is subject to fluctuations arising from random activity in sensory receptors and neurons. We can refer to this generally as *noise*. The goal of the decision-maker, therefore, is to accurately detect signals despite the presence of noise. The assumption is that to counteract noise, the observer *accumulates* evidence over time, until at some point a *threshold* level is reached (**Figure 1.5b**). At this point, the decision process is terminated and the response is triggered. Under the sequential sampling framework, understanding of noise is also important, as both evidence and noise determine RT. While the signal may be essentially identical between trials, the noise elicited can change depending on experimental conditions (**Figure 1.6a**).

The sequential sampling framework is well-regarded for multiple reasons. First, it is an accurate approximation of neural behaviour during decision making (e.g. Hanes & Schall, 1996; Shadlen & Newsome, 2001). Second, it captures and explains empirical effects observed in RTs (**Figure 1.6**). For instance, RT variability can be understood as fluctuations in evidence between trials, which affects the time taken to accumulate to a decision threshold (**Figure 1.6a**). Also, the rightward skew of RTs is explained by the model geometry (e.g. Smith & Ratcliff, 2004). For a simple illustration of this, the accumulation to threshold can be simplified as a linear process (**Figure 1.6b**). Symmetrically increasing and decreasing the mean rate of evidence accumulation results in asymmetrical RTs. For



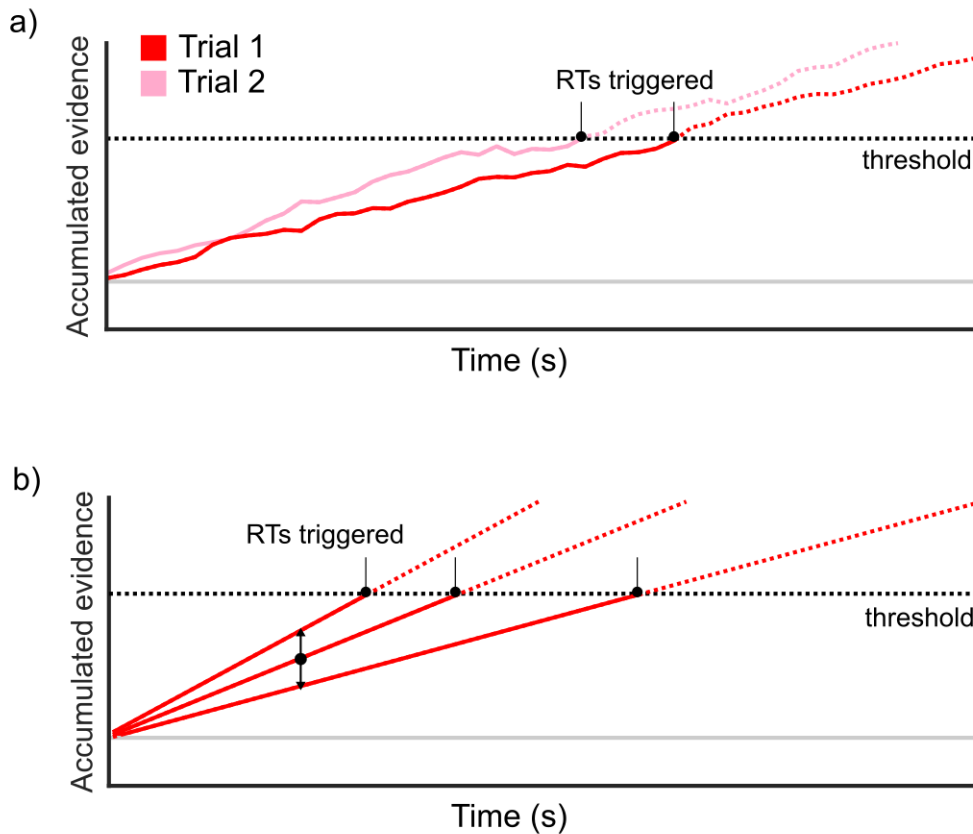
**Figure 1.5 The unisensory decision-making process**

**a)** Sampling evidence. The strength of evidence for the presence of a signal is shown over time. This is subject to random fluctuations in the sensory system (noise). The red line shows an example timecourse with a signal present. The blue line shows an example with no signal present. Based on any individual sample, it is difficult to correctly determine the presence of a signal. Considering sample point A, evidence for a signal was stronger when a signal was present, as expected. Considering sample point B, however, evidence for a signal was stronger when no signal was present, owing to noise.

**b)** Accumulating evidence. The same evidence is shown, but summed over time. By integrating evidence, it becomes easier to distinguish whether a signal was present or absent (evident by the clear divergence of the traces). A criterion level of accumulated evidence (or 'threshold') is also implemented. The decision to respond is made only when this criterion is met.

such reasons, the sequential sampling framework provides a useful link between behaviour and the underlying brain processes.

Though the common framework of sequential sampling is agreed upon, specific model implementations can differ considerably. One main difference is the level of complexity. In choice RT, for instance, one classic model is the Diffusion Decision Model, or DDM (Ratcliff, 1978; Ratcliff & McKoon, 2008). This model very precisely captures the evidence accumulation trace over time, modelling this as a random walk of accumulation between two thresholds (one for each response option). However, an alternative formulation of a choice RT model is the Linear Ballistic Accumulator, or LBA (Brown & Heathcote, 2008). In contrast to the DDM, the LBA simplifies accumulation of evidence to a linear rise to threshold (as in **Figure 1.6b**). It also employs two accumulators to a single

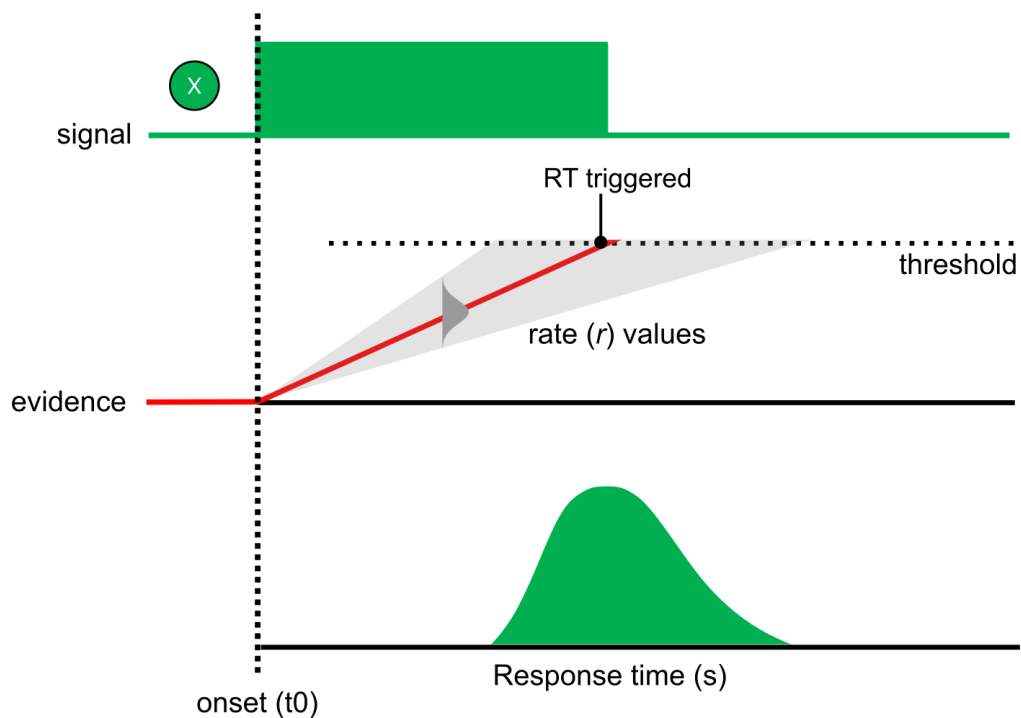


**Figure 1.6 Explaining empirical observations in response times**

**a)** Response time variability. The accumulation of the evidence over time for two different trials is shown. Because of the variability in the evidence over time, the time taken to reach threshold (black dots) varies between trials, which results in RT variability.

**b)** Response time skew. To demonstrate how skew arises, the accumulation process is simplified to a linear accumulation to threshold. In each trace (red lines) the mean rate of evidence accumulation has been incrementally altered by  $\pm 1$  unit (vertical arrows). The model geometry, however, means that the points at which the traces reach threshold (black dots) are asymmetrically distributed, and thus the resulting RT distribution will be right-skewed.

threshold, rather than a single accumulator between two thresholds. The important point here is that many models capture the important elements of decision-making framework, but each is useful in different scenarios. The DDM very precisely models within-trial variabilities in evidence accumulation. However, it cannot easily be extended to tasks with more than two choices. Conversely, the LBA disregards within-trial variability in evidence accumulation (which is often a useful simplification for behavioural work). However, it is easily extended to multiple response options. Overall, the model employed depends on the task and the level of detail needed. In the following section, I detail a model which is useful for the purposes of simple detection tasks.



**Figure 1.7 The LATER model**

RTs to signal X are determined by a linear rise-to-threshold. Values for the linear rate of accumulation are defined by a normal distribution with parameters  $\mu$  and  $\sigma$  (shown in grey). A single example accumulation is shown with the red line. The projection of the rate distribution onto the threshold results in a right-skewed RT distribution (probability density function shown in green). This figure is partially modelled after Noorani & Carpenter (2016; their Figure 4).

### 1.2.3.1. The LATER Model

A specific model implementation of the sequential sampling framework is the Linear Approach to Threshold with Ergodic Rate (LATER) model (Carpenter & Williams, 1995; Noorani & Carpenter, 2016). This model will be detailed in the General Methods (Section 2.4.4.1), but I describe it here briefly as a primer. As the name implies, the LATER model (Figure 1.7) simplifies within-trial variability in evidence accumulation down to a linear accumulation (as also in Figure 1.6b). Thus, to describe RT, the LATER procedure simply attempts to model between-trial variability in evidence accumulation (i.e. the rate of evidence accumulation on each trial). This between-trial variability is defined by two parameters: the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the rate distribution.

The LATER model is particularly appealing for its simplicity; with only two parameters, it provides a well-fitting, continuous description of the entire RT distribution. This is a parsimonious platform for further modelling work. The rise-to-threshold also approximates the behaviour of neurons prior to a response, for instance in the oculomotor network (e.g. Hanes & Schall, 1996). It is important, however, to be aware of the LATER model's limitations. First, it does not provide a way to

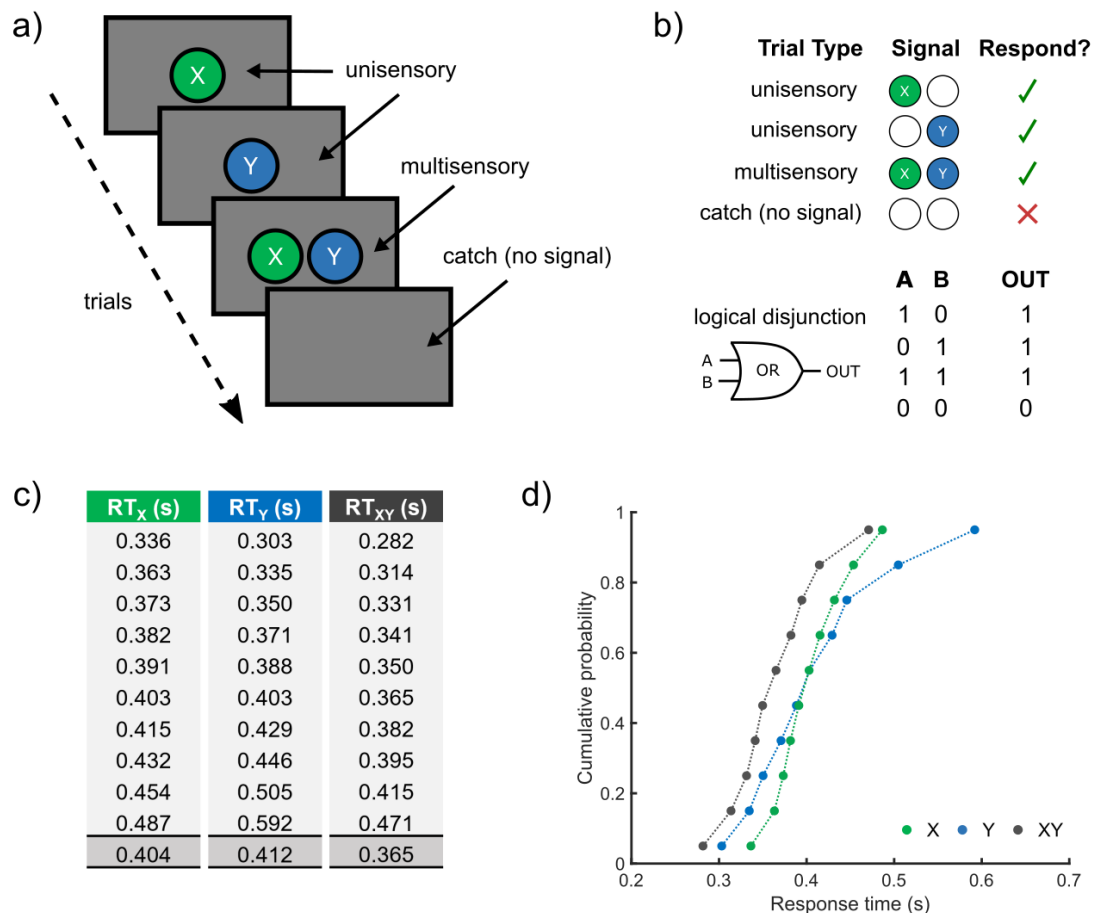
model error responses, as the entire RT distribution is modelled by a single accumulation process. Thus, while it might not be an appropriate model if performance is poor, LATER fits will generally be good in instances where performance is at ceiling (as in simple detection tasks presented here). Second, the LATER model does not explicitly model the non-decision components of RT, and considers entire RTs reflective of decisional processes. Thus, while using the LATER model may not be ideal in cases where non-decision processes are examined, there are few problems if non-decisional components are constant over conditions. Despite these limitations, the model is seen as appropriate and useful in many contexts (Gold & Shadlen, 2007).

### 1.3. Multisensory Response Times

Having reviewed the background on unisensory RTs, I will now turn to the thesis topic: multisensory RTs. Following a similar overall structure, I discuss the RT paradigm which is the focus of the thesis (the *redundant signal paradigm*). I will also detail the multisensory RT benefit observed within the task (the *redundant signal effect*). I will then describe extensions to the unisensory model framework which could account for the benefit: in contrast to unisensory RTs, however, there is no agreed-upon framework. I therefore detail the two most common model explanations.

#### 1.3.1. The Redundant Signal Paradigm (RSP)

Multisensory RTs can be evaluated in the *redundant signal paradigm*, or RSP (**Figure 1.8**). This is an extension of a simple detection task (see **Section 1.2.1**) but in this version there are more trial types. On some trials, one of two unisensory signals is presented, each in a different sensory modality (for example, an auditory beep on some trials, and a visual flash on others). For generalisability, I will refer to these unisensory signals as *signal X* and *signal Y*. On other trials, both unisensory signals are presented simultaneously for a multisensory, or *redundant*, signal (for example, the beep and the flash together). For generalisability, I refer to the redundant signal as *signal XY*. On additional trials, no stimulus is presented at all (*catch* trials). With these 4 trial types (**Figure 1.8a**), the task follows the same instruction as simple detection: respond as long as any signal is present (X, Y or XY) and refrain from responding on trials when no signal is presented (catch trials). Two points can be noted regarding this description of the RSP. First, redundancy is not meant here in the sense that each signal component necessarily provides an estimate of the same physical properties, as in other multisensory contexts (Ernst & Banks, 2002; Ernst & Bulthoff, 2004); rather, the multisensory signal is redundant in the sense that response to either signal component (X or Y) is sufficient for a correct response. Second, examining the task demands of the RSP (**Figure 1.8b**), it is clear that it matches a logical disjunction, or the behaviour of an OR-gate (Otto & Mamassian, 2017; Townsend & Eidels, 2011). Overall, by examining RTs in the RSP (**Figure 1.8c,d**), differences in the underlying unisensory and multisensory processing can be made apparent.



**Figure 1.8 The redundant signal paradigm (RSP)**

**a)** The four trial types in the RSP. Participants are presented with unisensory signals (X, Y), a multisensory/redundant signal (XY), or no signal (catch) on any given trial. These trials are equally likely (25% each), and are typically randomised within an experimental session.

**b)** Task demands. For any given trial, participants are required to respond if any signal is present, but withhold responses if no signal is present. This mirrors a logical disjunction (OR-gate).

**c)** Summarising RTs. The left columns show RTs collected for each trial type (10 simulated RTs, ranked, with mean in grey). RTs for the redundant signal ( $RT_{XY}$ ) are on average faster than either component ( $RT_X$ ,  $RT_Y$ ).

**d)** Visualising RTs. Each distribution, as in **c)**, is plotted as a CDF. It can be seen that the redundant distribution (XY) is faster than both of the unisensory distributions.

The RSP has become a standard task for evaluating multisensory decision-making, with application over a wide field of research including neuroimaging (e.g. Martuzzi et al., 2007), neuroelectrophysiology (e.g. Molholm et al., 2002; Murray, Thelen, Ionta, & Wallace, 2018; Wang et al., 2018; Wynn, Jahshan, & Green, 2014) and neurostimulation (e.g. Bolognini, Miniussi, Savazzi, Bricolo, & Maravita, 2009; Bolognini, Olgiati, Rossetti, & Maravita, 2010; Romei, Murray, Merabet, & Thut, 2007). It has been used as a tool to study difference in multisensory processing between chosen populations, for instance between genders (e.g. Collignon et al., 2010) and age groups (e.g. Couth,



Gowen, & Poliakoff, 2018; Downing, Barutchu, & Crewther, 2014; Murray, Eardley, et al., 2018; Peiffer, Mozolic, Hugenschmidt, & Laurienti, 2007; Ren, Yang, Nakahashi, Takahashi, & Wu, 2017; Wang et al., 2018). It has also been used to study how multisensory processing changes in clinical conditions such as Parkinson's Disease (e.g. Plat, Praamstra, & Horstink, 2000; Ren et al., 2018), schizophrenia (e.g. Williams, Light, Braff, & Ramachandran, 2010; Wynn et al., 2014), and dyslexia (e.g. Harrar et al., 2014). The paradigm has also been adapted for comparative research between humans and non-human animals (e.g. Juan et al., 2017; Lanz, Moret, Rouiller, & Loquet, 2013). In short, the continually-developing applications of the RSP are vast. As such, any tool which can help guide research by finding commonalities will almost certainly be beneficial.

### 1.3.2. The Redundant Signal Effect (RSE)

The classic empirical effect arising in the RSP is that RTs to the redundant signal combination are faster than to the individual unisensory components (**Figure 1.8c,d**). This benefit, first described by Todd (1912), is known as the *redundant signal effect*, or RSE (Kinchla, 1974; Miller, 1982). The reliability of the RSE is firmly established; it has been observed across most sensory pairings, including audio-tactile (AT; Gondan, Lange, Rosler, & Roder, 2004; Marinovic, Milford, Carroll, & Riek, 2015; Nava et al., 2014), tactile-visual (TV; Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002; Girard, Pelland, Lepore, & Collignon, 2013), olfactory-visual (e.g. Amsellem, Hochenberger, & Ohla, 2018) and gustatory-olfactory (e.g. Veldhuizen, Shepard, Wang, & Marks, 2010) signals, as well as with trisensory (i.e. auditory, tactile and visual) signals (e.g. Couth et al., 2018; Diederich & Colonius, 2004; Haggmann & Russo, 2016; Pomper et al., 2014). In addition, there is a body of research which has reported the RSE using multiple sensory signals within the same modality, such as vision (e.g. Corballis, 2002; Mishler & Neider, 2018; Mittelstadt & Miller, 2018; Moradi, Yankouskaya, Duta, Hewstone, & Humphreys, 2016; Mordkoff & Danek, 2011; Ridgway, Milders, & Sahraie, 2008; Ritchie, Bannerman, & Sahraie, 2014; Savazzi & Marzi, 2008; Vrancken, Vermeulen, Germeys, & Verfaillie, 2018). However, the most common signal arrangement, building on classic studies (Hershenson, 1962; Miller, 1982), is audio-visual (AV). This is also the main focus of this thesis.

Considering even just AV pairings, the range of signals which have been shown to elicit the RSE is vast. For example, in the majority of studies, the signal is sudden-onset, however the RSE has also been observed when signals are presented in continuous background stimulation (e.g. Otto, Dassy, & Mamassian, 2013; Otto & Mamassian, 2012). Signal qualities are also broad. In the auditory case, typical signals range from pure tones (e.g. Miller, 1982; Minakata & Gondan, 2018; Murray, Eardley, et al., 2018; Murray, Thelen, et al., 2018; Otto et al., 2013; Otto & Mamassian, 2012; Plat et al., 2000) to noise bursts (e.g. Bolognini et al., 2010; Harrar, Harris, & Spence, 2017; Harrar et al., 2014; Hershenson, 1962). In the visual case, typical signals range from simple pattern or shapes (e.g.

Bolognini et al., 2010; Murray, Eardley, et al., 2018; Murray, Thelen, et al., 2018; Plat et al., 2000), to more complex psychophysical stimuli such as Gabor patches (e.g. Harrar et al., 2017; Harrar et al., 2014; Minakata & Gondan, 2018) or random dot motion (e.g. Otto et al., 2013; Otto & Mamassian, 2012). More ‘realistic’ stimuli have also been examined, such as congruent real word images and sounds (e.g. Bailey, Mullaney, Gibney, & Kwakye, 2018; Collignon et al., 2010; Downing et al., 2014; Juan et al., 2017). Overall, therefore, there is huge potential for comparative investigation across stimuli to understand the factors which meaningfully influence the overall benefit.

Despite such a broad literature, however, it appears that there has been little progress in the development of a common modelling approach. In fact, in a recent review of the RSE literature, Gondan and Minakata (2016, p. 731) speculate that this may not be possible given the breadth of stimuli and contexts in which the task is used. To understand further the lack of a common framework, which ultimately prevents comparative analysis, it is important to understand candidate models of multisensory decision-making.

### 1.3.3. Models of Multisensory Decision-Making

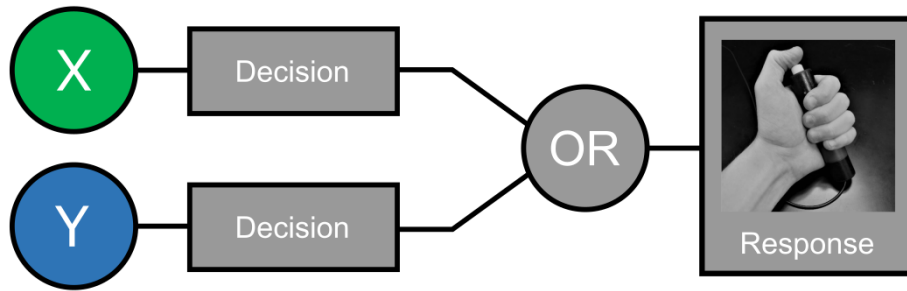
As the unisensory decision-making model is widely agreed-upon, expanding it to multisensory decisions is an obvious starting point. However, there are multiple alternate ways to extend the unisensory accumulation framework (i.e. one signal input) to the more complex case of multisensory decisions (two signal inputs). Each explanation has a different *architecture*, and thus a different explanation for how the RSE arises. The two dominant model classes are described here in detail. The first (race model framework) extends the unisensory framework by assuming there are two *decision-units* which process each signal in parallel. The second (pooling model framework) assumes that there is a single decision-unit into which evidence from both signals is combined.

#### 1.3.3.1. Race Models

The first explanation of the RSE is given by *race* models (e.g. Mordkoff & Yantis, 1991; Otto & Mamassian, 2012; Raab, 1962). The basic race model architecture (**Figure 1.9**) assumes that for the two unisensory signals (X, Y), separate unisensory decision-units accumulate evidence in parallel. Each unisensory decision-unit is independently capable of reaching a decision threshold. These two decision-units are then coupled by a logical ‘OR’ gate to a response component. Either decision-unit is able to trigger a motor response when presented with a multisensory signal.

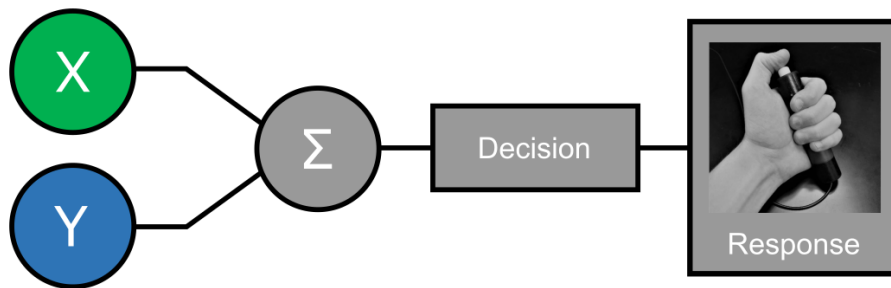
#### 1.3.3.2. Pooling Models

A second explanation of the RSE is given by *pooling models* (e.g. Diederich, 1995; Miller, 1982, 1986; W. Schwarz, 1989, 1994; Zehetleitner, Ratko-Dehnert, & Muller, 2015). The basic pooling model architecture (**Figure 1.10**) assumes that during multisensory signal presentations, evidence for each signal is summed together into a single decision-unit. When this decision-unit reaches threshold, a



**Figure 1.9 The basic race model architecture**

When presented with multisensory signals, evidence accumulates for signal components X and Y independently with individual decision-units working in parallel. Each decision-unit can reach a decision threshold. The decision-units are coupled by a logical OR gate, such that whichever reaches threshold first triggers the response (e.g. button press).



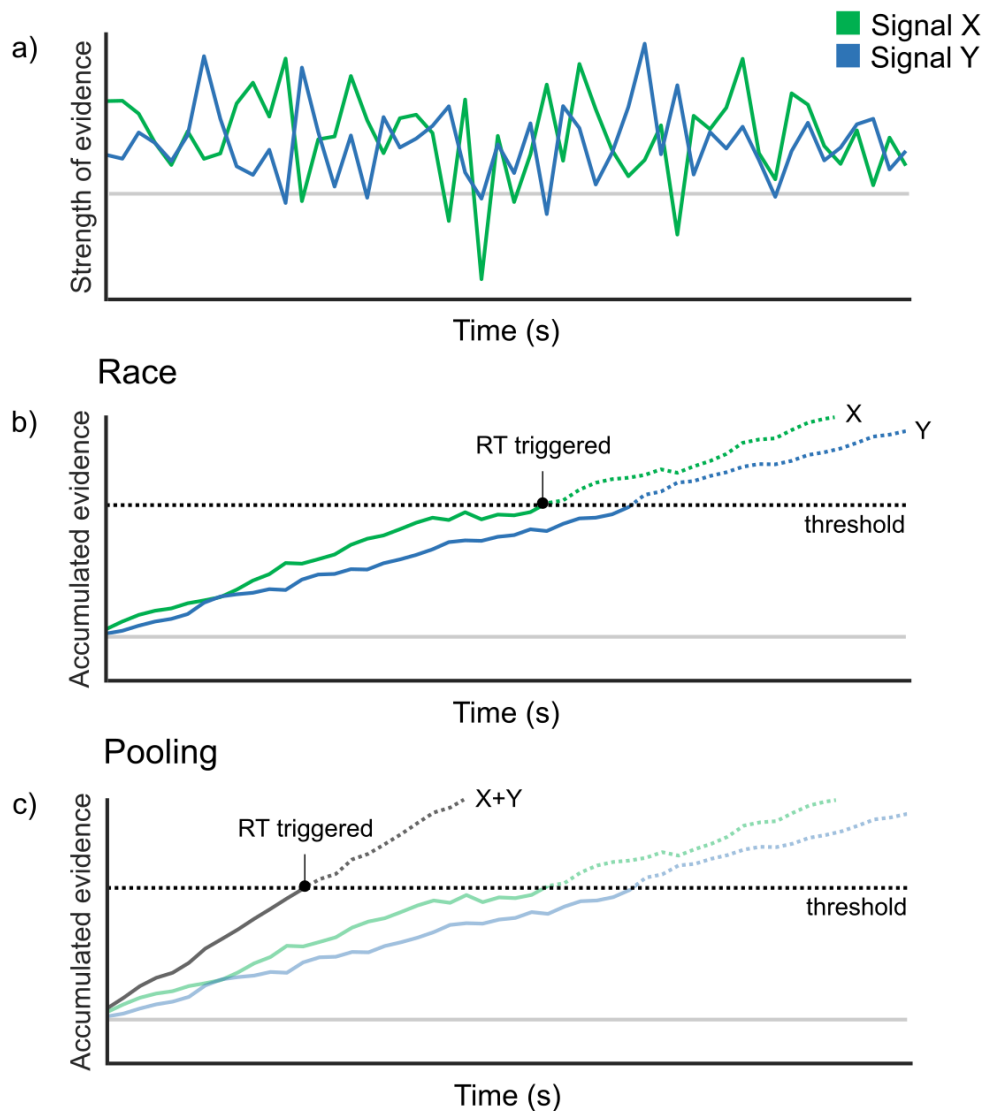
**Figure 1.10 The basic pooling model architecture**

When presented with multisensory signals, evidence for signal components X and Y is summed together (pooled) into a single decision-making process. When a threshold level of evidence is reached, this triggers the response.

response is triggered. Note that pooling models are also sometimes known as ‘co-activation’ models (Miller, 1982, 2016). However, as co-activation (meaning one decision-unit influences the decision time of the other) can be implemented within a race or a pooling architecture, I follow the terminology of Otto and Mamassian (2017) in this thesis (see also Mordkoff & Yantis, 1991, p. 537).

### 1.3.3.3. How Do Race and Pooling Models Explain the RSE?

Race and pooling models suggest two different combination rules. These are shown in terms of evidence accumulation in **Figure 1.11**. Assume that we have two evidence traces for a given redundant trial, one for each unisensory component signal (**Figure 1.11a**). According to race models (**Figure 1.11b**), the RSE arises as a result of *statistical facilitation* (Raab, 1962). When multisensory signals are presented, two decision-units gather evidence in parallel to reach a threshold level of evidence; whichever decision-unit reaches threshold first triggers the motor response (i.e. the *minimum decision time* of two accumulation processes). The distribution of these times is overall faster and less variable, which accounts for the benefit. For a simple illustration of this process using runners’ lap times, refer



**Figure 1.11 Race and pooling model explanations of the RSE**

**a)** Evidence traces for unisensory component signals (X, Y) on a redundant signal trial (XY).

**b)** The race model explanation. Multisensory RTs, under a race model, are triggered by the faster of two unisensory accumulators to reach threshold. As shown above, the accumulator for unisensory component X (green line) reaches threshold before the accumulator for unisensory component Y (blue line). On this trial, therefore, the RT is determined by X. According to race models, multisensory benefit arises from always choosing the faster (or minimum) decision time from two accumulation processes over multiple trials.

**c)** The pooling model explanation. Multisensory RTs, under a pooling model, are triggered by an accumulator which pools evidence from unisensory components X and Y together (grey line). According to a pooling model, multisensory benefit arises because threshold is reached faster when pooling two sources of evidence together compared to one. Note also, by comparison to panel **b)**, that pooling model benefits can very easily exceed those possible with a race model.

to the appendices (**Section 9.1**). According to pooling models (**Figure 1.11c**), the RSE arises as sensory evidence on multisensory trials is ‘pooled’ from two sources rather than from only one source, and thus threshold is reached faster.

#### 1.3.4. Deciding Between Multisensory Models: An Evaluation of Current Practice

Despite the fact that there are two main explanations for the RSE, there has been surprisingly little debate regarding which model explains the RSE for number of decades. The general consensus is that the race model architecture cannot explain the RSE. This is interesting, given the task demands map onto a logical disjunction (OR-gate), which perfectly matches the architecture of the race model. The reason for this rejection is discussed in the following section.

##### 1.3.4.1. Testing the Race Model

The rejection of race models dates back to a seminal paper by Miller (1982). In this paper, Miller developed a test with the goal of aiding model selection in RSE research. The validity of the test will be considered in much more detail later (**Section 1.7.2**). The basic idea, however, is this: a pure statistical facilitation mechanism, which is the core of the race model framework (**Figure 1.9**), has an *upper bound* regarding the size of the possible benefit. In other words, there is a limit to the redundant RT distribution in relation to the unisensory RT distributions. This is known as *Miller's bound*. If Miller's bound is *violated* (i.e. the redundant XY distribution is faster than the bound at any point), then according to the test's logic, race models cannot explain the RSE. In Miller's 1982 paper, and in many papers since, violations of this bound are indeed reported, which has led to a general rejection of all race models as an explanation for multisensory benefits (for a recent review, see Gondan & Minakata, 2016). In current practice, therefore, the race model is effectively a "nullmodel" (W. Schwarz, 1989, p. 498) against which the empirical RSE is evaluated. If violated, the results are taken in support of pooling models.

##### 1.3.4.2. Re-examining the Race Model Test

Since Miller (1982) and the widespread rejection of race models, it might be expected that the development of pooling models is well-advanced. However, as noted in the introduction, there has been little progress in terms of a common model framework. A likely reason for this is that most RSP studies neglect to apply a formal model to explain the RT data collected. This was recently highlighted by Gondan and Minakata (2016). In their review, the authors collated 83 papers which claimed to show violations of Miller's bound. Of these papers, only 9 actually followed up with a formal RT model of any kind. Thus, model development under the dominant pooling approach appears to be lacking.

This seems to remain largely the case. To evaluate publications from the previous year (2018), I followed the same journal search procedure as Gondan and Minakata (2016). From this search (see **Section 9.2** in the appendices for details) I collated 55 Journal articles, of which 22 papers measured human RTs and applied some evaluation of the race architecture (e.g. Miller's bound). Of the 22 papers included in the review, only 2 applied any formal model of RTs. It appears, therefore, that despite an

implicit consensus regarding how multisensory benefits could be quantified and explained, an explicit modelling approach is still not being applied.

Given this lack of a common model framework, it is crucial to return to the race model test and evaluate its validity. As the most basic race model (**Figure 1.9**) relies only on statistic facilitation, the deduction that this model is inadequate following a violation of Miller's bound is valid. However, the rejection of *all* possible race models is not valid. The main reason to question the general rejection is one of the assumptions of Miller's bound. It has since been shown that Miller's bound relies on the assumption of *context invariance* (Ashby & Townsend, 1986; Gondan & Minakata, 2016; Luce, 1986; Miller, 2016; Otto & Mamassian, 2017; C. T. Yang, Altieri, & Little, 2018). This assumption will also be considered in much more detail later (**Section 1.4.2**) but in essence, this is the assumption that the processing of one unisensory signal is unaffected by the presence or absence of the other unisensory signal. To return to our example of runners (see **Section 9.1**), this is equivalent to stating that the lap time of runner X is unaffected by the presence or absence of runner Y (and vice versa for runner Y). Context invariance is a useful simplification in the formulation of models, including the most basic race model; however, it is not clear if this is actually the case for how signals are processed, and there is no straightforward method to test the assumption experimentally (Luce, 1986). This is a crucial point for how we interpret the results of Miller's bound. Given that context invariance is assumed in the test, violations of Miller's bound can only show that the race model architecture and context invariance are not *both* true (Otto & Mamassian, 2017). An alternative explanation, which also allows for violations of Miller's test, is that the race architecture is correct, but the context of the unisensory signal is also important. Returning to the runner metaphor, runner X *may* be affected by the presence or absence of runner Y: it may make runner X faster, or more variable across laps. Similarly, in the case of race models, decision-units could either exchange evidence, for faster processing, or noise, for more variable processing (Otto & Mamassian, 2017; C. T. Yang et al., 2018).

Ultimately, as Miller's bound is unable to conclusively test architecture (Yang et al., 2018), violation of Miller's bound indicates two possible routes of progress. The first is to turn to the alternative pooling model architecture. The second is to assume there are additional processing interactions to be accounted for and incorporated by a race model architecture. Previous research has explored the first approach in detail, with a common comparative framework for studies yet to emerge. The potential of the second approach to provide a comparative framework, on the other hand, remains unexplored. In this thesis, I explore the potential of race models in offering a common framework for the RSE.

#### 1.4. Formulating the Basic Race Model

A common framework for the RSE must begin with the output of the basic race model architecture (**Figure 1.9**), because this provides the benchmark for whatever model explanation is subsequently adopted. For more complex race model approaches, it provides a baseline understanding for benefits and highlights where development is necessary. For pooling model approaches (**Figure 1.10**), it provides the “nullmodel” (W. Schwarz, 1989) which must be exceeded for justifying an alternative model architecture. Before attempting to develop a common framework, understanding of the basic race model must be clear in terms of how it is formulated.

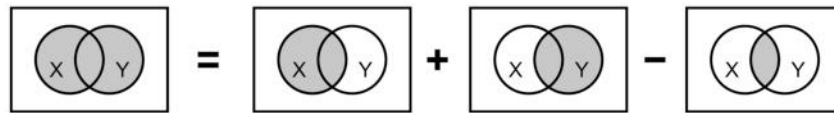
In its most correct form, the race model is formulated in terms of decision time ( $T$ ). We can represent these decision times as *random variables* for all signal distributions. Recall that in the RSP, there are typically two unisensory signals ( $X, Y$ ) and the redundant signal ( $XY$ ). Unisensory decision time variables are denoted as  $T_X$  and  $T_Y$ , and the redundant variable is represented by the union of unisensory component signals ( $T_{X \cup Y}$ ). For redundant signal trials,  $T$  is determined by the faster decision time of the two parallel decision-units which process the unisensory signal components ( $X, Y$ ). Thus, the redundant signal  $XY$  distribution is calculated by taking the minimum distribution of decision times for unisensory signals  $X$  and  $Y$ . This can be formalised for the decision time variables as follows:

$$T_{X \cup Y} = \min(T_X, T_Y). \quad (2)$$

To generate a *model* of the redundant distribution, we need to compute  $P_{XY}(T_{X \cup Y} \leq t)$ ; this is the probability ( $P_{XY}$ , the subscript denoting that both signals were presented) that the decision time for redundant signals ( $T_{X \cup Y}$ ) is smaller than or equal to a particular timepoint ( $t$ ). According to a race process, the model of the redundant distribution is formalised as:

$$P_{XY}(T_{X \cup Y} \leq t) = P_{XY}(T_X \leq t) + P_{XY}(T_Y \leq t) - P_{XY}(T_{X \cap Y} \leq t), \quad (3)$$

where  $P_{XY}(T_X \leq t)$  and  $P_{XY}(T_Y \leq t)$  are the unisensory cumulative probabilities, and  $P_{XY}(T_{X \cap Y} \leq t)$  is the cumulative probability that both responses have been triggered (i.e. the *joint probability*). The essential idea of the equation therefore, as shown by the Venn diagrams in **Figure 1.12**, is to sum unisensory probabilities together and then remove any of the overlap between the two.



**Figure 1.12 A Venn diagram illustration of the race model equation (Otto & Mamassian, 2017)**

To model the decision time for the multisensory signal (XY), the distributions for unisensory decision times (X and Y) are added together. In doing so, the overlapping of the distributions (i.e. the joint probability) has been included twice, so must be subtracted. Reprinted from Otto and Mamassian (2017; their Figure 3), which is freely available for redistribution under the Creative Commons licence Attribution 3.0 Unported (CC BY 3.0). No changes have been made.

In this form (**Equation 3**), however, the race model cannot be implemented. One reason is that the equation is formulated in terms of *decision* time, whereas in experiments we measure overall *response* time (the latter being the sum of both decision and non-decision components). However, as non-decision time is assumed to be a small, fixed component of overall RT, the empirical RT distribution is typically substituted as an estimation of the values of  $T$ , as standard procedure in RT analysis (Gondan & Minakata, 2016; Luce, 1986). Regardless of this substitution, however, there are also two *unknowns* in the equation, i.e. two values which cannot be measured from the corresponding RT, but are necessary to calculate the race model. To address these unknowns, two further assumptions can be made which simplify the model.

#### 1.4.1. Assumption 1: Statistical Independence

The first unknown in the race model (**Equation 3**) is the size of the joint probability,  $P_{XY}(T_{X \cap Y} \leq t)$ . The size of the joint probability is determined by any statistical dependencies in decision times (i.e. how one decision time affects another). A positive dependency will ultimately increase the joint probability, whereas a negative dependency will decrease the joint probability (e.g. Hughes et al., 1994). One simple approach, which avoids the issue of an unknown joint probability, is to assume that decision times to signals X and Y are actually independent events (e.g. Meijers & Eijkman, 1977). This is the assumption of *statistical independence*. To relate this back to the runner analogy (**Section 9.1**), this assumption is equivalent to stating that the time on each lap does not influence other lap times in any way. Following the multiplication law, the joint probability of two independent events is given by the product of two. The race model equation (**Equation 3**) can therefore be simplified to the case of independence as

$$P_{XY}(T_{X \cup Y} \leq t) = P_{XY}(T_X \leq t) + P_{XY}(T_Y \leq t) - P_{XY}(T_X \leq t) \times P_{XY}(T_Y \leq t) \quad (4)$$

I will refer to any race model which assumes statistical independence by the general term *independent race model*.



#### 1.4.2. Assumption 2: Context Invariance

The second unknown in the race model (**Equation 3**) is the probability associated with unisensory signals, X and Y, on presentation of redundant signal XY:  $P_{XY}(T_X \leq t)$  and  $P_{XY}(T_Y \leq t)$ . There is no way to estimate these probabilities directly, as we cannot record corollary RTs to both components of the redundant signal – in the RSP, we only have one RT for the redundant signal and no simple way of knowing whether signal X or Y triggered it.

One simple approach, in the absence of a direct corollary for these decision times, is to substitute the probabilities derived from unisensory conditions. In doing so, however, the implicit assumption is made that decision times to each unisensory signal are the same regardless of presence or absence of the other component signal. This assumption, which was encountered earlier (**Section 1.3.4.2**) is known as *context invariance* (Ashby & Townsend, 1986; Gondan & Minakata, 2016; Luce, 1986; Otto & Mamassian, 2017; C. T. Yang et al., 2018). To relate this back to the runner analogy (**Section 9.1**), this assumption would be equivalent to stating that the lap times for each runner are unaffected by whether they run alone or together. Under the context invariance assumption, **Equation 4** can be written as

$$P_{XY}(T_{XUY} \leq t) = P_X(T_X \leq t) + P_Y(T_Y \leq t) - P_X(T_X \leq t) \times P_Y(T_Y \leq t). \quad (5)$$

This model is the most straightforward implementation of the model architecture proposed by Raab (1962). It is interesting to note that this simple formulation of the race model is *parameter-free*, i.e. the redundant distribution can be predicted by providing only the empirical unisensory distributions, with no additional parameters needed. I will therefore refer to this specific, parameter-free formulation as the *simple race model*. A useful feature of the simple race model, as will be discussed in a following section, is that benefits can also be predicted based only on unisensory distributions.

#### 1.5. Developing a Comparative Approach

The goal of this thesis is to test the ability of race models to provide a common framework for understanding the RSE. Previously, evaluation of the race model framework has been limited to testing for violations of Miller's bound, and rejecting the race model if any are observed. The model itself, however, has considerable power to predict benefits, which has previously not been investigated. Beyond this, violations may also be the result of a violation of context invariance. The ability of the race model to account for such interactions in a more complex form has also remained relatively unexplored. To test the race model, a *comparative approach* was established. The essential idea of the comparative approach is to test the predictive and explanatory power of the race model framework.

Within this comparative approach, experimental factors are introduced to target benefits or interactions. Next, three analytical steps are applied for a comprehensive understanding of RTs in the RSP. At each stage, the race model is in some way evaluated. The first analytical step is to determine the *benefit* (i.e. the RSE). Given the basic race model (**Section 1.4**) can also predict the size of the benefit, a comparison between predicted and empirical benefit will give an indication of how strong the most basic race model (Raab, 1962) account is. The second analytical step is to quantify empirical processing *interactions*, such as violations of Miller's bound (Miller, 1982), which the basic race model cannot explain. If a race model framework is assumed to be correct, these indicate violations of the assumptions of Raab's basic race model, which must be accounted for. The third analytical step is to account for both benefits and interactions with a further model. In this thesis, as I evaluate the ability of race models to account for the RSE, I will apply a more complex race model with additional interaction parameters (Otto & Mamassian, 2012). Over the next three sections, I detail each step in full.

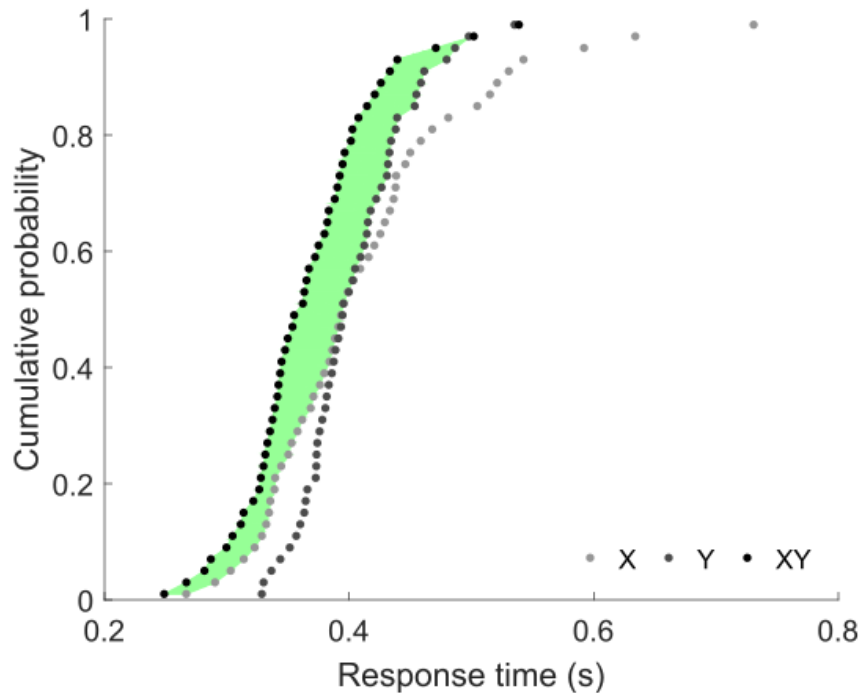
## 1.6. Comparative Approach Step 1: Understand Benefits

The first step of the comparative analysis is to quantify the benefit (i.e. the RSE). In the typical RSP, RTs are collected from two unisensory signals (X, Y) and the redundant signal (XY). Working on the distributions, we can calculate the benefit as the area between the redundant distribution and the faster of the two unisensory distributions at each comparison point (**Figure 1.13**). By comparing changes in benefit over different experimental manipulations, we can gain insight into the underlying processes which are important.

### 1.6.1. Race Model Principles for Multisensory RT Benefit

As in this thesis I am evaluating the power of race models to provide a common framework, it is important to understand the principles that race models offer for how multisensory benefits should change. This topic was first addressed by Raab (1962) alongside his proposal of the original race model. Here, it was noted that the race model distribution shifted more towards smaller values (i.e. larger benefit) as the overlap between the component distributions increased (Raab, 1962; his Figure 7). In terms of RTs, therefore, one prediction would be that the redundant distribution becomes faster on average as the overlap between component unisensory distributions increases. Until recently, however, such proposals had not been systematically investigated. In this section, an overview of recent work on race model predictions is presented.

A clear way to establish race model principles for multisensory benefits is to systematically manipulate unisensory distributions X and Y and examine changes in the predicted benefit. This was done by Otto et al. (2013) in a series of RT simulations. Using the simple race model (**Equation 5**), the same unisensory CDFs can also be used to generate a quantitative parameter-free prediction for the



**Figure 1.13 Calculating benefit**

50 simulated RTs are shown for each condition of the RSP. The benefit (in seconds) is calculated as the average difference between the redundant curve (XY) and the faster of the two unisensory curves (X, Y), considering all cumulative probability points. The benefit area is highlighted in green.

redundant CDF. This is done by substituting the simple race model CDF in place of the empirical redundant CDF. Based on their results (summarised in the following sections), two principles for maximising multisensory benefits were established: the principle of *equal effectiveness* and the *variability rule*.

#### 1.6.1.1. Principle 1: Equal Effectiveness

As a first step to understanding race model principles of multisensory benefit, Otto et al. (2013) posed the following question: given a fixed CDF for signal X, which CDF for signal Y would produce the most benefit? First, the authors worked towards a mathematical proof on the level of individual cumulative probability points (i.e. independent of any assumptions about the shape of RT distributions). According to the simple race model equation, for any arbitrary RT, maximum benefit occurs when the cumulative probability of Y is identical to X. Building on Raab's (1962) original observation, therefore, it can be concluded as a general principle that the largest benefit should occur when the unisensory distributions are identical.

Second, the authors worked on the same principle, but demonstrated it on the more concrete level of RT distributions. A fixed RT distribution was simulated for signal X. The distribution was summarised by two statistics: median and median absolute deviation (MAD). The authors then varied the median and MAD of RTs to signal Y according to a linear grid space. The predicted benefit for each

distribution of X and Y was then computed using the simple race model (**Figure 1.14; centre panel**). The largest predicted benefit occurred when distributions X and Y shared the same median and MAD i.e. when the unisensory RT distributions were the same. This principle was termed *equal effectiveness* (also known as *congruent effectiveness*), and can be summarised as follows: following a race model architecture, benefits increase as unisensory RT distributions become more similar.

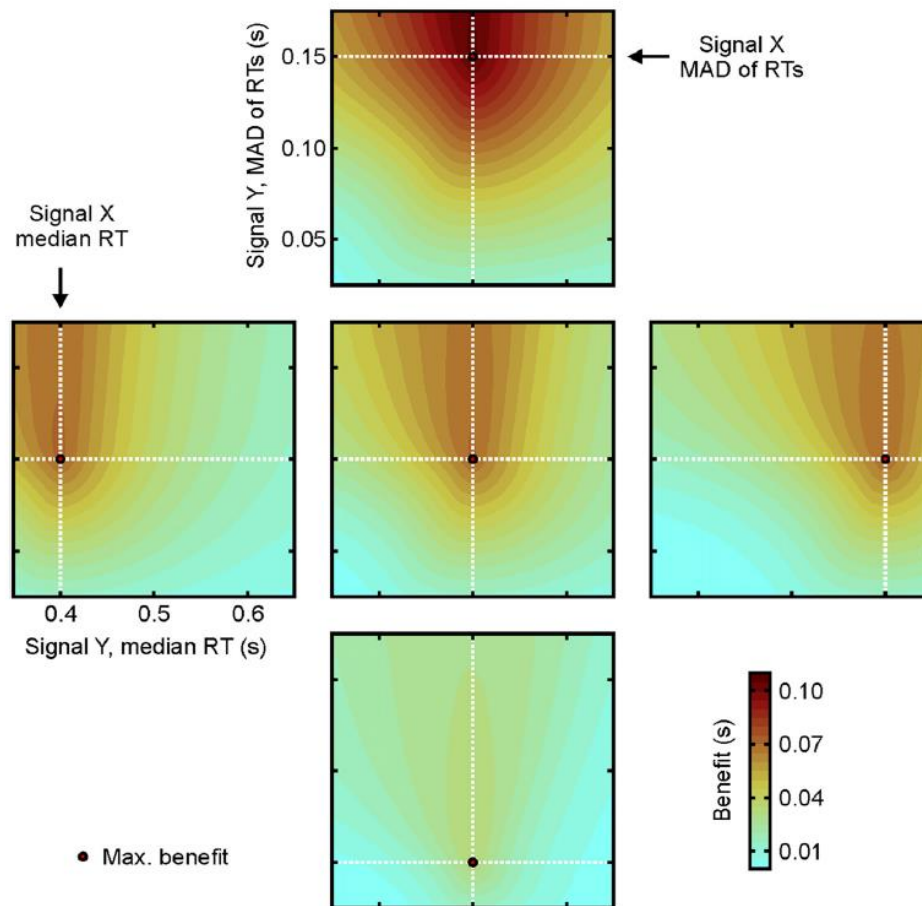
#### 1.6.1.2. Principle 2: Variability Rule

As a second step to understanding race model principles of multisensory benefit, Otto et al. (2013) repeated the previous simulation with an added level of complexity. As well as varying the properties of distribution Y within simulations, the properties of distribution X were also varied between simulations (**Figure 1.14**). As one manipulation, the median of X was increased or decreased while leaving the MAD fixed. These results are shown across the horizontal of **Figure 1.14**. It was shown that the principle of equal effectiveness holds across simulations, as in each plot the maximum benefit occurs when both distributions are the same. The overall benefit size was not particularly affected by changes in the median, as the benefits are similar across plots.

As a second manipulation, the MAD of X was increased or decreased while leaving the median fixed. These results are shown on the vertical plots. Once more, the principle of equal effectiveness holds across simulations. However, the overall benefits are no longer similar across plots. Benefit increases as variability (as assessed by MAD) increases. This demonstrates a key point: variability in the unisensory RTs, according to race models, is the driving force behind multisensory benefits. Specifically, increases in benefits are determined by the least variable of the two unisensory distributions. Note for instance, that benefit remains maximal at the point where the MAD of RTs to both X and Y is the same, and does not increase further as the MAD of RT to Y increases. This principle was termed the *variability rule* and can be summarised as follows: following a race model architecture, benefits increase as the unisensory RT distributions become more variable.

#### 1.6.1.3. An Analogy for Understanding Race Model Principles

The race model principles for multisensory benefit can be understood more intuitively by returning to the analogy with runners' lap times (see also **Section 9.1**). Recall that here, the faster time of two runners (X and Y) is taken as the winning time overall. First, to understand the equal effectiveness principle, consider their relative performance. If one runner (e.g. X) is always much faster than another (e.g. Y) on every lap, the winning time would always come from runner X. If this was true, then both the winning times and the lap times of runner X would be the same, and there would be no improvement by taking the faster time of X and Y. However, if runners X and Y are equally matched, then the opportunity for winning times to be faster than either individual is much larger. A similar mechanism explains the speedup of RTs. If the underlying unisensory decision times (as measured by



**Figure 1.14 Predicted benefits according to the simple race model (Otto et al., 2013)**

For signal X, the RT distribution is given a fixed median and MAD in each plot, shown by the dotted line. The RT distribution for signal Y is then allowed to vary along the axes of the grid. Each colour point on the grid represents the size of the predicted benefit for that combination of distributions for signals X and Y. Warmer colours indicate larger benefits. All plots demonstrate the principle of equal effectiveness i.e. the maximum benefit (red dot) always occurs at the point where the distributions for X and Y are equal. The horizontal row of plots show changes to the median of signal X (with a fixed MAD). The vertical plots show changes to the MAD of signal X (with a fixed median). The large increase in benefit as MAD increases demonstrates the variability rule. This figure is reproduced from Otto et al. (2013; their Figure 3), published in the *Journal of Neuroscience*. It is reprinted with permission of Society for Neuroscience in the format “republish in a thesis/dissertation” via Copyright Clearance Centre. No changes have been made.

RT) are very different, then there is no improvement when going from unisensory to redundant conditions; the faster unisensory process always determines the redundant trial RTs. However, if there is equal effectiveness, then the potential to see multisensory benefits is maximised because redundant RTs are always triggered by the faster of two unisensory processes.

Second, to understand the variability rule, consider the lap-to-lap variability of these runners. Assume again that we are taking the faster time of X and Y as the winning time, and in this case they have similar (equally effective) performance across laps. If these runners are both consistent in their

times, then the winning times on average will be very similar to the average times of each runner (i.e. they will not be much faster than either individual runner). If these runners are very *variable* in their times, however, then on average the winning times will be very different than the average times for each individual runner. On some laps, runner X will be fast, and on others very slow (the latter laps slowing down their individual average time). Similarly on different laps, runner Y will also be fast and on others slow. The winning times however only take into account the faster time of the two runners on each lap, and so will be impacted less often by any slow times. This means that on average, the winning times will be much faster than the average lap time for either runner. A similar process is at work on decision times, according to the race model; if unisensory decision times (as measured by RT) are not variable, then there will be little benefit going from unisensory to redundant conditions. However if there is large variability in decision times, then the multisensory benefit (i.e. the difference between redundant and unisensory performance) will be much larger.

#### 1.6.2. Key Analysis: Compare Predicted and Empirical Benefits

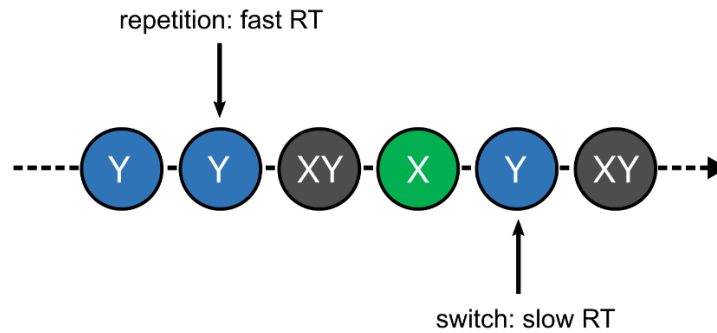
An important step to evaluating the race model framework is to compare predicted benefits with empirical benefits. In previous research, this important comparison is neglected: benefit is typically calculated on the level of central tendency, followed by an evaluation on the level of distribution (computing Miller's bound) to accept or reject the race model. The general interpretation from the latter is that race models predict benefits which are slower than empirical benefits. The goal of the comparative approach, however, is to evaluate this on a much more detailed level. For instance, it may be that the simple race model falls short on average, owing to simplified assumptions, but the general directional predictions offered by the race model principles hold true across factors. By comparing changes in predicted and empirical benefits on the level of the group average and the level of the individual participant, a clearer idea of how these principles account for benefits can be gained.

#### 1.7. Comparative Approach Step 2: Quantify Interactions

There are at least two empirical interactions observed in RSP studies which the simple race model is unable to account for. It is important that these are quantified for a complete explanation of multisensory behaviour. These also relate strongly to the assumptions which are made in formulating the race model, therefore understanding the interactions is important to know how they must be accounted for by any more complex model.

##### 1.7.1. Interaction 1: Statistical Dependence (Trial History Effects)

One key assumption in the formulation of the simple race model is *statistical independence* (**Section 1.4.1**). This simplification is inaccurate, however, as in experimental trial sequences there are often *sequential dependencies* across trials (Fischer & Whitney, 2014). In the specific case of the RSP, RTs on a given trial are influenced by RTs from previous trials (Gondan et al., 2004; Juan et al., 2017; Miller,



**Figure 1.15 Trial history effects**

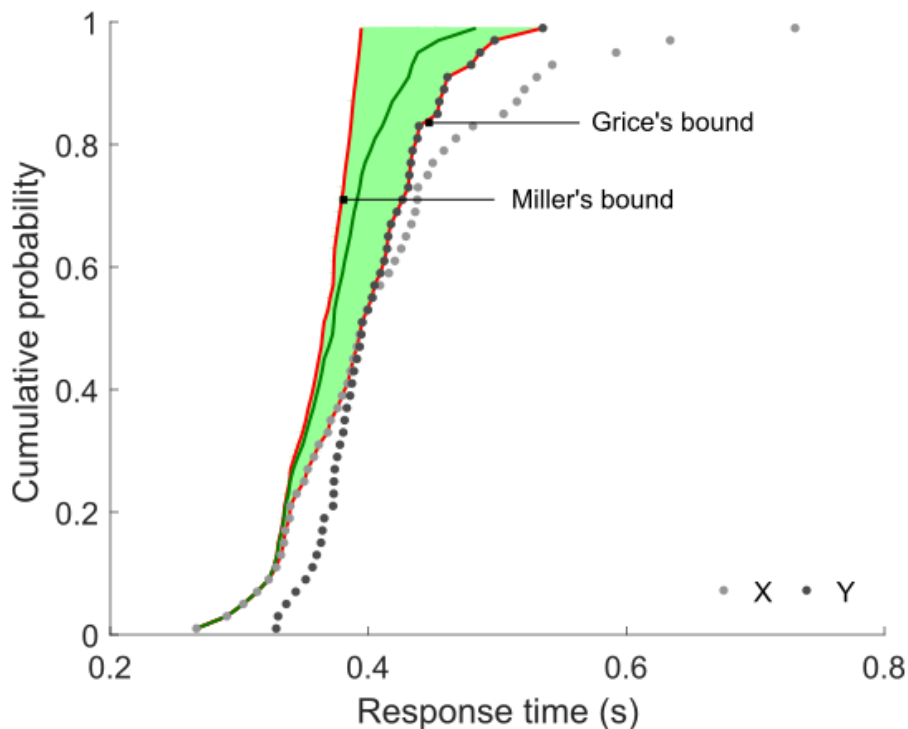
Examples of history effects in trial sequences. Unisensory repetition typically produces a fast RT. Unisensory switches, however, typically produce a slow RT.

1982, 1986; Otto & Mamassian, 2012). I refer to these as trial *history effects* (Otto & Mamassian, 2017). These are particularly observed on unisensory trials in the form of a ‘modality switch’ effect. For instance, if the current trial is a repetition of the previous unisensory trial, then the current RT is typically fast, and if the current trial is a switch from the previous unisensory trial, then RT is typically slow (**Figure 1.15**). A typical measure of this is to calculate the cost of the modality switch (i.e. the difference between switch and repetition trials).

The modality switch cost impacts the overall correlation between unisensory RTs. If the previous trial was auditory (A), then RT will be fast if the current trial is A, but slow if it is visual (V). If the previous trial was V, however, then RT will be fast if the current trial is V but slow if it is A. Because switches work in these opposing ways between unisensory RTs, the dependency overall is assumed to be negative (Miller, 1982). This negative dependency between unisensory RTs has been shown in experiments (Otto & Mamassian, 2012). As shown in the formulation of the race model (**Section 1.4**) a negative dependency will mean that the joint probability is smaller than the simple race model assuming independence. Overall, therefore, benefit will likely be larger than predicted by independent race models. Though potential correlations created by history effects cannot be accounted for by a simple race model, they do not challenge the pure race model architecture in and of themselves. This is because joint probability is not limited to the case of statistical independence; it can take on any value between 0 and 1. However, if independence is not assumed, then there is still the issue of the joint probability being unknown. The next interaction returns to this problem.

#### 1.7.2. Interaction 2: Context Variance (Violations of Miller’s Bound)

The second assumption in the formulation of the basic race model is context invariance (**Section 1.4.2**). This assumption is also made by Miller’s bound, which is typically used to test the race model. Assuming the race model framework is correct, violations of the race model bounds are indicative of context variance i.e. the unisensory signals interact in some way, which means that processing is



**Figure 1.16 Limits of the race model (assuming context invariance)**

The same simulated unisensory RTs (X, Y) as in **Figure 1.13** are shown. According to the basic race model equation under context invariance (**Equation 5**), the redundant curve (XY, not shown) must lie within the green area. This is defined by two bounds (red lines). Grice's bound represents the case of maximum positive correlation; if the redundant RT distribution exceeds this bound, it indicates a cost of redundant signals which a basic race process cannot account for. More crucially, Miller's bound represents the case of maximum negative correlation; if the redundant RT distribution exceeds this bound it indicates a benefit of redundant signals which a basic race process cannot account for. The simple race model (i.e. assuming statistical independence) is shown plotted for reference (green line).

different between unisensory and multisensory trials. An indication of context variance can then be given by quantifying the amount of violation observed.

#### 1.7.2.1. Bounds of the Race Model

Determining the bounds of the race model (**Figure 1.16**) under the context invariance assumption brings us back to the problem of the unknown joint probability. A classical line of reasoning, beginning with Miller (1982), simplifies this problem and allows us to compute bounds to the race model. As the joint probability can only take on values between 0 and 1, this leads to an interesting point: the joint probability can only be a positive number, and removing it from **Equation 3** means the left hand term,  $P_{XY}(T_{XUY} \leq t)$ , can only become larger (or remain unchanged, if the joint probability is 0).

Formally, this can be written as follows:



$$P_{XY}(T_{XUY} \leq t) \leq P_{XY}(T_X \leq t) + P_{XY}(T_Y \leq t). \quad (6)$$

This inequality represents the *upper bound* to the context invariant race model. It is also the case of a race model with maximal negative correlation between unisensory units, which gives rise the largest benefit (e.g. Colonius, 1990; Colonius, Wolff, & Diederich, 2017). However, as written here, it is not possible to compute, because of the unknown probabilities. To compute the bound, Miller (1982) substituted the unisensory probabilities into the equation (i.e. assumed context invariance; see **Section 1.4.2**). The inequality in **Equation 6** can then be rewritten as

$$P_{XY}(T_{XUY} \leq t) \leq P_X(T_X \leq t) + P_Y(T_Y \leq t). \quad (7)$$

Following this substitution, the upper bound to the race model (*Miller's bound*) is calculated as

$$Miller's\ bound = P_X(T_X \leq t) + P_Y(T_Y \leq t). \quad (8)$$

As a corollary, we can also compute a *lower bound* for the race model (Grice, Canham, & Gwynne, 1984). This assumes the maximum possible joint probability, and can be stated as follows:

$$\max(P_{XY}(T_X \leq t) + P_{XY}(T_Y \leq t)) \leq P_{XY}(T_{XUY} \leq t). \quad (9)$$

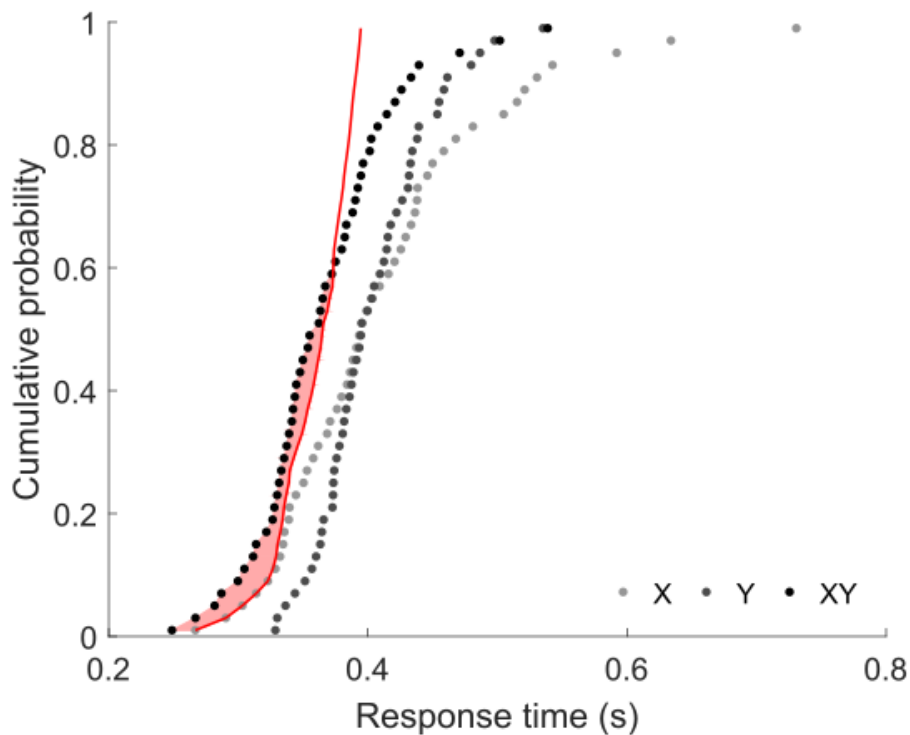
Following a similar substitution using the context invariance assumption, the lower bound to the race model (*Grice's bound*) is calculated as:

$$Grice's\ bound = \max(P_X(T_X \leq t), P_Y(T_Y \leq t)). \quad (10)$$

This bound is the race model in the case of a maximal positive correlation. Theoretically, this bound can be used to test whether RTs are slower than the race model prediction (Van Zandt & Townsend, 2013). However, as we typically observe benefits in the RSP (i.e. the redundant curve is faster than Grice's bound), this is generally not of interest. However, much like Miller's bound represents the case of maximal benefit, Grice's Bound indicates the case of no benefit, as the distribution follows the faster of the two unisensory distributions at all time points. Grice's bound, therefore, is actually useful in the computation of benefit across the entire RT distribution (see **Section 2.3.2.1**).

### 1.7.3. Key Analysis: Quantify History Effects and Violations

It is important to quantify both history effects and violations of Miller's bound (**Figure 1.17**) because they both contribute to benefit beyond the simple race model. Under a race model framework, this can point to processing interactions which go beyond the basic combination rule (statistical



**Figure 1.17 Violation of Miller's bound**

The same simulated RTs as in **Figure 1.13**. Miller's Bound (red line) is calculated according to **Equation 8** based on the unisensory (X, Y) distributions. The redundant curve (XY) exceeds, or violates, this bound at the fast tail (filled red area). This violation area represents facilitation unaccounted for by a pure race process.

facilitation). History effects, for instance, contribute to the overall variability of unisensory RTs. According to the variability rule, if history effects make RTs more variable, they should contribute to the overall benefit. As detailed later, it is possible to measure how much the history effect contributes to overall variability (**Section 2.3.3.1**). Further, violations of Miller's bound (by definition) represent the area of multisensory benefit which is not accounted for by race models. A race model framework attributes this to context variance i.e. changes in the processing of one unisensory component depending on the presence/absence of the other component signal. By comparing changes in interactions across conditions, we can understand their sources and hence which elements of stimulation contribute to benefits beyond the race mechanism.

### 1.8. Comparative Approach Step 3: Apply and Evaluate a Model

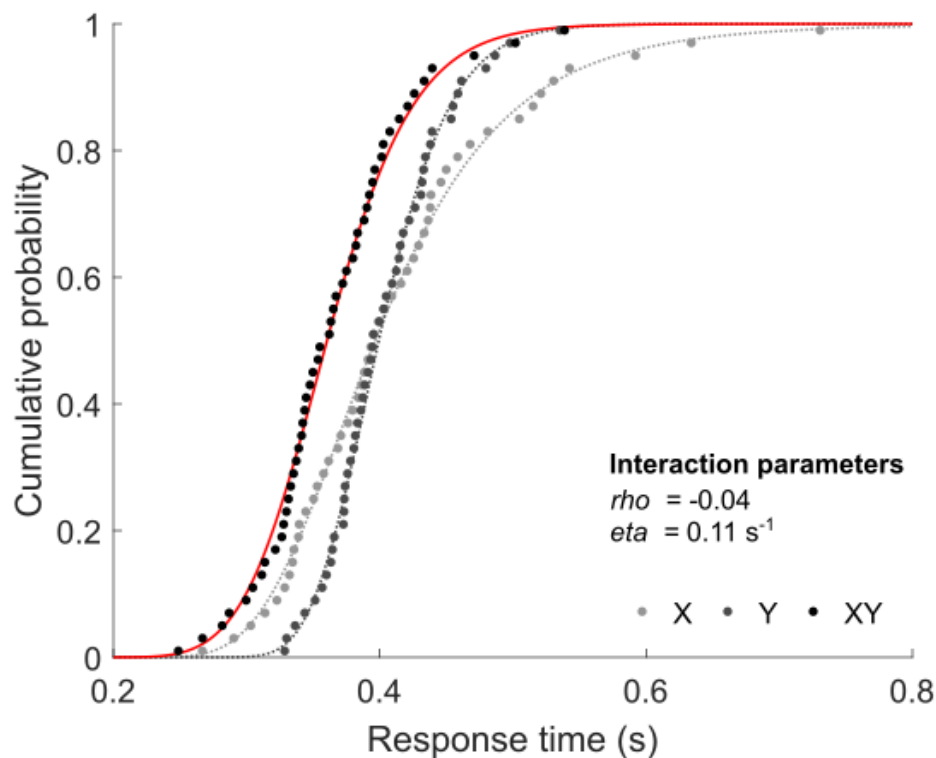
A final step, after quantifying benefits and interactions, is to attempt to fit a formal model to the data which can explain the effects. In comparison to the earlier quantification stages, this latter stage is more discretionary – as shown earlier (**Section 1.3.4.2**), violations may be accounted for by race or pooling models. In this thesis, my general aim is to evaluate the race model framework, so rather than turn to existing pooling models, I apply the context variant race model (Otto et al., 2013; Otto &

Mamassian, 2012, 2017). The reason for this is that, to my knowledge, it is the only published model which has shown the potential to account for benefits and interactions on the level of entire RT distributions.

### 1.8.1. The Context Variant Race Model

The context variant race model (Otto & Mamassian, 2012) is an extension of the basic race model architecture, which attempts to address the assumptions. As a first step, it establishes a basic race model. To do this, it makes use of the LATER model (**Section 1.2.3.1** and **Section 2.4.4.1**) to fit continuous distributions to each unisensory RT distribution. This provides an independent race model with just 4 parameters for the decision-units ( $\mu_x$ ,  $\mu_y$ ,  $\sigma_x$ ,  $\sigma_y$ ). As a second step, the modelling process adds two free parameters, one for each interaction (**Figure 1.18**). To address history effects in the data, the statistical independence assumption is dropped and a correlation parameter ( $\rho$ ) is modelled between unisensory decision-units. By explicitly modelling dependencies by correlation, there is no longer any issue of an unknown joint probability as the correlation indicates the size of this area; complete overlap of unisensory probabilities is the case of maximal positive correlation, and no overlap is the case of maximal negative correlation. Thus, the model can take on any shape between Miller's bound and Grice's bound. Typically, unisensory RTs are assumed to be negatively correlated (Miller, 1982; W. Schwarz, 1989); one reason for this is that trial history effects work in opposing ways (as evidenced by an overall switch cost). In accordance with this, Otto and colleagues (Otto et al., 2013; Otto & Mamassian, 2012) find that the correlation parameters for their group RT distributions are negative.

To address the issue of context variance, the context invariance assumption of the model is dropped, and a noise parameter ( $\eta$ ) is modelled for the accumulation process in redundant trials. This parameter is motivated by empirical observations of the redundant RT distribution. As noted by Otto and Mamassian (2017), violations of Miller's bound are usually interpreted as evidence of faster RTs than predicted by the basic race model. However, violations typically only occur at the faster tail of the redundant distribution; as the slower tail is generally not violated, it is not evaluated. Otto and Mamassian (2012) evaluated the residuals between the empirical redundant distribution and the best fitting race model (including the correlation parameter). In line with Miller's bound violations, empirical RTs were faster than the race model fit at the fast tail of the distribution. In addition, empirical RTs were slower than the race model fit at the slow tail. Their interpretation, therefore, is not that empirical CDF for redundant RTS is overall *faster* than an independent race model predicts, but rather *more variable*. To be clear, however, this does not mean that multisensory RTs are more variable than *unisensory* RTs; as shown previously, the race mechanism results in faster and less variable performance in multisensory conditions compared to unisensory conditions (see **Section 9.1**).



**Figure 1.18 Fitting the context variant race model**

The same simulated RTs as in **Figure 1.13** are shown here with model fits. The dashed lines indicate the underlying unisensory LATER model fits. The solid red line shows the context variant race model fit (including additional parameters ( $\rho$ ,  $\eta$ )). In this individual simulation, the recovered parameters show there was basically no correlation between unisensory RTs, but a large amount of additional noise in the accumulation process on redundant trials. This is in accordance with the large violation area shown in **Figure 1.17**.

This additional noise, therefore, is variability in redundant RTs in relation to the prediction of the 4 parameter independent race model. This is modelled as a constant added onto both unisensory  $\sigma$  parameters. In accordance with their interpretation, Otto and colleagues (Otto et al., 2013; Otto & Mamassian, 2012) found that additional noise was modelled beyond the basic race model to account for their group RT distributions.

#### 1.8.2. Key Analysis: Observe Changes in Model Parameters

Model parameters introduced by the context variant race model to account for the empirical processing interactions observed in Analysis Step 2 (**Section 1.7**). In the first instance, therefore, it is expected that there is a correspondence between changes in interactions and changes in the model parameters. In the case of the latter, however, these are modelled across entire distributions of RTs, whereas interactions are quantified on smaller samples. For example, the history effect is only calculated on unisensory repetition and switch trials. Violations, similarly, are observed only at the fast tail of the distribution. In this sense, model parameters are likely to be more sensitive to

underlying changes in RTs. By formalizing these changes in terms of a model, it also allows us to link the empirical effects to underlying processing rules (which this thesis aims to understand).

## 1.9. Thesis Chapter Outline

In this first chapter, I have covered the necessary background for RT frameworks to understand multisensory benefits in the RSP.

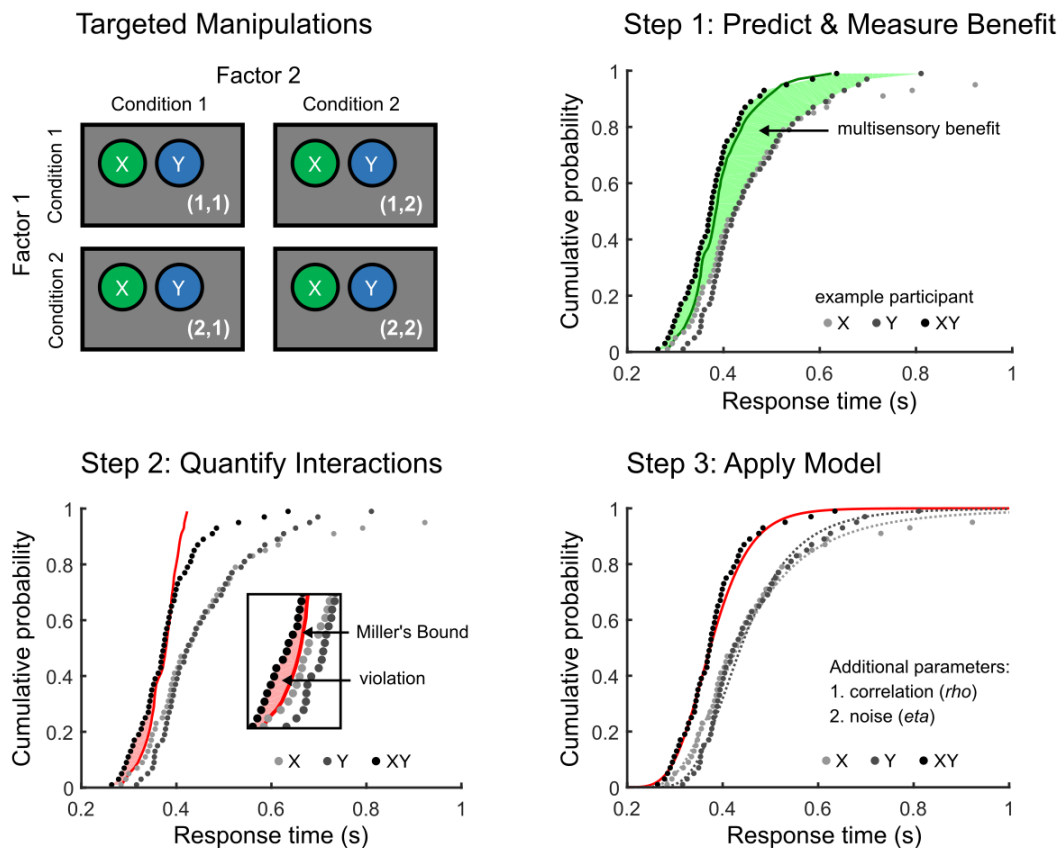
In the following chapter (**Chapter 2**), I describe the *General Methods* that I used across all experimental work presented in the remaining thesis. This includes procedures used to collect experimental data, steps for processing data and computing statistical measures for analysis, and a necessary background to understand the specific modelling and simulation work.

In the following three chapters (**Chapter 3-5**), I describe all experimental work which was carried out using the novel comparative approach (**Figure 1.19**). In each experiment, I manipulate factors according to a 2×2 within-subjects factorial design. In the first experimental chapter (**Chapter 3**), I establish the overall comparative approach. As such, I introduce factors which target both benefits and interactions. The first manipulation, *stimulus construction*, aims to clarify how basic elements of signals across RSP research contribute to the overall RSE. The second manipulation, *signal features*, explicitly targets history effects in an attempt to understand their underlying source.

Having established the comparative approach, the second experiment (**Chapter 4**) focuses on revealing sources of interactions. The first manipulation, *signal duration*, aims to understand the underlying source of history effects. The second manipulation, *task-irrelevant stimulation*, attempts to understand context variance, and how violations of Miller's bound might arise within a race model framework.

Having provided a comprehensive account of benefits and interactions in previous chapters, the last experimental chapter (**Chapter 5**) focuses on potential sources of multisensory benefits at different processing stages (i.e. decisional, non-decisional). The first manipulation, *signal strength*, aims to clarify how the strength of evidence for signals impacts the overall RSE. The second manipulation, *response effector*, aims to clarify how non-decisional elements (i.e. different methods of responding) contribute to the overall RSE.

Following the experimental chapters, I attempt to provide an overall assessment of how the basic principles of the comparative framework established here account for benefits. In an overall summary analysis chapter (**Chapter 6**), I use the data from experimental chapters to highlight key relationships between variables. I assess the overall predictive and explanatory power of race models, and look in more detail at how effective specific implementations of race models are as an overall explanation by using model comparisons methods.



**Figure 1.19 Overview of the comparative approach**

Experimental manipulations are developed which attempt to target the sources of processing interactions (e.g. a 2x2 within-subjects design). Three analytical steps are then applied to identify how the race model can account for the RSE across these conditions. First, benefit is predicted using the simple race model (green line) and compared to empirical benefit. Second, interactions beyond the simple race model are quantified, such as violations of Miller's bound (red line). Third, a more complex model is applied, which can account for benefits and interactions (the red line shows the context variant race model fit). This figure has been adapted from original material which was later included in Innes and Otto (2019). The latter is available for redistribution in any form according to the Creative Commons licence Attribution 4.0 International (CC BY 4.0).

In my final chapter (**Chapter 7**), I provide a *General Discussion* of the overall experimental work in relation to our current understanding of multisensory behaviour. I conclude that the comparative approach based on race models presented here is a promising start for a multisensory RT framework.

## 2. General Methods

### 2.1. Experimental Procedures

In this section, I describe the general procedures involved in setting up and conducting my experiments. All procedures were submitted to and approved by the University Teaching and Research Ethics Committee (UTREC, approval code: PS12181).

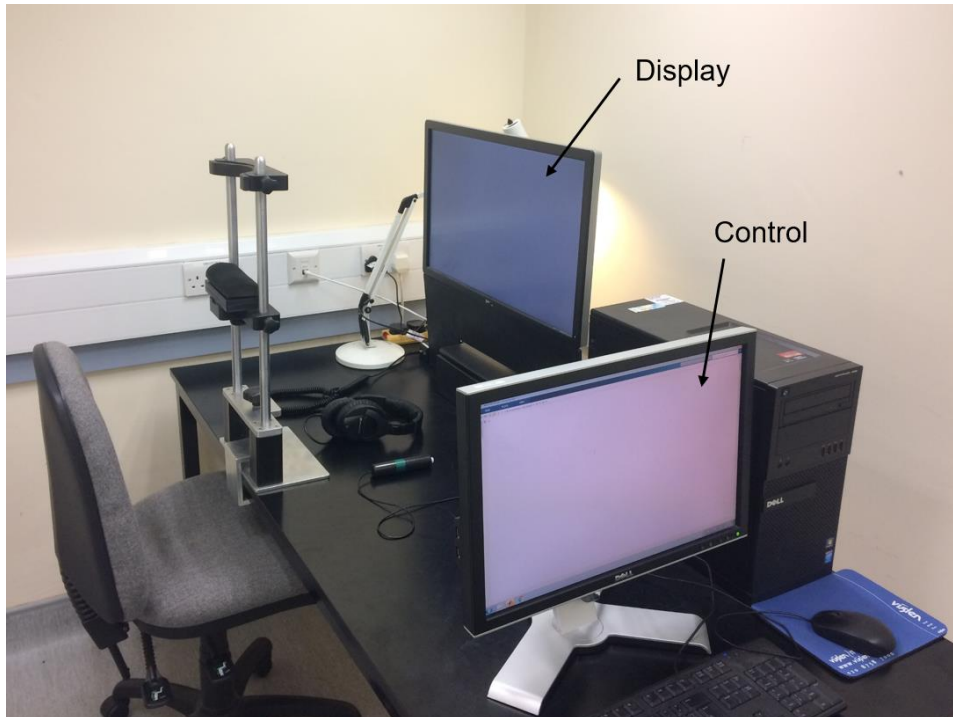
#### 2.1.1. Apparatus

All experiments (unless otherwise stated) were run on a Dell Optiplex XE2 (Windows 7) with an AMD Raedon HD 8490 graphics card. A dual-display setup (**Figure 2.1**) was used, with a Dell UltraSharp 2009W monitor (1680×1050 pixel resolution) functioning as the experimenter's control screen and a U2713HM monitor (1920×1200 resolution; 60 Hz refresh rate) functioning as the experimental display (**Figure 2.1**). Viewing distance was 57 cm from the experimental display, as maintained by a chinrest. Experiments were conducted in dim light with a small lamp behind the experimental display. Auditory stimulation was delivered using Sennheiser HD 280 Pro over-ear headphones at a sample frequency of 44.1 kHz. All stimuli were calibrated individually (see **Section 2.1.3**). In some experiments (see **Chapter 4**) tactile stimuli were also presented using a C-2 Tactor (Engineering Acoustics, Inc.). All experiments were coded in MATLAB (versions R2015b onwards) with standard toolboxes installed. Additional Psychtoolbox 3.0 functions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) were also used to generate stimuli (**Section 2.1.2**).

RTs were measured using an RTbox Version 5 (**Figure 2.2**) and the corresponding MATLAB toolbox (Li, Liang, Kleiner, & Lu, 2010). The RTbox records and stores the time of button-presses according to its own microcontroller, which is synchronised to the computer clock prior to data collection. This avoids any random delays introduced by USB data transfer (e.g. Woods et al., 2015), and allows for extremely accurate (millisecond precision) recording of RTs. Custom-built push button controllers were used for manual responses in all experiments. For responses by foot, a custom-built foot pedal was used. The additional photodiode and sound ports of this device were also used for calibration of sensory stimuli (**Section 2.1.4**).

#### 2.1.2. Stimulus Generation

Auditory stimuli were generated as vectors. The resolution of stimulus vectors was the same as the sample frequency (44.1 kHz), thus a stimulus lasting 1 s consisted of 44100 elements. Presentation of auditory stimuli was handled using the Psychtoolbox *PsychPortAudio* functions. The presentation area for all stimuli in this thesis was a notional annulus: the inner annulus radius was 1° and the outer radius was 4°. All stimuli were presented on a grey screen, with a central green fixation point (0.11° degrees visual angle). Presentation of visual stimuli was handled using the Psychtoolbox *Screen* functions.



**Figure 2.1 Setup of experimental equipment**

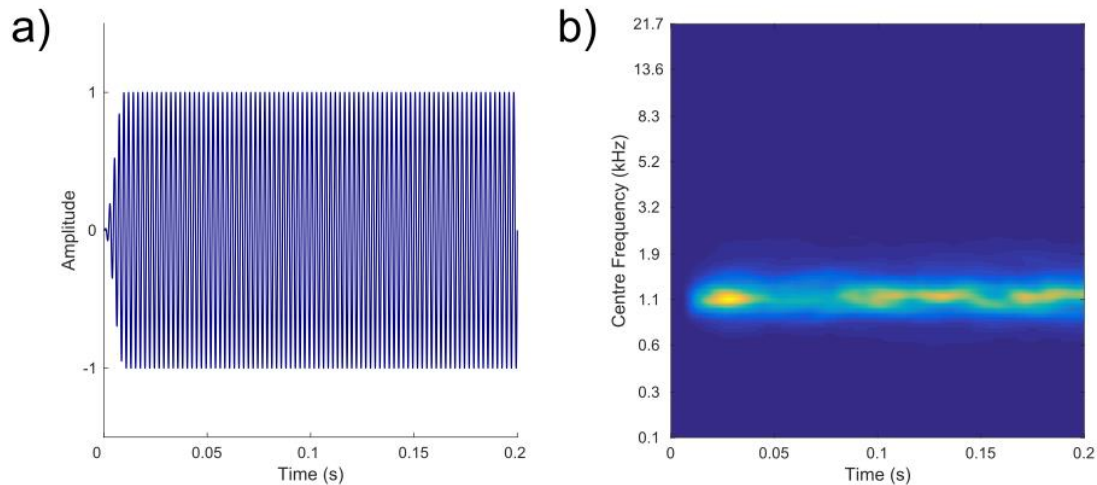
Participants sat at the chinrest and were presented with experimental stimuli via the headphones and display monitor. The experiment code was run from the control monitor.



**Figure 2.2 The RTbox**

The device (A) was connected to custom hand (B) and foot (C) response controllers via the TTL IN port. The LIGHT and SOUND ports were connected to a photodiode and audio jack respectively to record the onset of sensory stimuli prior to running experiments (these were not connected during experiments).

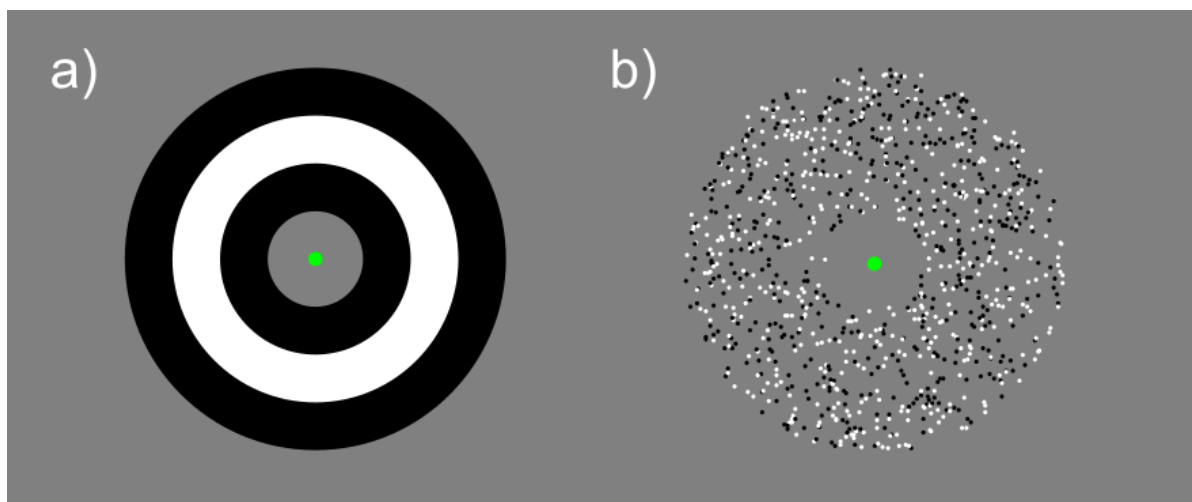




**Figure 2.3 Auditory signal examples**

**a)** Pure tone (simple) signal. The example shown here is 440 Hz tone.

**b)** Noise sound (complex) signal. The example shown here is a randomly-generated Gaussian noise with edge frequencies of 1.0/1.2 kHz. These complex signals were presented within continuous background noise with broader edge frequencies e.g. 0.5/2.4 kHz (not depicted). Both signals are shown played over 0.2 s with a 0.01 s ramp onset.

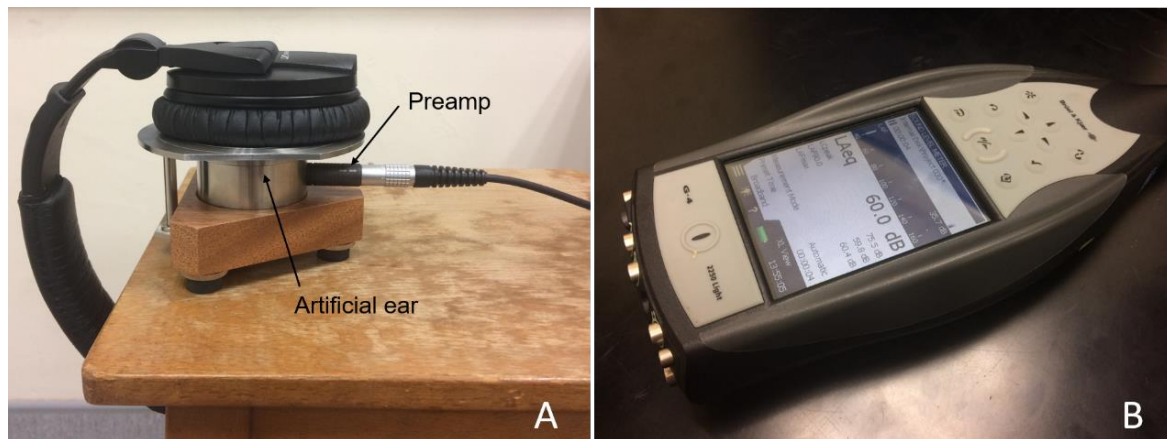


**Figure 2.4 Visual signal examples**

**a)** Concentric ring (simple) signal. On each trial, participants were presented with only the grey screen and the green fixation point. The signal was the sudden onset of the rings.

**b)** Random dot (complex) signals. On each trial, dots moved linearly with a randomised direction and speed. The signal was the coherent rotation of a certain percentage (e.g. 50%) of the dots.

Two kinds of audio-visual stimulation were constructed in this thesis, which I will refer to broadly as *simple* and *complex*. Simple auditory stimuli were pure tones (**Figure 2.3a**). The signal was the sudden onset of the tone. Pure tones were generated using the Psychtoolbox function *MakeBeep*. Complex auditory stimuli were filtered Gaussian noise (**Figure 2.3b**). The signal was a noise burst with narrower edge frequencies (e.g. 1.0/1.2 kHz) than the background noise (e.g. 0.5/2.4 kHz). Gaussian noise was generated using Butterworth filters with the MATLAB function *butter*.



**Figure 2.5 Calibrating volume of auditory stimuli**

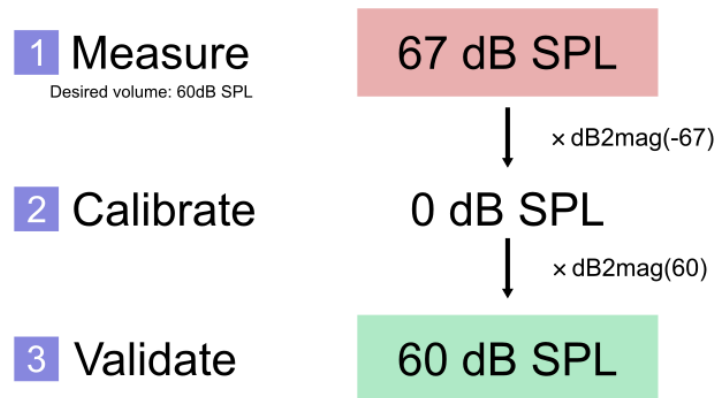
The headphones were secured to an artificial ear (A) and the stimulus sound was played continuously. The preamp was connected to a sound level meter (B) which measured the volume from the headphones (in dB SPL).

Simple visual stimuli were 3 concentric rings (**Figure 2.4a**). The rings were spaced in  $1^\circ$  increments and alternated black and white. The signal was the sudden onset of the rings. Complex visual stimuli were random dot kinematograms (**Figure 2.4b**). 1000 dots (each 1 pixel; either black or white) were distributed uniformly throughout the presentation area. Each dot had a lifetime of 0.1s, after which they were randomly replaced at a new location within the annulus. Each dot initially moved in a random linear direction with a speed of  $1$  (SD  $\pm 0.2$ )  $^\circ/s$ . The signal was the coherent rotation of a percentage (e.g. 50%) of the dots (random allocation). The mean rotation speed for signal dots was  $0.67$  (SD  $\pm 0.067$ ) rad/s.

### 2.1.3. Volume Setting & Calibration

Auditory stimulus volume was controlled in MATLAB by controlling the amplitude of the stimulus vector. The volume of each stimulus was individually calibrated with a Brüel and Kjær Type 2250 sound level meter equipped with a Type 4153 artificial ear (**Figure 2.5**). The artificial ear approximates the structure of the human ear, thus the pressure on the ear drum when the stimulus is played through the headphones can be estimated in dB SPL.

The procedure for calibration (**Figure 2.6**) follows 3 basic steps. As a first step, I measured the volume of the stimulus using the sound level meter (*recorded volume*), and compared this to the *desired volume* (as was programmed in MATLAB). Consider an example where I generated a stimulus to play at 60 dB SPL (*desired*), but measured the volume as 67 dB SPL (*recorded*). In this example, there is a 7 dB SPL discrepancy to correct for. As a second step, the stimulus was calibrated. To do this, I calculated a theoretical 0 dB SPL value from the recorded volume. This was done using the MATLAB function *db2mag*. This function converts decibels to magnitude such that



**Figure 2.6 Procedure for calibrating auditory stimulus volume**

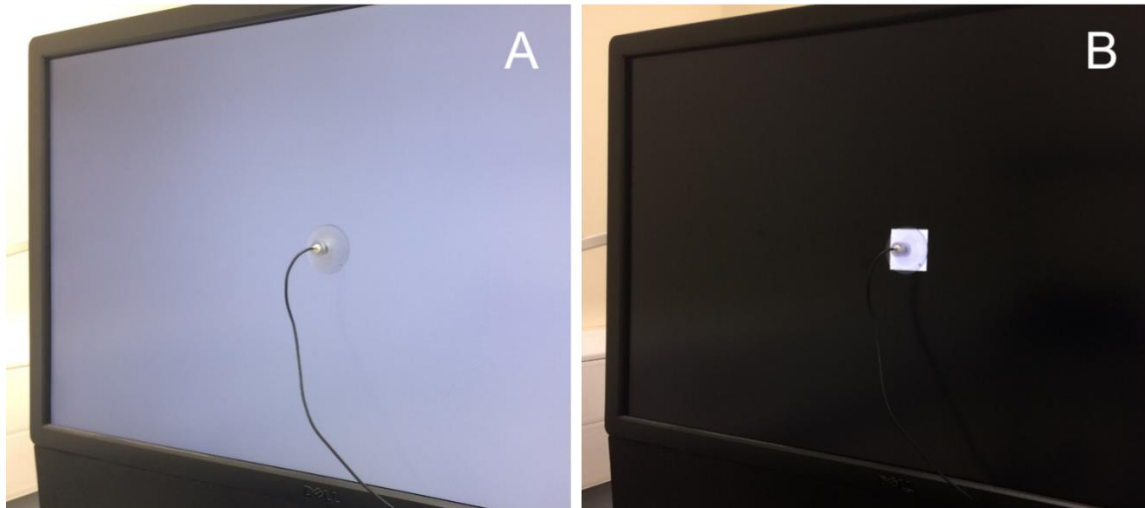
Stimulus vectors were manipulated in MATLAB using the function `db2mag`. First, the volume is measured with a sound level meter. Second, the stimulus is calibrated by finding a theoretical 0 dB SPL level. Third, the stimulus volume is validated by multiplying the 0 dB SPL value by the desired volume. If the recorded volume reliably matches the desired volume, the stimulus is calibrated.

$$y_{dB} = 20 \log_{10}(y), \quad (11)$$

where  $y_{dB}$  is the specified volume in dB SPL and  $y$  is the magnitude. To calculate the theoretical 0dB SPL value, I computed `db2mag(-recorded)`. In the above example, the recorded value is 67, so I would compute `db2mag(-67)`. To play the stimulus to the desired volume, I would multiply the theoretical 0 value by the desired volume: `0dB SPL × db2mag(desired)`. Considering the above example, the desired value is 60, so I would compute `0dB SPL × db2mag(60)`. Third, to validate the calibration, stimulus volume was recorded again; calibration was validated when the recorded volume reliably matched the desired volume.

#### 2.1.4. Calibrating Stimulus Timing

As one of the areas of focus of this thesis is RT variability, it is important that timing variability elsewhere i.e. coming from hardware and software, is measured and corrected for. Accurate calculation of RT is highly dependent on an accurate timestamp for signal onset. However, with uncalibrated equipment, onset timestamps can be inaccurate. This is because there is a delay between the command to present the stimulus (which provides the timestamp) and the physical onset of the stimulus (e.g. Woods et al., 2015). Onset delays affect both auditory and visual stimulus presentation, but are typically larger for visual stimuli. With reliable equipment, delays are consistent, and can simply be measured and subtracted from RTs (offline correction). With unreliable equipment, however, the delay can be unpredictable across trials, and each physical onset should also be recorded to provide an accurate timestamp (online correction).



**Figure 2.7 Measuring visual onset delay**

A photodiode is shown attached to the centre of the display monitor via suction cup. To measure visual onset delay, a test stimulus (white square) was flashed onscreen. The time difference between the display command timestamp (recorded in MATLAB) and the physical onset (recorded by the photodiode attached to the RTBox) was then calculated over many trials to give an estimate the average visual delay.

To first assess stimulus timing, onset delays for auditory and visual stimuli (i.e. the delay between the requested stimulus onset time and actual stimulus onset) were measured using the RTbox. A 3.5 mm audio cable (running from the experimental computer) was connected to the SOUND port of the RTbox to measure the onset delay for a test auditory stimulus (a pure tone). A photodiode was connected to the LIGHT port of the RTbox to measure the onset delay for a test visual stimulus (a white square on a black background; **Figure 2.7**). When measured, the auditory delay was 0.0042 ( $\pm 0.00029$ ) s and the visual delay was 0.0268 ( $\pm 0.00004$ ) s.

Next, onset delays were adjusted to ensure synchronous presentation of signals on redundant trials. As the auditory onset delay was smaller than the visual delay, an additional delay was added to auditory stimuli to create synchrony with visual stimuli. The additional delay was simply the difference between the audio and visual delay (i.e. 0.0226 s). To demonstrate the effectiveness of this approach, a synchronisation test was run using the RTbox. Here, the audio and visual test stimuli were presented simultaneously and the onset difference between the individual signals was recorded over 1000 trials. The mean difference between auditory and visual onset was 0.00011 s ( $\pm 0.000009$  s), indicating excellent synchrony. As the overall variability of stimulus onset was small ( $< 0.001$  s for both signals) an offline correction of RT was used (i.e. the synchronised stimulus onset delay was subtracted from all RTs).

## 2.2. Analytic Procedures

In this section, I describe the general procedures for processing and analysis of experimental data. All data handling and preparation was performed using MATLAB. All analyses, unless otherwise stated, were performed in IBM SPSS Statistics (versions 22 onwards).

### 2.2.1. Assessing General Performance

In the study of decision-making, there are known trade-offs between RT and accuracy (Forstmann et al., 2016; Gold & Shadlen, 2007). Unless performance is close to ceiling (100% accuracy), RT data needs to be analysed or adjusted in conjunction with accuracy data. In some cases, researchers are explicitly interested in such trade-offs, or in modelling both accuracy and RT together; however, this thesis is only concerned with RTs. For this reason, ceiling performance is often assumed in the measures used. This assumption must be assessed with each data set collected to ensure good quality.

Useful indicators of performance are the number of *false alarms* (i.e. responses when a signal was not present) and *misses* (failures to respond when a signal was present). A high percentage of false alarms, especially on trials with no signals (catch trials), suggests that RT distributions are contaminated by a large number of non-genuine anticipatory responses. A high percentage of misses, on the other hand, is more clearly indicative of poor accuracy. If a participant has a high miss rate, this is problematic as many of the CDF plotting procedures used assume an overall cumulative probability of 1 (100% accuracy). With a miss rate of 0.2 (20%), for example, the cumulative probability would in truth only sum to 0.8. If large discrepancies like this are simply ignored, it distorts the CDF of RTs, which can lead to errors in calculations. For these reasons, the percentage of false alarms and misses was always evaluated prior to analysis of RTs.

### 2.2.2. Outlier Correction

Human RT performance is never perfect, even in the absence of false alarms and misses. A brief lapse in attention, for instance, may produce an unusually slow RT. Similarly, an erroneous response may just happen to fall shortly after signal onset within the valid response window, producing an unusually fast RT. Broadly, these RTs are *outliers*, as they do not clearly reflect the genuine behaviour of interest. Prior to data analysis, we should conduct *outlier correction*, which removes these responses and conserves only genuine RTs.

The problem here is that there is no fool-proof way to detect outliers amongst valid responses. As such, there is no universally-applied outlier correction. In many experiments, simple cut-offs are used to trim the data based on reasonable assumptions about the nature of RTs. For example, one frequently applied cut-off excludes all RTs faster than 0.1 s, as this was estimated to be the minimum time to execute a manual response (Luce, 1986). However, there is no corollary cut-off for slow outliers, as RT is largely dependent on the task and the data set collected (Whelan, 2008). Another

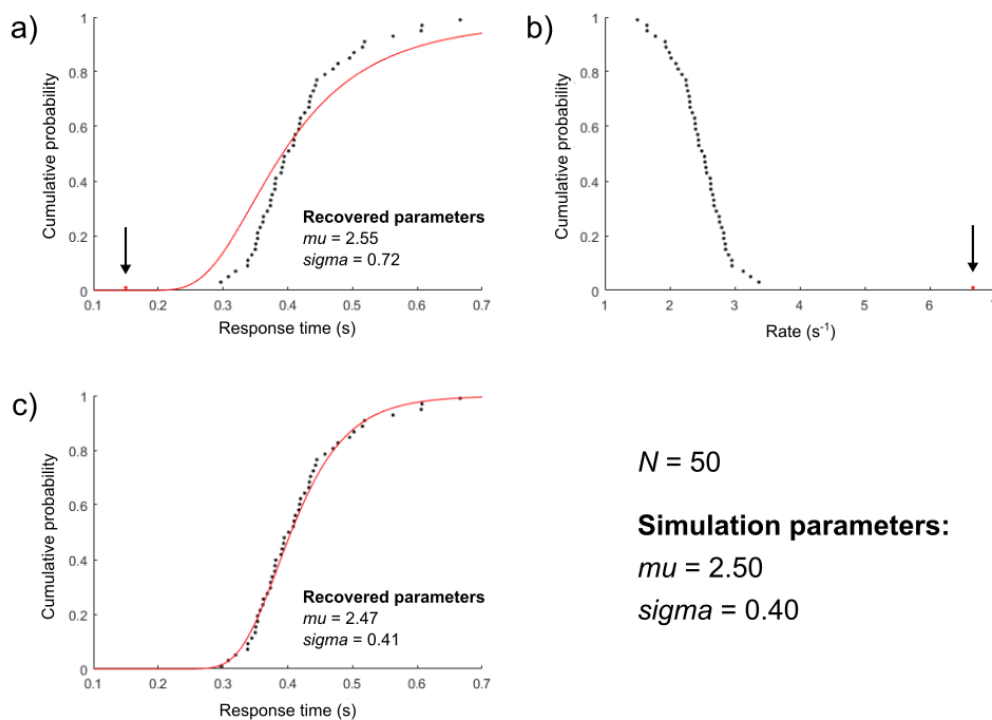
approach is to use central tendency based measures (e.g.  $\pm 3$  SD around the mean of the RT distribution). There are however some problems with such an approach. First, these methods assume normally-distributed data, but RTs are known to be right-skewed. Secondly, the standard deviation of RTs is itself strongly impacted by outliers, so the criterion may be ineffective at identifying RTs to remove.

In light of these considerations, the method of outlier correction used here was based on median and median absolute deviation (MAD), adapted from Leys, Ley, Klein, Bernard, and Licata (2013). Unlike using mean and SD, median and MAD are less impacted by outliers. In this thesis, outlier correction was performed on the  $1/RT$  transformed data. There were two main motivations for this. First, this addresses the issue of skew in outlier correction, as  $1/RT$  distributions are typically assumed to be normally distributed (Noorani & Carpenter, 2011, 2016). Second, model fitting was also performed on the  $1/RT$  space (see **Section 2.4.4**). Clear model deviations can be informative for identifying outliers which are not obvious from the data alone (Baayen & Milin, 2010). In the specific case of RTs, the fastest RTs in a distribution can sometimes reveal themselves as extreme outliers when transformed into  $1/RT$ . This issue can be identified by clear problems in model fitting (**Figure 2.8**).

To maintain an approach which retains as many valid RTs as possible, whilst also facilitating accurate modelling, we used a data-conservative criterion of  $\pm 1.4826 \times 3$  MADs around the median  $1/RT$ . Under the assumption of a normal distribution, this would be equivalent to  $\pm 3$  SDs with a corresponding exclusion of around 0.27% of RTs. As a general rule, we would expect that the number of excluded trials is not much larger than this ideal value if the data is good quality. Therefore, the percentage of outliers is always reported prior to RT analysis.

### 2.3. Creating RT Distributions

In order to compute benefits and interactions and analyse changes across participants, distributions of RT data need to be created. One way that this has been done previously is to create a *group distribution* of RTs, i.e. a single CDF representing the RTs of all participants for that condition. To estimate a group CDF, one technique is Vincent averaging (Ratcliff, 1979). Essentially, each individual participant's CDF is computed with a common number of cumulative probability points (e.g. 100). The RT values at each cumulative probability point are then averaged, creating a single CDF representing the whole group. The Vincent averaging approach has previously been used to create CDFs for applying the models used in this thesis (Otto et al., 2013; Otto & Mamassian, 2012). However, the present work is primarily interested in variability sources in RT data, which are linked to benefits. Averaging RT distributions introduces group variability into RT data, which is not the primary interest of this research. Further, recent simulation work by Otto (2019) has shown that even when data is



**Figure 2.8 The effect of fast outliers on modelling**

**a)** 50 RTs were first simulated. To approximate an anticipatory response falling in the valid response window, a single RT was randomly replaced with a fast outlier (according to the MAD criterion). The outlier is shown by the arrow. The presence of a fast outlier causes the model (red curve) to strongly deviate from the RT data.

**b)** The same data, but now plotted in the  $1/RT$  space where the model is fitted. Note that the outlier is even more pronounced following this transformation.

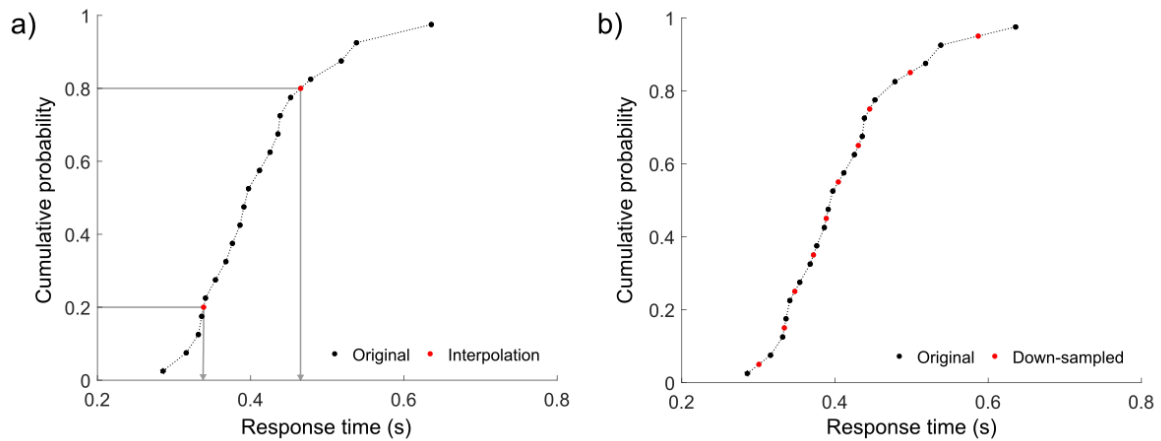
**c)** The same data after outlier correction by the data-conservative MAD criterion. The model-fit (and associated parameters) now closely resembles the simulated data (and associated parameters).

drawn from independent race models, the Vincent average process can create artificial violations of Miller’s bound. Therefore, in the present work, the calculation of empirical measures and modelling were always done on the *individual CDFs*.

### 2.3.1. Linear Interpolation of RTs

Linear interpolation was used to estimate continuous RT distributions based on empirical RT data from each participant. Such continuous distributions are needed later, for example in the computation of the simple race model (Raab, 1962). The advantage of using linear interpolation to acquire continuous distributions (as opposed to model fitting) is that it requires no distributional assumptions to be made about the underlying RTs. The linear interpolation process is shown in **Figure 2.9**.

Another specific use of linear interpolation is *down-sampling* of RT distributions. For instance, to measure the size of the multisensory benefit for each participant, we take the difference between redundant RT distribution and the minimum of the unisensory RT distributions for each cumulative



**Figure 2.9 Linear interpolation of RTs**

**a)** Estimating RTs. In linear interpolation, a straight line (red) is fitted between each point. Corresponding RT values (x-axis) can then be found for any cumulative probability value (y-axis) within the existing range. This example shows 20 simulated RTs with interpolation query points of 0.2 and 0.8 on the y-axis.

**b)** Down-sampling RT distributions. The original data (20 simulated RTs) are shown in black (as in panel **a**). Linear interpolation (shown by the solid red line) was used to down-sample this data. The down-sampled data (10 RTs) are shown in red.

probability point (see **Section 2.3.2**). This procedure requires equal numbers of trials in each condition to compare equivalent cumulative probability points. However, while these experiments were programmed to present equal numbers of trials in each condition, the number of remaining valid trials (following outlier correction) is often not equal. In such cases, linear interpolation can be used to create equal numbers of cumulative probability points. As a simple procedure in this thesis, I down-sampled all original RT distributions using linear interpolation to a smaller common sample size (50 data points per condition) where necessary.

### 2.3.2. Measures of Benefit (The RSE)

One of the overarching goals of this thesis is to establish a model framework which allows for a comprehensive understanding of multisensory processing in the RSP. The framework of this model is consistently applied in all experiments to understand changes in benefits and interactions (as well as modelled parameters for different factors). Here, the relevant calculations and procedures for each step are detailed.

#### 2.3.2.1. Empirical Benefit

In RSP research, multisensory benefit is typically assessed on central tendency measures e.g. the difference between the mean redundant RT and the smaller of the two means for unisensory RTs. This can be expressed as



$$benefit = \min(\bar{X}, \bar{Y}) - \overline{XY}, \quad (12)$$

where  $X$  and  $Y$  are the RT distributions for two unisensory signals and  $XY$  is the redundant RT distribution. RT measures based only on central tendency, however, neglect differences which only occur at the extremes of RT distributions (Noorani & Carpenter, 2011; Whelan, 2008).

A more precise estimate of benefit, therefore, would be gained by calculating benefit across the entire CDFs (**Figure 2.10**). For each cumulative probability point, the measure of benefit (as in **Equation 12**) can be applied. As the CDFs of RTs in unisensory conditions can cross over, the minimum term must be applied at all cumulative probability points. Note that this is equivalent to *Grice's bound* (see **Section 1.7.2.1**). *Grice's bound* is calculated as

$$Grice_i = \min(X_i, Y_i), \quad (13)$$

where  $i$  is an index of cumulative probability. To compute a measure of benefit based on the distribution (simulated from Otto et al., 2013 their Equation 3), we can essentially compute the difference between the redundant CDF and the faster of the two unisensory CDFs (*Grice*) at all cumulative probability points, and then average these values. This can be formalised as

$$empirical\ benefit = \frac{\sum Grice_i - XY_i}{N}, \quad (14)$$

where  $N$  is the total number of cumulative probability points.

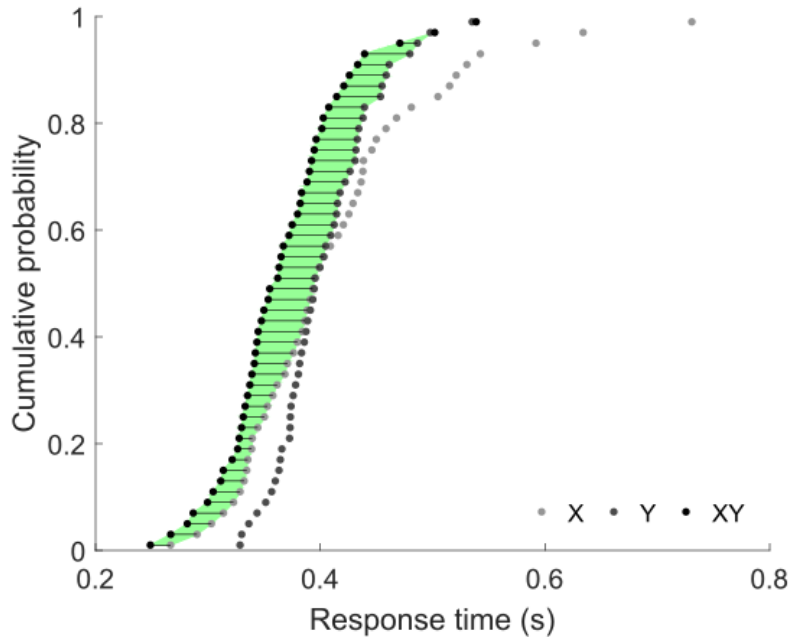
#### 2.3.2.2. Predicted Benefit

A prediction of the multisensory benefit can be made by calculating the simple race model (**Equation 5**). The CDF for the simple race model was computed from the unisensory CDFs ( $X$ ,  $Y$ ) using linear interpolation to obtain cumulative probability values for all unique timepoints of  $t$  (see **Section 2.3.1**):

$$P_{Raab}(t) = P_X(t) + P_Y(t) - P_X(t) \times P_Y(t), \quad (15)$$

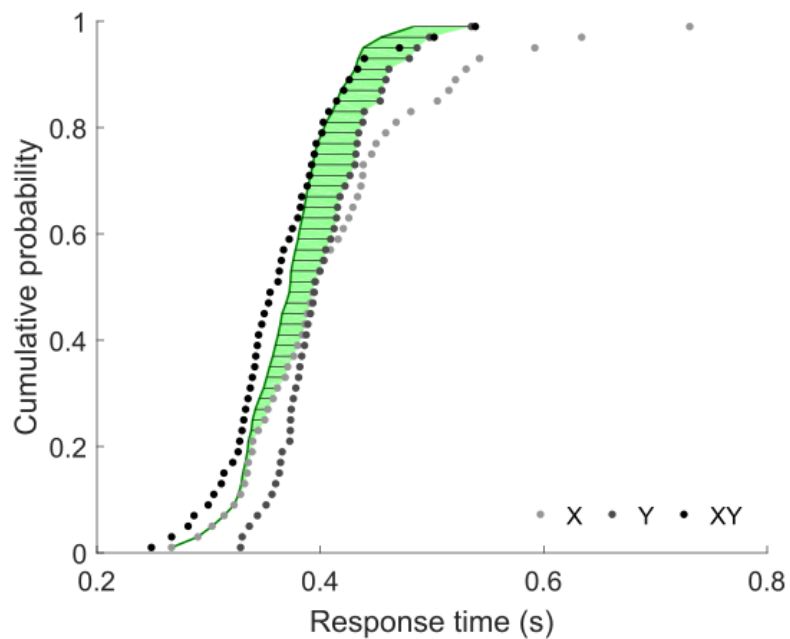
where  $P$  is cumulative probability and  $t$  is the associated timepoint. The formula for predicted benefit is the same as for empirical benefit (**Equation 14**), but each cumulative probability point of the empirical redundant CDF ( $XY_i$ ) is substituted for the CDF generated by the simple race model ( $Raab$ ):

$$predicted\ benefit = \frac{\sum Grice_i - Raab_i}{N}. \quad (16)$$



**Figure 2.10 Measuring multisensory benefit**

The data shows 50 simulated RTs in each of 3 conditions (X, Y, XY). The green area shows the overall benefit. The solid lines indicate the RT differences (s) between the redundant (XY) and minimum unisensory (X, Y) RT distributions at each cumulative probability point. These are averaged to measure the overall multisensory benefit.



**Figure 2.11 Predicting multisensory benefit**

The data shown is identical to **Figure 2.10**. The green area indicates the overall predicted benefit. The solid lines indicate the RT differences (s) between the simple race model (green line) and minimum unisensory (X, Y) RT distributions at each cumulative probability point. These are averaged to measure the overall predicted multisensory benefit.

### 2.3.2.3. Validation of Procedures

To obtain empirical CDFs with equivalent cumulative probability points, all RT distributions are downsampled to a common  $N$  (50 data points; **Section 2.3.1**). To assess whether this procedure provides an unbiased measure of the true multisensory benefit size, 100 RTs were simulated (see **Section 2.5.3**) for each condition of a typical redundant signal experiment (i.e. unisensory conditions  $X$  and  $Y$ , and redundant  $XY$ ). The calculated multisensory benefit (**Equation 14**) was compared for original and down-sampled RTs. This overall sampling procedure was repeated 1000 times (full parameter details are reported in the appendices in **Section 9.3**). A paired-samples  $t$ -test found a significant difference between the original ( $0.020 \pm 0.0005$  s) and down-sampled ( $0.020 \pm 0.0005$  s) benefit,  $t(999)=13.393$ ,  $p<0.001$ . However, the down-sampled benefit was on average only  $0.00006$  s ( $\pm 0.0001$ ) larger than original benefit. Given that this bias is extremely small, therefore, it is unlikely to affect the measure of multisensory benefit in practice. In addition, the correlation between original and down-sampled benefit closely follows identity (see **Figure 2.12**).

### 2.3.3. Measures of Interactions

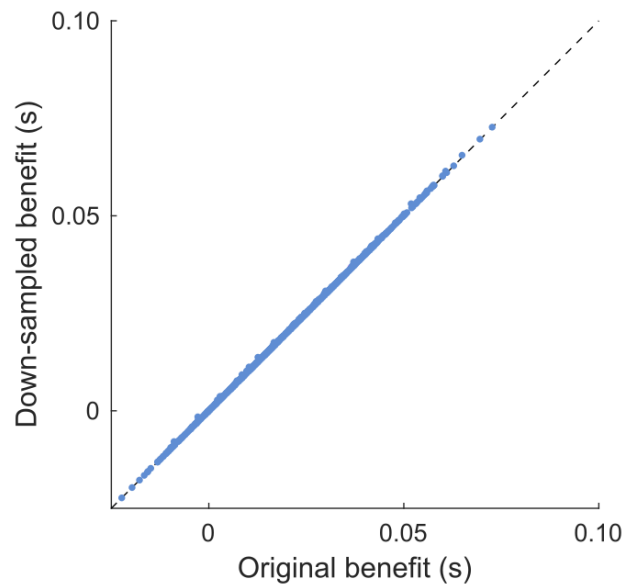
#### 2.3.3.1. History Effects

Sequential dependencies in RT, i.e. the influence of one trial on another, violate the simple race model assumption of statistical independence in RTs (that each observed RT is *not* influenced by the previous RT). Sequential dependencies are particularly pronounced between unisensory trials. A simple measure of sequential dependency is to compute the RT cost for unisensory signals (Miller, 1982; Spence, Nicholls, & Driver, 2001). This measure will be referred to as the *history effect* (Otto & Mamassian, 2017). First, all valid unisensory RTs whereby the previous trial ( $N-1$ ) was also a unisensory signal are indexed. These are then sorted into ‘repetitions’ and ‘switches’. For example, with unisensory signals  $X$  and  $Y$ , repetitions would be trials where the  $N-1 - N$  sequence was  $X-X$  or  $Y-Y$ . On the other hand, switches would be sequences  $X-Y$  or  $Y-X$ . The history effect can then be calculated as follows:

$$history\ effect = \overline{RT}_{Switch} - \overline{RT}_{Repetition}. \quad (17)$$

This history effect was calculated for each unisensory condition ( $X$ ,  $Y$ ) separately and then averaged into a single measure.

History effects are relevant because they contribute to the overall variability of unisensory RTs (which is linked to the size of multisensory benefit). To assess how much history effects contribute to the overall variability of unisensory RTs, therefore, we can also compute a measure which normalises the history effect by the variance of unisensory RTs. I refer to this measure as the *history index*. History index was calculated as



**Figure 2.12 Validating the down-sampling procedure for benefit calculation**

Multisensory benefit calculated from the original ( $N=100$ ) and down-sampled ( $N=50$ ) simulated RTs. The dashed line indicates identity. Each point shows one of 1000 simulations.

$$history\ index = \frac{history\ effect^2}{var(RT)}. \quad (18)$$

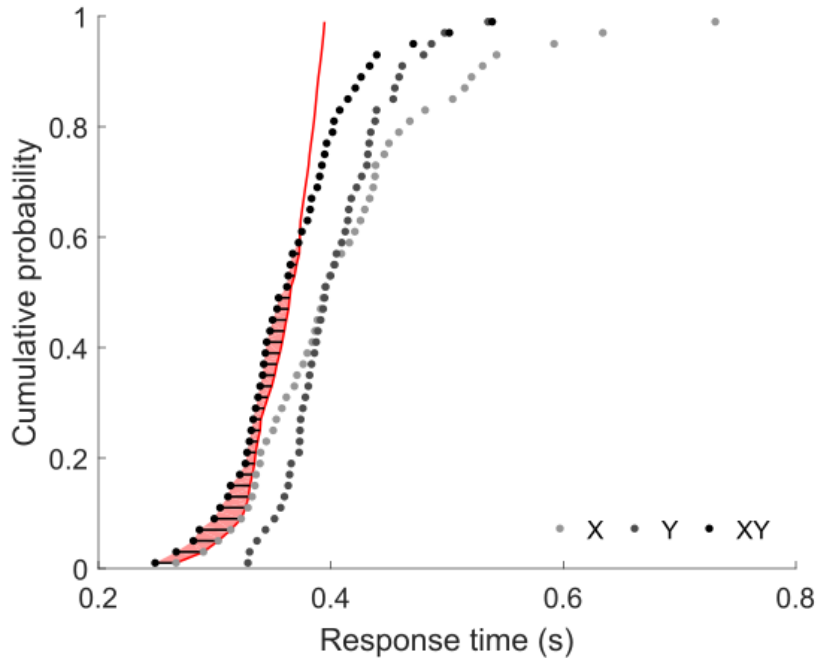
As for history effects, history index was calculated for each unisensory RT distribution separately and averaged into a single measure.

### 2.3.3.2. Violations of Miller's Bound

Violations of Miller's bound, following a race model framework, indicate violations of the context variance assumption (i.e. that the processing of each component signal is not affected by the presence or absence of another signal). This violation is typically calculated on the whole distribution of RTs (Miller, 1982). Miller's bound is defined by the equation:

$$P_{Miller}(t) = \min[P_X(t) + P_Y(t), 1]. \quad (19)$$

There are multiple methods for testing violations of Miller's bound. Originally, Miller implemented a method in which the redundant curve and Miller's bound are determined on the level of the individual participant. At predetermined comparison quantiles, the values of the redundant RT and Miller's bound are compared by  $t$ -test to see if there is a significant difference (violation). An alternative method, introduced by Colonius and Diederich (2006), uses integration to calculate the area between the redundant CDF and Miller's bound. Following a similar procedure for the calculation of benefit



**Figure 2.13 Quantifying violations of Miller's bound**

The data shown is identical to **Figure 1.13**. The red area shows the overall violation. The solid lines indicate the RT differences (s) between Miller's bound (red line) and redundant (XY) RT distributions at each cumulative probability point. These are averaged to estimate the overall violation area.

(Otto et al., 2013, their Equation 3), this reduces the size of the violation across the distribution to a single value. This can be expressed, following Otto (2019), as:

$$violation_{XY} = \int_T \max[P_{XY}(t) - P_{Miller}(t), 0] dt. \quad (20)$$

A clear appeal for using this latter approach is that it would place measures of benefit and violation in the same reference frame (i.e. measured in seconds). This would allow for direct comparison between the two, and allow a clearer investigation of how they relate.

Following a similar calculation for benefits (**Equation 14**), therefore, the CDF for Miller's bound (*Miller*) was computed from the unisensory CDFs (i.e. X and Y) using linear interpolation to obtain cumulative probability values for all unique timepoints of  $t$  (see **Section 2.3.1**). The difference between *Miller* and the empirical redundant CDF was then calculated at each cumulative probability point, and averaged over all comparison points (**Figure 2.13**). The formula for violation is:

$$violation = \frac{\sum \max(Miller_i - XY_i, 0)}{N}. \quad (21)$$

As we are only interested in comparison points where the redundant CDF (XY) exceeds *Miller*, we implement a maximum rule; this ensures that if the empirical curve falls behind Miller's bound, a value

of 0 s (no violation) is assigned to the comparison point. To obtain equal cumulative probability points, the same down-sampling procedure used for computing benefits was applied to the *Miller* CDF (see **Section 2.3.2.3**).

## 2.4. Model Fitting Procedures

In this section, I describe all techniques used to fit models to the RT data collected in experiments.

### 2.4.1. Parameter Search Methods

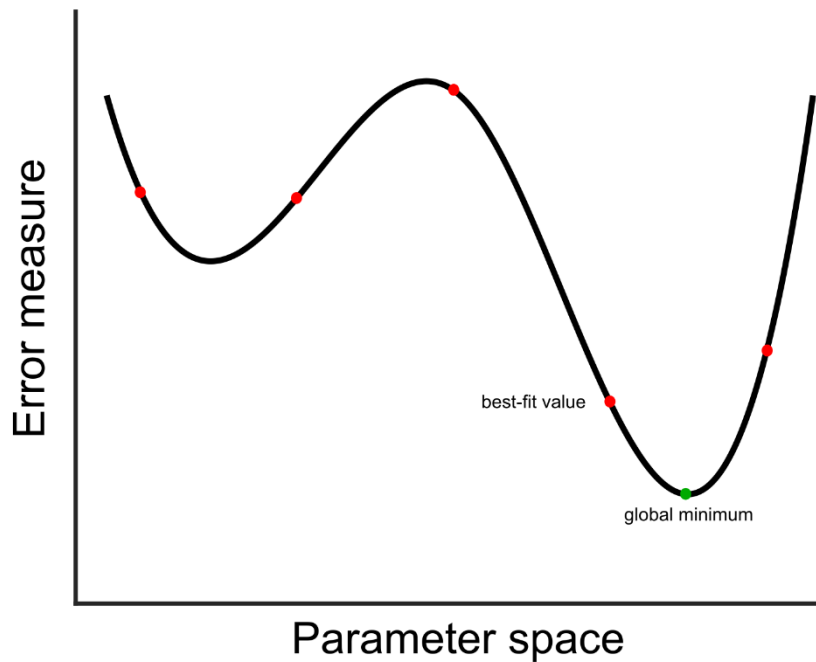
Modelling of behavioural data involves estimating the best fitting values for the available free parameters of the specified model. For instance in the case of the LATER model (see **Section 2.4.4.1**) this would involve finding the best *mu* and *sigma* parameter values which describe the 1/RT distribution. The basis for parameter estimation is *mathematical optimisation*, whereby the best values are found according to a specified measure of discrepancy between the empirical data distribution and the model-fit distribution (i.e. some *error measure*). In model fitting, the mathematical optimisation of parameter estimates is treated as a *minimisation problem*, i.e. finding parameter values which minimise the discrepancy captured by the error measure.

It is of course impossible to test all possible combinations of parameter values; therefore, search methods to sensibly limit the number of combinations tested are used. The goal of these search methods is to provide fit values which reflect or approximate the true *global minimum* of the error measure as closely as possible.

#### 2.4.1.1. Grid Search

One search method is grid search (**Figure 2.14**), which approaches the optimisation problem in three steps. First, a grid is constructed with multiple parameter combinations (e.g. *mu* and *sigma* combinations when fitting a LATER model). Second, the optimisation measure is computed for each combination of parameters in the grid. Third, the best-fitting parameter values, i.e. the combination of parameter values that minimised the error measure the most, are selected as the model parameters.

Grid search can be useful to provide a coarse-detail view of how the error measure changes over the whole parameter space. However, the model fit values returned are highly dependent on the resolution of the grid used. More precise estimates (gained by increasing the number of sample points for each parameter) become computationally expensive. This problem is made even greater the more free parameters are included in the model. Consider, for example, a simple two-parameter model. We can apply a grid search with 10 test values per parameter. To complete this grid search for just one experimental condition, the error measure for 100 parameter combinations must be evaluated. To improve our parameter estimation, we can increase the resolution of the search to 100 test values per parameter. This now means the error measure must be evaluated for 10000 parameter combinations.



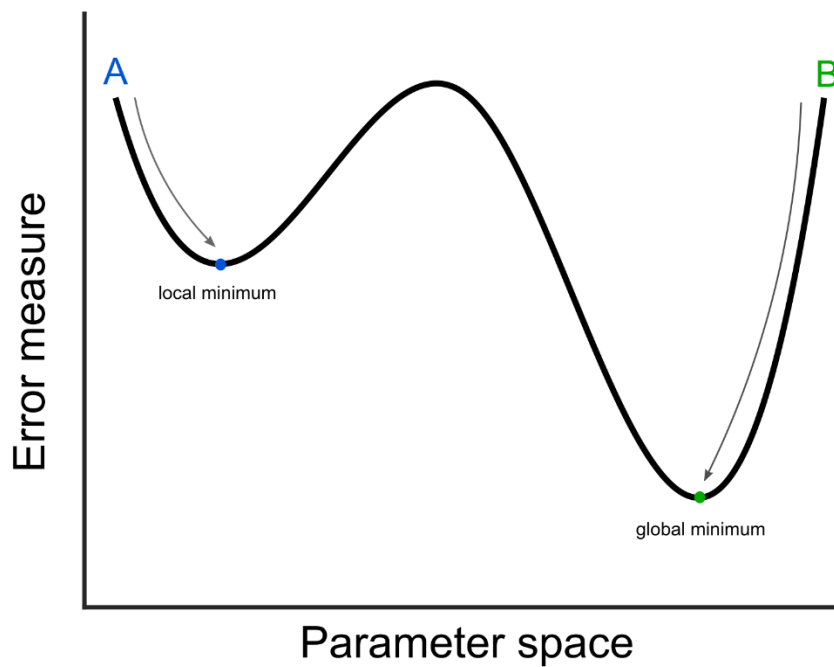
**Figure 2.14 The grid search method**

The figure shows a simple non-linear parameter space. The red dots represent 5 equidistant points at which the error measure is evaluated. The 'best-fit' point shows the grid search point at which the error measure was smallest. This is only somewhat close to the true global minimum (green dot). By further increasing the resolution of the grid, the best-fit point will more closely approximate the true global minimum, but at the expense of more computation time. This figure was inspired by Lacouture and Cousineau (2008; their Figure 3).

Adding an additional parameter to the model would then further increase the number of evaluations to 1000000. Given that computational models typically have more than three parameters, and experiments can have large numbers of participants and experimental factors, it can be incredibly time-consuming to use grid search for fitting models. Further, even by evaluating the error measure over many parameter combinations, the best-fitting values may still be far from the true global minimum.

#### 2.4.1.2. Optimisation Algorithms

An alternative search method is the use of an optimisation algorithm (**Figure 2.15**). Optimisation algorithms incrementally alter parameter values from given *start values* (an initial value chosen by the programmer). This is done to a point at which the error measure appears to be minimised. The main optimisation algorithm used in this work, for instance, is the Nelder-Mead SIMPLEX method (Nelder & Mead, 1965). The basic idea of this method is to construct a test space within the overall parameter space. Next, the algorithm adjusts the point of the test space which is least optimised on each iteration. Over iterations, the test area shrinks until the error appears to be minimised. This method is known to be effective for fitting RT distributions and can be implemented in MATLAB using



**Figure 2.15 The optimisation algorithm method**

The figure shows the same simple, non-linear parameter space as **Figure 2.14**. A and B represent two possible start values for the algorithm's search, which are then incrementally altered. If the algorithm starts from point A, the returned parameter value (blue dot) may be the result of a local minimum in the error measure. By contrast, with a different start point (B), the returned parameter value (green dot) is the result of the global minimum in the error measure. By using multiple start points, the possibility of falling into local minima is decreased. This figure was inspired by Lacouture and Cousineau (2008; their Figure 3).

*fminsearch* (Lacouture & Cousineau, 2008). Additionally, the MATLAB function *fmincon* can be used to implement optimisation algorithms under certain parameter constraints. This is useful when estimates are bounded (such as correlation parameters, which must be between -1 and 1). Though such optimisation algorithms are computationally less expensive, the non-linear nature of the parameter space and the incremental nature of the algorithms means the returned parameters values can depend on the choice of start values. In some cases, parameter values which would further reduce the error measure are not returned, as the optimisation algorithm falls into a *local minimum* (**Figure 2.15**). For this reason, multiple start points for parameter values are often selected, and convergence on the same parameter values following each starting point suggests the global minimum has been found. A straightforward way to select start values which are very likely to be close to the global minimum is to first conduct a coarse grid search (see **Section 2.4.1.1**) and select the combinations which minimise the error measure the most as starting points for optimisation algorithms.



## 2.4.2. Error Measures

In addition to the choice of search method, the returned parameters will depend upon the choice of error measure. Here I review two error measures, which are relevant to the models fitted in this thesis.

### 2.4.2.1. Root Mean Square Error

Root Mean Square Error (RMSE) uses residuals (i.e. the difference between empirical data and model fit data at each point) as the error measure. The RMSE is calculated

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (model_i - data_i)^2}{N}} \quad (22)$$

where  $i$  indicates individual data points and  $N$  is the total number of data points for the *model* and the empirical *data*. This provides a positive number which approaches 0 (indicating perfect fit to the data). RMSE is a straightforward method of parameter estimation, which does not necessarily require that the underlying structure of the distribution is known.

RMSE has been previously used to fit the models applied in this thesis (Otto et al., 2013; Otto & Mamassian, 2012). RMSE, however, does not directly lend itself to model selection methods based on likelihood (see **Section 2.4.3**). To facilitate model selection in this thesis, therefore, all model fitting was adapted from this previous approach to use a different error measure: Maximum Likelihood Estimation.

### 2.4.2.2. Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE) uses the log-likelihood of a model as the criterion for optimisation. Likelihood is the probability of obtaining data given the specified model parameters. In contrast to RMSE therefore, which attempts to find the model parameters which generate the best fit to the data, MLE attempts to find model parameters that are the most *likely* to give rise to data (Myung, 2003; Van Zandt & Townsend, 2013). Likelihood is calculated

$$Likelihood(\theta) = \prod_{i=1}^N f_i(data_i|\theta), \quad (23)$$

where  $i$  indicates individual data points,  $N$  is the total number of data points, and  $\theta$  indicates the model parameters. The overall model likelihood gives a value between 0 and  $\infty$ , with larger values indicating a higher likelihood that the model generated the data (i.e. better fit). Individual likelihood values however can be very small. When many such small values are multiplied, this can lead to computer underflow, i.e. the number requires more decimal places than the computer is capable of representing in its memory (Hélie, 2006; Lacouture & Cousineau, 2008). The solution is typically to log-transform the likelihood. As the logarithm of likelihood is a strictly increasing function, the

maximum likelihood value directly corresponds to the maximum log-likelihood value (e.g. Van Zandt & Townsend, 2013). Log-likelihood is calculated as

$$ll(\theta) = \log \prod_{i=1}^N f_i(\text{data}_i|\theta), \quad (24)$$

where  $ll$  is *log*-likelihood. The logarithm of a product is equal to the sum of the logarithms of the individual factors. Therefore instead of using the product, the sum can be used as

$$ll(\theta) = \sum_{i=1}^N \log f_i(\text{data}_i|\theta). \quad (25)$$

MLE in practice attempts to find parameter values which maximise the log-likelihood (or, when expressed as a minimisation problem, minimises the negative log-likelihood). Because MLE uses likelihood, it provides a basis for engaging in model selection methods (see **Section 2.4.3**). MLE and the optimisation algorithm were implemented in MATLAB using the function *mle*.

### 2.4.3. Model Selection

Many different candidate models can offer an explanation for the same data. Thus, some method of selecting the model which most appropriately summarises the data is necessary. For models with the same number of parameters, the best model may simply be the one which minimises the error measure. When comparing models with varying numbers of parameters, however, this is not as straightforward. Additional parameters always provide an equal or improved fit to the dataset. However, the addition of unnecessary additional parameters limits the ability of the model to generalise to novel datasets (*overfitting*). Model selection methods are ways of quantifying whether the addition of parameters is necessary to explain the data, with the goal of selecting the model with the smallest number of necessary parameters. This is the principle of *model parsimony* (Burnham & Anderson, 2002; Palminteri, Wyart, & Koechlin, 2017). In this section, different model selection methods are examined.

#### 2.4.3.1. Likelihood Ratio Test (LRT)

The likelihood ratio test (LRT) allows for direct selection between two competing *nested* models. A nested model is a simple (i.e. more constrained) version of a more general model (Hélie, 2006). For example, we can consider a hypothetical three parameter model. The full model, which includes all three free parameters, would be the *general* model. A nested version of this model can be created by constraining the general model to just two-parameters (e.g. by fixing one parameter to 0). This model is the *simple* model. The LRT is computed by comparing log-likelihood values for each version of the nested model. These log-likelihood values are then subjected to the test

$$LRT = 2 \times (-ll(\text{simple}) + ll(\text{general})) \quad (26)$$

whereby  $ll$  is log-likelihood (**Equation 25**) and *simple* and *general* are the simple and general versions of the same nested model. The null hypothesis assumes that the simple model is correct. Lower values of the LRT statistic suggest that the null hypothesis is true, whereas larger values favour the general model. To assess statistical significance of the result, the LRT statistic is submitted to a *chi*-square test with degrees of freedom equal to the difference in the number of parameters between the simple and general model. A significant result indicates that the general model is favoured.

While the LRT is able to compare nested models (e.g. two race models), it cannot compare models with different formulations or architectures (e.g. a race model and a pooling model). Even in the case of nested models, only two models can be compared at one time, which requires large numbers of comparisons. Additional measures, like AIC and BIC, can be used to compare multiple models and select an overall most-desirable model, regardless of whether they are nested or non-nested.

#### 2.4.3.2. Akaike's Information Criterion (AIC)

Similar to the likelihood ratio test, Akaike (1974) developed a test which compares likelihood for non-nested models. This also penalises by the addition of free parameters to avoid the problem of overfitting. AIC is calculated as follows:

$$AIC = 2[-ll(\text{model})] + 2d, \quad (27)$$

whereby  $d$  represents the number of free parameters. In a comparison of two or more models, the model with the smaller AIC value is selected as the more desirable model.

#### 2.4.3.3. Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC), introduced by G. E. Schwarz (1978), is very closely linked to the AIC. However, a stricter approach is followed which penalises not only by the addition of free parameters but also by the number of observations. BIC follows a very similar equation to AIC (**Equation 27**) and is calculated as follows:

$$BIC = 2[-ll(\text{model})] + \log(n)d, \quad (28)$$

whereby  $n$  is the number of observations. Both AIC and BIC were implemented in MATLAB using the function *aicbic*.

#### 2.4.4. Models of Response Time Distributions

In this section, the specific models fitted to RT distributions are detailed. Models of multisensory RT typically require continuous unisensory RT distributions. In contrast to an estimation with Linear Interpolation (see **Section 2.3.1**), it is necessary here to fit a model. To model unisensory RTs, the LATER model (Carpenter & Williams, 1995) was used. To model multisensory RTs, a race model was used which is similar to that introduced by Raab (1962), but with additional parameters to account for interactions (Otto & Mamassian, 2012).

##### 2.4.4.1. Unisensory RTs: The LATER model

The Linear Approach to Threshold with Ergodic Rate (LATER; Carpenter & Williams, 1995) model is a straightforward, two-parameter model of RT distributions. This model takes advantage of the observation that, while RTs themselves are not normally distributed, the *reciprocal* of RT ( $1/RT$ ) is normally distributed (Noorani & Carpenter, 2016). Such a distribution is sometimes called a *recinormal* distribution. In practice therefore, the RT distribution can be described by two parameters: the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the recinormal  $1/RT$  distribution.

Relating to the unisensory model framework (see **Section 1.2.3**), the  $1/RT$  values correspond to the *rate* of evidence accumulation on a given trial. Small rate values, therefore, correspond to long RTs (and vice versa). For this reason,  $1/RT$  values are also referred to as rate (or  $r$ ) values. As implied by the name, the LATER model simplifies what is assumed to be noisy accumulation to threshold to a linear accumulation to threshold (**Figure 1.7**). In behavioural work, however, this is of little concern as we are not particularly concerned with modelling the precise accumulation process. The  $\mu$  parameter describes the mean rate of rise, whereas  $\sigma$  describes the variability in the rate from trial to trial.

Neither a starting point for evidence accumulation nor a threshold level need to be specified to fit the model. This is because these two parameters can be simplified into one parameter (i.e. distance between start point and threshold). Even including this parameter is unnecessary, as it causes the model to become over-specified (i.e. multiple parameter combinations equally minimise the error measure). The threshold therefore is implicitly set as 1. To implement the LATER model, the MATLAB function *normfit* was used to fit a normal distribution using a Minimum Variance Unbiased Estimator (MVUE). In addition to providing a continuous estimation of the RT distribution, the LATER model was also used as the primary basis for *simulating* RTs (see **Section 2.5.3**).

##### 2.4.4.2. Multisensory RTs: The Context Variant Race Model

The model that was fitted to the redundant CDF to explain the RSE was the context variant race model (Otto & Mamassian, 2012). This model follows the same basic architecture of Raab's (1962) simple race model. However, two parameters are added to this model to explain empirical interactions which

violate the simple race model's assumptions. To account for violations of *statistical independence* (as evidenced by trial history effects) a correlation parameter  $\rho$  is modelled between unisensory distributions. To account for violations of *context invariance* (as evidenced by violations of Miller's bound), a noise parameter  $\eta$  is modelled on the redundant distribution. This noise parameter adds to the variability of unisensory distributions when modelling the redundant distribution. This model is described formally by Otto and Mamassian (2012; Supplementary Material). The complete equations are also reproduced here in **Section 9.4** of the appendices.

To fit the context variant race model to multisensory RTs, two LATER units are fit to the unisensory rate distributions. To fit the redundant rate condition according to a race model, the maximum rate distribution (corresponding to the minimum RT distribution after the reciprocal transform) must be computed from these units. The formula for the maximum distribution for two normal distributions is known (Nadarajah & Kotz, 2008), thus the redundant rate distribution can be directly modelled with a given correlation  $\rho$  between these two unisensory LATER units. To allow for noise  $\eta$ , the *sigma* parameters of the LATER units are adjusted by an additional constant in the fitting of the redundant rate distribution. Model fitting was done using MLE with the MATLAB function *mle*.

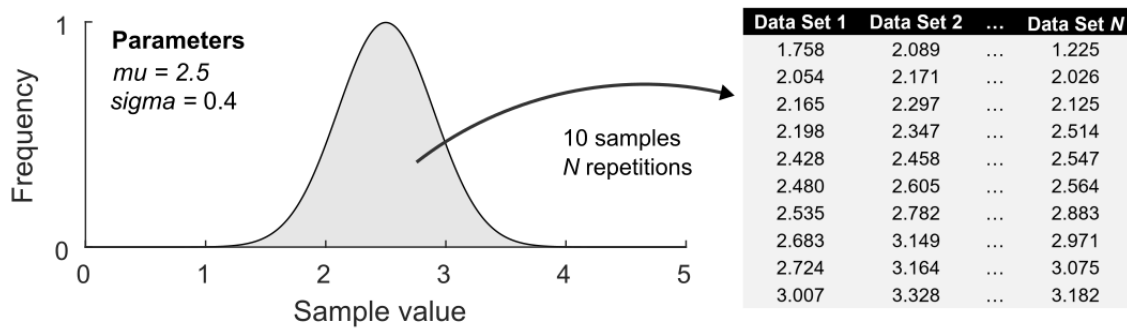
The context variant race model could be fit in two ways. One option, which has been used previously (Otto et al., 2013; Otto & Mamassian, 2012), is a two-step fitting procedure. According to this method, LATER units ( $\mu_x, \mu_y, \sigma_x, \sigma_y$ ) are first fit to the unisensory distributions. The best-fitting values are then passed to a second fitting process with two free parameters ( $\rho, \eta$ ). An alternative option is to fit all 6 parameters ( $\mu_x, \mu_y, \sigma_x, \sigma_y, \rho, \eta$ ) in a single step. The advantage of this method is that a single likelihood measure is returned, which opens the model up to evaluation and selection techniques (see **Section 2.4.3**).

## 2.5. Simulation Procedures

In this section, I describe techniques used to conduct simulation work. Novel insights, predictions and tests can be developed by simulating data sets and subjecting them to analysis. This can also help indicate the validity of certain analytic approaches before applying it to real data. Further, when real data has been collected, additional resampling methods can be applied to give estimates of the reliability of certain results. Here I describe two sampling methods used to simulate data, and the specific techniques used to generate simulated RTs in this thesis.

### 2.5.1. Monte Carlo Sampling

Monte Carlo sampling, broadly, is a method of randomly-generating many data points (e.g. Kroese, Brereton, Taimre, & Botev, 2014). This might include, for instance, generating data sets from a defined probability distribution or model (e.g. a normal distribution, as in **Figure 2.16**). One particularly useful



**Figure 2.16 Monte Carlo simulation**

A simple Monte Carlo simulation from a normal distribution. First, the distribution is defined (with the  $\mu$  and  $\sigma$  parameter values shown). Secondly, the desired number of samples (e.g. 10 per data set) is taken. These data are shown in the columns, ranked from smallest to largest values. Third, the sampling process is repeated until the desired number of data sets  $N$  is generated. These data were sampled using MATLAB's *normrnd* function.

application of Monte Carlo sampling for the present modelling work is the simulation of RTs (see **Section 2.5.3**). Provided a model is somehow defined (for example, the LATER model for RTs; see **Section 2.4.4.1**) then large numbers of simulated datasets can be generated from it (Burnham & Anderson, 2002). This is particularly useful if the model simulation parameters can be estimated from real subject data, as it can give the experimenter a clear prediction for how real data would change according to certain manipulations. Additionally, certain test statistics can be generated from Monte Carlo data and compared to real results. This is known as a Monte Carlo test (e.g. Manly, 2007).

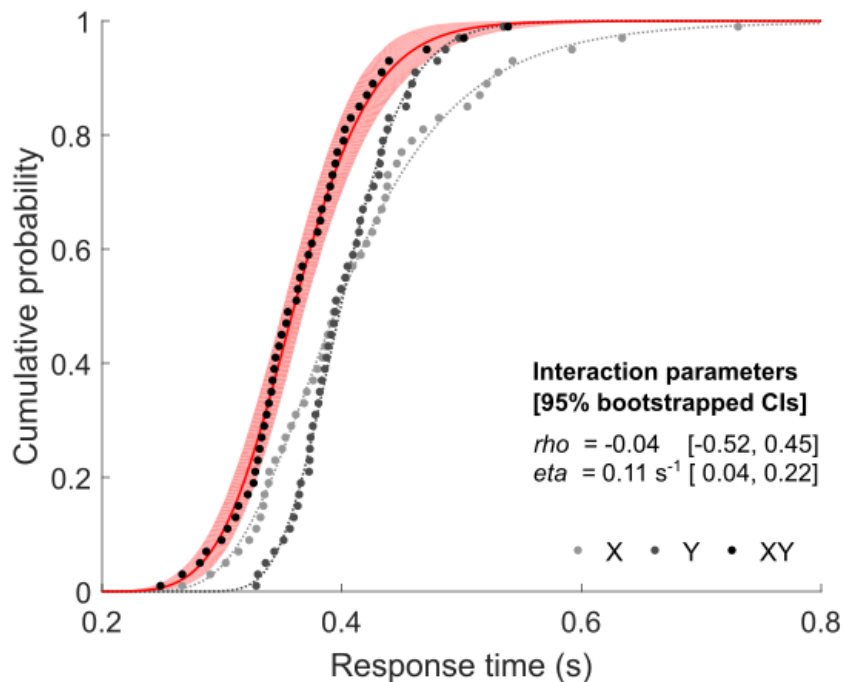
### 2.5.2. Bootstrapping

Bootstrapping is a method to generate new, replicate datasets from an existing data set (Efron & Tibshirani, 1993). This method is useful for estimating the error of certain measurements. To perform bootstrapping, data points are resampled (with replacement) from the existing data set to create a dataset of the equivalent size. Each sample point is referred to a bootstrap *sample*, whereas each new data set is a bootstrap *replicate*. This process is repeated a large number of times (typically 1000) to create many replicates. From each bootstrap replicate, a measure of interest can be calculated. For instance, if the data were RTs, a model could be fitted to each replicate to give 1000 bootstrap *estimates* of the model parameters. By ranking bootstrap estimates from lowest to highest, 95% confidence intervals can be calculated simply by taking the 25<sup>th</sup> and 975<sup>th</sup> ranking estimates (Efron & Tibshirani, 1993). Bootstrapping was primarily used to visually inspect individual model fit values (**Figure 2.18**), which are not analysed here; by plotting the bootstrapped confidence intervals for each cumulative probability point, however, this gave an indication of the reliability of the model fits.

Data		Rep 1	Rep 2	...	Rep N
1.758	10 samples (with replacement)	1.758	1.758	...	2.054
2.054		2.054	1.758	...	2.198
2.165		2.054	2.198	...	2.480
2.198		2.165	2.198	...	2.480
2.428		2.165	2.428	...	2.480
2.480		2.165	2.535	...	2.480
2.535		2.428	2.535	...	2.683
2.683		2.535	2.683	...	2.683
2.724		2.683	2.724	...	2.724
3.007		3.007	2.724	...	3.007

**Figure 2.17 Bootstrap sampling**

Bootstrapping relies on an existing data set (here, the data is identical to **Figure 2.16**, Data Set 1). From this data set, we randomly select the required number of samples (e.g. 10). Samples are replaced, i.e. the same data point can be sampled more than once. As such, numbers in the bootstrapped data sets (right columns; shown ranked smallest to largest) sometimes repeat. This process is repeated until the desired number of replicate data sets  $N$  is generated.

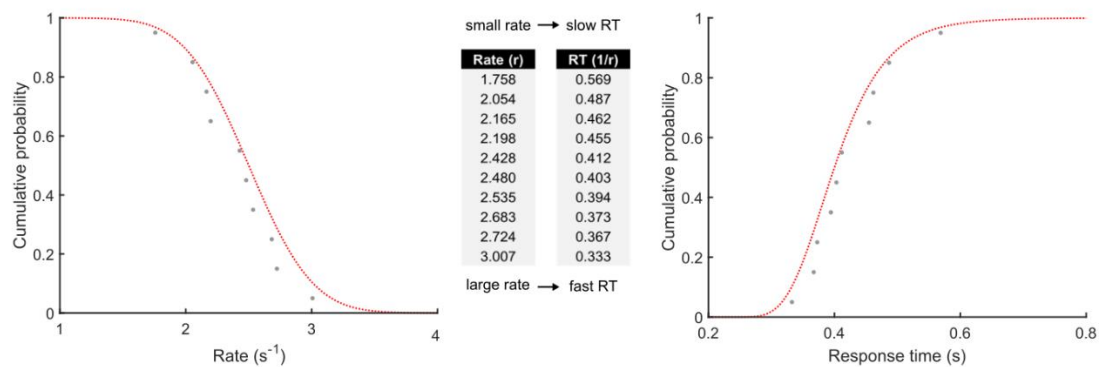


**Figure 2.18 Bootstrapping model fits**

The same simulated RTs as in **Figure 1.13** are shown here with bootstrapped model fits. The dashed lines indicate the underlying unisensory LATER model fits. The solid red line shows the context variant race model fit, including additional parameters ( $\rho$ ,  $\eta$ ). The red shaded area indicates the 95% bootstrapped confidence intervals (CIs) on the context variant race model fit.

### 2.5.3. Simulating Response Times Using Models

Simulation of unisensory RTs was done following the LATER model (see **Section 2.4.4.1**). First, rate ( $r$ ) parameters are specified; these correspond to  $\mu$  and  $\sigma$  parameters of a normal distribution. Some typical values, for instance, would be a  $\mu$  of  $2.5 \text{ s}^{-1}$  (corresponding to a median RT of 0.4 s) and



**Figure 2.19 Simulating unisensory RTs using a LATER model**

First, rate values are drawn from a normal distribution using Monte Carlo sampling. The rate data (left column) are identical to those in **Figure 2.16** (Data Set 1). Secondly, simulated RT values are generated by taking the reciprocal of the rate sample ( $1/r$ ). Note that following this transformation, smaller rate values correspond to slower RT values, and vice versa.

a  $\sigma$  of  $0.4 \text{ s}^{-1}$ . Second, the required number of samples ( $N$ ) is randomly taken from this distribution following Monte Carlo sampling (see **Section 2.5.1**). The MATLAB function `normrnd` was used to sample from a normal distribution. Finally, the reciprocal of the rate values ( $1/r$ ) is taken to transform the data into the RT space (**Figure 2.19**). As previously discussed, the rate relates to the average amount of evidence accumulated at each time point. As such, smaller rate values correspond to a longer time to reach threshold (resulting in slower RTs). Large rates, in turn, correspond to a shorter time to reach threshold (resulting in faster RTs).

Simulation of redundant RTs was done following the context variant race model (Otto & Mamassian, 2012). First, the two underlying unisensory distributions are defined using LATER models (as above). Second, rate value pairs are sampled from these two distributions with Monte Carlo sampling. Thirdly, the maximum rate value for each pair (corresponding to the minimum RT) was taken to form the redundant rate distribution. Finally, all rate values are transformed into RT space by taking the reciprocal of the rate ( $1/\text{rate}$ ). The above method generates redundant RTs which are consistent with the simple race model (i.e. assuming statistical independence and context invariance).

To simulate interactions between RTs, the two additional parameters can also be included in this process. To simulate a correlation ( $\rho$ ), the covariance matrix was adjusted to reflect the desired correlation between the unisensory distributions. The MATLAB function `mvnrnd` was used to sample rate pairs following a multivariate normal distribution. To simulate additional noise ( $\eta$ ) in the redundant RT distribution, a constant was added to the  $\sigma$  values of the unisensory LATER distributions when sampling rate pairs for the redundant distribution. Following the above simulation procedure, it is straightforward to simulate all conditions of a typical redundant signal experiment at



once  $(X, Y, XY)$ . This can be useful for validating certain approaches before applying them to real data (see for instance **Section 2.3.2.3**).

### 3. Experiment 1: Stimulus Construction vs. Signal Features

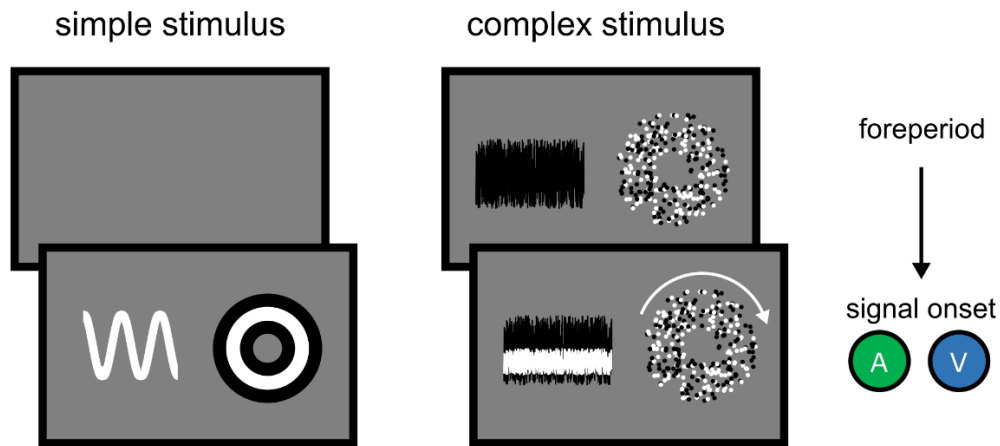
#### 3.1. Introduction

Assessment of the RSP literature in the General Introduction (**Section 1.3**) indicated a lack of a consistent, model-based approach for understanding the RSE. The broad aim of this thesis therefore, beginning with this experiment, is to introduce and evaluate a novel comparative approach, based on the previously-neglected race model framework. This approach will allow for an evaluation of potential sources of benefits and interactions, which have previously been unclear. As a brief reminder, this comparative approach follows three broad steps (see **Sections 1.6 to 1.8** for a complete outline). First, benefits are measured across conditions and compared to a benchmark prediction made by the parameter-free (simple) race model. Second, interactions which go beyond the simple race model are quantified. Third, a more complex model is applied which attempts to explain both benefits and interactions; to further explore the explanatory power of race models in this thesis, I apply a context variant race model with two additional parameters which account for the interactions. To evaluate the effectiveness of the approach, I introduce manipulations which are expected to target both benefits and interactions within a 2x2 within subjects design. These manipulations are motivated by the race model and the broader decision-making framework, and attempt to change RT distributions across factors. For benefits, the race model principles (see **Section 1.6.1**) provide a solid starting point for understanding underlying sources. For interactions, such as trial history effects (see **Section 1.7.1**), the wider literature suggests alternative hypotheses which can be tested across studies. It should be highlighted that a journal article corresponding to this chapter (Innes & Otto, 2019) has been published, and is also included at the end of this thesis.

##### 3.1.1. Factor 1: Stimulus Construction

The RSE has been replicated with a wide range of signals (**Section 1.3.2**). Elements of the effects observed in RSP studies, such as the overall size of the RSE (i.e. benefit), have been shown to change depending on elements of stimuli presented. For recent examples, Bailey et al. (2018) showed effects based on how 'rich' realistic stimuli were in virtual reality environments, whereas Juan et al. (2017) attempted to relate the size of benefits to physical properties of stimuli from a large databank. Given the lack of model framework, however, underlying explanations for *why* benefits change based on particular stimulus elements are not clear. A strong starting point for the comparative approach, therefore, would be to attempt to understand how basic stimulus elements can contribute to benefits.

As a first factor, I consider *stimulus construction*, which attempts to capture broad differences in stimuli used across experiments. Consider, for now, just auditory stimuli from previous experiments. The most frequently used auditory stimuli are sudden-onset signals. One example is *pure tones* (e.g.



**Figure 3.1 Stimulus construction**

Previous experiments have typically used sudden onset stimuli (left column), but the RSE has also been observed in experiments with randomly generated stimuli, where signals are presented in noise (right column). To determine the contribution of these stimulus elements to the size of the RSE, two levels of stimulus construction (simple, complex) were implemented.

Miller, 1982; for recent examples: Minakata & Gondan, 2018; Murray, Eardley, et al., 2018; Murray, Thelen, et al., 2018; Noel, Modi, Wallace, & Van der Stoep, 2018; Ren et al., 2018; Wang et al., 2018). Another example is *noise bursts* (e.g. Herschenson, 1962; for recent examples: Freeman, Wood, & Bizley, 2018; Harrar et al., 2017; Sürig, Bottari, & Röder, 2018; W. Yang & Ren, 2018). Overall, pure tones and noise bursts clearly differ in randomness (or *stochasticity*). For instance, the frequency of tones presented is typically identical both within and between auditory trials. This signal could be considered somewhat *simple*. The frequencies of noise bursts, however, are randomly sampled within (and possibly between) trials. This signal could be considered more *complex*. In addition to this, some auditory signals have also been presented within background noise, which was randomised between trials (Otto et al., 2013; Otto & Mamassian, 2012). A useful comparison therefore would be to compare simple stimuli (i.e. non-random, sudden onset) and complex stimuli (i.e. randomised, presented in background stimulation). By directly comparing stimulus construction (**Figure 3.1**), it will be possible to observe its effect on the size of the RSE.

Despite being largely neglected in past decades, the race model framework offers predictions for how stimulus construction should influence multisensory benefits. According to sequential sampling (see **Section 1.2.3**), variability of RT is linked to the variability in the evidence accumulation process; if evidence for a signal is more variable (as in complex compared to simple stimuli), then the overall RT distributions are also likely to be more variable. If this is the case, then the variability rule (see **Section 1.6.1.2**) would predict that benefits should also be larger for complex stimuli compared to simple stimuli. Regarding interactions, the potential effect of such stimulus construction is less

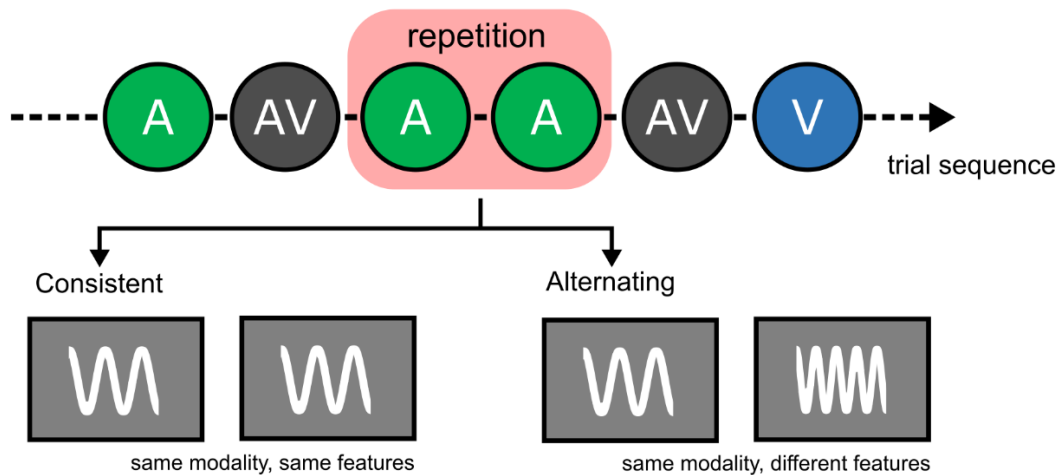
clear. Previous reports, however, suggest that transient onset elements of stimuli are important for multisensory processing (Jaekl, Perez-Bellido, & Soto-Faraco, 2014; Van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010; Werner & Noppeney, 2011). Some difference in interactions between simple and complex conditions might therefore be expected.

### 3.1.2. Factor 2: Signal Features

As for benefits, the lack of a consistent model framework means that the sources of interactions (i.e. trial history effects and violations of Miller's bound) remain unclear. For violations of Miller's bound, there at least exists a substantial literature; following Miller (1982), testing the race model has become standard procedure in RSP, thus the vast majority of RSP papers since have quantified violation in some way. In contrast, trial history effects are often neglected, being investigated in only a few papers (e.g. Gondan et al., 2004; Juan et al., 2017; Miller, 1982, 1986; Otto & Mamassian, 2012). As a starting point, therefore, the most advantageous target interaction for advancing our understanding of the RSE is trial history effects.

As a second factor, I consider *signal features* as a potential source of trial history effects. As described earlier (**Section 1.7.1**), one element of trial history effects is that the repetition of a unisensory signal leads to a faster RT compared to a switch. One simple explanation of this effect is that it arises due to repetition of low-level signal features. Continuing from the previous example, consider just auditory pure tone signals. Given the tonotopy of cells in auditory cortex (e.g. Saenz & Langers, 2014), a particular frequency (e.g. 440 Hz) will stimulate the same neural population every time it is played. On repetition trials, evidence accumulation may be facilitated by residual activity in these neurons, leading to faster RTs. There is already some suggestion that feature repetition may play a role in RT. In visual search tasks, repetition of features leads to facilitation of visual search (Kristjansson & Campana, 2010). Further, a comparison of auditory and visual search tasks suggests that similar processing might underlie repetition priming effects in both modalities (Klein & Stolz, 2015). Relating this to history effects, it may be that a similar feature-repetition effect is involved.

One way to test the signal feature explanation is to introduce multiple variants of each unisensory signal (**Figure 3.2**). Consider again the auditory pure tone signal (440 Hz). If history effects are linked causally to residual neural activity caused by this specific tone frequency, intermixing a second variant of the signal (a 660 Hz tone) across trials should reduce the overall effect, compared to conditions where there is just one signal variant across trials. The reasoning behind this is that the second variant would activate a different population of neurons, and thus residual activity from the previous trial would not contribute to accumulation on the current trial. By introducing multiple variants, therefore, it can be shown whether simple low-level features can explain trial history effects, or whether they are based in higher-level processes.



### Figure 3.2 Signal features

Repetition effects (leading to faster RTs) could be linked to repetition of low-level signal features (e.g. the specific frequency of an auditory pure tone) or to higher-level explanations (e.g. engagement of auditory attention generally). To tease apart these two classes of explanation, two levels of *signal features* can be implemented: only one tone variant (consistent) or two variants randomly intermixed (alternating). If the low-level signal features are important, the unisensory decision-making framework suggests that history effects should be smaller for alternating compared to consistent conditions.

#### 3.1.3. Experimental Hypotheses

The experiment proceeded with the following hypotheses. Regarding the first factor, *stimulus construction*, it was hypothesised that benefits should be larger in complex conditions compared to simple conditions. Further, based on the role of sudden onsets in multisensory processing, it was also expected that interactions would be different for simple and complex signals. Regarding the second factor, *signal features*, it was hypothesised that trial history effects would be larger when signal features were consistent compared to alternating, if they are based on low-level feature repetition.

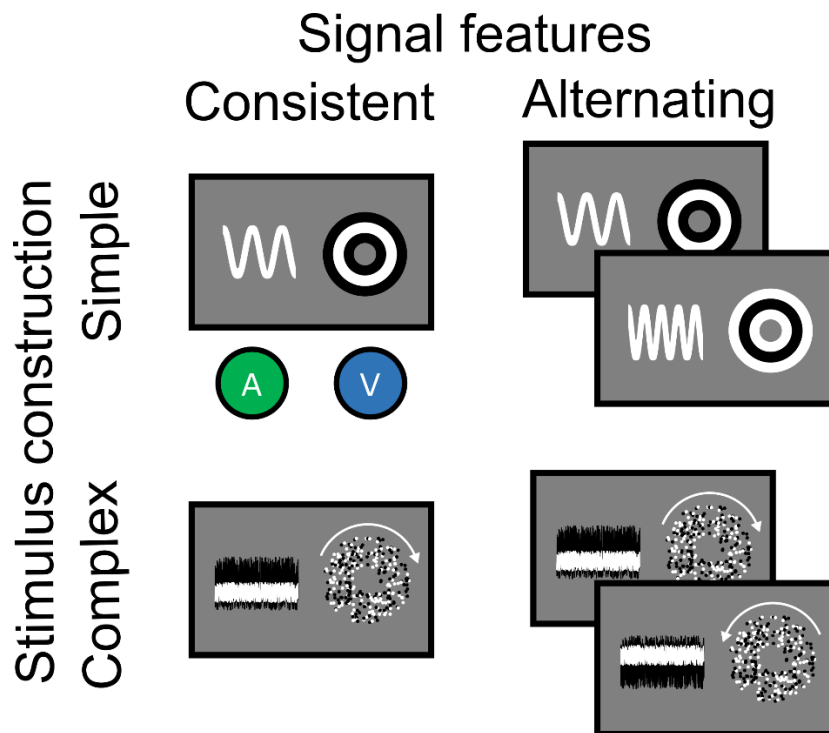
## 3.2. Methods

### 3.2.1. Participants

20 participants (14 female) were recruited at the University of St. Andrews. Age ranged from 18-29 years. All were naïve as to the aims of the experiment. Normal hearing and normal/corrected-to-normal vision was reported in all cases. Reimbursement was £9.

### 3.2.2. Stimulus Design

Stimuli were constructed according to a 2×2 design (**Figure 3.3**) with factors *stimulus construction* (simple, complex) and *signal features* (consistent, alternating). For stimulus construction, simple stimuli were identical across all trials and presented without background stimulation, whereas



**Figure 3.3 Stimulus design for Chapter 3**

Illustrations depict the redundant signal pairings. For more details on stimulus construction, see **Section 2.1.2**.

complex stimuli were randomly generated and presented within background noise (see **Section 2.1.2** in General Methods for more details). For signal features, consistent conditions presented one signal variant per modality across trials, whereas alternating conditions presented two signal variants per modality across trials.

#### 3.2.2.1. Auditory Stimuli

All signals were presented at 45 dB SPL with a ramp onset of 0.01 s. For simple conditions, the signal was a pure tone. In the simple-consistent condition, the tone frequency was 440 Hz. In the simple-alternating condition, the tone frequency was either 440 or 660 Hz (with equal numbers of each across trials). For complex conditions, the signal was filtered noise. Noise signals were generated randomly from Gaussian noise with a 2<sup>nd</sup> order Butterworth filter applied. In the complex-consistent condition, the edge frequencies were 1.0/1.1 kHz. In the complex-alternating condition, the edge frequencies were either 1.0/1/1 or 1.1/1.2 kHz (with equal numbers of each across trials). Noise signals were always presented within a background noise (50 dB SPL). This was also generated from Gaussian noise with a 1<sup>st</sup> order Butterworth filter applied. Edge frequencies were 0.5/2.4 kHz in all conditions.

#### 3.2.2.2. Visual Stimuli

For simple conditions, the signal was the onset of 3 concentric rings, which alternated between black and white. In the simple-consistent condition, the largest ring was black. In the simple-alternating

condition, the largest ring was black or white (with equal numbers of each across trials). For complex conditions, the signal was the coherent rotation of 50% of dots within a random dot stimulus. In the simple-consistent condition, the dots coherently rotated clockwise. In the simple-alternating condition, the dots coherently rotated clockwise or anti-clockwise (with equal numbers of each across trials).

### 3.2.3. Procedure

Each trial began by presenting a green fixation point. The audio-visual background was also presented in complex conditions. This was followed by a random foreperiod. The duration of the foreperiod was determined by two components: a fixed duration (1 s) and a random component drawn from an exponential distribution (mean: 0.75 s). An exponential distribution for the foreperiod was used as they contain no information about signal onset (Gondan & Minakata, 2016; Luce, 1986). On signal trials, one of the three signals (A, V, or AV) was then presented until the participant responded (maximum signal duration: 1.5 s). On catch trials, no signal was presented and the audio-visual background continued for the maximum signal duration. If any response was made, the fixation point changed from green to red, to indicate that the response had been recorded and the next trial would soon be initiated. All stimulation was frozen for 0.25 s before initiating the next trial. Responses recorded within 1.5 s of a signal onset were initially considered valid. If the response was an error (false alarm or miss), a feedback screen was then presented for 2 s which informed the participant of the error. Responses within the foreperiod (or at any point during catch trials) were considered false alarms. Signal trials in which no response was recorded within the maximum duration were considered misses. The next trial then began immediately.

Trials were randomly presented in blocks, which were initialised with 104 trials each (26 trials per signal type, 26 catch trials). A dummy trial was also inserted at the start of each block. Dummy trials always displayed AV consistent signal pairings (see **Figure 3.3**) and were not included in analyses. Trials in which a false alarm or miss was recorded were replaced within the trial order, which was reshuffled after every error. A dummy trial was also inserted before resuming experimental trials. A small pointer (attached to fixation) denoted progress through a block. This started in a vertical position, and after each trial the pointer was incremented clockwise. A complete rotation of the pointer indicated completion of a block. Each block lasted 4-5 minutes.

The total experiment consisted of 16 blocks (4 blocks for each of the 4 experimental conditions). Block order was randomised for each participant according to a Latin square procedure. Breaks were given between blocks as required. The whole experimental session lasted for around 105 minutes.

### 3.3. Results

For all repeated-measures ANOVAs presented in this section, I adhere to the following ordering of factors: stimulus construction (simple, complex) × signal features (consistent, alternating). In certain analyses, a third factor is included: signal modality (A, V, [AV]).

#### 3.3.1. General Performance

Before the main analysis of RT, general performance was assessed. This was first considered because ceiling performance is an assumption of the present procedures for RT analysis. If the data deviate from this assumption (e.g. by demonstrating a high overall percentage of misses), then these procedures must be adapted to also account for accuracy. However, if the ceiling performance assumption is met, no adaptations are necessary. As a first step, the number of error responses (false alarms and misses) was calculated. As a second step, the number of remaining outliers was calculated. Error responses and outliers were removed to clean the dataset for further RT analysis.

First, the percentage of false alarms was calculated for signal and catch trials. The mean percentage of false alarms was averaged across signal trials (A, V, AV); this is because no signal is actually presented on false alarm trials, as the response occurs in the foreperiod. For signal trials, false alarms occurred on 1.04% ( $\pm 0.19\%$ ) of trials. A 2×2 ANOVA revealed no significant main or interaction effects (all  $F \leq 4.110$ ,  $p \geq 0.057$ ,  $\eta p^2 \leq 0.178$ ). For catch trials, false alarms occurred on 1.53% ( $\pm 0.25\%$ ) of trials. A 2×2 ANOVA revealed a significant interaction between stimulus construction and signal features,  $F(1, 19) = 4.774$ ,  $p = 0.042$ ,  $\eta p^2 = 0.201$ . For simple conditions, the percentage of false alarms decreased from consistent conditions (1.86%  $\pm 0.39\%$ ) to alternating conditions (1.17  $\pm 0.32\%$ ). For complex conditions, the percentage of false alarms increased from consistent conditions (1.26  $\pm 0.32\%$ ) to alternating conditions (1.85  $\pm 0.44\%$ ). A possible explanation for this is that in complex-alternating conditions, it is overall more likely that random patterns in the background could be mistaken for a signal. For instance in audition, it is more likely that background noise mimicked a signal by chance if there were two possible signal variants (the signal window effectively ranging 1.0-1.2 kHz) compared to one (1.0-1.1 kHz). No other main effects were significant (all  $F \leq 0.024$ ,  $p \geq 0.878$ ,  $\eta p^2 \leq 0.001$ ).

Second, the percentage of misses was calculated. The percentage of misses averaged across signal modalities was only 0.46% ( $\pm 0.19\%$ ). A 2×2×3 ANOVA revealed a significant main effect of stimulus construction,  $F(1, 19) = 5.231$ ,  $p = 0.034$ ,  $\eta p^2 = 0.216$ . The percentage of misses was smaller for simple conditions (0.14  $\pm 0.08\%$ ) than complex conditions (0.78  $\pm 0.32\%$ ). This is expected as complex signals were presented in background noise, and thus the onset was less distinguishable. There was also a significant main effect of signal modality,  $F(1.211, 23.011) = 4.033$ ,  $p = 0.050$ ,  $\eta p^2 = 0.175$ . Pairwise comparisons revealed that the percentage of misses on visual trials (0.50  $\pm 0.18\%$ ) was larger than on



redundant trials ( $0.14 \pm 0.08\%$ ),  $p = 0.010$ . The percentage of misses on auditory trials ( $0.73 \pm 0.33\%$ ) however was not significantly different to visual or redundant trials (both  $p > 0.05$ ). Smaller misses on redundant trials are expected as there are two signals. As for auditory trials, the percentage of misses was also larger and more variable across participants, which is perhaps why no significant difference was observed. No other main or interaction effects were significant (all  $F \leq 3.689$ ,  $p \geq 0.061$ ,  $\eta p^2 \leq 0.163$ ).

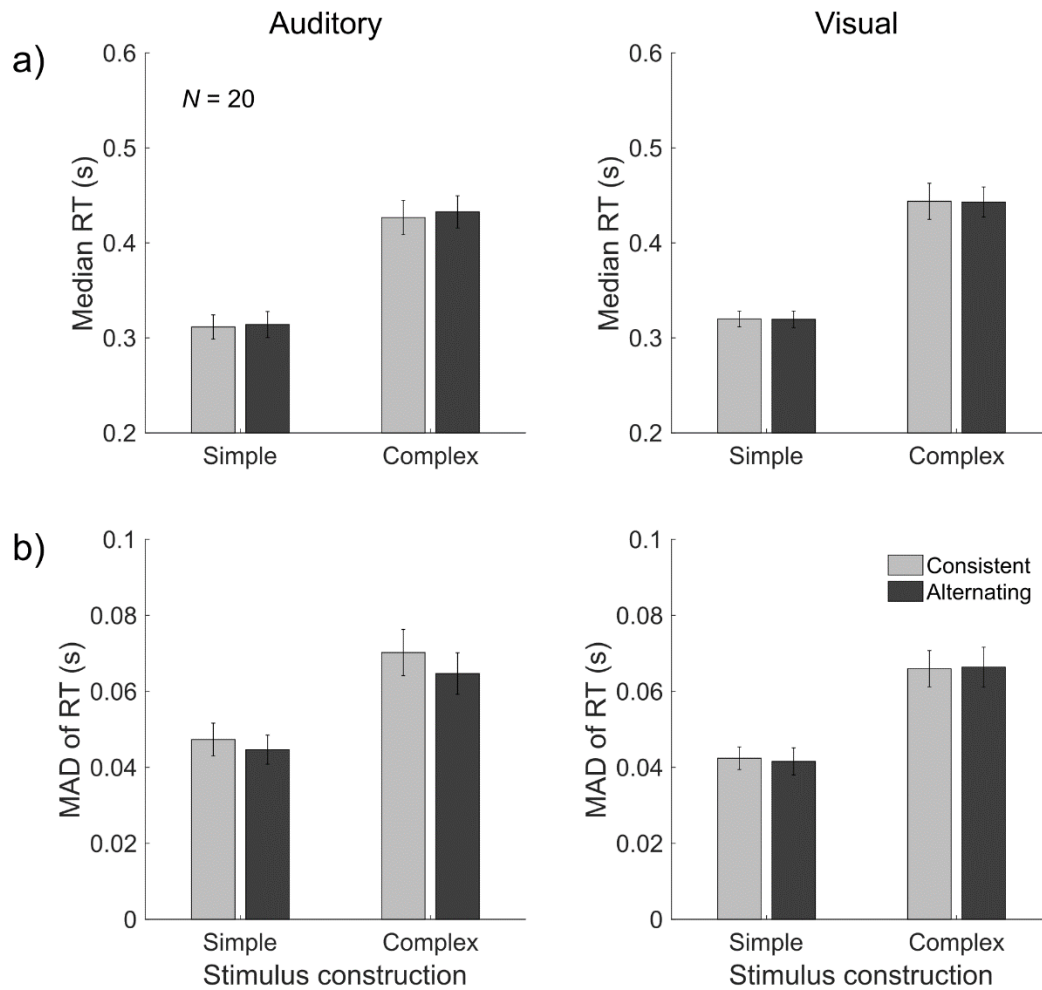
Third, the number of outliers was calculated according to the  $\pm 3 \times 1.4826$  MAD criterion (see **Section 2.2.2**). This was performed on the  $1/RT$  transformed data, as this is the data to which the model is eventually fitted. On average, fast outliers occurred on  $0.4\%$  ( $\pm 0.1\%$ ) of RTs, whereas slow outliers occurred on  $0.5\%$  ( $\pm 0.1\%$ ) of RTs. Thus, the average percentage of outliers ( $0.9\%$ ) is only slightly larger than what would be expected if the  $1/RT$  data followed a normal distribution ( $0.27\%$ ). Across all participants, 24720 valid RTs were included in the main analyses, with a mean of  $103$  ( $\pm 0.1$ ) valid RTs per participant for each condition (of a possible maximum of 104 RTs).

Overall, this analysis shows that the general performance of participants was very close to ceiling, with a low number of false alarms and misses ( $< 2\%$  signal trials on average). Further, the number of outliers was small, and thus the number of analysed trials was close to the maximum. For these reasons, accuracy was not considered in further RT analyses.

### 3.3.2. Unisensory RTs

Before applying the comparative approach (**Sections 1.6 to 1.8**), an analysis of unisensory RTs was conducted. This is because according to the race model principles (see **Section 1.6.1**), changes in unisensory RTs suggest corresponding changes in benefit. Ideally, therefore, the experimental factors should influence overall RT; this will give directional predictions for how benefits should change, which can be compared to empirical benefits.

As a first step, central tendency was evaluated to give a broad indication of how similar unisensory performance was; as RT distributions are right-skewed, this was done using the *median RT*. According to race models, benefit is maximised when unisensory RT distributions are the same (equal effectiveness). Though stimuli were not calibrated for each individual, stimulus values were selected following a pilot study, in which roughly similar performance in auditory and visual trials was observed across participants. As a second step, variability of unisensory RTs was assessed; corresponding to the median, this was done using *median absolute deviation (MAD)* of RTs. According to race models, the variability of unisensory RTs is the driving force of multisensory benefits (variability rule). Thus, changes in unisensory variability across conditions are indicative of changes in benefit, especially if the equal effectiveness principle holds. If unisensory RT variability increases for a particular condition, race models predict that benefit should increase correspondingly.



**Figure 3.4 Unisensory RT analysis**

**a, b)** Median RT and MAD of RT across conditions. The left column shows auditory RTs and the right column shows visual RTs. All bars show a mean of 20 participants ( $\pm 1$  SEM).

For median RT (**Figure 3.4a**), a  $2 \times 2 \times 2$  ANOVA revealed a significant main effect of stimulus construction,  $F(1, 19)=160.575$ ,  $p \leq 0.001$ ,  $\eta p^2=0.894$ . Median RT was smaller in simple conditions ( $0.316 \pm 0.010$  s) compared to complex conditions ( $0.437 \pm 0.016$  s). There was no main effect of signal features, and no other main or interaction effects were significant (all  $F \leq 2.258$ ,  $p \geq 0.149$ ,  $\eta p^2 \leq 0.106$ ). As there was no main effect of signal modality, median RT was similar for auditory and visual trials (i.e. auditory and visual performance was equally effective).

For MAD of RT (**Figure 3.4b**), a  $2 \times 2 \times 2$  ANOVA revealed a significant main effect of stimulus construction,  $F(1, 19)=70.808$ ,  $p \leq 0.001$ ,  $\eta p^2=0.788$ . MAD of RT was smaller in simple conditions ( $0.044 \pm 0.003$  s) compared to complex conditions ( $0.067 \pm 0.005$  s). No other main or interaction effects were significant (all  $F \leq 4.021$ ,  $p \geq 0.059$ ,  $\eta p^2 \leq 0.175$ ). As there was no main effect of signal modality, MAD of RT was also similar for auditory and visual signals (giving a further indication of equal effectiveness).

Overall, as unisensory RTs were similar for each factor, benefits will be maximised (according to the equal effectiveness principle). Further, as the variability of unisensory RTs was larger in complex conditions compared to simple conditions, this suggests that benefits will be larger in complex conditions compared to simple conditions (according to the variability rule).

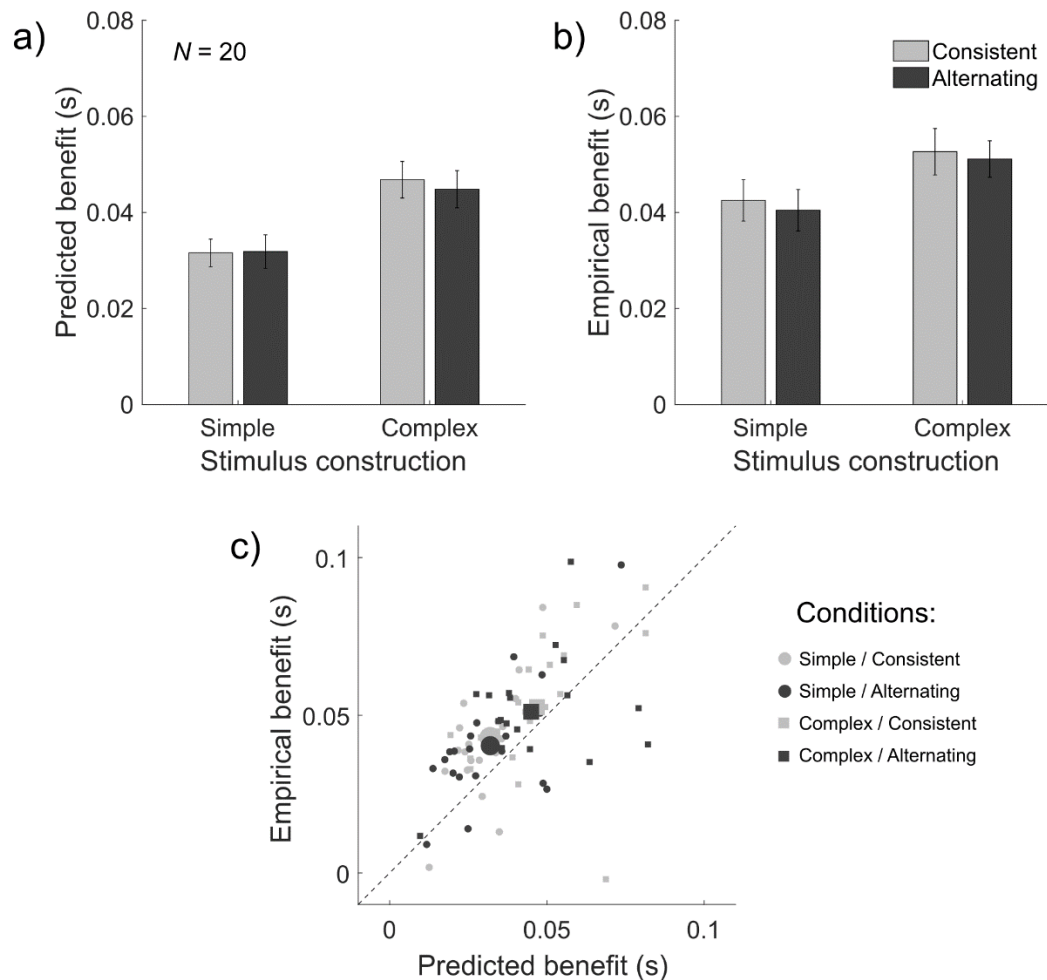
### 3.3.3. Comparative Approach Step 1: Multisensory Benefits

Following analysis of unisensory RTs, the next step is to understand multisensory RTs using the comparative approach. The first step is to understand *benefits* seen in multisensory RTs (i.e. the RSE). So far, analysis of unisensory RTs has provided broad, directional predictions for changes in benefits by considering the race model principles. To more precisely assess the predictive power of race models, however, a quantitative prediction of benefit can be obtained using the simple race model (see **Section 2.3.2.2**). This accounts for both principles in a single, parameter-free prediction of benefit size, which can be used as a benchmark for empirical benefits. To understand how well this simple race model can account for changes in the RSE across conditions, predicted benefits and empirical benefits were calculated and compared.

First, predicted benefits (**Figure 3.5a**) were calculated according to the simple race model (**Section 2.3.2.2**). A 2×2 ANOVA revealed a significant main effect of stimulus construction,  $F(1, 19)=41.537$ ,  $p\leq 0.001$ ,  $\eta p^2=0.686$ . The simple race model thus predicted that benefits would be smaller in simple conditions ( $0.032 \pm 0.003$  s) compared to complex conditions ( $0.046 \pm 0.004$  s). There were no other significant main or interaction effects (all  $F\leq 0.796$ ,  $p\geq 0.383$ ,  $\eta p^2\leq 0.040$ ).

Second, empirical benefits (**Figure 3.5b**) were measured (**Section 2.3.2.1**). A 2×2 ANOVA revealed a significant main effect of stimulus construction,  $F(1, 19)=5.866$ ,  $p\leq 0.026$ ,  $\eta p^2=0.236$ . In line with the results for predicted benefits, empirical benefits were smaller in simple conditions ( $0.041 \pm 0.004$  s) compared to complex conditions ( $0.052 \pm 0.003$  s). There were no other significant main or interaction effects (all  $F\leq 0.413$ ,  $p\geq 0.528$ ,  $\eta p^2\leq 0.021$ ). Thus, in terms of overall main effects, empirical benefits followed the predicted benefits.

To further investigate the simple race model's predictive power, a number of additional analytical steps were applied. First, the correlation between predicted and empirical benefits was evaluated on the individual participant level for each condition of the 2×2 design (**Figure 3.5c**). All correlation values were positive, and 3 of 4 were significant (**Table 3.1**). Second, differences between predicted and empirical benefits were assessed. As an initial broad indication of this difference, benefit was averaged across conditions and compared with a paired-samples *t*-test. Mean predicted benefit ( $0.039 \pm 0.003$ s) was smaller than mean empirical benefit ( $0.047 \pm 0.003$  s),  $t(19)=3.066$ ,  $p=0.006$ , two-tailed. To assess if this difference varied across conditions, predicted benefit was subtracted from empirical benefit for each condition. A 2×2 ANOVA revealed no significant main or



**Figure 3.5 Predicting and measuring multisensory benefits**

**a, b)** Predicted and empirical benefits across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

**c)** Empirical benefit as a function of predicted benefit. Each point represents data from one participant in one of the four conditions (80 data points total). Large symbols show the group mean.

interaction effects (all  $F \leq 0.791$ ,  $p \geq 0.385$ ,  $\eta p^2 \leq 0.040$ ). Thus, the difference between predicted and empirical benefit ( $0.008 \pm 0.003$  s) was consistent across conditions.

Overall, the simple race model predicted *changes* in empirical benefit well. However, it did not completely account for the *size* of the empirical benefit, which was consistently larger than

**Table 3.1 Pearson correlation coefficients (and  $p$  values) comparing predicted and empirical benefits**

Stimulus construction	Predicted vs empirical	
	Signal features	
	Consistent	Alternating
Simple	0.731 ( $p < 0.001$ )	0.690 ( $p = 0.001$ )
Complex	0.447 ( $p = 0.048$ )	0.332 ( $p = 0.152$ )

predicted. This is likely because the simple race model cannot account for empirical interactions in the RSP, which also contribute to benefits. These interactions are thus quantified in the next step.

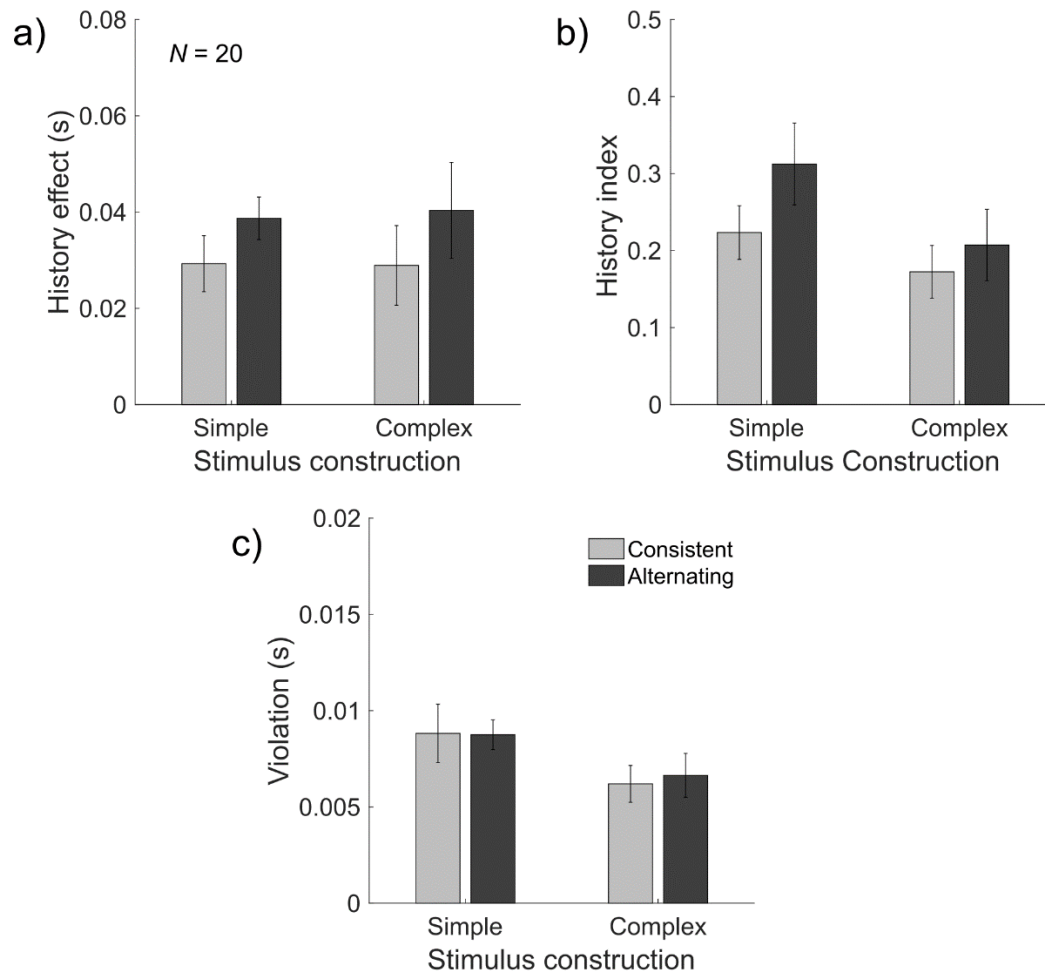
#### 3.3.4. Comparative Approach Step 2: Quantifying Interactions

The second step of the comparative approach is to understand *interactions* which are not considered by the simple race model. Two interactions in particular, which have been observed in previous empirical work, are investigated: trial history effects (**Section 1.7.1**), and violations of Miller's bound (**Section 1.7.2**). If these interactions are present, they must be quantified and accounted for by a further modelling approach.

First, trial history effects were considered (**Figure 3.6a**). As a first evaluation of this, the history effect was calculated (**Section 2.3.3.1, Equation 17**). A 2×2 ANOVA first revealed a significant intercept,  $F(1, 19)=75.624, p<0.001, \eta p^2=0.799$ . Thus, the overall history effect ( $0.034 \pm 0.004$  s) was significantly greater than 0, and must be accounted for. However, there were no significant main or interaction effects (all  $F \leq 4.134, p \geq 0.056, \eta p^2 \leq 0.179$ ). Therefore, contrary to the experimental hypothesis, the history effect did not change across conditions. It is important to note that the impact of the history effect on benefits depends on the overall variability of unisensory RTs. If unisensory RTs are generally not variable, then the contribution of the history effect will be relatively large in comparison to a case where unisensory RTs are highly variable. As a further evaluation, the history index was calculated (**Section 2.3.3.1, Equation 18**). This measure normalises the history effect by the variance of unisensory RTs. By doing so, it provides a measure of how history effects relate to the overall unisensory RT variability, and thus how much they are likely to contribute to multisensory benefits. For the history index (**Figure 3.6b**), as shown in the history effect, 2×2 ANOVA revealed no significant main or interaction effects (all  $F \leq 2.913, p \geq 0.104, \eta p^2 \leq 0.133$ ). Therefore, the overall contribution of the history effect to RT variability was consistent across conditions.

The second interaction considered was violations of Miller's bound (**Figure 3.6c**). A 2×2 ANOVA first revealed a significant intercept,  $F(1, 19)=130.701, p<0.001, \eta p^2=0.873$ . Thus, the overall violation area ( $0.008 \pm 0.001$  s) was significantly greater than 0, and must be accounted for. Further, there was a significant main effect of stimulus construction,  $F(1, 19)=4.564, p<0.046, \eta p^2=0.194$ . Violation was larger for simple conditions ( $0.009 \pm 0.001$ s) than for complex conditions ( $0.006 \pm 0.001$ s). There were no further significant main or interaction effects (all  $F \leq 0.054, p \geq 0.819, \eta p^2 \leq 0.003$ ).

Overall, analysis revealed the presence of two interactions: trial history effects and violations of Miller's bound. Contrary to the experimental hypothesis, the size of history effects did not change across conditions. Further, violations of Miller's bound were larger for complex conditions compared to simple conditions. As neither of these interactions can be accounted for by the basic race model architecture alone, these must be explained by applying an alternative model.



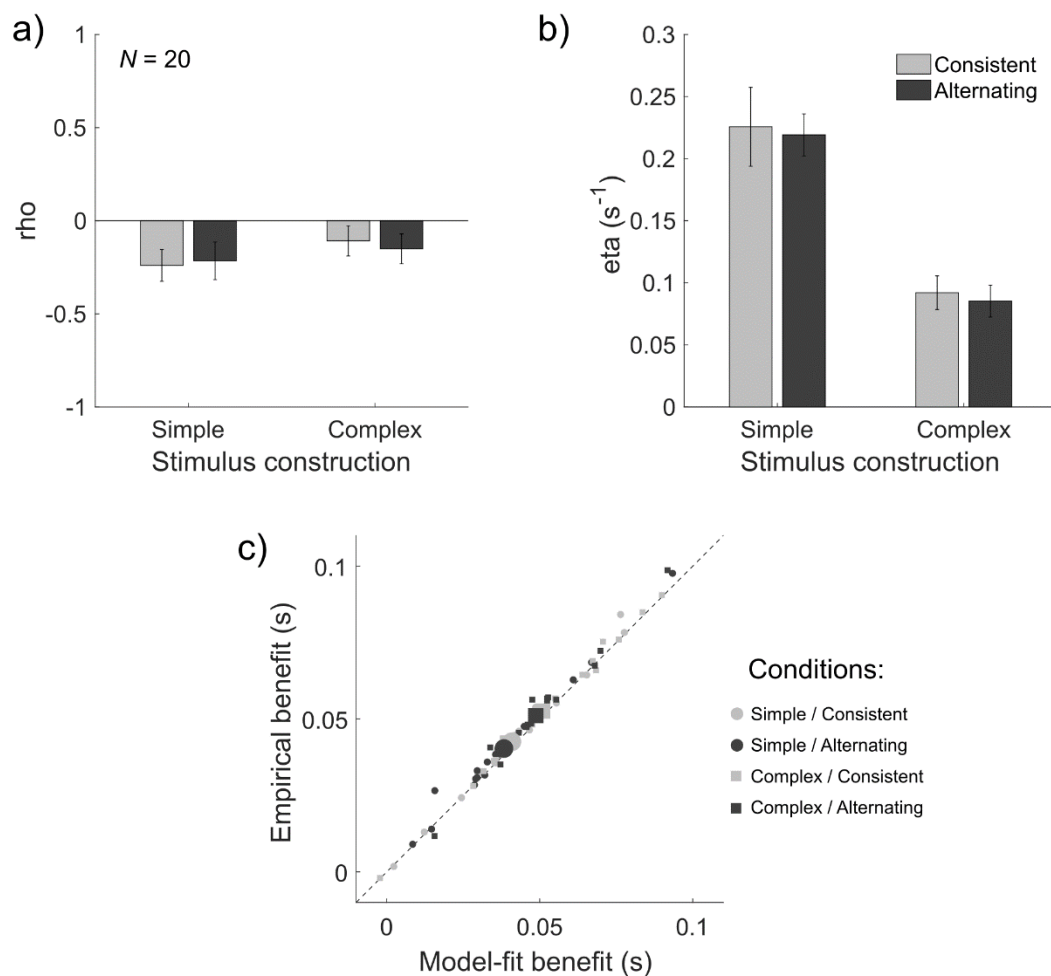
**Figure 3.6 Quantifying processing interactions**

**a, b)** History effect and history index across conditions.

**c)** Violation of Miller's bound across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

### 3.3.5. Comparative Approach Step 3: Applying the Context Variant Race Model

The third (and final) step of the comparative approach is to apply a model to understand both the benefits and interactions observed in previous steps. Here, the context variant race model (**Section 2.4.4.2**) was fitted. Broadly, this follows the same architecture as an independent race model, but two parameters are added to account for interactions. First, a correlation parameter  $\rho$  is added between unisensory distributions. This can account for trial history effects. Second, a noise parameter  $\epsilon$  models additional variability in the redundant distribution compared to the independent race model. This can account for violations of context invariance, and therefore violations of Miller's bound. To fit the model, the best-fitting LATER model parameters for each unisensory distribution were first determined. This is equivalent to fitting an independent race model. Next, the two additional parameters of the context variant race model ( $\rho$ ,  $\epsilon$ ) were determined. Average fit values for all 6 model parameters ( $\mu_{A_i}$ ,  $\mu_{V_i}$ ,  $\sigma_{A_i}$ ,  $\sigma_{V_i}$ ,  $\rho$ ,  $\epsilon$ ) across conditions are found in the appendices.



**Figure 3.7 Fitting the context variant race model**

**a,b)** Best fitting  $\rho$  and  $\eta$  values across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

**c)** Empirical benefit as a function of model-fit benefit. Each point represents data from one participant in one of the four conditions (80 data points total). Large symbols show the group mean.

Analysing the  $\rho$  parameter (**Figure 3.7a**), a  $2 \times 2$  ANOVA revealed a significant intercept,  $F(1, 19) = 7.722$ ,  $p = 0.012$ ,  $\eta^2 = 0.289$ . Thus, there was an overall negative correlation modelled between unisensory rate distributions ( $-0.178 \pm 0.064$ ). However, there were no significant main or interaction effects (all  $F \leq 1.240$ ,  $p \geq 0.279$ ,  $\eta^2 \leq 0.061$ ). Overall, therefore, the results for  $\rho$  closely follow those observed for the history effect and history index.

Analysing the  $\eta$  parameters (**Figure 3.7b**), a  $2 \times 2$  ANOVA also revealed a significant intercept,  $F(1, 19) = 176.056$ ,  $p < 0.001$ ,  $\eta^2 = 0.903$ . Thus, additional variability was added to unisensory rate distributions to account for variability in the redundant rate distribution ( $0.156 \pm 0.012 \text{ s}^{-1}$ ). There was also a main effect of stimulus construction,  $F(1, 19) = 44.616$ ,  $p < 0.001$ ,  $\eta^2 = 0.701$ . The  $\eta$  parameter was larger for simple conditions ( $0.222 \pm 0.019 \text{ s}^{-1}$ ) than for complex conditions ( $0.089 \pm 0.010 \text{ s}^{-1}$ ).

However, there were no further significant main or interaction effects (all  $F \leq 0.145$ ,  $p \geq 0.708$ ,  $\eta p^2 \leq 0.008$ ). Overall therefore, the results of *eta* closely follow those observed for violations of Miller's bound.

To assess model fit, benefit was calculated using the model-fit redundant RT distribution in place of the empirical redundant RT distribution. To obtain equivalent quantile points for comparison, the model-fit distribution was down-sampled using linear interpolation (**Section 2.3.1**). First, model-fit and empirical benefits were correlated as in Step 1. All 4 values were strongly positive and highly significant (**Table 3.2**). Second, differences between predicted and empirical benefits were assessed. Benefit was averaged across conditions and compared with a paired-samples *t*-test. Mean model-fit benefit ( $0.045 \pm 0.003s$ ) was smaller than mean empirical benefit ( $0.047 \pm 0.003 s$ ),  $t(19)=6.190$ ,  $p < 0.001$ , two-tailed. To assess whether this difference was consistent across conditions, empirical benefit was subtracted from model-fit benefit for each condition. A 2x2 ANOVA revealed no significant main or interaction effects (all  $F \leq 1.097$ ,  $p \geq 0.308$ ,  $\eta p^2 \leq 0.055$ ). Thus, the difference between model-fit and empirical benefit ( $0.002 \pm 0.0003 s$ ) was consistent across conditions.

Overall, applying the context variant race model provided a clear account of multisensory RTs. Interactions beyond the race model were accounted for by the model's parameters, which also closely followed interactions in terms of main effects. Further, as shown in **Figure 3.7c**, benefits were almost perfectly accounted for by taking account of interactions.

### 3.4. Discussion

The overall aim of this experiment was to test the comparative approach as a tool for understanding sources of benefits and interactions. To this end, two factors were introduced. First, different signal types which mimic the diverse signals employed across the RSP literature were presented (*stimulus construction*, with levels simple and complex). It was found that complex stimuli, which resulted in more variable RTs, produced larger benefits than simple stimuli (which gave rise to less variable RTs). Second, a potential source of history effects was tested by manipulating *signal features* (with levels consistent and alternating). Counter to the experimental hypothesis, history effects were not smaller

**Table 3.2 Pearson correlation coefficients (and *p* values) comparing model-fit and empirical benefits**

Stimulus construction	Model-fit vs empirical	
	Signal features	
	Consistent	Alternating
Simple	0.995 ( $p < 0.001$ )	0.992 ( $p < 0.001$ )
Complex	0.996 ( $p < 0.001$ )	0.986 ( $p < 0.001$ )



for alternating conditions compared to consistent conditions. Rather, history effects were constant across conditions.

It was expected that the factors in the design would elicit differences in overall unisensory RTs. Though there was no main effect of signal features, the factor stimulus construction has a substantial effect on unisensory RTs, with simple stimuli eliciting faster and less variable RTs than complex stimuli. Further, the auditory and visual stimuli presented here were designed (based on pilot studies) to elicit roughly similar RT distributions, as benefits are maximised when unisensory RT distributions are similar (**Section 1.6.1.1**). Accordingly, no main effect of signal modality was found for median or MAD of RT, suggesting signals successfully elicited similar unisensory distributions.

The first step of the comparative approach focused on multisensory benefits (i.e. the RSE) across the factors of the design. This was first predicted using a simple race model then calculated empirically. In both cases, a main effect of stimulus construction was found, whereby simple stimuli elicited smaller benefits than complex stimuli. This supports the variability rule (**Section 1.6.1.2**); as complex stimuli elicited more variable unisensory RTs than simple stimuli, larger benefits were expected. This suggests that stimulus qualities contribute to overall multisensory benefits. In addition, not only were the main effects predicted by the simple race model, but a general correspondence was shown on the level of individual subjects. As seen in **Table 3.1**, positive correlations between the size of predicted benefits and empirical benefits were significant in 3 of 4 conditions. Notably, empirical benefit was consistently larger than predicted benefit (around 0.008 s across conditions). Overall, as a benchmark comparison, the race model accounts well (but not completely) for changes in benefits across conditions and on the level of individual participants.

The second step of the comparative approach focused on interactions (i.e. history effects and violation of Miller's bound). Considering trial history effects, there was a modality switch cost (0.034 s) across all factors. This is broadly in agreement with previous studies which have quantified these effects (e.g. Gondan et al., 2004; Miller, 1982; Otto & Mamassian, 2012). Following a similar reasoning to the literature of repetition priming, it was expected that alternating low-level stimulus features between unisensory trials would reduce history effects, as the RT benefits of signal repetition would be diminished. Contrary to this expectation, no changes in history effects were found across conditions. This therefore suggests that history effects cannot be targeted by low-level signal features. History effects may therefore have a basis in higher-level processes, and be more similar to costs observed in attentional (Spence et al., 2001) and task-switching (Kiesel et al., 2010; Monsell, 2003) paradigms. One previous fMRI study of task switching showed that costs of switching between face and word processing (which activate distinct processing pathways) are related to the level of residual activity in the pathway associated with the switched behaviour (Yeung, Nystrom, Aronson, & Cohen, 2006).

History effects may work in a similar way, whereby general activation of the auditory and visual pathways is important, rather than the specific properties of the stimulus presented. The fact that both simple and complex stimuli, which have very different properties, elicit essentially the same size history effect would seem to support this proposal.

Considering violation of Miller's bound, significant violations (around 0.008 s) were observed across all conditions. This corresponds approximately to the difference between empirical benefits and those predicted by the simple race model (which by definition cannot violate Miller's bound). Further, a significant effect of stimulus construction was observed, whereby simple stimuli elicited larger violations than complex stimuli. One possible explanation for this effect is the sudden onset element of the simple stimuli. Previous reports have suggested that onset transients have an important role in multisensory processing (Jaekl et al., 2014; Van der Burg et al., 2010; Werner & Noppeney, 2011). The sudden onset of a signal also likely produces a greater change in *context* between unisensory and redundant conditions. In simple redundant trials, two processing pathways become suddenly activated compared to just one on unisensory trials. In contrast, during complex conditions, both processing pathways are active during all trial types because of background noise. The onsets, therefore, cause a smaller change in context. Similar changes to context are explored in **Chapter 4**.

Alternatively, previous papers have reported similar effects on violation by manipulation of signal strength (or 'saliency'): larger violations are produced by stronger stimuli (Minakata & Gondan, 2018). It is possible the effects here also reflect a difference in signal strength. Typically, however, signal strength is manipulated by changes to single property of the signal, e.g. volume (auditory) or contrast (visual), whereas the differences in the stimuli presented here are multifaceted. Nonetheless, given that RTs in the present experiment were overall slower for more complex stimuli, a difference in signal strength between simple and complex stimuli is a distinct possibility. A more direct manipulation of signal strength is explored in **Chapter 5**.

The third step of the comparative approach was to apply a model to account for benefits and interactions, namely the context variant race model. First, benefit was accounted for much more precisely than the simple race model prediction, evidenced by much stronger positive correlations (**Table 3.2**) and a reduction in the overall difference between model-fit and empirical benefit (0.002 s, compared to the 0.008 s difference between predicted and empirical benefit). Further, the main effects observed in *rho* (corresponding to history effects) and *eta* (corresponding to violations) are the same, suggesting that the model is overall able to account for them. Most interesting here is the large reduction in *eta* modelled in complex conditions compared to simple conditions. This suggests that for complex stimuli, comparatively less noise is introduced as a result of both decision-units

accumulating evidence (redundant trials). This supports the idea that complex signals produced smaller changes in context variance between unisensory and redundant trials with complex stimuli.

Overall, the current data provide a validation of the comparative approach, and provide novel insights into multisensory behaviour. One instance of a novel insight from the present data comes from the relationship between multisensory benefits and violations of Miller's bound. Under the approach detailed here, both benefit and violation are distinct (though related) effects. In current practice, however, these two effects are considered somewhat interchangeable measures of 'multisensory integration' (e.g. Stevenson et al., 2014). In contrast to this assumption, a dissociation is observed here on the group level analyses; benefits *increase* from simple to complex stimuli, whereas violations *decrease* correspondingly. The interpretation of experimental results, therefore, could be very different depending on which measure is chosen as a measure of multisensory integration. This should be considered carefully in future RSP experiments.

## 4. Experiment 2: Signal Duration vs. Task-Irrelevant Stimulation

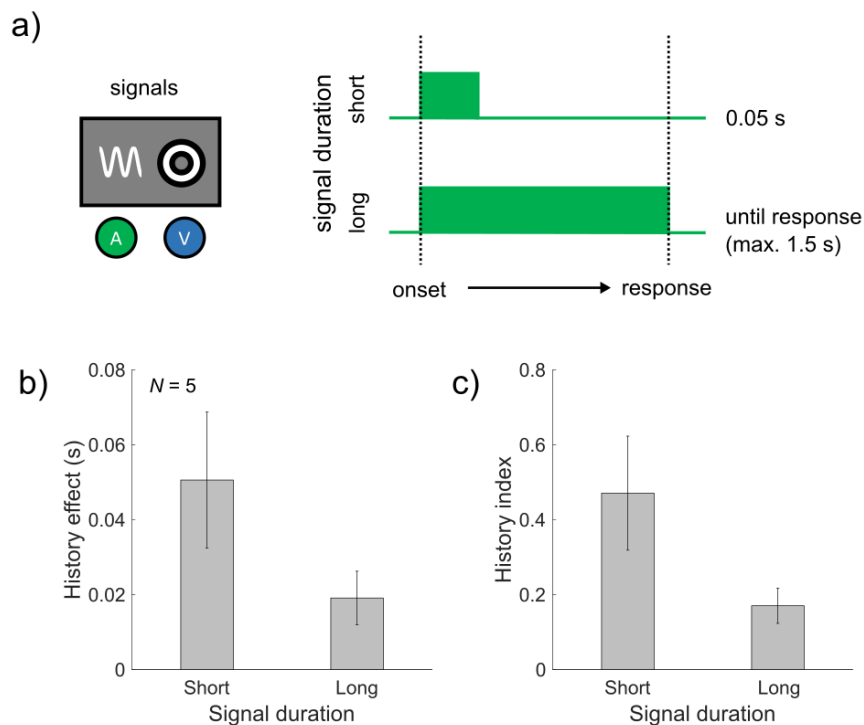
### 4.1. Introduction

The previous chapter demonstrated that the novel comparative approach based on the race model framework is an effective investigative tool for studying benefits and interactions in the RSP. By applying the approach across different experimental factors, it was possible to make effective predictions for changes in benefits (based on race model principles), observe changes in interactions, and provide an overall account of multisensory RTs by applying an extended model. This has provided a clearer understanding of the overall RSE. What remains to be understood, however, are sources of interactions. This is particularly true for history effects, which are frequently neglected within RSP studies. The overall goal of this chapter, therefore, is to again apply the comparative approach with a 2×2 within-subjects design, with each factor designed to target a particular interaction.

#### 4.1.1. Factor 1: Signal Duration

As a first factor, I consider *signal duration* as a method of manipulating history effects. The motivation for this factor comes from an examination of the few RSP studies which have explicitly quantified modality switch costs. In Miller's classic study, for instance (Miller, 1982; their Table 4), the modality switch cost for AV signals was as large as 0.074 s for detection tasks and 0.118 s for a discrimination task. Similarly, Gondan et al. (2004) reported history effects of up to 0.069 s for AT signals. In more recent work by Otto, however, the modality switch costs reported have been more variable, ranging from around 0.050 s (Otto & Mamassian, 2012) to 0.020-0.030 s (Otto & D'Souza, 2015) for AV signals (see also Harrar et al. (2014), who recorded a cost of around 0.025 s). These latter results are more comparable to the history effects observed in this thesis (e.g. 0.034 s in **Section 3.3.4**). Interestingly, the studies which report using short signals report the largest history effects: for instance, signal durations of 0.02 s (Gondan et al., 2004) and 0.15 s (Miller, 1982). In keeping with this, studies which present longer signals (essentially until the participant responds) on average report smaller history effects: signal durations of 0.34 s (Harrar et al., 2014), 0.5 s (Otto & Mamassian, 2012), 1 s (Otto & D'Souza, 2015), or up to 1.5 s (see **Chapter 3**).

One potential explanation of why signal duration would change the size of the history effect is provided by attentional mechanisms, which have been investigated in similar multisensory paradigms (e.g. Spence et al., 2001). To understand this effect, first consider the sequential sampling framework (Forstmann et al., 2016; Gold & Shadlen, 2007; Smith & Ratcliff, 2004), where RTs are explained by accumulation of a decision variable to threshold. The decision variable takes account of sensory evidence, but it can also include other factors; importantly, this decision variable can accumulate based on a physical stimulus, as well as the memory trace of one (Gold & Shadlen, 2007).



#### Figure 4.1 Signal duration

**a)** Illustrating manipulations to signal duration. Audiovisual signals (left) were identical to the simple-consistent stimuli in **Chapter 3 (Figure 3.3, top left quadrant)**. The green lines illustrate the signal onset/offset for a given trial. In short signal conditions, the signal lasted for 0.05 s. In long signal conditions, the signal lasted until the participant responded, or until the maximum valid response time allowed by the response window (1.5 s).

**b, c)** History effect and history index from a pilot study. All bars show the mean of 5 participants ( $\pm 1$  SEM).

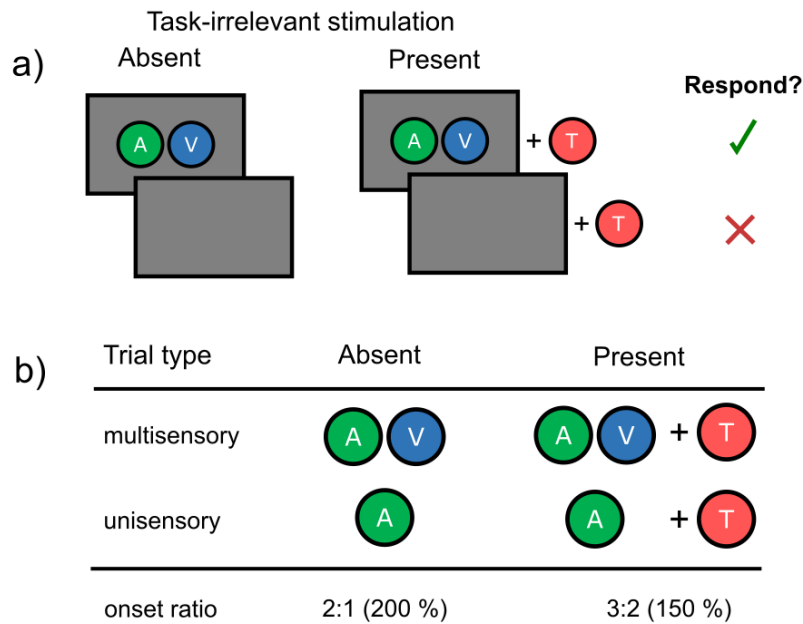
Now, consider how signal duration and attention might interact to impact accumulation. First, assume a signal is long (i.e. presented until a response is made, as in **Chapter 3**). If the current trial is a unisensory switch (e.g. audition to vision), the observer experiences a time-cost of switching attention from one modality to another. Once this switch is made, however, accumulation can continue unaffected, as the signal is still physically present. Next, consider a short signal duration (e.g. 0.05 s, as in Spence et al., 2001). Here, the observer experiences the same time-cost of switching attention between modalities. However, in addition, the signal will likely have been offset by the time attention has moved between modalities. Accumulation of the decision variable must therefore proceed on a memory trace, rather than activity directly linked to the physical signal. Assuming that the memory trace of the stimulus can only be less reliable, it is expected that the corresponding accumulation will also be less reliable, with additional time-costs of retrieving the signal from sensory memory. Following this reasoning, it would be expected that short signals elicit a greater overall cost of switching between modalities.

Unfortunately, as few studies have explicitly reported modality switch costs in the RSP, it is difficult to say whether this is more than just speculation. To further investigate the possible role of signal duration therefore, and to provide a clear motivation for including this factor, I conducted a small pilot study ( $N=5$ ; **Figure 4.1**). Here I used the simple signals from **Chapter 3** (see **Figure 3.3**; top left). Both the history effect (**Figure 4.1b**) and history index (**Figure 4.1c**) were observed to be larger with short signals (max. duration: 0.05 s) than long signals (max. duration: 1.5 s). Signal duration, therefore, was included as a factor which was expected to manipulate history effects.

#### 4.1.2. Factor 2: Task-Irrelevant Stimulation

A second interaction to target is context variance. In the RSP, context variance is imposed because there is always a difference in stimulation between unisensory and redundant trials. In unisensory trials, only one decision-unit accumulates alone. On redundant trials, however, both decision-units accumulate, each alongside the other. This change in context likely affects the accumulation process related to each unisensory signal component; one possibility is that accumulation by one decision-unit decreases the reliability of the other (i.e. introduces more *noise* into accumulation on redundant trials). Following a race model framework, a measure of context variance is provided by violations of Miller's bound; this is the upper limit to race models assuming context invariance. Larger violations, therefore, suggest larger context variance between unisensory and redundant trials. Accordingly, any factor which reduces context variance between unisensory and redundant trials (i.e. makes stimulation more similar over all trials) should reduce violations of Miller's bound.

As a second factor, I consider *task-irrelevant stimulation* as a potential manipulation of context variance. The motivation here is that an additional task-irrelevant stimulus will make trials more similar overall. For instance, in **Chapter 3**, this was achieved by adding background stimulation on all trials; signals with background stimulation (complex) were associated with smaller violations than signals without background stimulation (simple). In this experiment, however, I present task-irrelevant stimulation in a third modality. To demonstrate how this might work, recall the typical RSP (**Section 1.3.1**) with AV signals. Four trial types are presented throughout: two unisensory (A, V), redundant (AV), and catch (no signal). The change in stimulation between redundant and unisensory trials (i.e. the ratio of stimulus onsets) is 2:1. Now, consider what happens when a task-irrelevant stimulus in a third modality is added across conditions (tactile; T). Again, four trial types are presented throughout: two unisensory (A+T, V+T), redundant (AV+T), and catch (T). There is no change in the number of task-relevant signals; the presence of A or V determines whether a participant responds, whereas the presence of T does not (**Figure 4.2a**). However, the change in stimulation between redundant and unisensory trials (i.e. the ratio of stimulus onsets) is now 3:2 (**Figure 4.2b**). Overall, therefore, with the



**Figure 4.2 Task-irrelevant stimulation**

**a) Task demands.** In the left column (absent), typical RSP trials are presented. The participant responds so long as either A or V is present, and withholds responses when neither is present. In the right column, the same trials are presented with an additional tactile stimulus (T). As in the left column, participants respond only when A or V is present and withhold responses when neither is present. The tactile stimulus is thus irrelevant to correctly complete the task.

**b) Trial similarity.** When stimulation is absent, there is a larger dissimilarity, and thus a larger context variance, between multisensory and unisensory trials than when stimulation is present (as shown by the larger onset ratio).

same number of task-relevant signals, stimulation is more similar across trials. Additional task-irrelevant stimulation, therefore, should reduce the size of violations by reducing context variance.

Additionally, it might also be expected that history effects change according to task-irrelevant stimulation. For instance, a similar manipulation was introduced by Gondan, Vorberg, and Greenlee (2007) to remove modality switch effects seen in neuroelectrophysiological recordings. Drawing also on the attention literature, previous research has shown non-relevant auditory cues can facilitate attention-switching within a visual task (Doyle & Snowden, 2001); the original authors hypothesise that additional stimulation acts to ‘reset’ attentional biases, which allows for quicker attention-switching within one modality. If such effects extend to switching *between* modalities, then history effects may be smaller when task-irrelevant stimulation is present compared to absent.

#### 4.1.3. Experimental Hypotheses

I began with the following hypotheses. Regarding the first factor, *signal duration*, it was hypothesised based on the pilot data (**Figure 4.1**) that history effects should be larger for short signals compared to long signals. Regarding the second factor, *task-irrelevant stimulation*, it was hypothesised that (if a

task-irrelevant stimulus reduces context variance) violations of Miller's bound would be smaller when task-irrelevant stimulation was present compared to absent. To further test whether any results confirming these hypotheses were reliable, I conducted two versions of the same experiment. In the first instance (Part 1) I used audio-visual (AV) signals with a tactile (T) task-irrelevant stimulus, as described in the introduction. In the second instance (Part 2) I switched the stimulation to produce tactile-visual (TV) signals with an auditory (A) task-irrelevant stimulus.

## 4.2. Methods (Part 1)

### 4.2.1. Participants

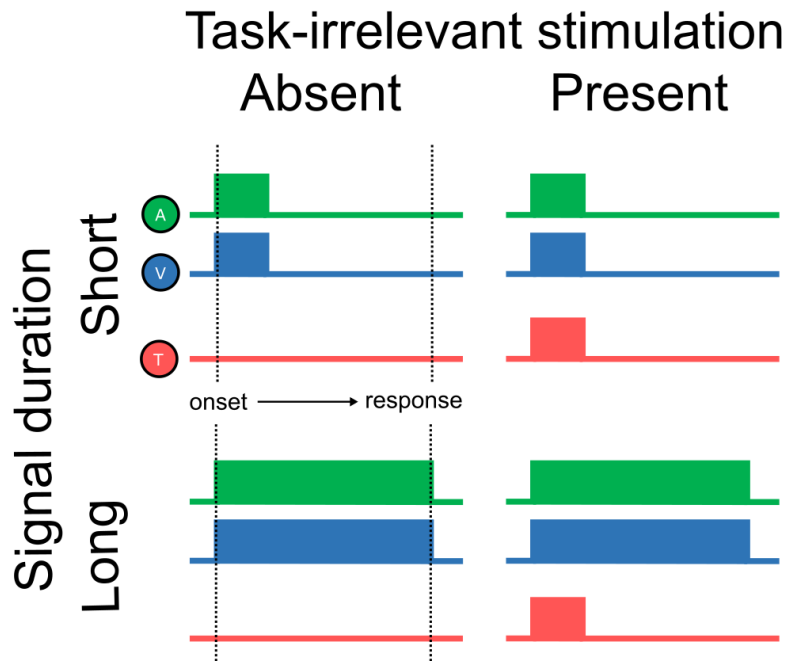
In contrast to **Chapter 3**, a miss rate exclusion criterion (>10% missed trials in any condition) was also implemented. This was to further ensure the ceiling performance assumption was met for each participant included in the analysis. 21 participants were initially recruited at the University of St. Andrews. Of these participants, one met the miss rate exclusion criterion. The data from 20 participants (15 female) were thus analysed. The age of these participants ranged from 19-38 years. All were naïve as to the aims of the experiment. Normal hearing and normal/corrected-to-normal vision was reported in all cases. Reimbursement was £10.

### 4.2.2. Stimulus Design

Stimuli were constructed according to a 2×2 design (**Figure 4.3**) with factors *signal duration* (short, long) and *task-irrelevant stimulation* (absent, present). Signals were identical to those presented in **Chapter 3** (simple-consistent condition). Simple signals were used as they previously elicited the largest interactions overall. For signal duration, short signals were presented for 0.05 s, whereas long signals were presented until a response was recorded (maximum duration: 1.5 s). For task-irrelevant stimulation, absent conditions included no additional stimulation (i.e. standard RSP), whereas present conditions also included a task-irrelevant tactile signal on all trials (always 0.05 s).

Tactile stimulation was delivered via the C-2 Tactor (Engineering Acoustics, Inc.). This device delivers stimulation through vibration of a central contactor which is placed in proximity to the skin. The device is optimised to deliver stimuli between 200-300 Hz, therefore stimulation was set to 250 Hz. Unfortunately, a precise measure of amplitude for stimulation could not be determined using an oscilloscope, due to artefacts created by the amplifier. This device was secured to the participant's wrist (**Figure 4.4**). The device was always placed on the opposite wrist to the hand which provided responses. As the tactile device was driven by the same soundcard as the auditory stimulus, the same auditory delay was applied to all tactile stimulation to achieve synchrony (see **Section 2.1.4**). To remove the possibility that participants responded to sounds made by the vibration of the tactor, background noise with edge frequencies covering the stimulation range (200-300 Hz; 2<sup>nd</sup> order





**Figure 4.3 Stimulus design for Chapter 4 (Part 1)**

In each plot, bars indicate the onset/offset of signals (A, V) and the task-irrelevant stimulus (T). Illustrations depict a redundant signal trial.

Butterworth filter applied) was played continuously throughout trials. The volume of the background noise was 45 dB SPL.

#### 4.2.3. Procedure

Each trial began by presenting the green fixation point, followed by a random foreperiod. Duration of the foreperiod was determined by two components: a fixed duration (1.9 s) and a random component drawn from an exponential distribution (mean: 0.25 s). In contrast to **Chapter 3**, the fixed component here was larger and the exponential component was smaller. This was introduced to allow participants more time between responses within a 2 hour experiment run-time. On signal trials, one of the three signals (A, V, AV) was then presented until the participant responded (maximum duration: 1.5 s). On catch trials, no signal was presented but the trial continued for the maximum signal duration (1.5 s). To indicate the end of a trial, the fixation point turned to red and all stimulation was halted for 0.1 s. The next trial was then initiated.

In contrast to **Chapter 3**, the first 1 s of the foreperiod was considered a 'grace period' in which a response by the participant would not trigger a false alarm and end the trial. This was introduced as it was noted that participants in the previous experiment sometimes erroneously responded to the onset of the end-of-trial feedback (i.e. the fixation point colour change). As the feedback duration was very brief, these responses were actually executed very early into the foreperiod of the following trial.



**Figure 4.4 Placement of the tactor**

Participants wore the C-2 tactor on their wrist secured via a velcro strap. The tactor was placed with the central contactor face down (towards the skin) between two layers of the wrist strap fabric, as shown above.

This meant that these responses were recorded as genuine false alarms. Though the number of these feedback responses was assumed to be very small (given the low number of false alarms in **Chapter 3**), they may still have inflated the overall percentage of false alarms. By introducing this grace period, therefore, the error response was still recorded (and thus could be analysed separately) but the participant was allowed to continue the trial. As expected, the average number of these responses for each participant ( $11.70 \pm 1.89$ ) was very low in comparison to the 1600 RSP trials presented overall (i.e.  $< 1\%$  of all trials). Further, the average time of these presses was consistent ( $0.250 \pm 0.016$  s into the foreperiod) and very early into the trial. As these responses occurred so early, and with little variability between participants, this indicated that they were likely erroneous responses to the fixation feedback. Further, because they occurred so early into the trial, there was still a considerable foreperiod duration remaining until the next possible signal onset, so any RTs on these trials are unlikely to be affected. As such, these responses were not any considered further in this thesis.

Responses recorded within 1.5 s of a signal onset were initially considered valid. If the response was in error, a feedback screen was then presented for 2 s which informed the participant of the error. Responses recorded within the foreperiod (or at any point during catch trials) were considered false alarms. Signal trials in which no response was recorded within the valid response window were considered misses.

Trials were randomly presented in blocks with 100 trials each (25 trials per signal type, 25 catch trials). A dummy trial (AV) was also inserted at the start of each block. Trials in which a miss was recorded were replaced within the trial order, which was reshuffled after every error. A dummy trial was also inserted before experimental trials following a miss. In contrast to **Chapter 3**, these were drawn randomly from each of the 4 trial types (with equal probability). The proportion of different

trials in the experiment, therefore, was now constant. Dummy trials were not included in analyses. As described in **Chapter 3**, a small pointer (attached to fixation) denoted progress through a block. Each block lasted for around 5-6 minutes.

The entire experiment consisted of 16 blocks (4 blocks for each of the 4 experimental conditions). Block order was randomised (intermixing all conditions) for each participant according to a Latin square procedure. Breaks were given between blocks as required. The whole experimental session lasted for around 120 minutes.

#### 4.3. Results (Part 1)

For all repeated-measures ANOVAs presented in this section, I adhere to the following ordering of factors: signal duration (short, long)  $\times$  task-irrelevant stimulation (absent, present). In some cases an additional third factor was included: signal modality (A, V, [AV]).

##### 4.3.1. General Performance

The performance of participants was assessed before RT analysis to ensure the ceiling performance assumption was met. This was also used to provide an indication of whether participants' responses were affected by the additional task-irrelevant stimulation. First, the percentage of false alarms was calculated for signal and catch trials. The percentage of false alarms was averaged across all 3 signal trials (A, V, AV). This was because, by definition, false alarms on signals trials occur in the foreperiod (i.e. before the onset of any stimulation). For signal trials, false alarms occurred on only 0.070% ( $\pm 0.023\%$ ) of trials. A  $2 \times 2$  ANOVA revealed no significant main or interaction effects (all  $F \leq 1.479$ ,  $p \geq 0.239$ ,  $\eta p^2 \leq 0.072$ ). For catch trials, false alarms occurred slightly more often, on 1.843% ( $\pm 0.449\%$ ) of trials. A  $2 \times 2$  ANOVA revealed a significant main effect of task-irrelevant stimulation,  $F(1, 19) = 14.464$ ,  $p = 0.001$ ,  $\eta p^2 = 0.432$ . The percentage of false alarms was smaller for absent conditions (0.173%  $\pm 0.065\%$ ) than present conditions (3.513%  $\pm 0.886\%$ ). No other main effects were significant (all  $F \leq 3.382$ ,  $p \geq 0.082$ ,  $\eta p^2 \leq 0.151$ ). Overall, therefore, participants made more false alarms on catch trials when task-irrelevant stimulation was present. These false alarms are likely erroneous responses to the task-irrelevant stimulus onset.

Second, misses were evaluated. The percentage of missed trials averaged across signal modalities was only 0.392% ( $\pm 0.131\%$ ). A  $2 \times 2 \times 3$  ANOVA revealed a significant main effect of signal duration,  $F(1, 19) = 10.176$ ,  $p = 0.005$ ,  $\eta p^2 = 0.349$ . The percentage of misses was larger for short conditions (0.684  $\pm 0.221\%$ ) than long conditions (0.099  $\pm 0.047\%$ ). This is likely because the signal was presented for less time, therefore it was easier for participants to miss (e.g. by blinking, for visual signals). There was also a significant main effect of signal modality,  $F(1.184, 22.497) = 7.823$ ,  $p = 0.008$ ,  $\eta p^2 = 0.292$ . Pairwise comparisons revealed that the percentage of misses on visual trials (0.814  $\pm 0.273\%$ ) was larger than on auditory trials (0.323  $\pm 0.125\%$ ,  $p = 0.041$ ) and on redundant trials (0.038

$\pm 0.021\%$ ,  $p=0.010$ ). The percentage of misses on auditory trials however was not significantly different to redundant trials ( $p>0.05$ ). One possibility is that participants were more likely to miss visual signals by blinking (particularly in short signal conditions). Finally, there was a significant interaction between signal duration and signal modality,  $F(1.408, 26.745)=7.070$ ,  $p=0.007$ ,  $\eta p^2=0.271$ . For short conditions, there was a larger difference between the percentage of misses on auditory trials ( $0.621 \pm 0.250\%$ ) and visual trials ( $1.382 \pm 0.445\%$ ), both of which were larger than redundant trials ( $0.050 \pm 0.034\%$ ). For long conditions, there was a smaller difference between the percentage of misses on auditory trials ( $0.025 \pm 0.025\%$ ) and visual trials ( $0.245 \pm 0.115\%$ ), both of which were much closer to the percentage of misses on redundant trials ( $0.025 \pm 0.025\%$ ). No other main or interaction effects were significant (all  $F \leq 2.628$ ,  $p \geq 0.085$ ,  $\eta p^2 \leq 0.121$ ). Overall, the redundant condition consistently showed a low percentage of misses regardless of signal duration; given that two signals were always presented on these trials, it is likely that longer durations did not help participants much in terms of improving detection as performance was already so close to ceiling. On unisensory trials however, where there is only one signal, the larger percentage of misses for short signals compared to long would be expected, as a longer duration would make the signals harder to miss (e.g. if attention was initially directed to the wrong modality).

Third, the number of outliers was calculated according to the  $\pm 3 \times 1.4826$  MAD criterion (see **Section 2.2.2**). This was done on the  $1/RT$  transformed data, to which a modelling approach is later applied. On average, fast outliers occurred on  $0.39\%$  ( $\pm 0.05\%$ ) of RTs, whereas slow outliers occurred on  $0.55\%$  ( $\pm 0.10\%$ ) of RTs. The overall percentage of outliers ( $\sim 0.94\%$ ), therefore, is only slightly larger than what would be expected if the  $1/RT$  data followed a typical normal distribution ( $0.27\%$ ). Across all participants, 23664 valid RTs were included in the main analyses, with a mean of  $98.6$  ( $\pm 0.17$ ) valid RTs per participant for each condition (of a possible maximum of 100 RTs).

Overall, though there were small differences in false alarms and misses across conditions, the data indicate close to ceiling performance in all conditions. Further, the number of valid trials included in the analysis was close to maximum. In general, therefore, accuracy was not further considered. One exception to this is the observation that task-irrelevant stimulation leads to more false alarms as measured in catch trials, which will be reconsidered in a later section (see **Section 4.8**).

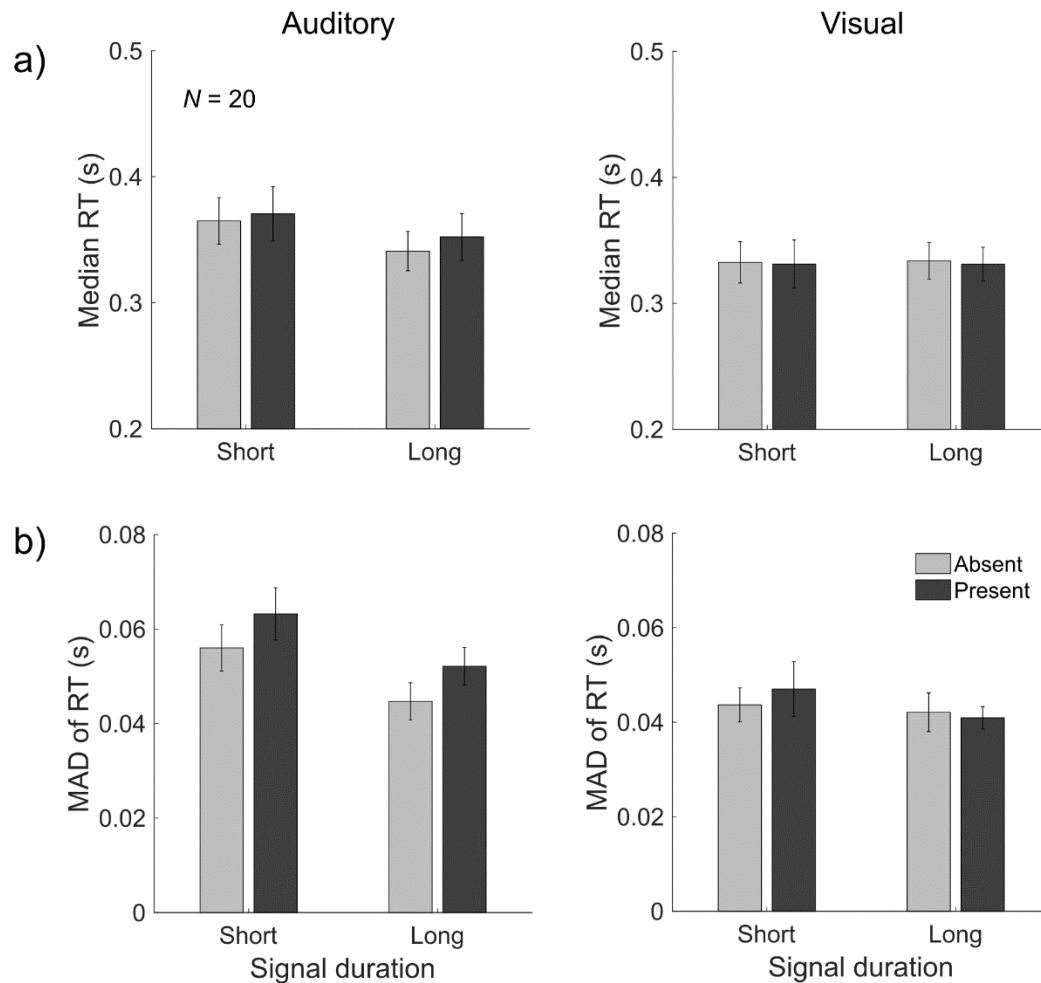
#### 4.3.2. Unisensory RTs

Prior to applying the comparative approach, unisensory RTs were analysed to give an indication of how the experimental factors affected RTs. This was also used to give an indication of possible changes in benefits. According to the race model principles (see **Section 1.6.1**), benefits should increase if unisensory RTs have similar distributions (equal effectiveness). This was assessed by comparing the central tendency (median) and variability (MAD) of unisensory RTs. Further, race model principles

state that benefits should increase as RT variability increases (variability rule). This was assessed by examining the MAD of unisensory RTs.

For median RT (**Figure 4.5a**), a 2×2×2 ANOVA revealed a significant main effect of signal modality,  $F(1, 19)=8.153$ ,  $p=0.010$ ,  $\eta p^2=0.300$ . Median RT was larger for auditory trials ( $0.357 \pm 0.018$  s) compared to visual trials ( $0.332 \pm 0.016$  s). Next, there was a significant interaction between signal duration and signal modality,  $F(1, 19)=15.280$ ,  $p=0.001$ ,  $\eta p^2=0.446$ . For short conditions, there was a larger difference in the median RT for auditory ( $0.368 \pm 0.020$  s) and visual ( $0.332 \pm 0.018$  s) trials. For long conditions, however, the median RT for auditory ( $0.347 \pm 0.017$  s) and visual ( $0.332 \pm 0.014$  s) trials was more similar. There was also a significant interaction between task-irrelevant stimulation and signal modality,  $F(1, 19)=6.608$ ,  $p=0.019$ ,  $\eta p^2=0.258$ . For auditory conditions, there was a difference in median RT between absent ( $0.353 \pm 0.017$  s) and present ( $0.361 \pm 0.020$  s) trials. For visual conditions, however, median RT for absent ( $0.333 \pm 0.015$  s) and present ( $0.331 \pm 0.016$  s) was similar. There were no further significant main or interaction effects (all  $F \leq 3.492$ ,  $p \geq 0.077$ ,  $\eta p^2 \leq 0.155$ ). Overall, as there is a main effect of signal modality, this suggests that unisensory performance is not equally effective. There was also no main effect for either of the two experimental manipulations, which suggests the pattern of unisensory RTs was broadly similar across conditions. However, as auditory and visual trials had a similar median in conditions where task-irrelevant stimulation was present, the corresponding predicted benefit for this condition might also be larger.

For MAD of RT (**Figure 4.5b**), a 2×2×2 ANOVA revealed a significant main effect of signal duration,  $F(1, 19)=11.238$ ,  $p=0.003$ ,  $\eta p^2=0.372$ . MAD of RT was larger in short conditions ( $0.052 \pm 0.004$  s) compared to long conditions ( $0.045 \pm 0.003$  s). There was also a significant main effect of task-irrelevant stimulation,  $F(1, 19)=10.959$ ,  $p=0.004$ ,  $\eta p^2=0.366$ . MAD of RT was smaller in absent conditions ( $0.047 \pm 0.004$  s) compared to present conditions ( $0.051 \pm 0.004$  s). There was also a significant main effect of signal modality,  $F(1, 19)=11.746$ ,  $p=0.003$ ,  $\eta p^2=0.382$ . MAD of RT was larger for auditory trials ( $0.054 \pm 0.004$  s) compared to visual trials ( $0.043 \pm 0.004$  s). Next, there was a significant interaction between signal duration and signal modality,  $F(1, 19)=4.440$ ,  $p=0.049$ ,  $\eta p^2=0.189$ . In short conditions, MAD of RT for auditory ( $0.060 \pm 0.005$  s) and visual ( $0.045 \pm 0.005$  s) trials was quite different. Considering long conditions, however, MAD of RT for auditory ( $0.048 \pm 0.004$  s) and visual ( $0.042 \pm 0.003$  s) trials was more similar. There was also a significant interaction between task-irrelevant stimulation and signal modality,  $F(1, 19)=7.672$ ,  $p=0.012$ ,  $\eta p^2=0.288$ . Considering auditory conditions, MAD of RT for absent ( $0.050 \pm 0.004$  s) and present ( $0.058 \pm 0.004$  s) conditions were slightly different. Considering visual conditions, however, MAD of RT for absent ( $0.043 \pm 0.004$  s) and present ( $0.044 \pm 0.004$  s) conditions was more similar. No other main or interaction effects were significant (all  $F \leq 0.553$ ,  $p \geq 0.466$ ,  $\eta p^2 \leq 0.028$ ). Overall, as for median RT, there were differences in the



**Figure 4.5 Unisensory RT analysis (Part 1)**

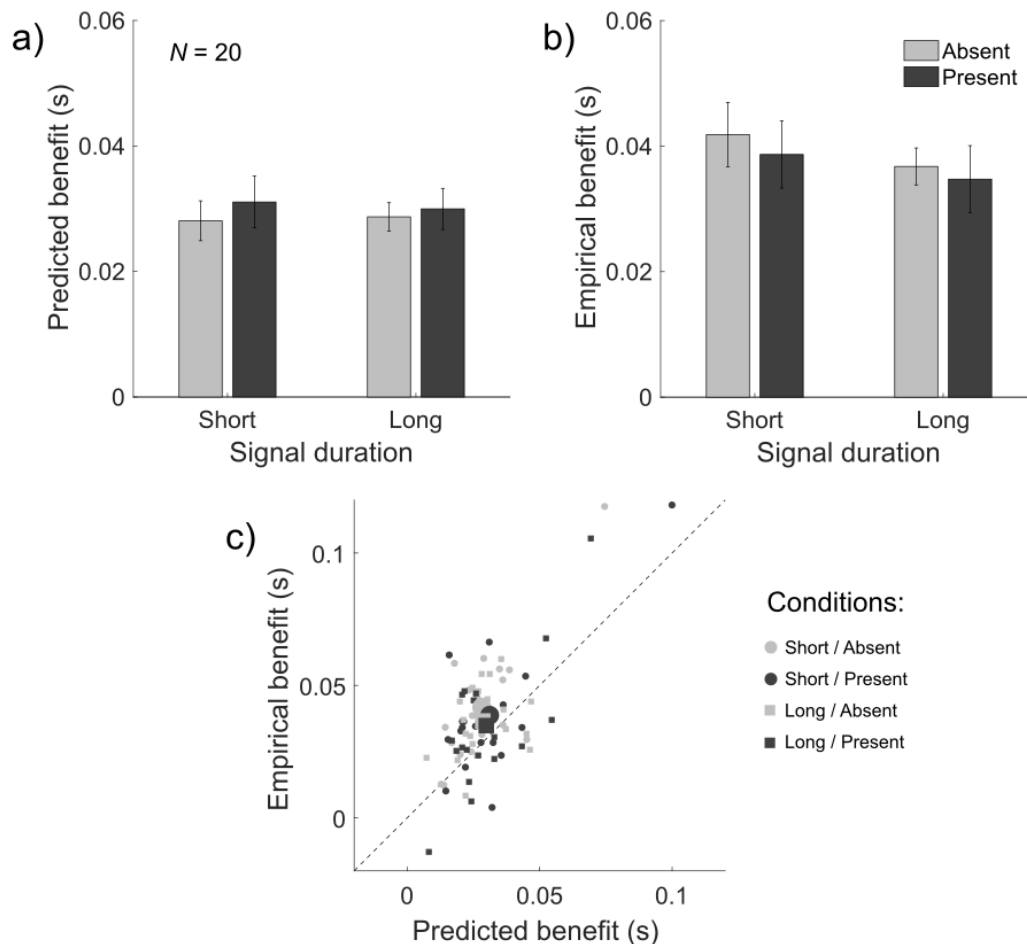
**a, b)** Median RT and MAD of RT across conditions. The left column shows auditory RTs and the right column shows visual RTs. All bars shown the mean of 20 participants ( $\pm 1$  SEM).

MAD of RT between unisensory conditions, suggesting that performance is not equally effective. As MAD was overall larger for short conditions and present conditions, however, it may be that these conditions show corresponding increases in benefit.

Overall, as there are interactions between factors for unisensory RTs, it is difficult to make straightforward directional predictions for benefits at a glance. In addition, for median RT, there was a main effect of signal modality. Unisensory performance was therefore not equally effective in general; auditory RTs were slower and more variable than visual trials. According to a race model framework, this means that changes in benefits could be small.

#### 4.3.3. Comparative Approach Step 1: Multisensory Benefits

Following the analysis of unisensory RTs, the first step of the comparative approach is to understand multisensory RTs. So far, consideration of unisensory RTs in relation to the race model principles has not suggested a clear directional prediction for benefits. To obtain a straightforward quantitative



**Figure 4.6 Predicting and measuring multisensory benefits (Part 1)**

**a, b)** Predicted and empirical benefits across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

**c)** Empirical benefit as a function of predicted benefit. Each point represents data from one participant in one of the four conditions (80 data points total). Large symbols show the group mean.

prediction, I applied the simple race model, which accounts for both principles in a single, parameter-free estimate of benefit. I then calculated the empirical benefit, and compared this to the simple race model prediction.

First, benefits were predicted (**Figure 4.6a**) according to the simple race model (see **Section 2.3.2.2**). A  $2 \times 2$  ANOVA revealed no main or interaction effects (all  $F \leq 1.325$ ,  $p \geq 0.264$ ,  $\eta p^2 \leq 0.065$ ). Therefore, benefit was predicted not to change across conditions. This is broadly in line with the lack of equal effectiveness seen in unisensory RTs.

Next, benefits were measured empirically (see **Section 2.3.2.1**) to see if the results followed the pattern predicted by the simple race model (**Figure 4.6b**). A  $2 \times 2$  ANOVA revealed no main or interaction effects (all  $F \leq 1.271$ ,  $p \geq 0.274$ ,  $\eta p^2 \leq 0.063$ ). Therefore, in line with the predictions of the simple race model, empirical benefits did not change across conditions.

Following the previous chapter (**Chapter 3**), I then applied additional analyses to further investigate predicted and empirical benefits. First, the correspondence between predicted and empirical benefits was evaluated on each participant's data. To do this, predicted and empirical benefits were correlated for each condition of the 2x2 design. All correlation values were positive, and 3 of 4 were highly significant (**Table 4.1**). Second, differences between predicted and empirical benefit were assessed for each participant. For an initial coarse understanding of the difference, predicted and empirical benefits were each averaged across conditions and compared with a paired-samples *t*-test. The mean predicted benefit ( $0.030 \pm 0.003s$ ) was overall smaller than the mean empirical benefit ( $0.038 \pm 0.004 s$ ),  $t(19)=2.898$ ,  $p=0.009$ , two-tailed. To assess if this difference varied across conditions, predicted benefits were then subtracted from empirical benefits for each condition. Interestingly, a 2x2 ANOVA revealed a significant main effect of task-irrelevant stimulation,  $F(1, 19)=14.309$ ,  $p=0.001$ ,  $\eta p^2=0.430$ . The difference between empirical and predicted benefit was larger for absent conditions ( $0.011 \pm 0.003s$ ) compared to present conditions ( $0.006 \pm 0.003s$ ). There were no further main or interaction effects (all  $F \leq 3.182$ ,  $p \geq 0.090$ ,  $\eta p^2 \leq 0.143$ ).

Overall, this analysis has shown that the simple race model predicts empirical benefits well. However, it has also shown that predicted benefits fall short of empirical benefits. An important finding, however, is that the presence of task-irrelevant stimulation appears to result in benefits which more closely follow the simple race model prediction. This can be understood in more detail by considering changes in interactions.

#### 4.3.4. Comparative Approach Step 2: Quantifying Interactions

The second step of the comparative approach is to quantify interactions. These cannot be accounted for by the basic race model architecture, therefore changes in interactions may eventually suggest a reason for any discrepancy between predicted and empirical benefit. The overall aim of this chapter was to target both interactions. The first factor, signal duration, was introduced to manipulate history effects. The second factor, task-irrelevant stimulation, was introduced to manipulate context variance (as measured by violations of Miller's bound).

**Table 4.1 Pearson correlation coefficients (and *p* values) comparing predicted and empirical benefits (Part 1)**

Signal duration	Predicted vs empirical	
	Task-irrelevant stimulation	
	Absent	Present
Short	0.790 ( $p < 0.001$ )	0.739 ( $p < 0.001$ )
Long	0.274 ( $p = 0.243$ )	0.725 ( $p < 0.001$ )

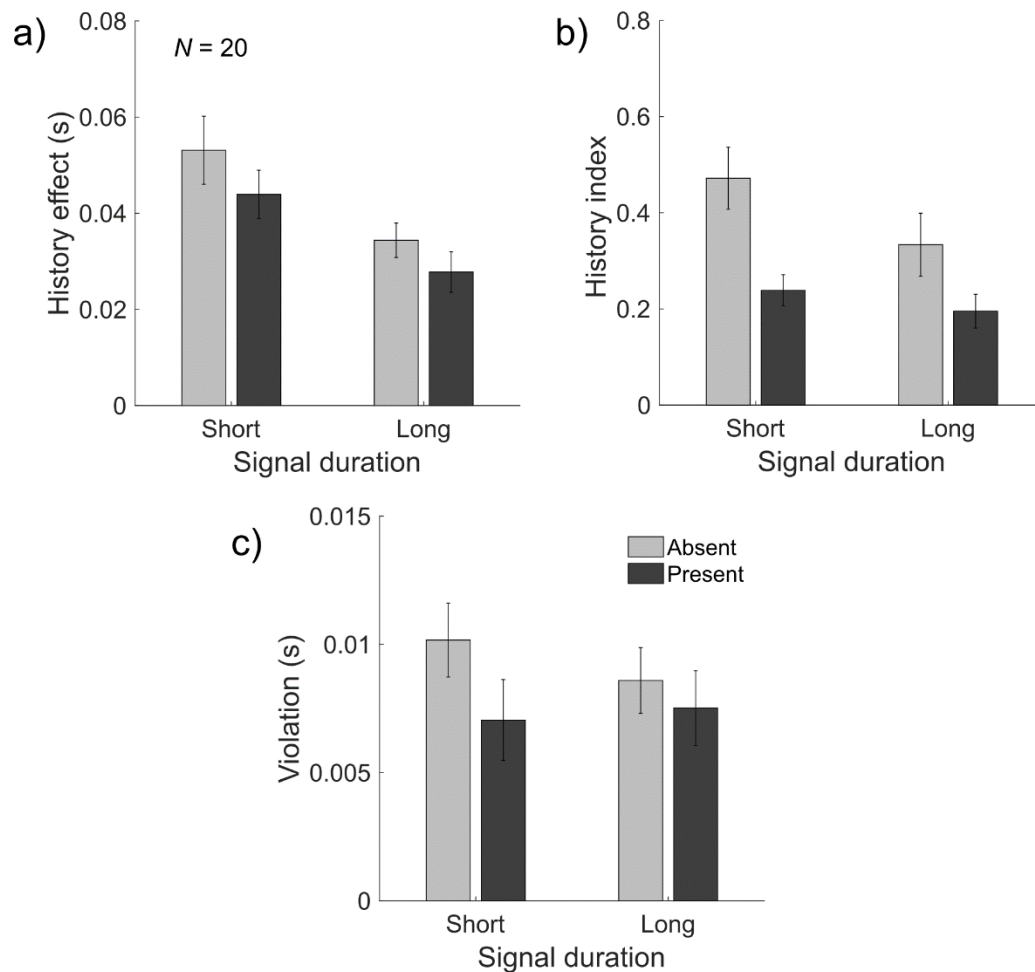


First, history effects (**Figure 4.7a**) were calculated (see **Section 2.3.3.1**). A 2×2 ANOVA revealed a significant intercept,  $F(1, 19)=140.644$ ,  $p<0.001$ ,  $\eta p^2=0.881$ . Thus, the overall history effect ( $0.040 \pm 0.003$  s) was significantly greater than 0, and must be accounted for. As expected, there was a significant main effect of signal duration,  $F(1, 19)=8.345$ ,  $p=0.009$ ,  $\eta p^2=0.305$ . As hypothesised, and as found in the pilot study (**Figure 4.1**), the history effect was larger in short conditions ( $0.049 \pm 0.006$  s) compared to long conditions ( $0.031 \pm 0.003$  s). Further, there was also a marginal (i.e.  $p=0.05$ ) main effect of task-irrelevant stimulation,  $F(1, 19)=4.374$ ,  $p=0.050$ ,  $\eta p^2=0.187$ . The history effect was larger in absent conditions ( $0.044 \pm 0.005$  s) compared to present conditions ( $0.036 \pm 0.003$  s). There were no further significant main or interaction effects (all  $F \leq 0.160$ ,  $p \geq 0.694$ ,  $\eta p^2 \leq 0.008$ ).

As a further evaluation, the history index (**Figure 4.7b**) was calculated (see **Section 2.3.3.1**). As a brief reminder, the history index normalises the history effect by the variance of unisensory RTs. This measures how much the history effect contributed to RT variability (and thus overall benefit) in each condition. As for the history effect, there was also a significant main effect of signal duration,  $F(1, 19)=5.973$ ,  $p=0.024$ ,  $\eta p^2=0.239$ . As expected from the pilot study (**Figure 4.1**), the history index was larger in short conditions ( $0.355 \pm 0.040$ ) compared to long conditions ( $0.264 \pm 0.044$ ). In short conditions, therefore, the history effect contributed more to overall variability than long conditions. There was also a significant main effect of task-irrelevant stimulation,  $F(1, 19)=15.568$ ,  $p=0.001$ ,  $\eta p^2=0.450$ . The history index was larger in absent conditions ( $0.403 \pm 0.056$ ) compared to present conditions ( $0.217 \pm 0.029$ ). The history effect, therefore, contributed to overall variability more in absent compared to present conditions. There were no further significant main or interaction effects (all  $F \leq 1.574$ ,  $p \geq 0.225$ ,  $\eta p^2 \leq 0.077$ ).

Second, violations of Miller's bound (2.3.3.2) were considered (**Figure 4.7c**). A 2×2 ANOVA revealed a significant intercept,  $F(1, 19)=56.206$ ,  $p<0.001$ ,  $\eta p^2=0.747$ . Thus, the overall violation area ( $0.008 \pm 0.001$  s) was significantly greater than 0, and must be accounted for. As expected, there was a significant main effect of task-irrelevant stimulation,  $F(1, 19)=8.020$ ,  $p=0.011$ ,  $\eta p^2=0.297$ . As hypothesised, violation was larger for absent conditions ( $0.009 \pm 0.001$ s) than for present conditions ( $0.007 \pm 0.001$ s). There were no further significant main or interaction effects (all  $F \leq 0.637$ ,  $p \geq 0.435$ ,  $\eta p^2 \leq 0.032$ ).

Overall, this analysis revealed the presence of both interactions. As hypothesised, history effects were reduced by making signals longer, and violations of Miller's bound were reduced by introducing task-irrelevant stimulation. Interestingly, an effect of task-irrelevant stimulation occurred for both history effects and violation; both of which were weaker when task-irrelevant stimulation was present compared to absent. As these two interactions contribute to benefits beyond the basic



**Figure 4.7 Quantifying interactions (Part 1)**

**a, b)** History effect and history index across conditions.

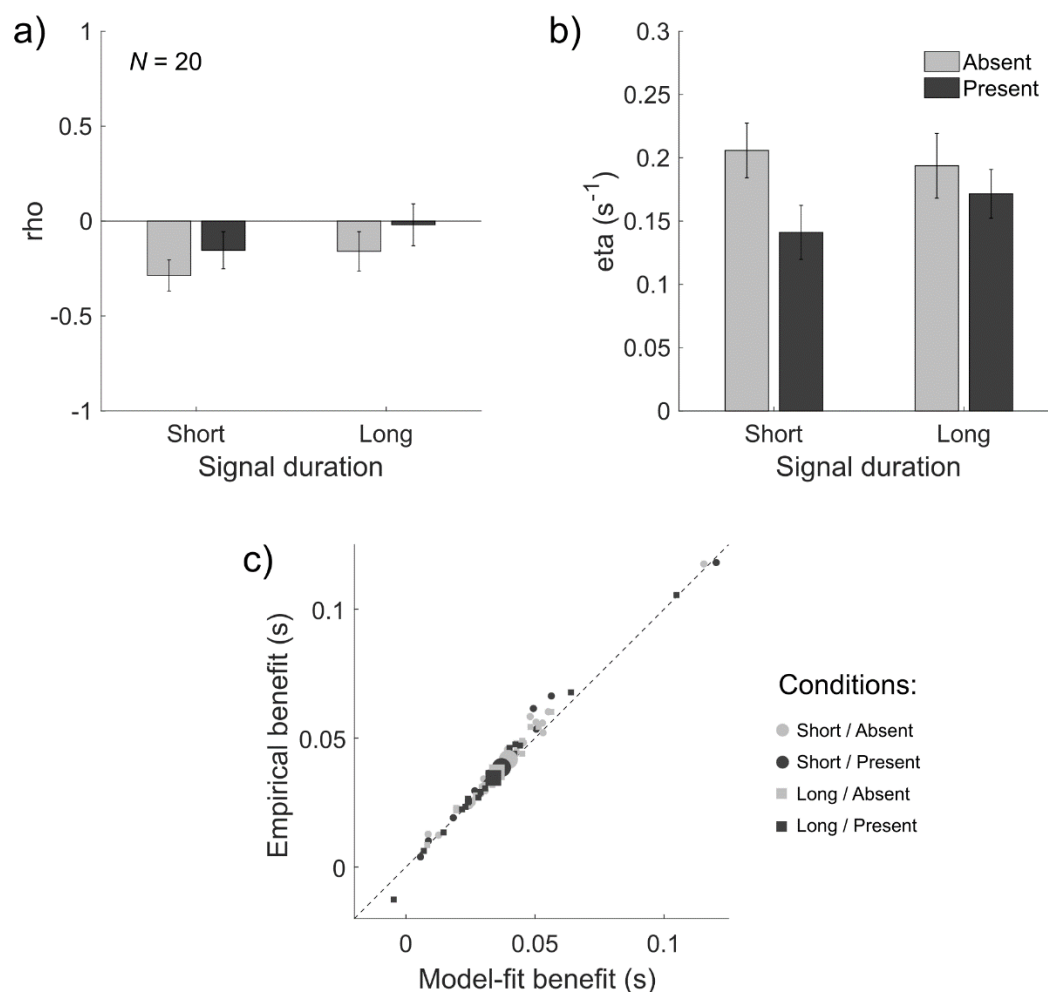
**c)** Violation of Miller's bound across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

race mechanism, this explains why empirical benefits also come closer to predicted benefits when task-irrelevant stimulation is present.

#### 4.3.5. Comparative Approach Step 3: Applying the Context Variant Race Model

For the third step of the comparative approach, I applied the context variant race model to account for changes in benefits and interactions. This includes two interaction parameters: *rho* (which accounts for history effects) and *eta* (which accounts for violations of Miller's bound). Broadly, therefore, it is expected that the changes in interactions are also observed in the corresponding model parameters.

Analysing the *rho* parameters (**Figure 4.8a**), a  $2 \times 2$  ANOVA did not reveal a significant intercept,  $F(1, 19) = 3.309$ ,  $p = 0.085$ ,  $\eta p^2 = 0.148$ . Thus, the overall correlation modelled between unisensory rate distributions ( $-0.155 \pm 0.085$ ) was not significantly different to 0. However, there was a significant main effect of task-irrelevant stimulation,  $F(1, 19) = 14.412$ ,  $p = 0.001$ ,  $\eta p^2 = 0.431$ . The correlation was more



#### Figure 4.8 Fitting the context variant race model (Part 1)

**a, b)** Best fitting  $\rho$  and  $\eta$  values across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

**c)** Empirical benefit as a function of model-fit benefit. Each point represents data from one participant in one of the four conditions (80 data points total). Large symbols show the group mean.

negative in absent conditions ( $-0.223 \pm 0.082$ ) than present conditions ( $-0.087 \pm 0.091$ ). Interestingly, there was no significant main effect of signal duration as seen for history effect. There were also no further interactions (all  $F \leq 3.777$ ,  $p \geq 0.067$ ,  $\eta p^2 \leq 0.166$ ). Overall, the main effects observed for  $\rho$  do not fully correspond to those seen for the history effect or the history index. Notably, however, the overall pattern in  $\rho$  (Figure 4.8a) follows that of the history index (Figure 4.7b) as expected; a stronger negative correlation is found in conditions with a larger history index.

Analysing the  $\eta$  parameters (Figure 4.8b), a  $2 \times 2$  ANOVA revealed a significant intercept,  $F(1, 19) = 128.152$ ,  $p < 0.001$ ,  $\eta p^2 = 0.871$ . Thus, additional variability was added to unisensory rate distributions to account for the variability of redundant rate distributions ( $0.178 \pm 0.016 s^{-1}$ ). There was also a main effect of task-irrelevant stimulation,  $F(1, 19) = 9.004$ ,  $p = 0.007$ ,  $\eta p^2 = 0.322$ . The  $\eta$

parameter was larger for absent conditions ( $0.200 \pm 0.019 \text{ s}^{-1}$ ) than for present conditions ( $0.156 \pm 0.015 \text{ s}^{-1}$ ). There were no further significant main or interaction effects (all  $F \leq 0.850$ ,  $p \geq 0.368$ ,  $\eta p^2 \leq 0.043$ ). Overall, the main effects observed for *eta* are equivalent to those seen for violations of Miller’s bound.

Following the model-fit assessment in **Chapter 3**, benefit was calculated using the interpolated model-fit distribution in place of the empirical redundant RT distribution (**Figure 4.8c**). First, model-fit and empirical benefits were correlated. All 4 correlation values were positive correlated and highly significant (see **Table 4.2**). Second, differences between predicted and empirical benefit were assessed. Benefit was averaged across conditions and compared with a paired-samples *t*-test. The mean model-fit benefit ( $0.036 \pm 0.004 \text{ s}$ ) was smaller than the mean empirical benefit ( $0.038 \pm 0.004 \text{ s}$ ),  $t(19)=3.194$ ,  $p=0.005$ , two-tailed. To assess whether this difference was consistent across conditions, empirical benefit was subtracted from model-fit benefit for each condition. A 2x2 ANOVA revealed no significant main or interaction effects (all  $F \leq 2.789$ ,  $p \geq 0.111$ ,  $\eta p^2 \leq 0.128$ ). Thus, the difference between model-fit and empirical benefit ( $0.002 \pm 0.001 \text{ s}$ ) was consistent across conditions.

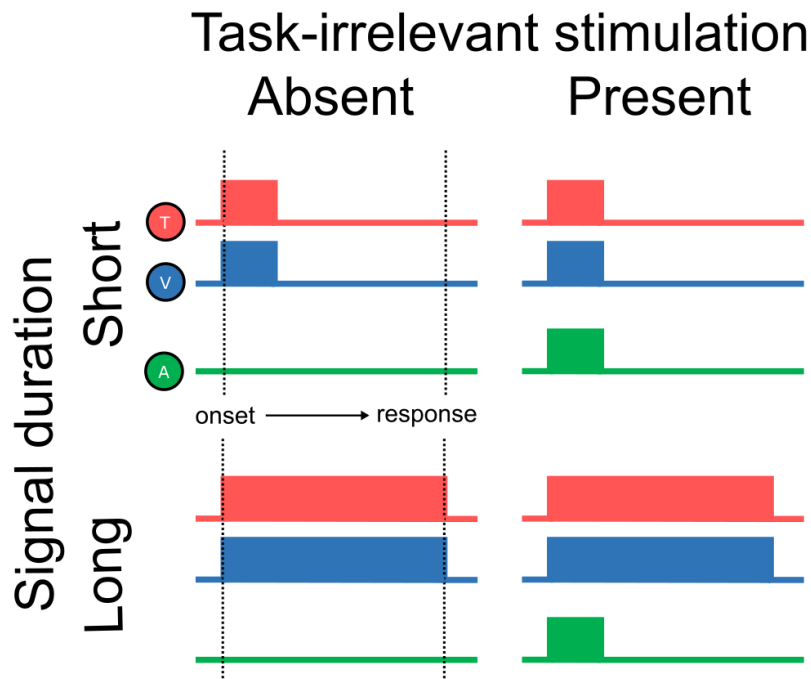
Overall, the context variant race model came much closer to approximating the empirical benefit by accounting for interactions. Concerning the model parameters, the correspondence between *eta* and violation was evident on the level of main effects, but the correspondence between *rho* and history measures was only observed in general trends.

#### 4.4. Interim Summary (Part 1)

The comparative approach was first applied to AV signals. For Step 1, predicted benefits followed empirical benefits well on the individual participant level, further suggesting that the race model is a strong predictor of the empirical effect. However, no main or interaction effects were significant for either predicted or empirical benefit. For Step 2, the sensory interactions quantified were significant. For trial history effects, in line with the experimental hypothesis, there was a main effect of signal duration, whereby short signals elicited a larger history effect than long signals. Interestingly, a main effect of task-irrelevant stimulation was also observed for history effects, which were larger when task-irrelevant stimulation was absent compared to present. For violations of Miller’s bound, in line

**Table 4.2 Pearson correlation coefficients (and *p* values) comparing model-fit and empirical benefits (Part 1)**

Signal duration	Model-fit vs empirical	
	Task-irrelevant stimulation	
	Absent	Present
Short	0.992 ( $p < 0.001$ )	0.989 ( $p < 0.001$ )
Long	0.989 ( $p < 0.001$ )	0.994 ( $p < 0.001$ )



**Figure 4.9 Stimulus design for Chapter 4 (Part 2)**

In each plot, the lines indicate the onset/offset of signals (T, V) and the task-irrelevant stimulus (V). Illustrations depict a redundant signal trial.

with the hypothesis, there was a main effect of task-irrelevant stimulation, whereby violation was larger in absent compared to present conditions. For Step 3, the context variant race model was applied. This approximated the empirical benefit much more closely, and (though not always producing identical main effects) there was overall a good correspondence between the interactions and the corresponding model parameters. Overall, the results of this experiment closely follow the hypotheses made at the start of this chapter. To ensure that the results observed here for the interactions were reliable and generalised across the senses, a second experiment was conducted. This was essentially the same as described in **Section 4.2**, but with 1) a different participant sample and 2) different modalities allocated to signals and task-irrelevant stimulation.

#### 4.5. Methods (Part 2)

The overall experimental procedure was virtually identical to Part 1 (**Section 4.2**). The only difference was that auditory and tactile stimulus vectors were switched (**Figure 4.9**). This experiment, therefore, used tactile-visual (TV) signals and the task-irrelevant stimulation was auditory (A).

##### 4.5.1. Participants

This part of the experiment was carried out with a different sample of participants. 29 participants were initially recruited at the University of St. Andrews. Of these, there was one drop-out during

testing, and eight met the miss rate exclusion criterion. Notably, all excluded participants met the miss rate criterion in the short-absent condition with tactile signals. This indicates that it was relatively difficult to detect the tactile stimulus in this condition. The data from 20 participants (15 female) were therefore analysed. Age of these participants ranged from 18-55 years. All were naïve as to the aims of the experiment. Normal hearing and normal/corrected-to-normal vision was reported in all cases. Reimbursement was £10.

#### 4.6. Results (Part 2)

For all repeated-measures ANOVAs presented in this section, factors adhere to the following ordering: signal duration (short, long) × task-irrelevant stimulation (absent, present). In some analyses, an additional factor was added: signal modality (T, V, [TV]).

##### 4.6.1. General Performance

Participant performance was assessed before RT analysis to ensure ceiling performance. The percentage of false alarms for signal trials was averaged across the three trial types (T, V, TV). For signal trials, false alarms occurred on only 0.045% ( $\pm 0.018\%$ ) of trials. A 2×2 ANOVA, however, revealed a significant interaction between signal duration and task-irrelevant stimulation,  $F(1, 19)=6.234$ ,  $p=0.022$ ,  $\eta p^2=0.247$ . For short signal trials, false alarms occurred on 0.082% ( $\pm 0.040\%$ ) of trials when task-irrelevant stimulation was absent, and on 0% of trials when task-irrelevant stimulation was present. For long signals, false alarms occurred on 0.049% ( $\pm 0.036\%$ ) of trials when task-irrelevant stimulation was absent, and on 0.050% ( $\pm 0.027\%$ ) of trials when task-irrelevant stimulation was present. The interaction therefore seems to be caused by the larger difference between the number of false alarms in short-absent and short-present conditions. However, false alarms are very rare in each condition. There were no other significant main or interaction effects (all  $F \leq 0.991$ ,  $p \geq 0.332$ ,  $\eta p^2 \leq 0.050$ ).

For catch trials, false alarms occurred on 2.978% ( $\pm 0.361\%$ ) of trials. A 2×2 ANOVA revealed a marginal (i.e.  $p=0.05$ ) main effect of signal duration,  $F(1, 19)=4.361$ ,  $p=0.050$ ,  $\eta p^2=0.187$ . The percentage of false alarms was larger for short conditions (3.388%  $\pm 0.411\%$ ) than long conditions (2.567%  $\pm 0.411\%$ ). This may be expected if the short signals were slightly harder to detect, as participants might have been unsure if a signal was presented or not. There was also a significant main effect of task-irrelevant stimulation,  $F(1, 19)=82.524$ ,  $p<0.001$ ,  $\eta p^2=0.813$ . The percentage of false alarms was smaller for absent conditions (0.466%  $\pm 0.130\%$ ) than present conditions (5.489%  $\pm 0.630\%$ ). There was also a significant interaction between signal duration and task-irrelevant stimulation,  $F(1, 19)=4.783$ ,  $p=0.041$ ,  $\eta p^2=0.201$ . For short signal trials, false alarms occurred on 0.491% ( $\pm 0.151\%$ ) of trials when task-irrelevant stimulation was absent, and on 6.284% ( $\pm 0.760\%$ ) of trials when task-irrelevant stimulation was present. For long signals, false alarms occurred on 0.441%

( $\pm 0.180\%$ ) of trials when task-irrelevant stimulation was absent, and on 4.694% ( $\pm 0.688\%$ ) of trials when task-irrelevant stimulation was present. The interaction therefore seems to come from a larger difference between short-absent and short-present conditions than between long-absent and long-present conditions.

Second, the percentage of misses averaged across signal modalities was only 0.708% ( $\pm 0.093\%$ ). A  $2 \times 2 \times 3$  ANOVA revealed a significant main effect of signal duration,  $F(1, 19) = 46.438$ ,  $p < 0.001$ ,  $\eta p^2 = 0.710$ . The percentage of misses was larger for short conditions ( $1.268 \pm 0.166\%$ ) than long conditions ( $0.149 \pm 0.058\%$ ). There was also a significant main effect of task-irrelevant stimulation,  $F(1, 19) = 32.992$ ,  $p < 0.001$ ,  $\eta p^2 = 0.635$ . The percentage of misses was larger for absent conditions ( $1.010 \pm 0.129\%$ ) than present conditions ( $0.407 \pm 0.079\%$ ). There was also a significant main effect of signal modality,  $F(1.312, 24.921) = 18.438$ ,  $p < 0.001$ ,  $\eta p^2 = 0.492$ . Pairwise comparisons revealed that the percentage of misses on tactile trials ( $1.522 \pm 0.260\%$ ) was larger than on visual trials ( $0.580 \pm 0.141\%$ ,  $p = 0.022$ ) and on redundant trials ( $0.024 \pm 0.024\%$ ,  $p < 0.001$ ). Visual and redundant trials were also significantly different ( $p = 0.003$ ). As almost all further interactions of this analysis were significant, the remaining results are reported in the appendices (**Section 9.5**). To summarise, as seen in **Table 4.3**, these interactions likely arise because there was a disproportionately large number of misses for tactile signals in the short-absent condition. This corresponds to the large number of participants excluded from this analysis based on a high percentage of misses in this condition.

Finally, the number of outliers was calculated based on the  $1/RT$  transformed data (see **Section 2.2.2**). On average, fast outliers occurred on 0.31% ( $\pm 0.05\%$ ) of RTs, whereas slow outliers occurred on 0.51% ( $\pm 0.11\%$ ) of RTs. The overall percentage of outliers ( $\sim 0.82\%$ ), therefore, was small and only slightly larger than would be expected if the  $1/RT$  data followed a normal distribution (0.27%). Across all participants, 23623 valid RTs were included in the main analyses, with a mean of 98.4 ( $\pm 0.16$ ) valid RTs per participant for each condition (of a possible maximum of 100 RTs).

Overall, as for Part 1, there were small differences in errors across conditions but the data indicate close to ceiling performance. One exception to this is performance in the short-absent

**Table 4.3 Percentage of misses ( $\pm$  SEM) for each condition by signal modality**

Condition	Misses (%)		
	Signal Modality		
	Tactile (T)	Visual (V)	Redundant (TV)
Short / Absent	4.145 (0.663)	1.125 (0.257)	0.094 (0.094)
Short / Present	1.596 (0.372)	0.647 (0.324)	0.000 (0.000)
Long / Absent	0.297 (0.146)	0.396 (0.168)	0.000 (0.000)
Long / Present	0.050 (0.082)	0.151 (0.050)	0.000 (0.000)

condition for tactile trials. Here, there was a slightly larger number of misses, indicating that the signal was harder to detect. Even in this condition, however, the average number of included trials was still fairly close to maximum, with 94.6 ( $\pm 0.93$ ) of a possible 100 trials included. As such, accuracy was not formally considered. Importantly, as for Part 1, task-irrelevant stimulation leads to more false alarms as measured in catch trials in this experiment as well. This is considered later (**Section 4.8**).

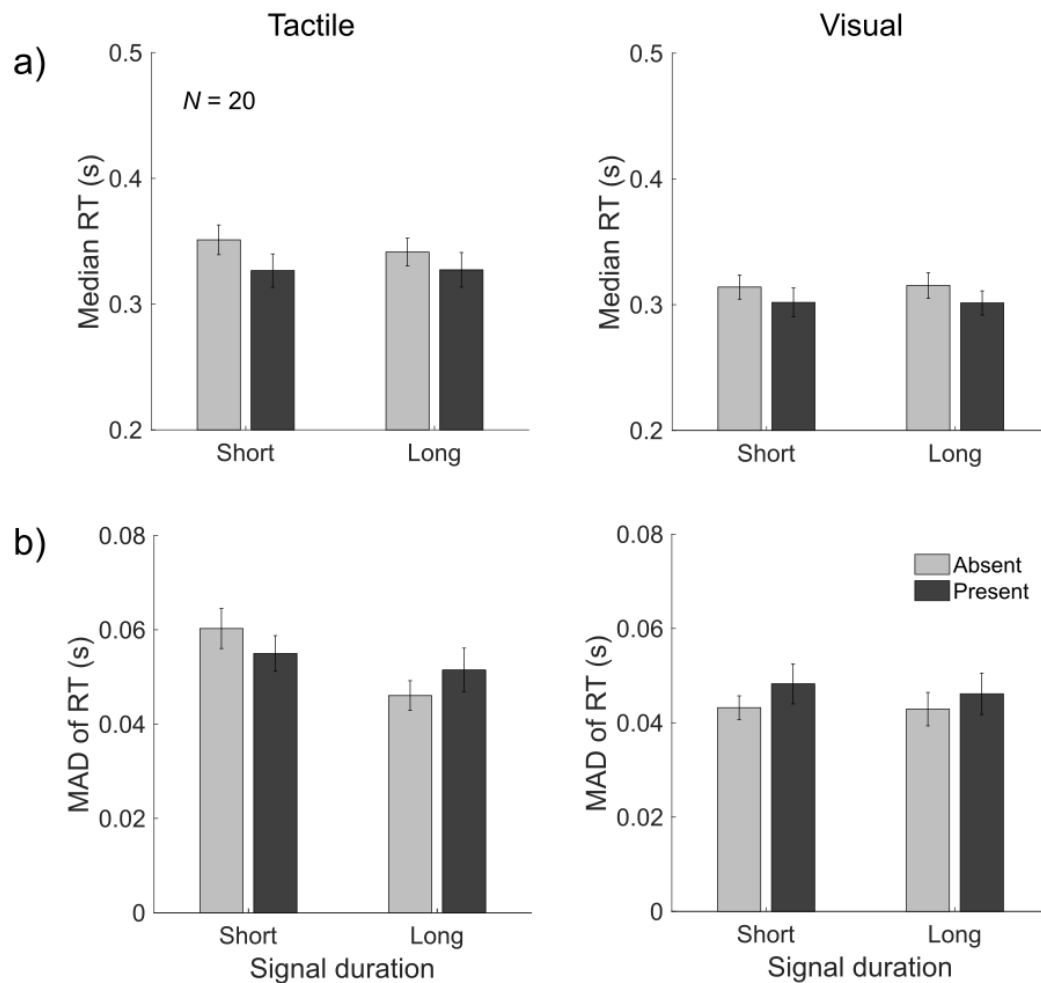
#### 4.6.2. Unisensory RTs

Prior to applying the comparative approach, unisensory RTs were analysed to indicate the general effect of experimental factors and to indicate potential changes in benefit. For median RT (**Figure 4.10a**), a  $2 \times 2 \times 2$  ANOVA revealed a significant main effect of task-irrelevant stimulation,  $F(1, 19)=18.413$ ,  $p<0.001$ ,  $\eta p^2=0.492$ . Median RT was larger for absent trials ( $0.330 \pm 0.010$  s) compared to present trials ( $0.314 \pm 0.011$  s). This might be explained by the task-irrelevant stimulus providing information about when a signal may or may not have been presented (which may have aided detection, particularly for short tactile trials). There was also a significant main effect of signal modality,  $F(1, 19)=29.892$ ,  $p<0.001$ ,  $\eta p^2=0.611$ . Median RT was larger for tactile trials ( $0.337 \pm 0.012$  s) compared to visual trials ( $0.308 \pm 0.010$  s). No other main or interaction effects were significant (all  $F \leq 3.207$ ,  $p \geq 0.089$ ,  $\eta p^2 \leq 0.144$ ). Overall, as there was a significant main effect of modality, this indicates that unisensory signals were not equally effective.

For MAD of RT (**Figure 4.10b**), a  $2 \times 2 \times 2$  ANOVA revealed a significant main effect of signal duration,  $F(1, 19)=8.465$ ,  $p=0.009$ ,  $\eta p^2=0.308$ . MAD of RT was larger in short conditions ( $0.052 \pm 0.003$  s) compared to long conditions ( $0.047 \pm 0.003$  s). There was also a significant main effect of signal modality,  $F(1, 19)=23.807$ ,  $p<0.001$ ,  $\eta p^2=0.556$ . MAD of RT was larger in tactile conditions ( $0.053 \pm 0.003$  s) compared to visual conditions ( $0.045 \pm 0.003$  s). There was a significant interaction between signal duration and signal modality,  $F(1, 19)=5.165$ ,  $p=0.035$ ,  $\eta p^2=0.214$ . For short conditions, there was a difference between tactile ( $0.058 \pm 0.003$  s) and visual ( $0.046 \pm 0.003$  s) MAD. For long conditions, however, tactile ( $0.049 \pm 0.003$  s) and visual ( $0.045 \pm 0.004$  s) MAD was more similar. This is in line with the performance analysis, which suggested short tactile signals were more difficult to detect. No other main or interaction effects were significant (all  $F \leq 4.359$ ,  $p \geq 0.051$ ,  $\eta p^2 \leq 0.187$ ).

Overall, as there are interactions between factors for unisensory RTs it is difficult to make straightforward directional predictions for benefits. However for both median and MAD of RT, there is a main effect of signal modality. This means (as for Part 1) that signals were not equally effective; tactile RTs were slower and more variable than visual RTs. This means that changes in predicted benefits may be small.





**Figure 4.10 Unisensory RT analysis (Part 2)**

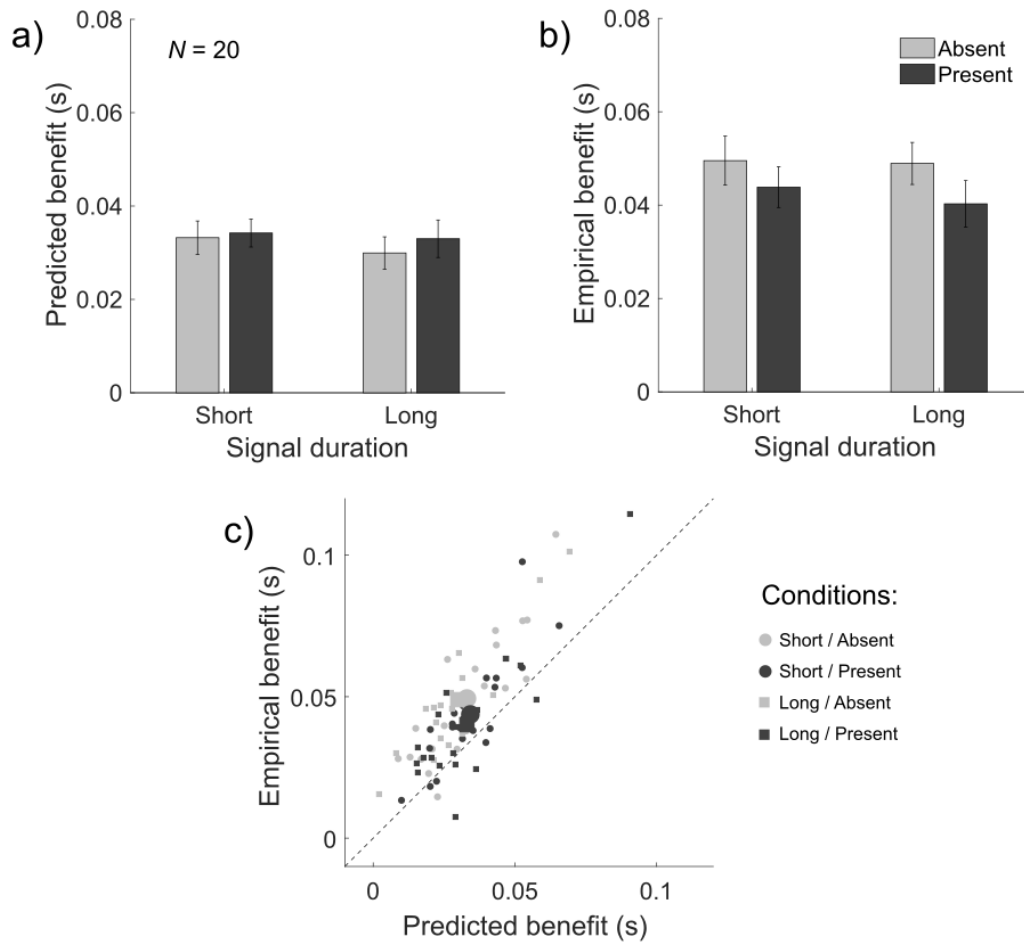
**a, b)** Median RT and MAD of RT across conditions. The left column shows tactile RTs and the right column shows visual RTs. All bars show the mean of 20 participants ( $\pm 1$  SEM).

#### 4.6.3. Comparative Approach Step 1: Multisensory Benefits

As for Part 1, the analysis of unisensory RTs does not provide a straightforward overall indication of how benefits should change. To obtain a straightforward quantitative prediction, I applied the simple race model. Empirical benefits are then measured and compared to the predicted benefit.

First, benefits were predicted (**Figure 4.11a**) according to the simple race model (see **Section 2.3.2.2**). A 2x2 ANOVA revealed no main or interaction effects (all  $F \leq 1.676$ ,  $p \geq 0.211$ ,  $\eta p^2 \leq 0.081$ ). Therefore, benefit was predicted not to change across conditions. This was in line with the lack of equal effectiveness observed in the unisensory RT analysis.

Next, benefits were measured empirically (see **Section 2.3.2.1**) to see if the results followed the pattern predicted by the simple race model (**Figure 4.11b**). Interestingly, a 2x2 ANOVA revealed a significant main effect of task-irrelevant stimulation,  $F(1, 19) = 4.806$ ,  $p = 0.041$ ,  $\eta p^2 = 0.202$ . Benefits were larger in absent conditions ( $0.049 \pm 0.005$  s) compared to present conditions ( $0.042 \pm 0.004$  s). No



**Figure 4.11 Predicting and measuring multisensory benefits (Part 2)**

**a, b)** Predicted and empirical benefits across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

**c)** Empirical benefit as a function of predicted benefit. Each point represents data from one participant in one of the four conditions (80 data points total). Large symbols show the group mean.

other main or interaction effects were significant (all  $F \leq 0.623$ ,  $p \geq 0.440$ ,  $\eta p^2 \leq 0.032$ ). Therefore, in contrast to the predictions of the simple race model, empirical benefits changed across conditions. This is considered in detail later.

To further investigate the predictive power of the simple race model, additional analytical steps were applied. First, the correspondence between predicted and empirical benefits was evaluated for each participant for each condition of the  $2 \times 2$  design. All 4 correlation values were positive and highly significant (see **Table 4.4**). Second, differences between predicted and empirical benefit were assessed. Benefit was averaged across conditions and compared with a paired-samples  $t$ -test. Mean predicted benefit ( $0.033 \pm 0.003$ s) was smaller than mean empirical benefit ( $0.046 \pm 0.004$  s),  $t(19)=7.348$ ,  $p < 0.001$ , two-tailed. To assess if this difference varied across conditions, predicted benefit was subtracted from empirical benefit for each condition. A  $2 \times 2$  ANOVA revealed a significant main effect of task-irrelevant stimulation,  $F(1, 19)=21.156$ ,  $p < 0.001$ ,  $\eta p^2=0.527$ . The difference

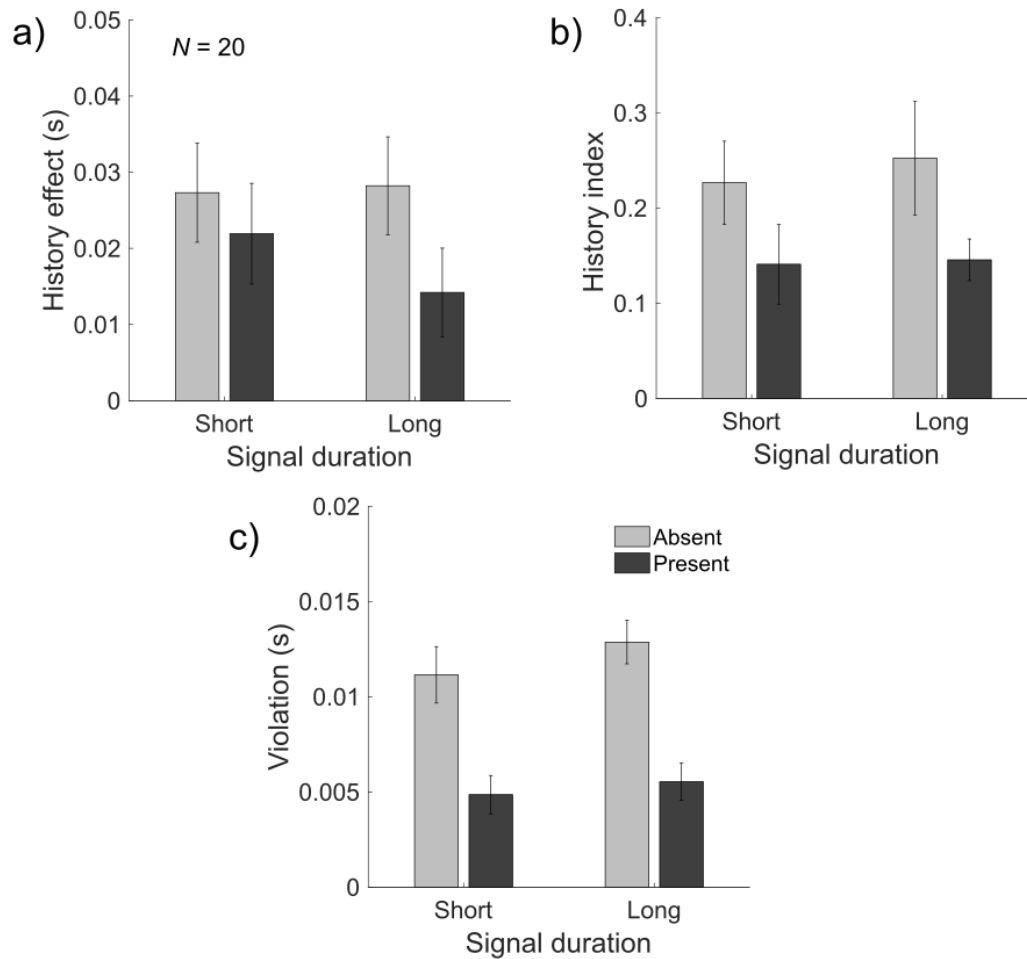
between empirical and predicted benefit was larger for absent conditions ( $0.018 \pm 0.002s$ ) compared to present conditions ( $0.008 \pm 0.002s$ ). There were no further main or interaction effects (all  $F \leq 1.751$ ,  $p \geq 0.201$ ,  $\eta p^2 \leq 0.084$ ). As for Part 1, therefore, the presence of task-irrelevant stimulation results in benefits which are closer to the simple race model prediction.

Overall, the results of Step 1 of the comparative approach show a good correspondence between predicted and empirical benefits across individual participants. However, empirical benefits showed a main effect of task-irrelevant stimulation which was not found in the predicted benefits: empirical benefits were larger when stimulation was absent compared to present. Further, the difference between empirical and predicted benefits was larger when stimulation was absent compared to present. These effects can be understood by considering the interactions, which go beyond the simple race model prediction. This is done in the next step.

#### 4.6.4. Comparative Approach Step 2: Quantifying Interactions

The second step of the comparative approach is to quantify interactions which are not accounted for by the basic race model architecture. Based on experimental hypotheses (and the results of Part 1), the following main effects were expected to replicate. For history effects, a main effect of signal duration was expected, whereby short signals elicit larger history effects than long signals. For violations of Miller's bound, a main effect of task-irrelevant stimulation was expected, whereby violation is larger when stimulation is absent compared to present.

For history effects (**Figure 4.12a**), a  $2 \times 2$  ANOVA revealed a significant intercept,  $F(1, 19) = 20.101$ ,  $p < 0.001$ ,  $\eta p^2 = 0.514$ . Thus, the overall history effect ( $0.023 \pm 0.005 s$ ) was significantly greater than 0, and must be accounted for. There was also a significant main effect of task-irrelevant stimulation,  $F(1, 19) = 4.682$ ,  $p = 0.043$ ,  $\eta p^2 = 0.198$ . The history effect was larger in absent conditions ( $0.028 \pm 0.006 s$ ) compared to present conditions ( $0.018 \pm 0.006 s$ ). This replicated a similar result found in Part 1. However, contrary to the experimental hypothesis and results of Part 1, the expected main effect of signal duration was not significant, as there were no further significant main or interaction effects (all  $F \leq 0.840$ ,  $p \geq 0.371$ ,  $\eta p^2 \leq 0.042$ ). As a further evaluation of trial history effects, the history index (**Figure 4.12b**) was calculated (see **Section 2.3.3.1**). Here, there was also a significant main effect of task-irrelevant stimulation,  $F(1, 19) = 5.090$ ,  $p = 0.036$ ,  $\eta p^2 = 0.211$ . The history index was larger in absent conditions ( $0.240 \pm 0.045$ ) compared to present conditions ( $0.143 \pm 0.026$ ). As for Part 1, the history effect therefore contributed to overall variability more in absent compared to present conditions. Again, however, there was no effect of signal duration, and no further interactions (all  $F \leq 0.345$ ,  $p \geq 0.564$ ,  $\eta p^2 \leq 0.018$ ).



**Figure 4.12 Quantifying interactions (Part 2)**

**a, b)** History effect and history index across conditions.

**c)** Violation of Miller's bound across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

For violations of Miller's bound (**Figure 4.12c**), a 2x2 ANOVA revealed a significant intercept,  $F(1, 19)=140.010$ ,  $p<0.001$ ,  $\eta p^2=0.881$ . Thus, the overall violation area ( $0.009 \pm 0.001$  s) was significantly greater than 0, and must be accounted for. As expected, there was a significant main effect of task-irrelevant stimulation,  $F(1, 19)=41.972$ ,  $p<0.001$ ,  $\eta p^2=0.688$ . Following the experimental hypothesis and results of Part 1, violation was larger for absent conditions ( $0.012 \pm 0.001$  s) than for

**Table 4.4 Pearson correlation coefficients (and  $p$  values) comparing predicted and empirical benefits (Part 2)**

Signal duration	Predicted vs empirical	
	Task-irrelevant stimulation	
	Absent	Present
Short	0.857 ( $p<0.001$ )	0.842 ( $p<0.001$ )
Long	0.902 ( $p<0.001$ )	0.850 ( $p<0.001$ )

present conditions ( $0.005 \pm 0.0001$  s). There were no further significant main or interaction effects (all  $F \leq 1.630$ ,  $p \geq 0.217$ ,  $\eta p^2 \leq 0.079$ ).

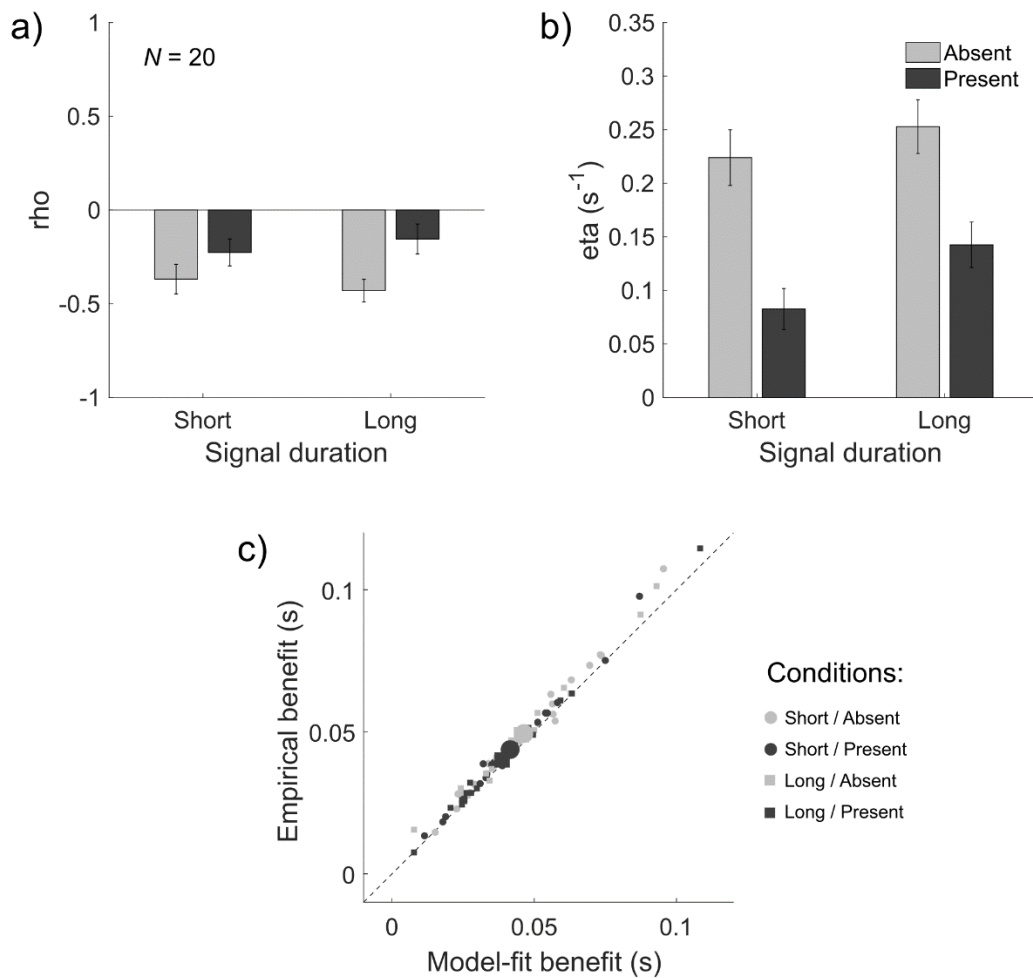
Overall, both interactions were again significantly present in the data. The main effect of signal duration on the history effect found with AV signals was not replicated with TV signals, however a main effect of task-irrelevant stimulation did replicate; the history effect was larger for absent compared to present conditions. The expected main effect of task-irrelevant stimulation on violation also replicated; violation was again larger when stimulation was absent compared to present. This can directly account for why empirical benefit was closer to predicted benefit when task-irrelevant stimulation is present.

#### 4.6.5. Comparative Approach Step 3: Applying the Context Variant Race Model

As a third step, I applied the context variant race model to explain benefits and interactions. As such, it is expected that the changes observed for interactions in the previous section are also observed in the corresponding model parameters.

Analysing the *rho* parameters (**Figure 4.13a**), a 2x2 ANOVA revealed a significant intercept,  $F(1, 19)=37.975$ ,  $p<0.001$ ,  $\eta p^2=0.667$ . Thus, the overall correlation modelled between unisensory rate distributions ( $-0.295 \pm 0.048$ ) was significantly different to 0. There was also a significant main effect of task-irrelevant stimulation,  $F(1, 19)=9.753$ ,  $p=0.006$ ,  $\eta p^2=0.339$ . The correlation was more negative in absent conditions ( $-0.399 \pm 0.061$ ) than present conditions ( $-0.191 \pm 0.056$ ). There were no significant main or interaction effects (all  $F \leq 1.675$ ,  $p \geq 0.211$ ,  $\eta p^2 \leq 0.081$ ). Overall, the analysis of *rho* corresponds to the main effects seen in the history effect and the history index.

Analysing the *eta* parameters (**Figure 4.13b**), a 2x2 ANOVA revealed a significant intercept,  $F(1, 19)=153.472$ ,  $p<0.001$ ,  $\eta p^2=0.890$ . Thus, additional variability was added to unisensory rate distributions to account for the variability of redundant rate distributions ( $0.175 \pm 0.014$  s<sup>-1</sup>). Interestingly, there was also a main effect of signal duration,  $F(1, 19)=7.975$ ,  $p=0.011$ ,  $\eta p^2=0.296$ . The *eta* parameter was smaller for short conditions ( $0.153 \pm 0.016$  s<sup>-1</sup>) than for long conditions ( $0.198 \pm 0.016$  s<sup>-1</sup>). There was also a main effect of task-irrelevant stimulation,  $F(1, 19)=33.951$ ,  $p<0.001$ ,  $\eta p^2=0.641$ . The *eta* parameter was larger for absent conditions ( $0.238 \pm 0.020$  s<sup>-1</sup>) than for present conditions ( $0.113 \pm 0.015$  s<sup>-1</sup>). There were no further significant main or interaction effects (all  $F \leq 0.395$ ,  $p \geq 0.537$ ,  $\eta p^2 \leq 0.020$ ). Overall, the analysis of *eta* shows the expected main effect of task-irrelevant stimulation found in violations of Miller's bound. However, an additional effect of signal duration was also present. This suggests that more noise was introduced when two long signals were accumulated in parallel compared to two short signals.



### Figure 4.13 Fitting the context variant race model (Part 2)

**a,b)** Best fitting  $\rho$  and  $\eta$  values across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

**c)** Empirical benefit as a function of model-fit benefit. Each point represents data from one participant in one of the four conditions (80 data points total). Large symbols show the group mean.

Model fit was assessed by calculating benefit from the best-fitting model distribution and comparing it to empirical benefit (Figure 4.13c). For correlations values, all 4 were strongly positive and highly significant (see Table 4.5). Second, differences between predicted and empirical benefits were assessed. Benefit was averaged across conditions and compared with a paired-samples  $t$ -test. The mean model-fit benefit ( $0.043 \pm 0.004$  s) was smaller than the mean empirical benefit ( $0.046 \pm 0.004$  s),  $t(19)=6.160$ ,  $p<0.001$ , two-tailed. To assess whether this difference was consistent across conditions, empirical benefit was subtracted from model-fit benefit for each condition. A  $2 \times 2$  ANOVA revealed a significant main effect of task-irrelevant stimulation,  $F(1, 19)=4.953$ ,  $p<0.038$ ,  $\eta p^2=0.207$ . The difference between model fit and empirical benefit was larger in absent conditions ( $0.003 \pm 0.001$  s) compared to present conditions ( $0.002 \pm 0.0004$  s). There was also a significant interaction between

signal duration and task-irrelevant stimulation,  $F(1, 19)=5.127$ ,  $p<0.035$ ,  $\eta p^2=0.212$ . The difference between model fit and empirical benefit was similar for short-absent, short-present and long-absent conditions ( $0.003 \pm 0.001$  s), but was smaller for long-present conditions ( $0.002 \pm 0.0004$  s). There were no further significant main or interaction effects (all  $F \leq 0.307$ ,  $p \geq 0.586$ ,  $\eta p^2 \leq 0.016$ ).

Overall, therefore, the context variant race model estimated the benefit well. Though there was still some difference between model-fit and empirical benefit, this was reduced in conditions where task-irrelevant stimulation was present, and was smallest in the long-present condition. Further, the model parameters generally followed effects observed in interactions.

#### 4.7. Interim Summary (Part 2)

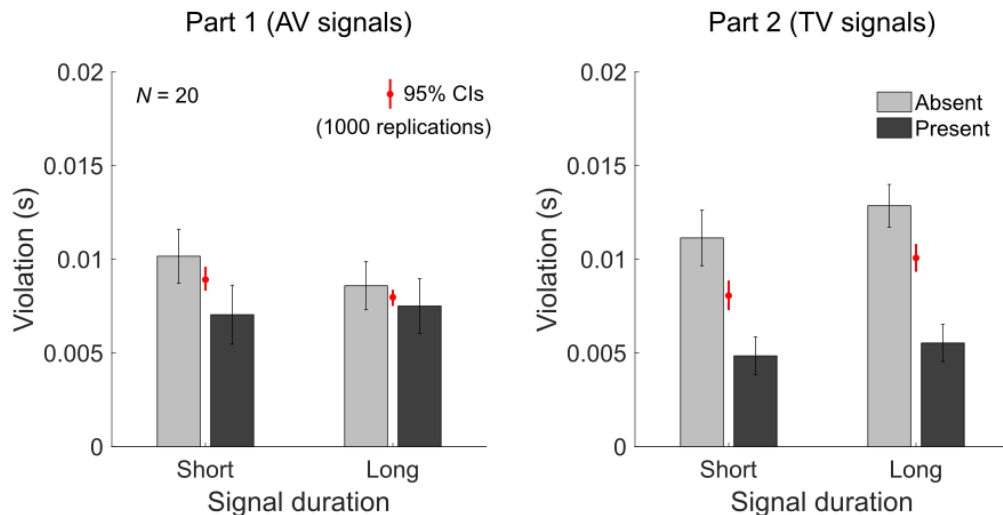
Following the AV version of the experiment in Part 1, the comparative approach was applied again here with TV signals with the overall goal of replicating the effects observed. In Step 1, predicted and empirical benefits were strongly correlated on the individual subject level. However, empirical benefits were larger when task-irrelevant stimulation (A) was absent compared to present. There was no corresponding main effect in predicted benefits. In Step 2, significant interactions were observed. As for Part 2, it was expected that history effects would show a main effect of signal duration. However, no effect of signal duration was observed. Interestingly, however, the main effect of task-irrelevant stimulation observed in Part 1 was replicated. It was also expected that for violations, there would be a main effect of task-irrelevant stimulation. This was replicated, as the violation area was larger for absent conditions compared to present conditions. In Step 3, the context variant race model provided a close account of benefits, and the model parameters closely followed the main effects seen in the interactions. Beyond this, *eta* demonstrated a main effect of signal duration; less noise was required in short conditions compared to long conditions.

#### 4.8. Can Increased False Alarms Account For Reductions in Violation?

In Parts 1 & 2, a reduction in violation was observed when task-irrelevant stimulation is present. This was interpreted, in line with experimental hypothesis, as evidence of a reduction in context variance by adding a third modality task-irrelevant stimulus. An alternative explanation, however, is that the reduction is a result of increases in anticipatory responses (as suggested by the larger number of false

**Table 4.5 Pearson correlation coefficients (and *p* values) comparing model-fit and empirical benefits (Part 2)**

Signal duration	Model-fit vs empirical	
	Task-irrelevant stimulation	
	Absent	Present
Short	0.993 ( $p<0.001$ )	0.993 ( $p<0.001$ )
Long	0.993 ( $p<0.001$ )	0.998 ( $p<0.001$ )



**Figure 4.14 Simulating the effect of false alarms on violation area**

The bars show the empirical violation data from **Figure 4.7c** (Part 1) and **Figure 4.12c** (Part 2). The red line indicates the 95% confidence intervals calculated from 1000 simulation values. The solid red dot shows the average violation calculated from the simulation results. Overall, the simulated values represent where the violation values for the present conditions (dark bars) should lie based on false alarms alone. If the dark bars are smaller than these expected values (shown in red), then false alarms cannot account for the reduction in violation.

alarms on catch trials). Across both experiments in this chapter, false alarms on catch trials were higher when task-irrelevant stimulation was present compared to absent. If anticipatory false alarms contaminate signal RT distributions, we would expect violations to be smaller merely as an artefact (e.g. Gondan & Minakata, 2016; Miller & Lopes, 1991). To rule out this possibility, therefore, an additional simulation was run. First, RTs on catch trials were calculated for each individual for both of the conditions where task-irrelevant stimulation was *present* (short-present, long-present). Second, these catch trial RTs were inserted into signal RT distributions for both conditions where task-irrelevant stimulation was *absent* (long-absent, long-present). Following a Monte Carlo procedure, a random selection of valid RTs in each signal trial RT distribution was randomly replaced with the catch trial RTs. This simulation therefore should provide an estimate of how much we would expect violation to decrease based on the increased false alarm contamination alone.

This simulation was run for each participant for both levels of signal duration. To estimate 95% confidence intervals (**Figure 4.14**), the simulation was repeated 1000 times (randomly replacing different signal RTs each time). Violation was calculated as in the experiment, such that the results for each participant ( $N=20$ ) were averaged across each of the 1000 replications. This created 1000 estimations of the mean violation area with increased false alarms. For Part 1 (AV signals) the mean empirical violation area for short-present conditions did not fall within the 95% confidence intervals, whereas the mean empirical violation area for long-present conditions was close to the edge of the



confidence intervals. For Part 2 (TV signals) the mean empirical violation area was well outside the 95% confidence intervals in both conditions which were simulated. This suggests that, while false alarms do indeed contribute to a reduction in violation area, they cannot account for the reduction in violation when task-irrelevant stimulation is present.

#### 4.9. Discussion

The overall aim of this chapter was to apply the comparative approach to identify sources of interactions. Two factors were introduced in a 2×2 design, each aiming to target a specific interaction. First, history effects were targeted by *signal duration* (short, long). It was hypothesised that short signals would elicit larger history effects than long signals. This result was found in Part 1 (AV signals) but not Part 2 (TV signals). Second, context variance (measured by violations of Miller's bound) was targeted by *task-irrelevant stimulation* (absent, present). It was hypothesised that task-irrelevant stimulation would reduce context variance (and thus violations of Miller's bound) when present compared to absent. This result was found in both Part 1 (AV signals) and Part 2 (TV signals). Overall, therefore, both interactions were successfully manipulated in at least one version of the experiment.

Regarding trial history effects, it is interesting that for TV signals the main effect of signal duration did not replicate. There are several factors which may have contributed to this null result. First, though it is difficult to compare results between experiments with different samples, the overall history effect is larger for AV (Part 1) than TV (Part 2) signals. This has also been found in a previous experiment which tested combinations of A, V, and T signals (Gondan et al., 2004). Second, the tactile signal was generally weaker and elicited slightly poorer performance, particularly in the short-absent condition. This may have somehow interfered with the size of the history effect in this condition. As further evidence of this, individual history effects for each modality are shown in **Section 9.6** of the appendices. Observing trends across conditions for both experiments, tactile signals are the only signals which fail to show a larger history effect in short-absent conditions than long-absent conditions. It might be therefore that if the tactile signal was delivered at a higher amplitude (to elicit stronger activation and overall more accurate performance) there may have been an overall main effect of signal duration. However, the goal here was to maintain the same level of stimulation for each modality across both parts of the experiment, to allow for comparisons between both parts. Thus the amplitude of the tactile stimulus was not altered from Part 1, and the problem was not observed until data collection for Part 2 had begun.

Continuing with the history effect, an interesting observation which replicated across both experiments was a main effect of task-irrelevant stimulation: history effects were larger when task-irrelevant stimulation was absent as compared to present. Here, task-irrelevant stimulation was primarily introduced to target context variance and thus manipulate the violation area. However,

previous work has introduced similar manipulations to remove modality switching artefacts from neurophysiological data (Gondan et al., 2007). Thus, the observed effect of task-irrelevant stimulation on history effects within the present data is somewhat expected. Further, this effect fits with previous behavioural observations. First, history effects are typically not observed following redundant trials; one reason for this would seem to be that both modalities are engaged, and as such there is less of a switch cost to the following unisensory trial. Following a similar logic, if an additional third-modality stimulus is always present, it may somehow help participants divide attention between modalities such that there is a smaller cost of switching between them. Second, effects of irrelevant stimulation have been observed previously in attention experiments. For instance, Doyle and Snowden (2001) reported that an irrelevant auditory tone (i.e. one which is non-informative for the spatial location of a visual target) nonetheless seems to reduce invalid visual cueing costs. The authors explain this effect by proposing that the irrelevant stimulation helps observers disengage attention, making it easier to shift to the task-relevant location. It is possible that similar attentional mechanisms may therefore contribute to history effects.

Regarding violations, a task-irrelevant stimulus in a third modality reduced the overall violation area. The logic behind including this factor was that an irrelevant stimulus creates a greater similarity of stimulation across trials (i.e. reduced context variance, and correspondingly, violations of Miller's bound). This main effect replicated over Parts 1 & 2. Further, an alternative explanation, which suggests that this may be due to an increase in false alarms associated with the task-irrelevant stimulus, was ruled out in a simulated experiment. The main effect of task-irrelevant stimulation was also mirrored in the *eta* parameter of the context variant race model. Overall, this is direct evidence to suggest that context variance introduces *noise* into the evidence accumulation process; when both decision-units are active simultaneously (as in redundant conditions), accumulation is less reliable than when only one is active (as in unisensory conditions). Overall, this suggests that context variance is an explanation for violations of Miller's bound, which (following Miller, 1982) has typically been interpreted as evidence in favour of the alternative pooling model architecture.

Regarding the model parameters, it was anticipated that there would be a good correspondence between the main effects for *rho* and history effects. Overall, this appeared to be the case according to comparisons of trends by condition. Interestingly, however, this was not always the case for main effects observed in the ANOVA results; for example, in Part 1 (AV signals) there was a main effect of signal duration in the history effect, which was not found in *rho*. Importantly, however, though we expect a broad correspondence there may be differences in main effects between these measures as they are calculated on different trials. The trials which are used to calculate history effects, for example, account for only a small portion of the overall unisensory distributions, whereas

*rho* is modelled based on entire unisensory distributions. As such, the *rho* parameter may be sensitive to other effects which can influence correlations between unisensory distributions; one example may be any positive dependencies introduced between all RTs, such as by practice and fatigue effects. Similarly for *eta*, we generally expected that main effects would map onto those for the violation area. However, violation area is usually only observed across a small percentage of the overall distribution (i.e. at the fast tail), whereas *eta* is modelled across the whole distribution. It is possible, therefore, that additional effects reflected in *eta* (as seen in Part 2) are reflecting this additional sensitivity. The correspondence between interactions and model parameters is evaluated in detail in **Chapter 6**.

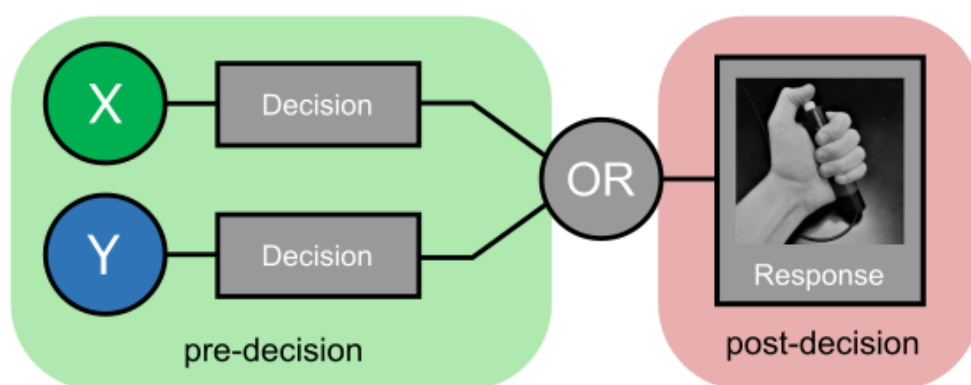
Though not the main focus of the present manipulations, the comparative approach again shows that race models provide a useful framework for predicting and explaining benefits, with a strong correlation between predicted and empirical benefits across Parts 1 & 2. No main effects were predicted in benefits across Parts 1 & 2, as neither experimental factor strongly influenced unisensory RT performance. One additional reason for the lack of effects may be that unisensory signals were not equally effective in either Parts 1 or 2 (particularly for short signals). Changes in predicted benefits across these factors may have been found if signals for both modalities were calibrated to elicit similar performance.

With TV signals, a main effect of task-irrelevant stimulation was also observed for empirical, but not predicted benefits. The comparative approach allows for a simple explanation of this discrepancy – for TV signals, there was a stronger main effect of task-irrelevant stimulation on violation. Violation contributes to empirical benefits, and is by definition unaccounted for in the simple race model prediction. Generally, changes in violation (though significant) are somewhat small. For example, the main effect of stimulus construction in **Chapter 3** caused a reduction in violation of around 0.003 s, whereas the main effect of task-irrelevant stimulation in Part 1 of this chapter caused a reduction of around 0.002 s. These small changes may not have been reflected as significant changes in overall benefit. However, the reduction observed in Part 2 was around 0.007 s (substantially larger than **Chapter 3**, or in Part 1 of this chapter). This larger reduction in violation possibly arose because the auditory (A) task-irrelevant stimulation in Part 2 was stronger than the tactile (T) task-irrelevant stimulation in Part 1. Note, for instance, that the task-irrelevant T stimulation (0.05 s) in Part 1 was identical to the short T signal in Part 2 (with the latter signal condition demonstrating relatively high percentage of misses). This large reduction in violation following the use of a stronger task-irrelevant stimulus likely contributed to the main effect observed for empirical benefit. This highlights that there is an important relationship between benefits and interactions, which can be understood by quantifying both in a comparative analysis.

## 5. Experiment 3: Signal Strength vs. Response Effector

### 5.1. Introduction

Application of the comparative approach in **Chapters 3 and 4** has suggested that race models are a convincing candidate for explaining the benefits and interactions observed across experiments. In this chapter, I examine whether this approach is able to shed light on which underlying *processing stages* contribute to the benefits observed. Previous research in the RSP has long been interested in the underlying ‘locus’ of the RSE, proposing relatively discrete stages of processing at which benefits could potentially arise (e.g. Cavina-Pratesi, Bricolo, Prior, & Marzi, 2001; Diederich & Colonius, 1987; Giray & Ulrich, 1993; Miller, 1982; Minakata & Gondan, 2018; Savazzi & Marzi, 2008). Typically, these include stages for sensory processing and evidence accumulation (which I refer to the *pre-decision* stage), and a stage for executing the response when the decision is made (which I refer to the *post-decision* stage). As previously demonstrated (**Section 1.6.1**), the race model already offers guiding principles for predicting and explaining changes in benefits based on properties of unisensory RTs. In particular, the race model implicates sources of RT variability as the underlying driving force behind benefit (the variability rule). In addition, however, examination of the race model architecture (**Figure 5.1**) is also able to provide hypotheses for which *sources* of variability should contribute to benefit, and which should not. In this chapter, I introduce factors which target pre-decisional and post-decisional variability within a 2x2 within-subjects design. This allows for an examination of the impact on benefit in relation to hypotheses made by the race model architecture (**Figure 5.1**).



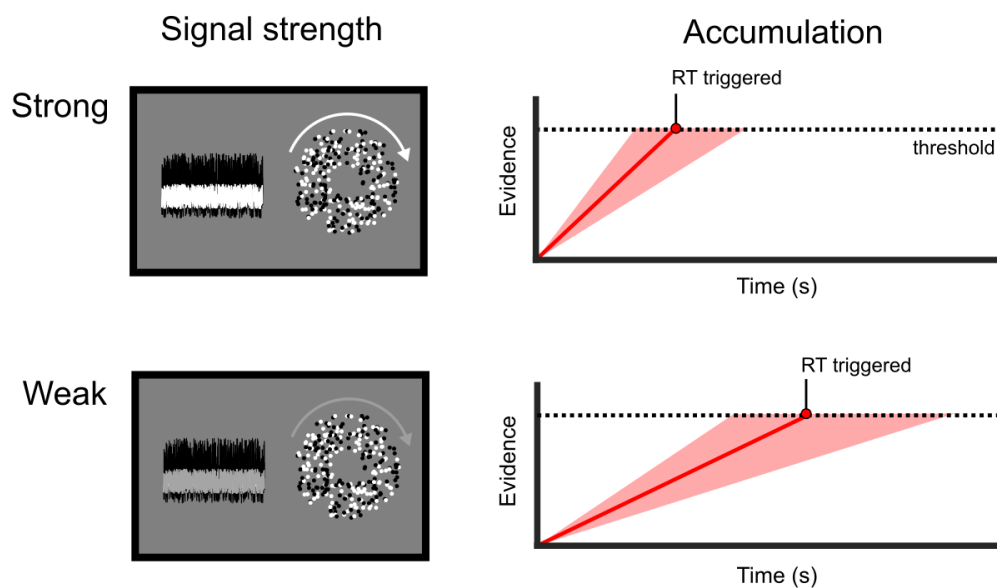
**Figure 5.1 Pre-decisional and post-decisional stages in the race model architecture**

The race model explains the RSE by a race between two parallel decision-units, which produces statistical facilitation. Variability which contributes to pre-decisional race processes (green area) should therefore contribute strongly to benefits. Post-decisional variability (red area) however occurs after statistical facilitation of decision times, and so should not strongly contribute to benefits.

### 5.1.1. Factor 1: Signal Strength

The first factor, *signal strength*, aimed to target pre-decisional RT variability by varying the overall strength of evidence for both signal components. Previous authors (Otto et al., 2013; Otto & Mamassian, 2017) have noted that signal strength is a useful manipulation, as weakening the strength of signals leads to slower and more variable RTs (Wagenmakers & Brown, 2007). This allows the principles of race models to be tested. Previous studies have found conflicting results as to whether weaker signal strength decreases or increases benefits (e.g. Chandrasekaran, Lemus, Trubanova, Gondan, & Ghazanfar, 2011; Diederich & Colonius, 2004; Senkowski, Saint-Amour, Hofle, & Foxe, 2011). Importantly, the race model offers the caveat that weaker signals should reliably produce larger benefit, but only if unisensory signals elicit similar performance (Otto et al., 2013). Based on the race model architecture (**Figure 5.1**), it is possible to derive a clear prediction for how this occurs and how it affects benefit. Here, the RSE is explained by a race process between two decision-units; benefits arise for redundant trials as a result of taking the faster of two decision times. Thus, in order for any variability source to contribute to benefit, it must impact this race process, which terminates when a decision is made. As signal strength affects the speed and variability of the accumulation process pre-decision (**Figure 5.2**), changes in benefit are expected according to race models (provided there are not substantially large differences in unisensory RT distributions). In line with the variability rule, larger benefits should be seen for weak signals, as these give rise to slower and more variable unisensory RTs.

Turning now to interactions, however, it is more difficult to have firm hypotheses. This is because previous research on signal strength has rarely quantified them (especially history effects). In the first experiment (**Chapter 3**), it was found that history effects were not manipulated by a change in stimulus construction, despite this factor eliciting large changes in the average and variability of unisensory RTs. These changes could be evidence of a change in signal strength, thus for the present experiment we might also expect no change. Alternatively, if previous unisensory trial acts to direct attention to a particular modality, it might be expected that the history effect is larger for weak signals. The idea here is that if attention is directed to the wrong modality, it may take longer to detect a signal and begin accumulation. This time-cost is likely to be larger if it is already more difficult to detect the signal. Regarding violations of Miller's bound, recent research has indicated that weaker signals produce smaller violations of Miller's bound (Minakata & Gondan, 2018). In addition, in the first experiment (**Chapter 3**), the size of violation was manipulated by stimulus construction, with complex signals producing smaller violations than simple signals. As above, if this result is interpreted as an effect of signal strength, smaller violations might be expected for weaker signals in this experiment.



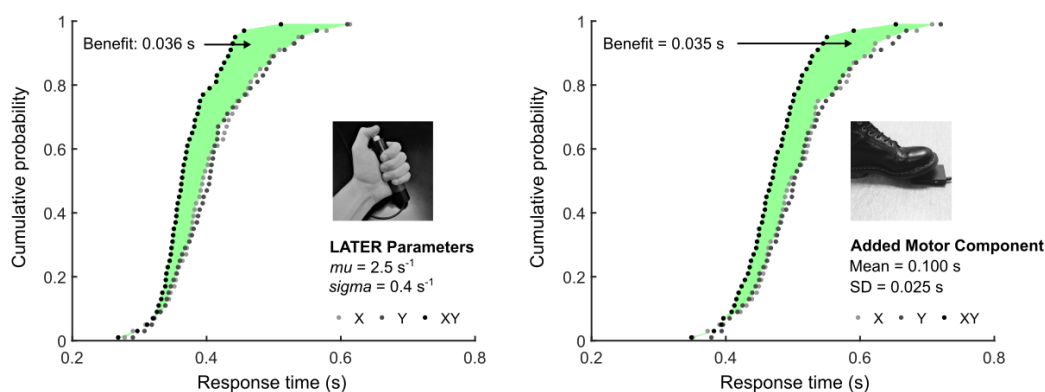
**Figure 5.2 Signal strength**

Considering complex stimuli (Section 2.1.2), auditory signals can be made weaker by reducing the volume, whereas visual signals can be made weaker by reducing the percentage of signal dots which rotate coherently. The right column shows a simplification of the evidence accumulation process. The shaded area shows the spread of a linear accumulation process (as in the LATER model, for example). The dashed line indicates threshold for decision. Overall, for each weak signal there is less evidence available at each time point. This leads to a shallower rise-to-threshold and thus longer and more variable decision times (shown by the larger spread of accumulation traces at the threshold).

### 5.1.2. Factor 2: Response Effector

The second factor, *response effector*, aimed to target post-decisional components of RT by changing the way participants responded. Different effectors are associated with different time courses for responses within the same task, as they rely on different pathways through the brain and body. Saccadic responses made with the eye, for example, are typically much faster than manual responses (Bompas, Hedge, & Sumner, 2017). Similarly, manual responses are slower than pedal responses (Pfister et al., 2014; Wiggett & Tipper, 2015). Though studies rarely present data on intra-individual RT variability (as analysed here by MAD, for instance), we might also expect slower effectors to be more variable as well. Speeded responses by foot, for example, are likely to be more variable than speeded responses by hand, because individuals are usually much more practiced at making manual responses. A promising method for manipulating post-decisional variability, therefore, would be to simply change the effector the participant uses with a block of trials, from hand to foot.

Based on the race model architecture (Figure 5.1), a very different prediction is made for how variability in the post-decision effector stage would affect benefit. In contrast to signal strength



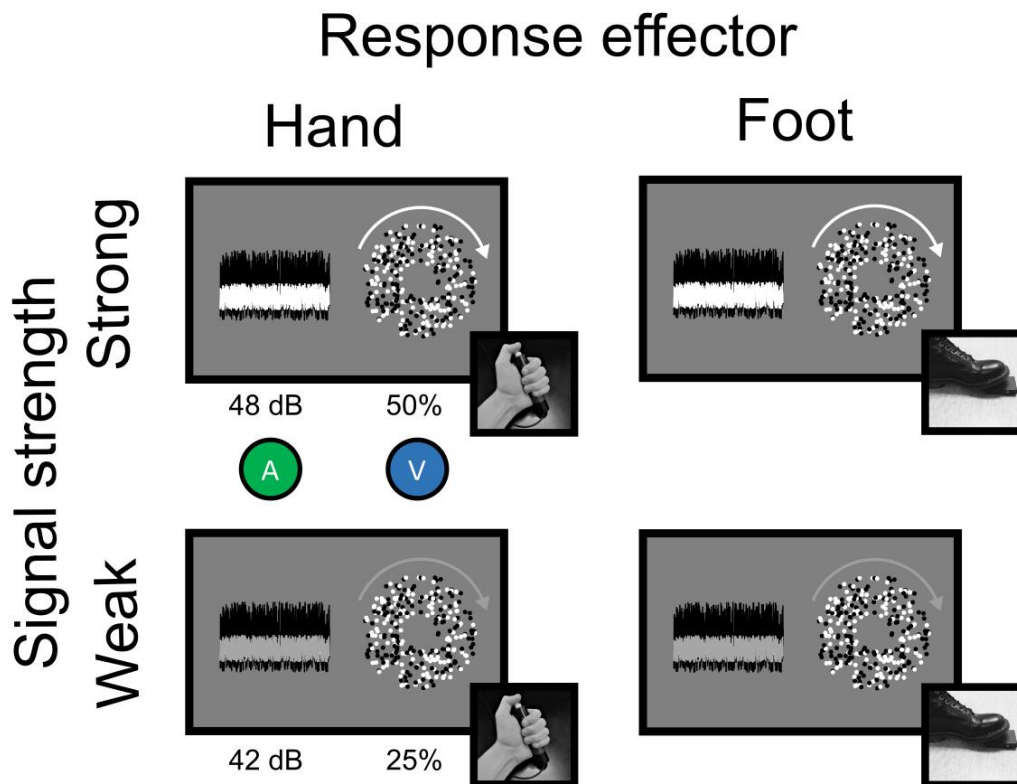
**Figure 5.3 Response effector**

A pedal response component is predicted to be slower and more variable than the manual response component. However, as this variability is post-decisional, any change in variability on this component should have only a small impact on benefits. To demonstrate this, RTs for all 3 signal conditions (100 trials, shown downsampled to 50) were simulated using an independent race model (left panel). To these same trials, an additional ‘motor component’ delay was randomly added from a normal distribution, to simulate slower and more variable responses made with the foot (right panel). Despite the overall changes in RTs, the benefit is ultimately similar in both conditions. This suggests that post-decisional components of RT should not contribute strongly to benefits.

(affecting pre-decisional variability), a change in response effector manipulates post-decisional (or non-decisional) variability. As this comes after the termination of the race process, all variability is ultimately common to signal trial RTs (both unisensory and redundant). As such, no substantial effect on benefits is expected according to race models (**Figure 5.3**). Furthermore, there is no clear reason to expect changes in interactions, as from a race model perspective (and following observations from **Chapters 3 and 4**) these are expected to have a largely pre-decisional locus.

### 5.1.3. Experimental Hypotheses

Overall it was expected that both main factors would significantly manipulate unisensory RTs to allow for an evaluation of the race model principles. For the first factor, *signal strength*, RTs were expected to be slower and more variable when the signal was weak, compared to strong. For the second factor, response effector, RTs were expected to be slower (and eventually more variable) when made with the foot, compared to the hand. Two contrasting experimental hypotheses were made for benefits, based on the race model architecture. For signal strength, the assumption was that pre-decisional RT variability would strongly contribute to benefits. Therefore, it was hypothesised that there would be a main effect of signal strength on benefit, with larger benefits in weak conditions compared to strong conditions. For the second factor, *response effector*, the assumption was that post-decisional components of RT do not contribute strongly to benefits. Therefore, no main effect of response effector for benefit was expected; benefits should be similar for hand and foot conditions.



**Figure 5.4 Stimulus design for Chapter 5**

Large rectangles depict the redundant signal pairings. Signal strength is indicated by colour, with strong signals shown in white and weak signals shown in grey. Small squares indicate the response effector used throughout trials within each condition.

## 5.2. Methods

### 5.2.1. Participants

25 participants were initially recruited at the University of St. Andrews. Of these, there was one drop-out during testing, and a further four participants met the miss rate exclusion criterion (>10% missed trials in any condition). The data from 20 participants (17 female) were thus analysed. Of these, age ranged from 18-24 years. All were naïve as to the aims of the experiment. Normal hearing and normal/corrected-to-normal vision was reported in all cases. Reimbursement was £10.

### 5.2.2. Stimulus Design

Stimuli were constructed according to a 2x2 design with factors *signal strength* (strong, weak) and *response effector* (hand, foot). Stimuli (**Figure 5.4**) were essentially the same as complex-consistent signals in **Chapter 3 (Figure 3.3, bottom left)** with the exceptions noted in the following sections.



#### 5.2.2.1. Auditory Stimuli

The auditory signal was a noise sound. This was presented within background stimulation (50 dB SPL, edge frequencies 0.5/2.4 kHz). Edge frequencies for the signal were 1.0/1.2 kHz. In strong conditions, the signal was presented at 48 dB SPL. In weak conditions, the signal was presented at 42 dB SPL. All noise stimuli were filtered with a 2<sup>nd</sup> order Butterworth filter.

#### 5.2.2.2. Visual Stimuli

The visual signal in all conditions was the coherent clockwise rotation of dots within random linear dot motion. In strong conditions, 50% of the dots rotated, whereas in weak conditions, 25% of the dots rotated.

#### 5.2.3. Procedure

Each trial began by presenting the fixation point (green). The audio-visual background was also presented in complex conditions. Duration of the foreperiod was determined by two components: a fixed duration (1.9 s) and a random component drawn from an exponential distribution (mean: 0.25 s). On signal trials, one of the three signals (A, V, or AV) was then presented until the participant responded (maximum duration: 1.5 s). On catch trials, no signal was presented and the audio-visual background continued for the maximum signal duration (1.5 s). To indicate the end of a trial, the fixation point turned red and all stimulation was halted for 0.1 s. The next trial was then initiated.

Similarly to **Chapter 4**, the first 0.5 s of the foreperiod was considered a 'grace period' in which responses would not trigger the end of a trial. This was because responses during this period were assumed to be erroneous responses to the onset of the end-of-trial feedback, rather than genuine false alarms. The duration was shortened from 1 s to 0.5 s based on the consistent timing of these error responses in **Chapter 4** (see **Section 4.2.3**). If a response was made after this 0.5 s grace period, the fixation point changed from green to red. This indicated that the response had been recorded and the next trial would soon be initiated. All stimulation was halted for 0.1 s before initiating the next trial. Responses recorded within 1.5 s of a signal onset were initially considered valid. If the response was in error, a feedback screen was then presented for 2 s which informed the participant of the error. Responses recorded within the foreperiod (or at any point during catch trials) were considered false alarms. Signal trials in which no response was recorded within the maximum duration were considered misses.

The whole experiment consisted of 16 blocks (4 blocks for each of the 4 experimental conditions). The order of these blocks was randomised for each participant according to a Latin square (all conditions intermixed). Breaks were given between blocks as required by the participant. The whole experiment lasted around 120 minutes.

### 5.3. Results

For all repeated-measures ANOVAs presented in this section, factors adhere to the following ordering: signal strength (strong, weak) × response effector (hand, foot). In some analyses, an additional factor was added: signal modality (A, V, [AV]).

#### 5.3.1. General Performance

Prior to assessing RT, general performance was considered to ensure that the ceiling performance assumption was met. First, false alarms were considered. The percentage of false alarms was averaged across all three signal types (A, V, AV); this is because on false alarm trials, no signal is presented (as the response occurs in the foreperiod). For signal trials, false alarms occurred on 0.42% ( $\pm 0.11\%$ ) of trials. A 2×2 ANOVA revealed a significant main effect of signal strength,  $F(1, 19)=5.849$ ,  $p=0.026$ ,  $\eta p^2=0.235$ . False alarms occurred less frequently in strong conditions (0.252  $\pm 0.079\%$ ) compared to weak conditions (0.582  $\pm 0.159\%$ ). This is expected as weaker signals are more similar to the background noise. Thus, participants are more likely to have mistaken the random background for a signal. There was also a significant interaction between signal strength and response effector,  $F(1, 19)=8.788$ ,  $p=0.008$ ,  $\eta p^2=0.316$ . For hand conditions, the percentage of false alarms increased from strong (0.190  $\pm 0.078\%$ ) to weak (0.734  $\pm 0.211\%$ ) conditions. This was also true of foot conditions, but the increase from strong (0.314  $\pm 0.091\%$ ) to weak (0.431  $\pm 0.123\%$ ) conditions was smaller. No other main effects were significant (all  $F \leq 1.436$ ,  $p \geq 0.246$ ,  $\eta p^2 \leq 0.070$ ). For catch trials, false alarms occurred on 3.531% ( $\pm 0.744\%$ ) of trials. The larger percentage of false alarms on catch trials compared to signal trials is explained by the fact that on catch trials, false alarms can occur over a longer time window (an additional 1.5 s, equivalent to the maximum signal duration). A 2×2 ANOVA revealed a significant main effect of signal strength,  $F(1, 19)=16.780$ ,  $p=0.001$ ,  $\eta p^2=0.469$ . False alarms were occurred less frequently in strong conditions (1.705  $\pm 0.382\%$ ) compared to weak conditions (5.358  $\pm 1.166\%$ ). This is also expected, as if signals are consistently strong it would be more difficult to confuse them with noise. No other main effects were significant (all  $F \leq 3.481$ ,  $p \geq 0.078$ ,  $\eta p^2 \leq 0.155$ ).

Second, misses were considered. The percentage of misses averaged across signal modalities was only 0.873% ( $\pm 0.142\%$ ). A 2×2×3 ANOVA revealed a significant main effect of signal strength,  $F(1, 19)=47.028$ ,  $p < 0.001$ ,  $\eta p^2=0.712$ . The percentage of misses was smaller for strong conditions (0.216  $\pm 0.087\%$ ) than weak conditions (1.530  $\pm 0.226\%$ ), as expected. There was also a significant main effect of signal modality,  $F(2, 38)=15.041$ ,  $p < 0.001$ ,  $\eta p^2=0.442$ . Pairwise comparisons revealed that the percentage of misses on auditory trials (1.505  $\pm 0.268\%$ ) was larger than on redundant trials (0.137  $\pm 0.046\%$ ),  $p < 0.001$ . The percentage of misses on visual trials (0.977  $\pm 0.223\%$ ) was also larger than on redundant trials,  $p=0.003$ . The percentage of misses on auditory trials was not significantly different to visual trials ( $p > 0.05$ ). This is also expected, as redundant trials present two signals (and thus increase

the chances of detection in at least one modality). There was also a significant interaction between signal strength and signal modality,  $F(2, 38)=13.886$ ,  $p<0.001$ ,  $\eta p^2=0.422$ . For strong conditions, the percentage of misses was largest for auditory trials ( $0.325 \pm 0.137\%$ ), followed by visual trials ( $0.249 \pm 0.137\%$ ) then redundant trials ( $0.075 \pm 0.055\%$ ). For weak conditions, percentage of misses was also largest for auditory trials ( $2.686 \pm 0.446\%$ ), followed by visual trials ( $1.705 \pm 0.389\%$ ) then redundant trials ( $0.199 \pm 0.056\%$ ). Thus, the interaction likely comes from a larger difference in unisensory misses between strong and weak conditions. This may indicate that weak auditory signals were slightly more difficult to detect than visual signals. No other main or interaction effects were significant (all  $F \leq 3.283$ ,  $p \geq 0.086$ ,  $\eta p^2 \leq 0.147$ ).

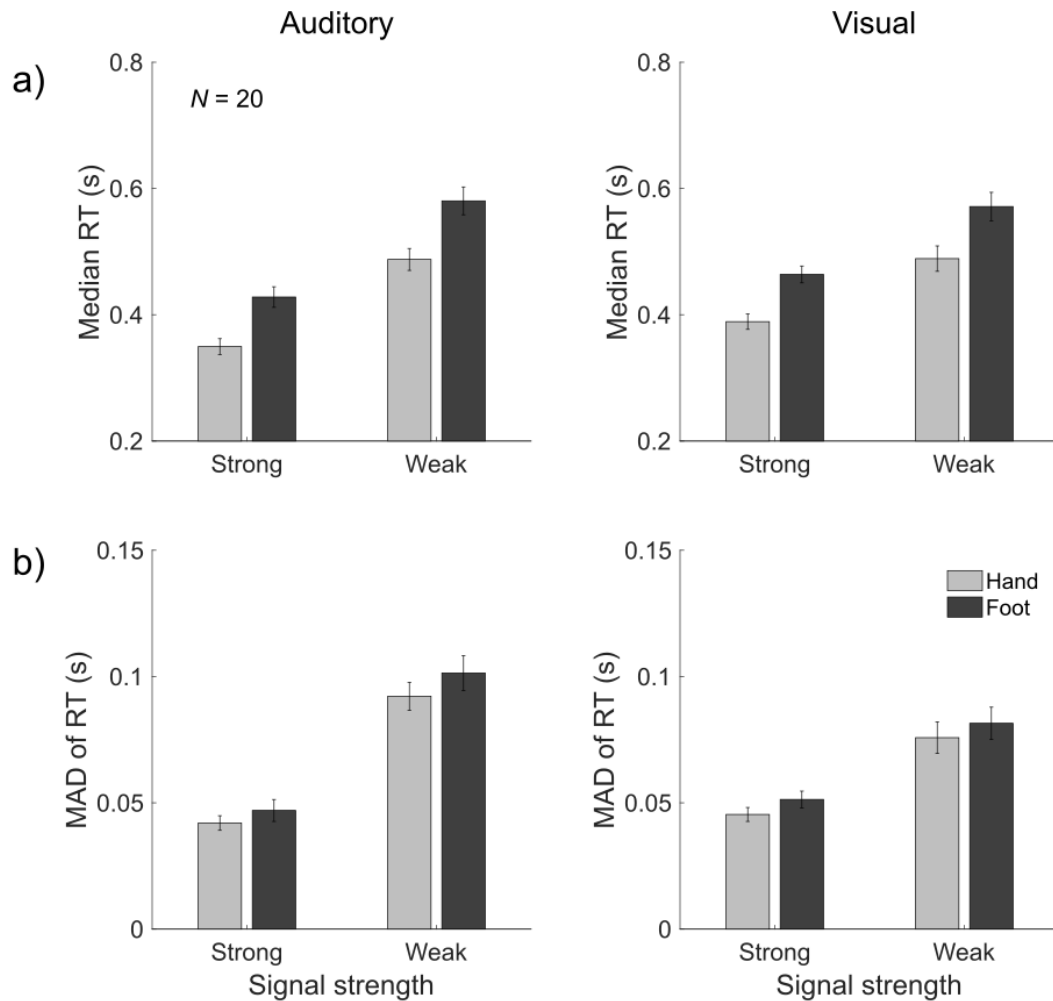
Third, all signal RTs were filtered for outliers according to the MAD criterion (**Section 2.2.2**). On average, fast outliers occurred on 0.69% ( $\pm 0.11\%$ ) of RTs, whereas slow outliers occurred on 0.69% ( $\pm 0.10\%$ ) of RTs. Thus, the overall percentage of outliers ( $\sim 1.38\%$ ) is low, though slightly larger than would be expected if the 1/RT data followed a typical normal distribution (0.27%). As this experiment included weak signals, however, it is possible that slightly more false alarms fell in the early RT window, creating fast outliers. Similarly, participants were likely less practiced in the use of a foot pedal compared to a push button controller, which may have led to more frequent slow outliers. Across all participants, 23360 valid RTs were included in the main analyses, with a mean of 97.3 ( $\pm 0.3$ ) valid RTs per participant for each condition (of a possible maximum of 100 RTs).

Overall, though there were small differences in false alarms and misses between conditions, largely owing to signal strength, performance was close to ceiling. Additionally, very few trials were removed as outliers, and so the number of trials in each condition was close to maximum. For these reasons, accuracy was not considered further in the analyses.

### 5.3.2. Unisensory RTs

Prior to applying the comparative approach, the effect of experimental factors on unisensory RT were analysed. It was expected that each factor would have a significant effect on unisensory RT distributions, both in terms of central tendency (evaluated by median) and variability (evaluated by MAD). These changes can then be interpreted in terms of the race model principles (**Section 1.6.1**) to make directional predictions for benefits. According to these principles, benefits should be larger when unisensory distributions are more similar (equal effectiveness principle) and more variable (variability rule). This gives an indication of the benefits that can be expected prior to applying the comparative approach.

For median RT (**Figure 5.5a**), a  $2 \times 2 \times 2$  ANOVA revealed a significant main effect of signal strength,  $F(1, 19)=213.287$ ,  $p \leq 0.001$ ,  $\eta p^2=0.918$ . As expected, median RT was smaller in strong conditions ( $0.408 \pm 0.013$  s) compared to weak conditions ( $0.532 \pm 0.018$  s). There was also a significant



**Figure 5.5 Unisensory RT analysis**

**a,b)** Median RT and MAD of RT across conditions. The left column shows auditory RTs and the right column shows visual RTs. All bars show the mean of 20 participants ( $\pm 1$  SEM).

main effect of response effector,  $F(1, 19)=141.633$ ,  $p \leq 0.001$ ,  $\eta^2=0.882$ . As expected, median RT was smaller in hand conditions ( $0.429 \pm 0.014$  s) compared to foot conditions ( $0.511 \pm 0.017$  s). There was also a significant interaction between signal strength and signal modality,  $F(1, 19)=11.551$ ,  $p=0.003$ ,  $\eta^2=0.378$ . For strong conditions, median RT for auditory signals ( $0.389 \pm 0.014$  s) was smaller than median RT for visual signals ( $0.427 \pm 0.012$  s). For weak conditions, median RT for auditory signals ( $0.534 \pm 0.019$  s) was slightly larger than median RT for visual signals ( $0.530 \pm 0.021$  s), but overall these weak signals elicited basically the same performance. No other main or interaction effects were significant (all  $F \leq 2.185$ ,  $p \geq 0.156$ ,  $\eta^2 \leq 0.103$ ). Overall, auditory and visual signals elicited more dissimilar performance for strong conditions. This may suggest, according to the equal effectiveness principle, that there will be smaller benefits for strong conditions. Most importantly, however, both factors were successful in strongly manipulating median RT, with a difference between the strong-hand and weak-foot conditions of around 0.2 s.

Second, variability was evaluated by MAD of RT (**Figure 5.5b**). A 2×2 ANOVA revealed a significant main effect of signal strength,  $F(1, 19)=218.046$ ,  $p\leq 0.001$ ,  $\eta p^2=0.920$ . As expected, MAD of RT was smaller in strong conditions ( $0.046 \pm 0.003$  s) compared to weak conditions ( $0.088 \pm 0.005$  s). There was also a significant main effect of response effector,  $F(1, 19)=6.810$ ,  $p=0.017$ ,  $\eta p^2=0.264$ . MAD of RT was smaller in hand conditions ( $0.064 \pm 0.003$  s) compared to foot conditions ( $0.070 \pm 0.004$  s). Notably, this change in variability is much smaller (around 0.007 s) than the change elicited by signal strength (around 0.042 s). There was also a significant interaction between signal strength and signal modality,  $F(1, 19)=10.203$ ,  $p=0.005$ ,  $\eta p^2=0.349$ . For strong signals, MAD of RT for auditory signals ( $0.044 \pm 0.003$  s) was smaller than MAD of RT for visual signals ( $0.048 \pm 0.003$  s). For weak signals, MAD of RT for auditory signals ( $0.097 \pm 0.006$  s) was larger than MAD of RT for visual signals ( $0.079 \pm 0.006$  s). There were no other significant main or interaction effects (all  $F\leq 3.371$ ,  $p\geq 0.082$ ,  $\eta p^2\leq 0.151$ ). Thus, both factors were also successful in manipulating RT variability. In particular, signal strength had the largest effect, almost doubling the RT variability from strong to weak conditions.

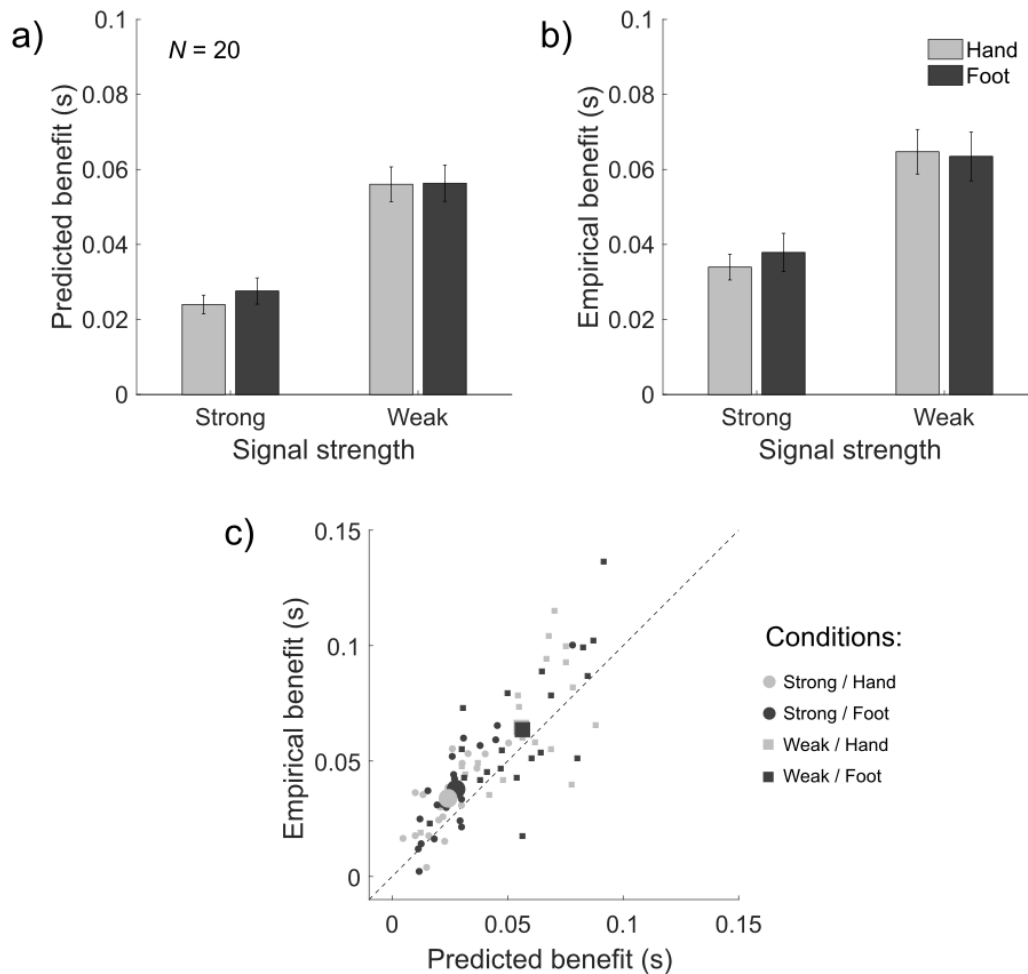
Overall, analysis of unisensory RTs indicated that both factors were successful in significantly manipulating unisensory RT distributions. Both measures showed a significant effect of signal strength: unisensory RT was slower and more variable for weak signals compared to strong signals. Further, both measures showed a significant effect of response effector: unisensory RT was slower and more variable for foot responses compared to hand responses. Following the directional predictions given by race model principles (particularly the variability rule) it is expected that benefits should generally follow the increases in variability across these factors. However, according to the race model architecture, the variability introduced by changes in signal strength should largely impact benefits, whereas variability introduced by changes in response effector should not impact benefits.

### 5.3.3. Comparative Approach Step 1: Multisensory Benefits

Having shown that each factor significantly manipulated unisensory RTs, the first step of the comparative approach looks to evaluate how these factors impacted multisensory RTs. First, the simple race model was applied to give a quantitative prediction for empirical benefit (see **Section 2.3.2.1** and **Section 2.3.2.2**). Second, empirical benefit was calculated. Finally, predicted and empirical benefits were compared.

For predicted benefits (**Figure 5.6a**), a 2×2 ANOVA revealed a significant main effect of signal strength,  $F(1, 19)=52.728$ ,  $p\leq 0.001$ ,  $\eta p^2=0.735$ . The simple race model predicted that benefits would be smaller in strong conditions ( $0.026 \pm 0.003$  s) compared to weak conditions ( $0.056 \pm 0.004$  s). There were no other significant main or interaction effects (all  $F\leq 0.930$ ,  $p\geq 0.347$ ,  $\eta p^2\leq 0.047$ ).

For empirical benefits (**Figure 5.6b**), a 2×2 ANOVA revealed a significant main effect of signal strength,  $F(1, 19)=32.491$ ,  $p\leq 0.001$ ,  $\eta p^2=0.631$ . Empirical benefits were smaller in strong conditions



**Figure 5.6 Predicting and measuring multisensory benefits**

**a, b)** Predicted and empirical benefits across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

**c)** Empirical benefit as a function of predicted benefit. Each point represents data from one participant in one of the four conditions (80 data points total). Large symbols show the group mean.

( $0.036 \pm 0.004$  s) compared to weak conditions ( $0.064 \pm 0.005$  s). There were no other significant main or interaction effects (all  $F \leq 0.585$ ,  $p \geq 0.454$ ,  $\eta p^2 \leq 0.030$ ). Overall, therefore, empirical benefits matched predicted benefits. In addition, the lack of a main effect of response effector supports the hypothesis made based on the race model architecture.

Predicted and empirical benefits were further compared to evaluate their overall correspondence (**Figure 5.6c**). First, predicted and empirical benefits were correlated for each condition of the  $2 \times 2$  design. Of these correlation values, 4 of 4 were positive and highly significant (see **Table 5.1**). Second, differences between predicted and empirical benefit were assessed. Benefit was averaged across conditions and compared with a paired-samples  $t$ -test. The mean predicted benefit ( $0.041 \pm 0.003$  s) was smaller than the mean empirical benefit ( $0.050 \pm 0.004$  s),  $t(19) = 3.697$ ,  $p = 0.002$ , two-tailed. To assess whether this difference was consistent across conditions, predicted benefit was

subtracted from empirical benefit for each condition. A 2×2 ANOVA revealed no significant main or interaction effects (all  $F \leq 0.486$ ,  $p \geq 0.494$ ,  $\eta p^2 \leq 0.025$ ). Thus, the difference between predicted and empirical benefit ( $0.009 \pm 0.002$  s) was consistent across conditions.

Overall, therefore, there is a strong correspondence between predicted and empirical benefits on both the group and individual levels. However, as also seen in **Chapters 3 and 4**, the simple race model consistently underestimates empirical benefit. This is because the simple race model does not consider interactions, which can also contribute to benefits.

#### 5.3.4. Comparative Approach Step 2: Quantifying Interactions

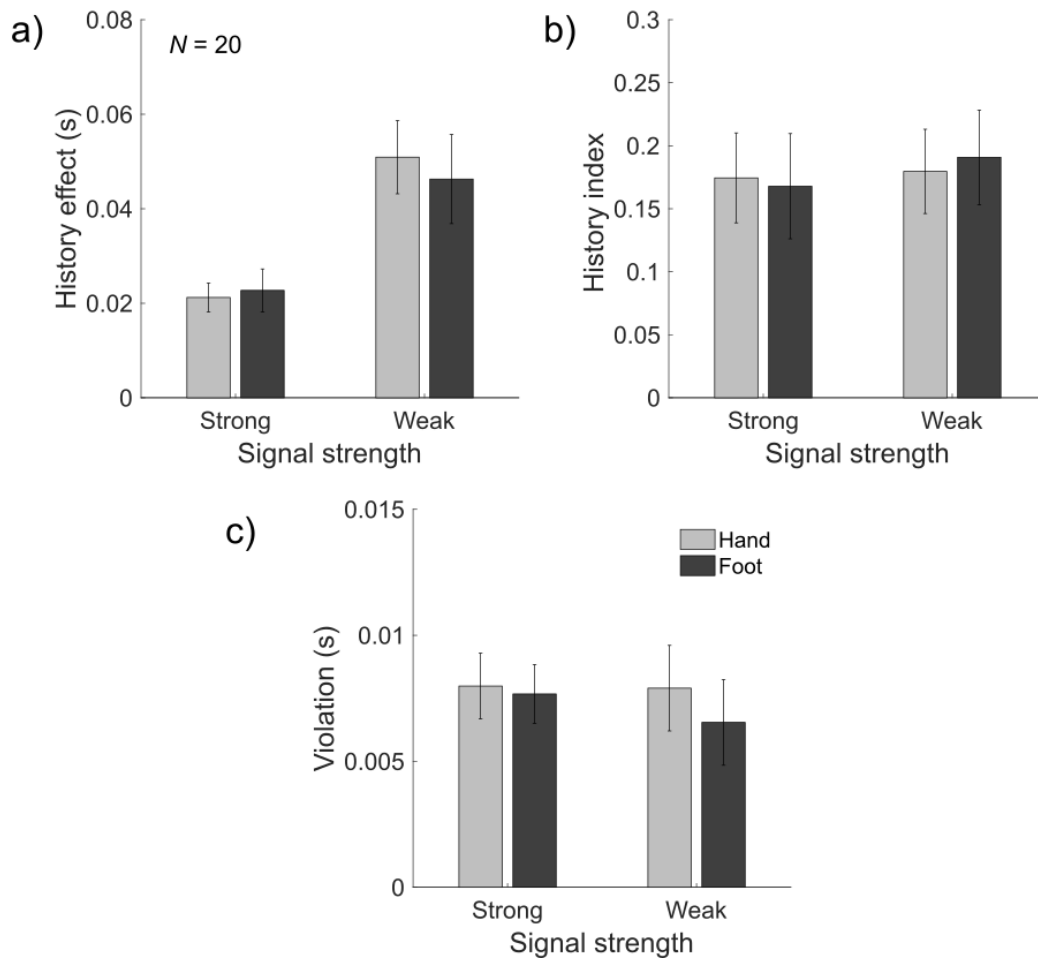
The second step of the comparative approach is to quantify two interactions which are not accounted for by the basic race model architecture. First, the history effect (**Figure 5.7a**) was quantified (see **Section 2.3.3.1**). A 2×2 ANOVA revealed a significant intercept,  $F(1, 19)=111.178$ ,  $p < 0.001$ ,  $\eta p^2 = 0.854$ . Thus, the overall history effect ( $0.035 \pm 0.003$  s) was greater than 0, and must be accounted for. There was also a significant main effect of signal strength,  $F(1, 19)=13.048$ ,  $p = 0.002$ ,  $\eta p^2 = 0.407$ . The history effect was smaller for strong conditions ( $0.022 \pm 0.003$  s) than weak conditions ( $0.049 \pm 0.006$  s). There were no further significant main or interaction effects (all  $F \leq 0.227$ ,  $p \geq 0.639$ ,  $\eta p^2 \leq 0.012$ ). As a further evaluation of trial history effects, the history index (**Figure 5.7b**) was quantified (see **Section 2.3.3.1**). This provides a measure of how much the history effect contributed to overall unisensory RT variability, and therefore to benefits. Interestingly, a 2×2 ANOVA revealed no significant main or interaction effects (all  $F \leq 0.160$ ,  $p \geq 0.694$ ,  $\eta p^2 \leq 0.008$ ). Therefore, though weak signals elicit a larger history effect, the relative contribution of the history effect to variability of unisensory RTs is proportional for both strong and weak signals.

Second, violation of Miller’s bound (**Figure 5.7c**) were quantified (see **Section 2.3.3.2**). A 2×2 ANOVA revealed a significant intercept,  $F(1, 19)=73.836$ ,  $p < 0.001$ ,  $\eta p^2 = 0.795$ . Thus, the overall violation area ( $0.008 \pm 0.001$  s) was greater than 0, and must be accounted for. However, there were no significant main or interaction effects (all  $F \leq 0.328$ ,  $p \geq 0.574$ ,  $\eta p^2 \leq 0.017$ ).

Overall, both interactions were significantly present in the data. The history effect was larger when signal strength was larger, but the overall contribution to benefit (over experimental factors) did not change according to the history effect. The average violation ( $0.008$  s) also roughly corresponds

**Table 5.1 Pearson correlation coefficients (and  $p$  values) for predicted and empirical benefits**

Signal strength	Predicted vs empirical	
	Response effector	
	Hand	Foot
Strong	0.746 ( $p < 0.001$ )	0.890 ( $p < 0.001$ )
Weak	0.671 ( $p = 0.001$ )	0.693 ( $p = 0.001$ )



**Figure 5.7 Quantifying interactions**

**a, b)** History effect and history index across conditions.

**c)** Violation of Miller's bound across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

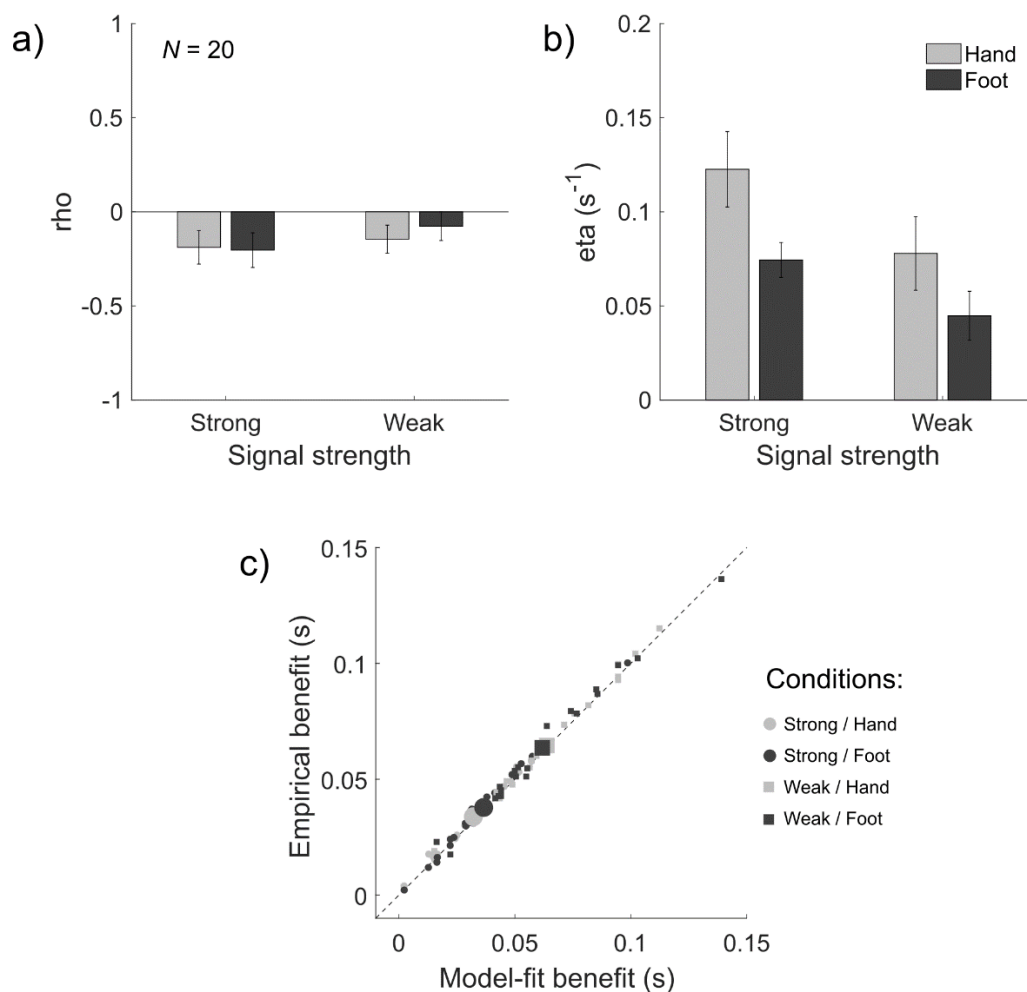
to the difference observed between predicted and empirical benefits (0.009 s); however, violations did not change according to either experimental factor.

### 5.3.5. Comparative Approach Step 3: Applying the Context Variant Race Model

A third step of the comparative approach was to account for benefits and interactions by fitting the context variant race model (see **Section 2.4.4.2**). Broadly, the model parameters were expected to exhibit similar changes across conditions as the corresponding interactions. Therefore,  $\rho$  (based on history index) and  $\eta$  (based on violation) were not expected to change substantially across conditions.

Analysing the  $\rho$  parameters (**Figure 5.8c**), a 2x2 ANOVA revealed a significant intercept,  $F(1, 19)=5.625$ ,  $p=0.028$ ,  $\eta p^2=0.228$ . Thus, the overall correlation modelled between unisensory rate distributions ( $-0.154 \pm 0.065$ ) was significantly different to 0. However, there were no significant main





**Figure 5.8 Fitting the context variant race model**

**a,b)** Best fitting  $\rho$  and  $\eta$  values across conditions. All bars show the mean of 20 participants ( $\pm 1$  SEM).

**c)** Empirical benefit as a function of model-fit benefit. Each point represents data from one participant in one of the four conditions (80 data points total). Large symbols show the group mean.

or interaction effects (all  $F \leq 1.213$ ,  $p \geq 0.284$ ,  $\eta p^2 \leq 0.060$ ). Thus,  $\rho$  (much like the corresponding history index) did not change across conditions.

Analysing the  $\eta$  parameters (**Figure 5.8d**), a  $2 \times 2$  ANOVA revealed a significant intercept,  $F(1, 19) = 59.001$ ,  $p < 0.001$ ,  $\eta p^2 = 0.756$ . Thus, additional variability was added to unisensory rate distributions to account for the variability of redundant rate distributions ( $0.080 \pm 0.010 \text{ s}^{-1}$ ). However, contrary to the analysis of violation, there was also a main effect of signal strength,  $F(1, 19) = 5.797$ ,  $p = 0.026$ ,  $\eta p^2 = 0.234$ . The  $\eta$  parameter was larger for strong conditions ( $0.098 \pm 0.013 \text{ s}^{-1}$ ) than for weak conditions ( $0.061 \pm 0.013 \text{ s}^{-1}$ ). There was also a main effect of response effector,  $F(1, 19) = 7.011$ ,  $p = 0.016$ ,  $\eta p^2 = 0.270$ . The  $\eta$  parameter was larger for hand conditions ( $0.100 \pm 0.016 \text{ s}^{-1}$ ) than for foot conditions ( $0.060 \pm 0.009 \text{ s}^{-1}$ ). There were no further significant main or interaction effects (all  $F \leq 0.431$ ,

$p \geq 0.519$ ,  $\eta p^2 \leq 0.022$ ). Interestingly, therefore, *eta* showed changes across both factors, which were not found in violations.

As an assessment of model fit, benefit was calculated using the interpolated model-fit redundant RT distribution in place of the empirical redundant RT distribution. First, model-fit and empirical benefits were correlated as in Step 1. Of these correlations, all 4 were strongly positively correlated and highly significant (see **Table 5.2**). Second, differences between predicted and empirical benefit were assessed. Benefit was averaged across conditions and compared with a paired-samples *t*-test. Mean model-fit benefit ( $0.048 \pm 0.004$  s) was smaller than mean empirical benefit ( $0.050 \pm 0.004$  s),  $t(19) = 5.435$ ,  $p < 0.001$ , two-tailed. To assess whether this difference was consistent across conditions, empirical benefit was subtracted from model-fit benefit for each condition. A 2x2 ANOVA revealed no significant main or interaction effects (all  $F \leq 0.593$ ,  $p \geq 0.451$ ,  $\eta p^2 \leq 0.030$ ). Thus, the difference between model-fit and empirical benefit ( $0.002 \pm 0.0003$  s) was consistent across conditions.

Overall, the context variant race model provided a good fit, allowing for a much better account of the benefit. Further, the lack of changes in *rho* correspond well to the history index results. For *eta* however, there were main effects not observed in the corresponding violation analysis. Possible reasons for this are discussed in the next section.

#### 5.4. Discussion

In this chapter, I applied the comparative approach to test the effects of variability in two processing stages on multisensory benefit. Both factors (signal strength and response effector) were successful in significantly manipulating unisensory RTs in terms of both average (median) and variability (MAD). However, based on observations of the race model architecture (**Figure 5.1**) very different hypotheses were made for the ultimate effect these manipulations would have on benefits. For the first factor, *signal strength*, a main effect was expected on the analysis of benefits, whereby weak stimuli elicit larger benefits than strong stimuli. This was observed in both predicted and empirical benefit. However for the second factor, *response effector*, no main effect was expected in the analysis of benefit, as hand and foot conditions should give equivalent benefits. Correspondingly, there was no

**Table 5.2 Pearson correlation coefficients (and *p* values) comparing model-fit and empirical benefits**

Signal strength	Model-fit vs empirical	
	Response effector	
	Hand	Foot
Strong	0.993 ( $p < 0.001$ )	0.997 ( $p < 0.001$ )
Weak	0.997 ( $p < 0.001$ )	0.993 ( $p < 0.001$ )

change in predicted or empirical benefits for this condition. Overall, therefore, the results on benefits are in line with predictions made based on the race model architecture.

An important consideration for the above is that, although the race model architecture predicts that response variability should not contribute to benefits, the simple race model (as a predictive tool) is blind to the source of variability. A more convincing result, therefore, might have been to observe a main effect of response effector in predicted benefits, but not empirical benefits. The fact that this main effect in predicted benefits was not observed may alternatively suggest that the size of the variability added by changing from hand to foot responses was simply not strong enough to influence the size of the benefit, regardless of whether the source was pre-decisional or post-decisional. Indeed, the change in unisensory MAD between hand and foot (though significant) was actually relatively small ( $\sim 0.007$  s). Nonetheless, *because* this change in variability is so small (in comparison to substantial changes in median RT), this suggests that (in contrast to previous pooling approaches which have posited a response stage locus of the RSE) targeting the response component is ultimately not useful for manipulating the size of the RSE.

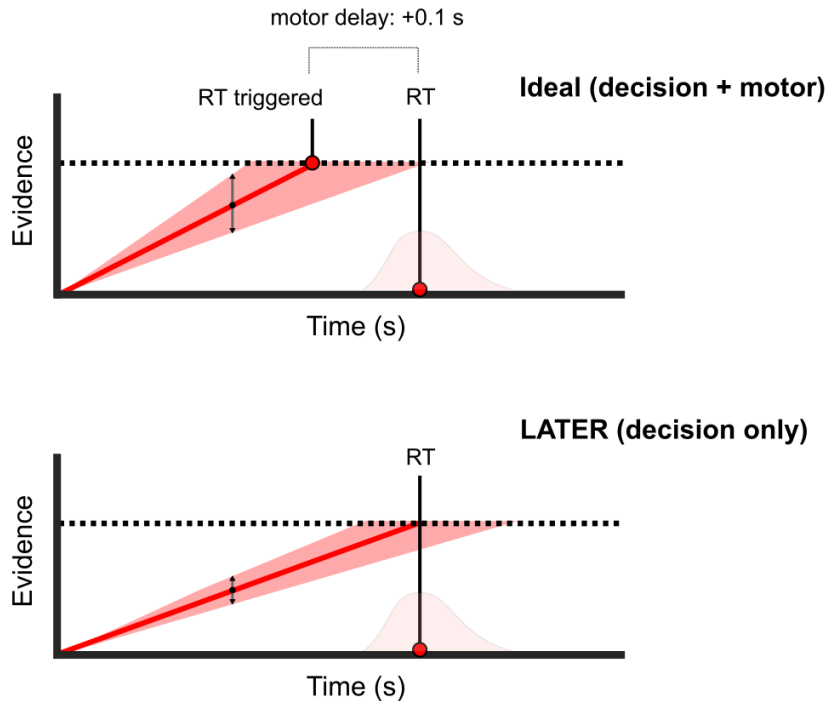
The present experiment corroborates previous research that suggests that signal strength significantly contributes to multisensory benefits (e.g. Diederich & Colonius, 2004). Previously, this effect has been explained in terms of the neurophysiological principles (e.g. Stein & Stanford, 2008; Stein et al., 2014), which suppose that benefits should increase when signal strength decreases (discussed further in **Section 7.1.2**). However, previous research has shown that this broad rule can fail (Chandrasekaran et al., 2011; Juan et al., 2017; Otto et al., 2013). As an alternative, the race model principles of equal effectiveness and the variability rule have been shown to explain this effect in terms of response time distributions; weak signals elicit more variable RTs. Further to this, the correspondence between race-model predicted benefit and empirical benefit was shown to extend to the individual participant level.

Regarding interactions, it is interesting to note that violation was not affected by either factor. This would seem to contrast with a previous study which observed that violation was manipulated by signal strength, with weaker signals eliciting smaller violations (Minakata & Gondan, 2018). There are however experimental differences between these two studies which may account for this result. For instance, these authors intermixed strong and weak signals in a go/no-go task, such that in some blocks participants only responded to strong signals and ignored weak signals (and vice versa). In addition, these authors computed the difference on the group level rather than the individual level. An alternative possible explanation of this discrepancy, however, concerns the role of context explored in **Chapters 3 and 4**. In the present experiment, we presented continuous background noise, which was equivalent in strong and weak conditions. Thus, there was little change in the context

between unisensory and redundant trials depending on signal strength. In Minakata and Gondan's study, however, signals were sudden-onset without background stimulation. In this case, the sudden onset of two strong signals compared to one (as for redundant compared to unisensory trials) is a larger change in context, whereas the onset of two weak signals compared to one is a smaller change in context.

Regarding the model parameters, a general correspondence with the interactions was expected. In line with this, there was no change in  $\rho$  across factors (which was in accordance with the lack of main effects for the history index). For  $\eta$  however, main effects of signal strength and response effector were observed, whereas violation did not change across conditions. Regarding signal strength, it is possible that stronger signals elicit more noise than weaker signals (as shown in **Chapter 3** for sudden-onset stimuli) as  $\eta$  is simply more sensitive to such effects than violation. Regarding the main effect of response effector, however, there is no clear reason that noise should be smaller for foot responses than hand responses. In fact, this likely points to problems in the assumptions made by the context variant race model. The context variant race model builds the interaction parameters on top of LATER units (see **Section 2.4.4.1**), which are fitted to unisensory distributions. The LATER model is a fairly simplistic model of RT, which assumes that all of the RT is decision time. As such, it does not explicitly model any non-decision components of RT, such as delays coming from the response component. Though this simplification likely causes little issue when comparing across responses made with the same effector, it may cause issues when comparing different effectors (e.g. hand and foot). This is because, as we have seen, large delays of around 0.082 s were added by responding with the foot compared to the hand. While we can safely attribute this to non-decisional processing, the LATER model can only interpret this change as a change in decisional processing, (i.e. slower evidence accumulation). As such, less noise will be required to produce the same interference simply because of changes in the LATER parameters (**Figure 5.9**).

In order to further investigate this possibility, I ran a simulation using the context variant race model (see **Section 9.7** in the appendix for full details). I simulated 100 RTs from the context variant race model, and repeated the simulation 1000 times to obtain different samples of RTs. Next, I added an additional motor component (0.1 s) to all simulated sets of RTs. Fitting the context variant race model back to this data revealed a reduction in the recovered values for both  $\mu$  and  $\sigma$ . Further, the recovered  $\eta$  parameter was essentially halved ( $0.051 \text{ s}^{-1}$ ) compared the value used to initialise the simulation ( $0.100 \text{ s}^{-1}$ ). Thus, the reduction in  $\eta$  observed from hand to foot responses likely arises from an issue in fitting the LATER model across conditions with large changes in non-decisional components, rather than a genuine effect in decisional processes.



**Figure 5.9 Problems modelling changes in non-decisional RT components with LATER models**

In this experiment, the factor *response effector* produced a large change in median RT, which can safely be attributed to changes in *non-decisional* components of RT (i.e. the post-decisional response stage). Ideally, such non-decisional RT components would be modelled separately from decisional components when fitting the RT distribution (top panel). The LATER model however, as a simple approximation of the sequential sampling framework, has only two parameters ( $\mu$ ,  $\sigma$ ), which both reflect changes in *decisional* components. As such, the increased delay with foot responses will be interpreted as a change in decision time, and thus will be fitted by different  $\mu$  and  $\sigma$  values (bottom panel). Compared to hand responses,  $\mu$  values for foot responses became smaller to account for the increase in median (see **Section 9.8** for average model parameter values by condition). Similarly,  $\sigma$  values for foot responses were smaller because, as  $\mu$  becomes smaller, less trial-to-trial variability in rate values is needed to capture the same degree of variability in RT (vertical arrows). This is due to the geometry of how rates project on the threshold.

One solution to this issue would be to include a non-decision component into the model, which explicitly models the post-decisional response element of RT. The pre-decisional component (modelled by the existing 6 parameters of the model) is expected to be the same for hand and foot conditions, and should change only between strong and weak conditions. Conversely, the post-decisional component is expected to be the same for strong and weak conditions, and change only between strong and weak conditions. The overall model parameters could be fit to the data for all conditions in the design simultaneously, according to these assumptions. Though it was not possible to implement such a model using current fitting procedures (and is thus unfortunately beyond the scope of the present thesis), developments towards this goal would allow firmer interpretation on the underlying changes in processing between these factors.

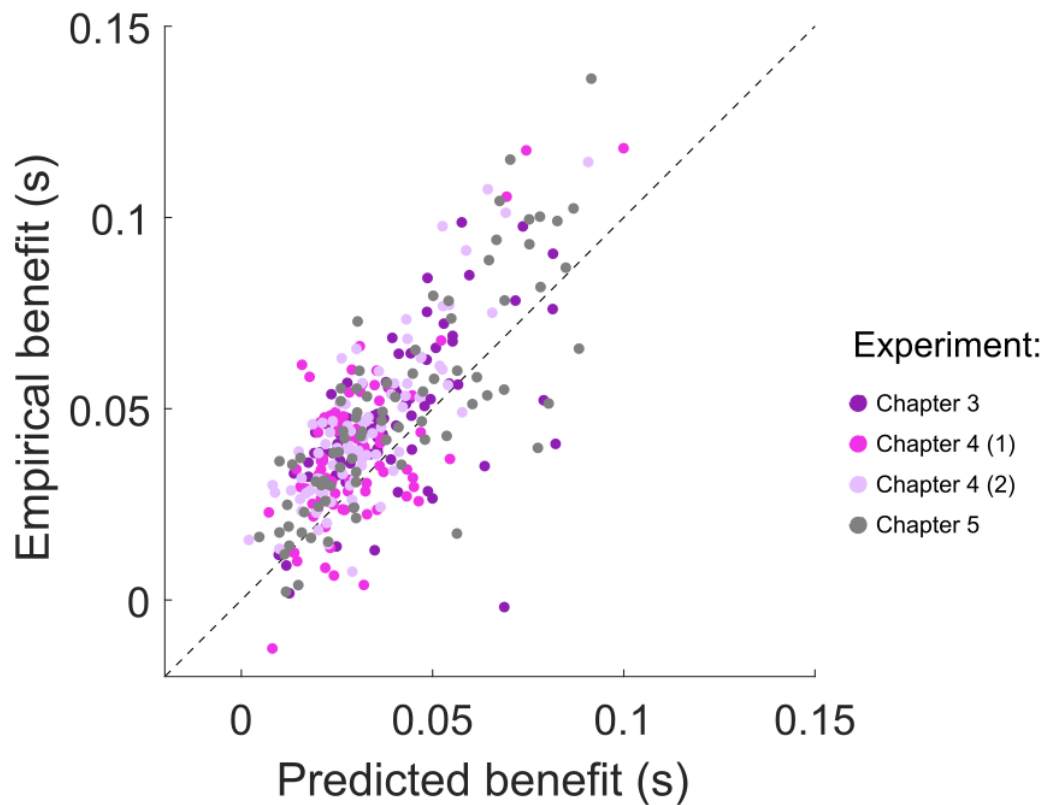
## 6. Examining Trends in Experimental Data

Across the experimental chapters presented in this thesis, four complete redundant signal experiments have been reported. Together, these experiments provide a substantial body of data, with 95367 valid RTs collected in total. In addition, each of these four experiments followed the same 2x2 within-subjects design and applied the same three analytical steps according to the comparative approach (see **Figure 1.19** for a review). Therefore, each experimental dataset (20 participants, 4 conditions) provides 80 unique values for each benefit and interaction measure, as well as each model parameter fitted. As a final exploratory step, therefore, these data values (320 in total) were evaluated together to reveal overall trends.

### 6.1. Race Model Principles and Empirical Benefits

One of the overarching goals of this thesis was to evaluate the power of the race model architecture to predict multisensory benefits. Given that the race model is typically rejected, following the standard interpretation of the seminal paper by Miller (1982), there has been little work on evaluating the model's predictive principles. These principles were only recently highlighted by Otto et al. (2013). To date, only a small body of work has evaluated the two principles (**Section 1.6.1**), however this has been largely supportive of their propositions. For instance, Harrar et al. (2017) tested different stimulus onset asynchronies, and found that benefits were largest when the resulting unisensory RT distributions were most similar. This supports the principle of equal effectiveness. Otto et al. (2013; see their Figure 6) also noted that, across signal strength conditions, empirical benefits follow predicted benefits on the level of group (i.e. Vincent averaged) RT distributions. To contribute to this growing body of work, I evaluated the predictions of the race model (and the individual principles) over the *individual* RT distributions collected here.

In the previous experimental chapters, the race model's predictive power has been assessed by measuring the correlation between individuals' empirical benefits and the benefits predicted by the simple race model. In the majority of conditions, these correlations were positive and significant (14/16 conditions). This suggests that the simple race model is a robust and effective tool for estimating empirical benefit for individual participants. To further highlight this here, I collated all values of predicted and empirical benefit from the previous experimental chapters (**Figure 6.1**). Overall, despite very different experimental manipulations across these experiments, empirical benefits clearly follow the simple race model prediction. This suggests that, if nothing else, the simple race model is an easy-to-implement, parameter-free 'rule of thumb' for estimating empirical benefit. More importantly, this strongly suggests that the race model may provide the basic combination rule for responses to multiple signals.



**Figure 6.1 Evaluating the quantitative predictions of the simple race model**

Empirical benefit shown as a function of the benefit predicted by the simple race model. Each data point represents a participant. Each experiment (colour-coded) contained 20 participants, and consisted of 4 within-subjects experimental conditions (individual conditions not colour coded). The overall plot therefore shows 320 measures of benefit across 80 participants.

What drives the simple race model's prediction (and, assuming the framework is correct, the empirical benefit)? As noted in the introduction (**Section 1.6.1**), two principles determine the overall quantitative prediction (Otto et al., 2013). For each principle, a corresponding measure can be computed and evaluated. The first principle, *equal effectiveness*, states that benefit increases as the difference between unisensory RT distributions decreases. Thus, a *negative* relationship is expected between empirical benefit and the difference between unisensory distributions. To assess this principle, one straightforward evaluation would be to compute the difference between the averages for unisensory RTs. As a simple measure, the absolute difference between medians of unisensory RTs was calculated as

$$\text{median difference} = |\bar{X} - \bar{Y}|, \quad (29)$$

where  $X$  and  $Y$  are the unisensory RT distributions.

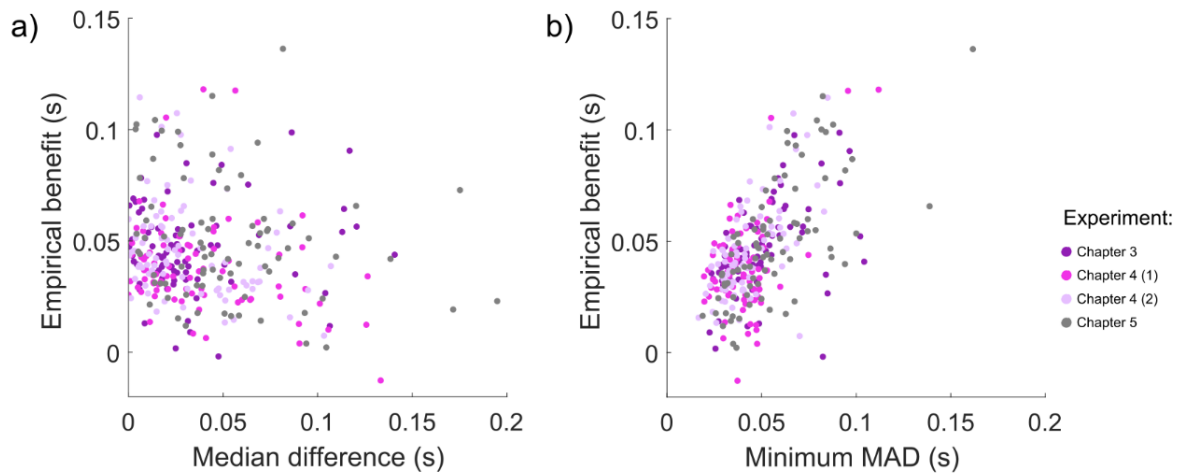
The second principle, the *variability rule*, states that benefit increases as unisensory RT variability also increases. Specifically, benefit is determined by the unisensory distribution with the smallest variability; as shown by Otto et al. (2013) in RT simulations (reproduced here in **Figure 1.14**), further increasing the variability of the most-variable unisensory distribution does not add to benefits. Thus, a *positive* relationship is expected between the empirical benefit and the smaller of the two unisensory RT variabilities. To assess this, one straightforward method is to simply select the smallest unisensory variability. As a simple measure, the minimum unisensory MAD was calculated as

$$\text{minimum MAD} = \min(\text{MAD}(X), \text{MAD}(Y)). \quad (30)$$

Before exploring the relationship between these predictive measures and empirical benefits (**Figure 6.2**), it is worth highlighting again how each principle works to increase benefit. According to the race mechanism, the benefit is determined by the size of the overlap in unisensory RTs; it is only where unisensory RTs overlap that a race mechanism can produce statistical facilitation. Benefits become larger, therefore, as the overlapping area between unisensory RT distributions increases. Conversely, if there is very little to no overlap in unisensory RTs, then no benefit is expected. Each principle suggests the extent of this overlap in a different way. First, consider equal effectiveness. If both distributions are completely equally effective, then there is a perfect overlap between them; thus, benefits should be maximised for those particular signals. Second, consider the variability rule. If both unisensory RT distributions are very variable, then the area of overlap can potentially be large because both distributions are wide. Notably, variability has the larger potential to increase benefits. For instance, consider two distributions demonstrating equal effectiveness. If both distributions are not very variable, then there will be only a small benefit of a race mechanism. This is because even though there is complete overlap, the area of this overlap is small. However, if both distributions are very variable, there will be a large benefit of a race mechanism as the overlapping area is much larger (as a demonstration of this, compare the top and bottom panels of **Figure 1.14**). Further, even in cases where equal effectiveness does not hold entirely (i.e. RT distributions have differences in their medians), then there may still be a substantial overlap if distributions are sufficiently variable. Overall, therefore, the variability rule is the main driver of benefit according to race models principles.

Having reviewed how the race model principles work, the relationships shown in **Figure 6.2** become clear. Considering the median difference (**Figure 6.2a**), there appears to be a negative relationship overall as expected; however, this relationship is rather weak. One reason for the weak relationship is that in the majority of cases, the median difference is actually quite small (i.e. <0.05 s). This is expected given that the signals used in these experiments were designed to elicit roughly similar





**Figure 6.2 Evaluating the race model principles**

**a, b)** Empirical benefit as a function of the median difference (**Equation 29**) and minimum MAD (**Equation 30**) of unisensory RTs. Each data point represents a participant. Each experiment (colour-coded) contained 20 participants, and consisted of 4 within-subjects experimental conditions (individual conditions not colour coded). The overall plot therefore shows 320 estimates across 80 participants.

unisensory performance. A second reason is that, with small differences in median performance, large differences in benefits can be observed as the RT variability more strongly determines the area of overlap. Considering the minimum MAD (**Figure 6.2b**), a strong positive relationship is shown, as expected. This demonstrates, in line with the evaluation of median difference, that unisensory RT variability is a much stronger determinant of benefits. Ultimately, therefore, the equal effectiveness principle can be understood as a prediction for *maximising* benefits given two predetermined signals. Variability, however, is a much better indicator for the size of the benefit provided unisensory RTs show roughly similar performance.

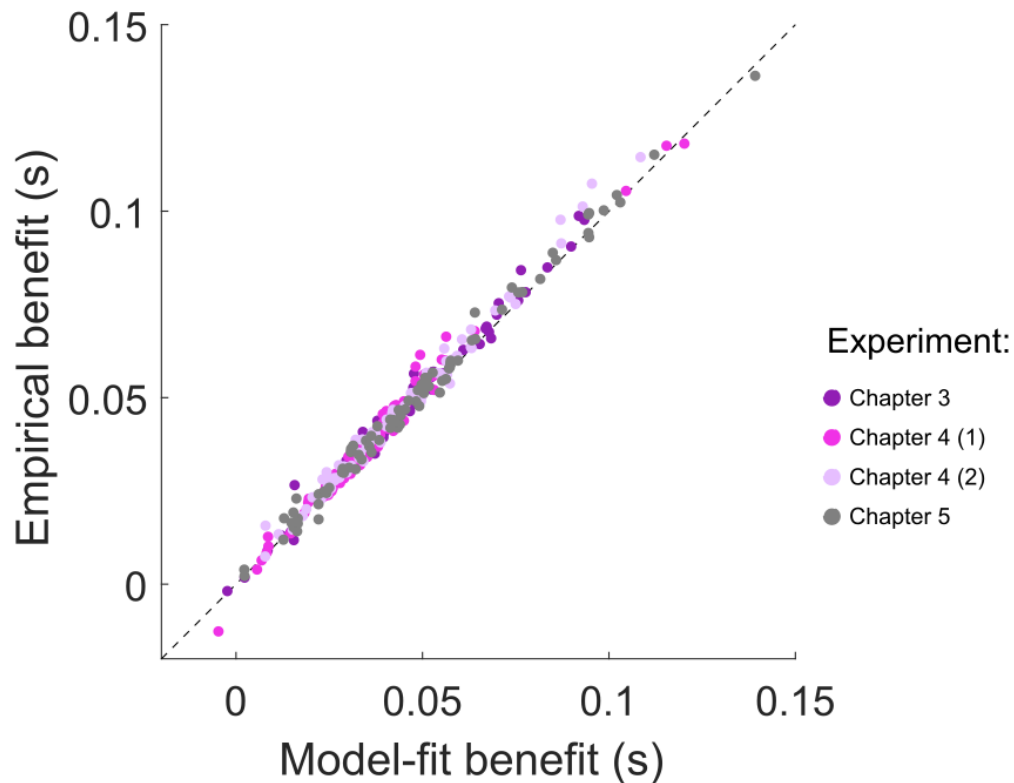
Overall, this evaluation strongly supports the race model principles as an explanation of overall changes in benefits. This adds to previous studies (Harrar et al., 2017; Otto et al., 2013) which have evaluated these principles in some way. In particular, this analysis has provided strong support for the variability rule with a large scale analysis on the level of individual participants. Given that unisensory performance was not strictly calibrated here, further experiments may wish to add to this evaluation by investigating the equal effectiveness principle in detail. By following similar signal onset manipulations to the authors above (Harrar et al., 2017; Otto et al., 2013), for instance, experiments could evaluate the extent to which calibrating signal timing for each individual (to produce maximally overlapping RT distributions) maximises the overall multisensory benefit observed.

## 6.2. Modelling Empirical Benefits and Interactions

The previous section demonstrated that the most basic race model provides a strong account of changes in benefit, suggesting a basic *combination rule* for multiple signals. However, it is important to note that this is not a complete account. For instance, in **Figure 6.1**, the simple race model consistently underestimates the empirical benefit. This is because the race model does not account for additional *processing interactions* which are observed empirically. This thesis has attempted to provide a more complete account of the RSE observed in each experiment by quantifying these interactions and subsequently accounting for them in the race model. In this section, I evaluate the effectiveness of this model's account.

As a reminder, both interactions contribute to multisensory benefits beyond the basic race mechanism. First, trial history effects contribute to benefits by increasing the variability of unisensory RTs. When a unisensory trial is repeated, RTs are typically fast, and when a unisensory signal is switched, RTs are typically slow (**Section 1.7.1**). This is an additional source of variability, which contributes to benefits according to the variability rule (**Section 1.6.1.2**). This interaction therefore violates the assumption of statistical independence (**Section 1.4.1**) in the basic race model. Second, violations of Miller's bound represent the area of multisensory benefit which is not accounted for by the basic race mechanism (**Section 1.7.2**). Under the race framework, this interaction suggests context variance (the processing of each unisensory component signal is different between unisensory and redundant trials). This interaction also violates the assumption of context invariance (**Section 1.4.2**) in the basic race model. Ultimately, both interactions can be incorporated into the race model framework, as demonstrated by the context variant race model (Otto & Mamassian, 2012). To account for history effects, a correlation parameter *rho* is included which allows statistical dependencies. To account for violations, the noise parameter *eta* is added to the variability of each decision-unit on redundant trials, which allows context variance.

First, I consider the context variant race model's ability to account for individual *benefits*. In experimental chapters, this was done by computing the correlation between model-fit and empirical benefits. It was expected that if interactions were successfully accounted for, then the model would be able to capture the empirical benefit much more effectively than the simple race model alone. In support of this, correlations in all 16 conditions were strongly positive and highly significant. This suggests that the context variant race model is able to account for empirical benefit across very different experimental manipulations. To further highlight this here, I collated all values of model-fit and empirical benefit from experimental chapters (**Figure 6.3**). All values closely follow the identity line, which suggests that the context variant race model provides a near-perfect explanation of the empirical benefit on the individual level.

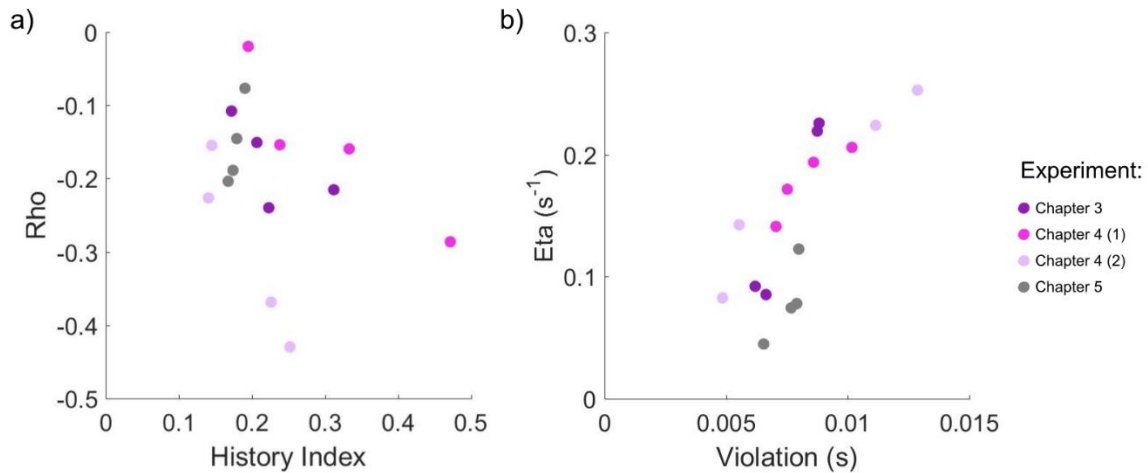


**Figure 6.3 Evaluating the context variant race model account of benefits**

Empirical benefit shown as a function of the model-fit benefit. Each data point represents a participant. Each experiment (colour-coded) contained 20 participants, and consisted of 4 within-subjects experimental conditions (individual conditions not colour coded). The overall plot therefore shows 320 measures of benefit across 80 participants.

Second, I consider the context variant race model's ability to account for *interactions*. In experimental chapters, this has been more broadly evaluated by the correspondence in the main effects observed in ANOVAs, and the general trends in group means. The consideration here is that if the model parameter properly accounts for the corresponding interaction, then both should change between experimental conditions in corresponding ways. In contrast to benefits, comparison on the group level here allows for greater clarity; this is because interactions are based on fewer trials than benefits. The history effects, for instance, takes only a small number of unisensory trials (in which repetitions or switches occurred) into account. Similarly, violations typically occur over a smaller area of the whole distribution than benefits (i.e. the fast tail). The larger variability in these measures, therefore, makes it difficult to evaluate overall trends on the individual level. In the following sections, therefore, I have computed the group average for each interaction and each model parameter (**Figure 6.4**).

First, the relationship between trial history effects and *rho* was considered (**Figure 6.4a**). Because trial history effects typically result in negative correlations, it is expected that there will be a *negative relationship* between these measures (i.e. a larger trial history measure corresponds to a



**Figure 6.4 Relating interactions to model parameters**

**a)** History index as a function of the  $\rho$  parameter.

**b)** Violation as a function of the  $\eta$  parameter. In both plots, each point shows the average of 20 participants.

more negative  $\rho$  value). In particular, it is useful to compare history index and  $\rho$ , as both are indicative of contributions to overall benefit. Indeed, observing the overall trend (**Figure 6.4a**), though there is some variability, there appears to be a weak negative relationship, as expected. This suggests that  $\rho$  broadly captures empirical trial history effects in the data.

Second, the relationship between violations of Miller's bound and  $\eta$  was considered (**Figure 6.4b**). Violations, under a race model framework, are indicative of greater context variance. One possible manifestation of this, as proposed by the context variant race model, is that each decision-unit is corrupted by additional noise ( $\eta$ ) when another signal component is also present. Therefore, a *positive* relationship is expected between violation and  $\eta$ , as greater context invariance would produce more noise. Observing the trend between these measures (**Figure 6.4**), the relationship appears to be positive, as expected. This suggests that  $\eta$  captures the violation area well. Further, this supports the context variant race model's explanation of this violation i.e. context variance arising from increased noise in redundant conditions.

Overall, this section suggests that race models can provide a complete account of the RSE by considering processing interactions (statistical dependencies, context variance) which impact benefits beyond the basic combination rule (statistical facilitation). Furthermore, the effectiveness of the context variant race model to account for both benefits and interactions is demonstrated; this model captures effects observed in interactions, and provides a clear explanatory framework for how such interactions arise. By employing this model, the RSE can be accounted for near perfectly (**Figure 6.3**).

It is important to be aware, however, that more complex models always have the potential to offer better explanations, because adding more parameters almost always improves fitting. For this

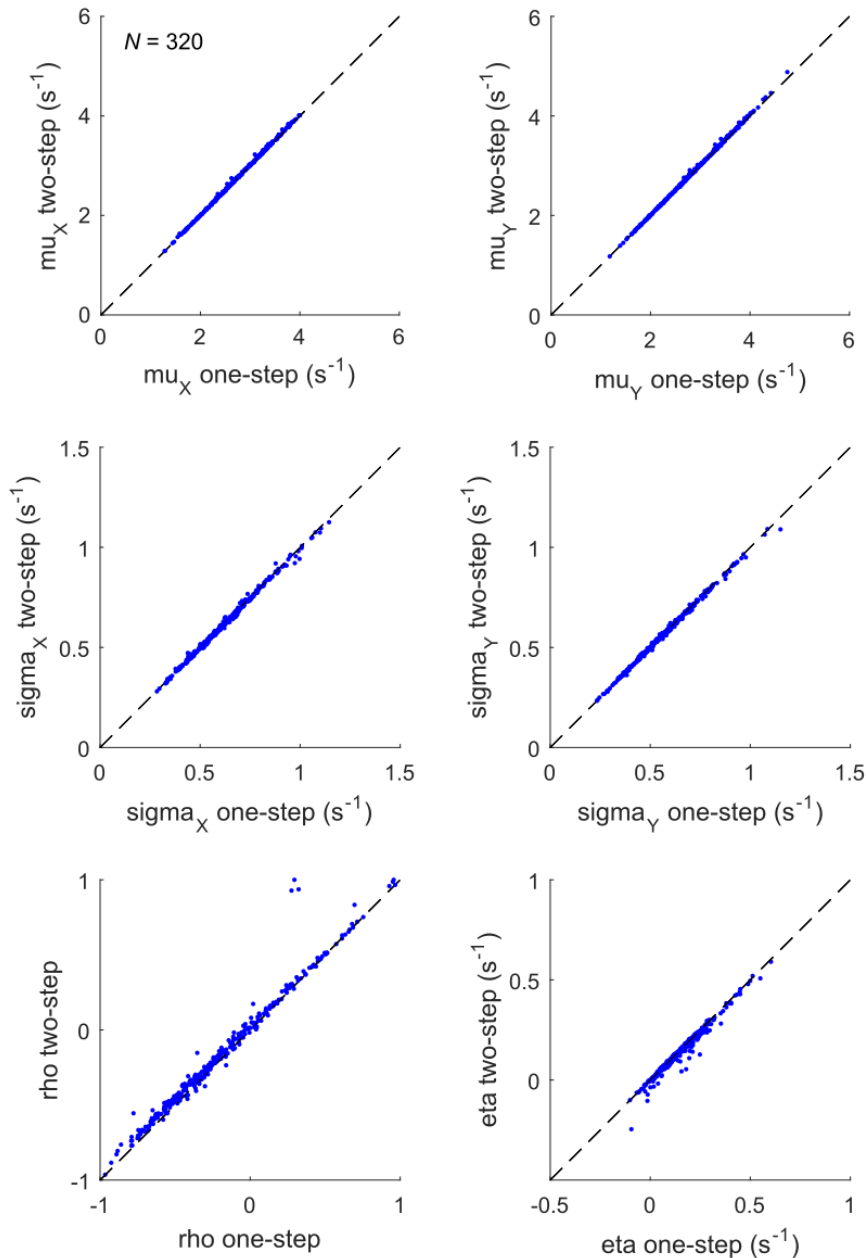
reason, the *parsimony* of models is also important (Burnham & Anderson, 2002; Palminteri et al., 2017). In the following model-based section, therefore, I evaluate the context variant race model fitting in more detail.

### 6.3. Evaluating the Context Variant Race Model

In the previous section, I noted the context variant race model's ability to provide an excellent account of benefits and interactions. This suggests that a clear framework for the RSE is in fact possible using race models. In this section, I evaluate the specific formulation of the context variant race model proposed by Otto and Mamassian (2012). In a first evaluation, I examine an alternative fitting procedure. This allows for a test of whether different fitting procedures produce essentially the same results. The alternative fitting procedure also provides a basis for further model selection. In a second evaluation, I compare nested versions of the context variant race model using model selection tools (**Section 2.4.3**). This allows for a test of whether the additional interaction parameters are useful for explaining behaviour, and broadly suggests whether the context variant race model is a parsimonious account of the RSE. As a final evaluation, I examine small discrepancies between the context variant race model fit and benefits. This suggests room for further improvement in modelling studies beyond the present thesis.

#### 6.3.1. Comparing Fitting Procedures

An important consideration for the context variant race model is whether the precise fitting procedure is important. Since Miller (1982), common practice in evaluating the race model has followed two steps. First, the unisensory RT distributions are used to define a comparison model. Second, this model is evaluated in light of the redundant RT distribution. This is also implicit in Otto & Mamassian's (2012) original fitting procedure. First, two LATER models (with parameters  $\mu_x$ ,  $\mu_y$ ,  $\sigma_x$ ,  $\sigma_y$ ) are used to describe the two unisensory distributions and provide a basic (i.e. independent) race model. Second, two additional parameters ( $\rho$ ,  $\eta$ ) are added onto this basic model to provide a fit to the redundant distribution. Though this two-step fitting procedure has evidently been effective (see **Section 6.2**), alternative fitting procedures, in which all modelling is completed in one step, are often desirable. To engage in model comparison, for instance, a single log-likelihood value must be returned for all data which goes into fitting the model. This essentially means that a two-step fitting procedure is not valid for model comparison; a one-step fitting procedure, which fits unisensory and redundant RTs simultaneously, must be used.



**Figure 6.5 Comparing fitting procedures for the context variant race model**

Each plot shows one of the 6 context variant race model parameters. The two-step fitting procedure value is plotted as a function of the one-step procedure value (dashed line indicates identity). Each point shows a single participant ( $N=80$ ) in one of four within-subjects experimental condition (320 points total).

To evaluate whether the one-step fitting procedure returns equivalent results for the present data, I used a novel version of context variant race model where all 6 parameters ( $\mu_x$ ,  $\mu_y$ ,  $\sigma_x$ ,  $\sigma_y$ ,  $\rho$ ,  $\eta$ ) were fit simultaneously. As an initial quality check, I evaluated whether the established two-step fit method (applied in all experiments here) and the novel one-step fit method produced equivalent results. Overall, there was no difference in the main or interaction effects in the experimental chapters when the analyses were re-run using the one-step fitted  $\rho$  and  $\eta$  values.

This indicates a broad equivalence in the fitting procedures, and a general stability in model fitting (see **Section 9.8** in the appendix for average fit values for all parameters according to both procedures). To further demonstrate this, I compared each individual parameter fit value using the two-step fitting procedure with the equivalent parameter from the one-step fitting procedure (**Figure 6.5**). The vast majority of points fall on the identity line, indicating the two methods are basically equivalent. In summary, this suggests that the one-step fitting procedure is a useful method for future modelling work, and also a platform for further model comparisons.

### 6.3.2. Nested Model Comparison

The previous section established the one-step fitting procedure as an equivalent method of applying the context variant race model. Next, I used the likelihood values returned from the one-step fitting procedure in a nested model comparison. For the first time, this will allow for a model comparison of context variant race models, and an evaluation of the model's overall parsimony.

As reviewed earlier (**Section 2.4.3**), nested models (e.g. restricted versions of the context variant race model) can be compared in various ways. The simplest way would be to use the LRT (**Section 2.4.3.1**). The limit to this test, however, is that it only allows two nested models to be compared at once. To compare multiple models across the 320 datasets here would thus require a large number of tests, and it would be difficult to establish a clear impression of how each performs relative to the others. Similar methods such as AIC (**Section 2.4.3.2**) and BIC (**Section 2.4.3.3**), however, allow multiple candidate models (nested or non-nested) to be compared at once. Such procedures also declare one overall 'winning' model (i.e. the most parsimonious model). The latter tools have been used by Zehetleitner et al. (2015), one of the few recent papers to fit a formal pooling model to RSP data. Here, these authors compared nested pooling models using the BIC as their model selection criterion. For these reasons, I use AIC and BIC to conduct the following nested model comparison.

To conduct the model comparison, I established three limited versions of the context variant race model (see **Table 6.1** for the parameters in each model). For the purposes of simplification, I will call the full context variant race model the *Full* model in the following section. The three limited versions were created by eliminating one or both of the interaction parameters of the Full model. This can be done by fixing the parameter value to zero: a *rho* of 0 indicates statistical independence, and an *eta* of 0 indicates context invariance. The three limited versions thus included: a version where the *eta* parameter was excluded (the *Rho* model), a version where the *rho* parameter was excluded (the *Eta* model), and finally a version in which both *rho* and *eta* were excluded. This latter model includes only the LATER model fits to the unisensory RT distributions, and is thus equivalent to an Independent Race Model (*IRM*).

### 6.3.2.1. Model Selection Using Minimum AIC & BIC

As an initial step, I quantified the overall percentage that each model won (i.e. returned the minimum criterion value) according to both AIC and BIC. This was evaluated across all 80 participants tested in 4 conditions (for 320 comparisons total). As a reminder, AIC penalises by the number of model parameters as a means to select the most parsimonious model. BIC penalises not only by model parameters but the number of data points. In the present datasets for each individual, there are approximately 300 RTs, with 100 RTs in each signal condition (X, Y, XY). Relatedly, a correction for AIC is generally recommended when the number of data points divided by the number of parameters is small (i.e.  $< 40$ ; Burnham & Anderson, 2002; Wagenmakers & Farrell, 2004). However, as the most complex model here (**Table 6.1**) has only 6 parameters ( $300/6 = 50$ ), no AIC correction was necessary. AIC and BIC were therefore calculated as described in a previous section (**Section 2.4.3**).

First, I considered the AIC criterion (**Figure 6.6**). According to this comparison, the Full model was selected the most frequently, followed closely by the Eta model, the IRM model, and finally the Rho model. Thus, while the model which includes both parameters is overall the most frequently selected model, a limited model which includes only the *eta* parameter also performs well. Further, the independent race model which includes *no* interaction parameters outperformed a limited model which included only the correlation parameter *rho*. Overall, therefore, it would seem that while both parameters can be important, *eta* alone is generally a more useful parameter than *rho*. This result is in some sense expected, given that only *eta* can capture violations of Miller’s bound, which is a large deviation from the basic race model seen in the vast majority of individual participants.

Second, I considered the BIC criterion (**Figure 6.7**). With this criterion, which applies a harsher penalty, the results are strikingly different. According to this comparison, the most selected model is the IRM, which does not model any interaction parameters (and is effectively the most basic race model). This is followed by the Eta model, the Rho model and finally the Full model. That the Full model is least selected links directly to the increased penalty for more complex models with BIC. However, this comparison corroborates the finding of the AIC comparison that the *eta* parameter is the more useful interaction parameter of the two. Further, it is interesting that by following this criterion, it is not considered necessary to model interactions at all. This highlights the explanatory power of even the basic race model architecture alone.

**Table 6.1 Nested models included in the comparison (and corresponding parameters)**

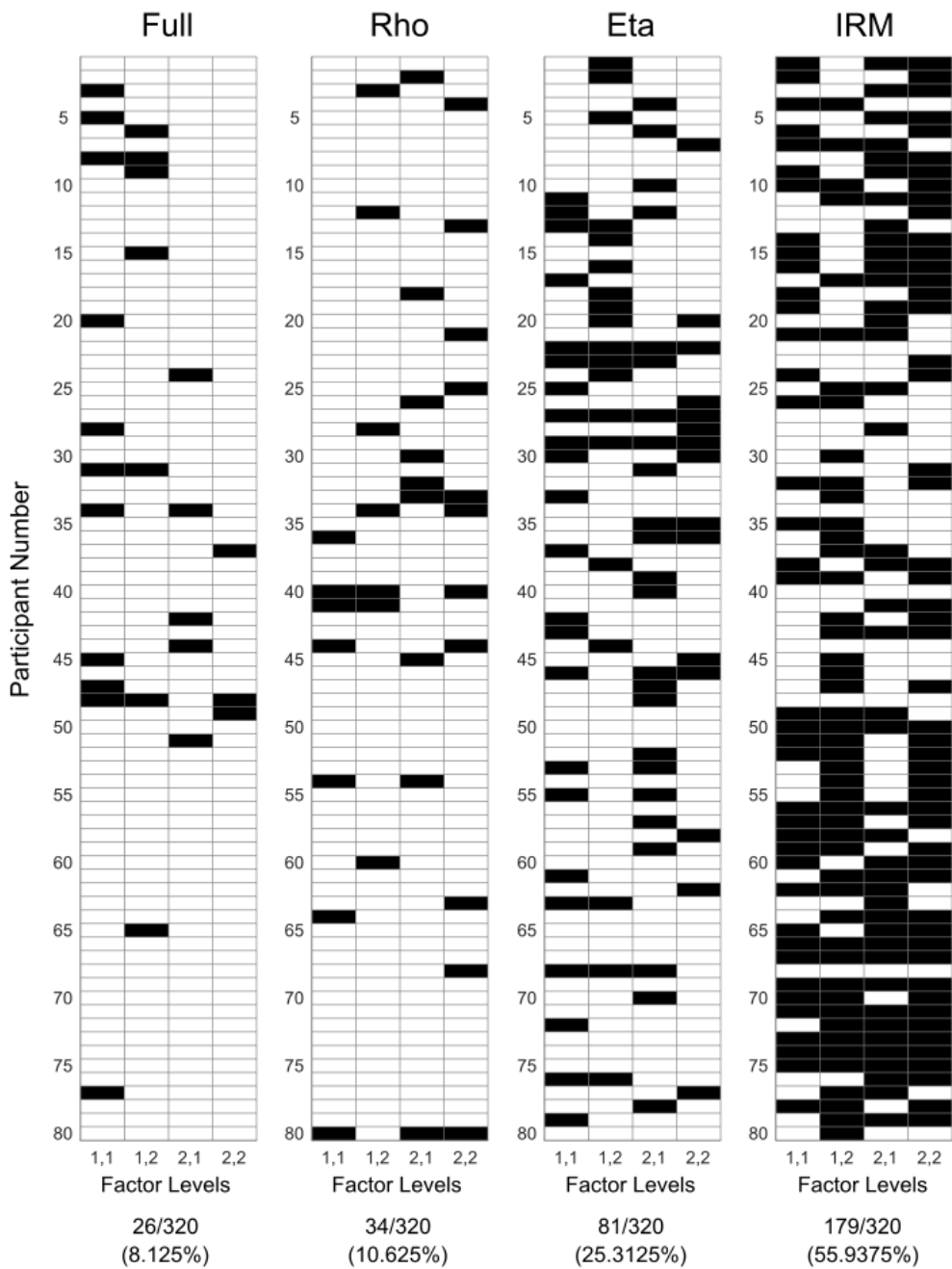
Model	Parameters Included					
<b>Full</b>	$\mu_X$	$\mu_Y$	$\sigma_X$	$\sigma_Y$	$\rho$	$\eta$
<b>Rho</b>	$\mu_X$	$\mu_Y$	$\sigma_X$	$\sigma_Y$	$\rho$	-
<b>Eta</b>	$\mu_X$	$\mu_Y$	$\sigma_X$	$\sigma_Y$	-	$\eta$
<b>IRM</b>	$\mu_X$	$\mu_Y$	$\sigma_X$	$\sigma_Y$	-	-





**Figure 6.6 Nested model comparison using AIC**

Each of the four plots represents a model included in the comparison (see **Table 6.1** for the parameters included in each model). Each cell represents a participant in one of the four experimental conditions (80 participants, for 320 comparisons in total). A black cell indicates that the model had the lowest AIC value (i.e. was the winner of the comparison). The summary statistics below indicate the frequency (and overall percentage) that the model won according to the AIC criterion. Note that in this comparison, as there is only one winner selected from all 4 models, all percentages sum to 100% across plots. Factor levels (factor 1 level, factor 2 level) can be used in conjunction with the participant number to determine the experiment the data came from. Each experiment contained 20 participants, which were entered in order. For instance, Participant 61 (Condition Label 2, 1) is the first participant for Chapter 5 (in the Weak-Hand condition).



**Figure 6.7 Nested model comparison using BIC**

Each of the four plots represents a model included in the comparison (see **Table 6.1** for the parameters included in each model). Each cell represents a participant in one of the four experimental conditions (80 participants, for 320 comparisons in total). A black cell indicates that the model had the lowest BIC value (i.e. was the winner of the comparison). The summary statistics below indicate the frequency (and overall percentage) that the model won according to the BIC criterion. Note that in this comparison, as there is only one winner selected from all 4 models, all percentages sum to 100% across plots. Factor levels (factor 1 level, factor 2 level) can be used in conjunction with the participant number to determine the experiment the data came from. Each experiment contained 20 participants, which were entered in order. For instance, Participant 61 (Condition Label 2, 1) is the first participant for Chapter 5 (in the Weak-Hand condition).

In summary, initial comparison with minimum AIC and BIC reveals an interesting discord. AIC most often selected the context variant model (which models both interactions) whereas BIC selected an independent race model (which models neither interaction). The BIC, however, employs a harsher penalty for complex models. This latter result is interesting, as one of the few nested pooling model comparisons (Zehetleitner et al., 2015) utilised BIC for comparisons. Further research may wish to work towards a non-nested comparison of race and pooling models. As the pooling models used by Zehetleitner (5-9 parameters) were more complex than the independent race model used here (4 parameters), it may be that in many cases BIC would also select simple race models over simple pooling models.

The choice between AIC and BIC for model comparison is not straightforward. Both are valued for different reasons, and a review of these reasons is far beyond the scope of this small section. For this reason, both were presented for a complete initial overview. For more detailed evaluation, however, I evaluated just one method (AIC) in particular. One reason for this, as noted by Burnham and Anderson (2002, p. 33), is that BIC “[tends] in realistic situations, to select models that are too simple (i.e. underfitted).” This would certainly fit with the empirical observation of processing interactions in the majority of participant datasets, which are mostly ignored in the BIC comparison. I therefore follow Burnham & Anderson’s approach and further investigate models using the AIC.

#### 6.3.2.2. AIC Difference

To further explore the effectiveness of the different models, it is useful to consider not just the minimum AIC value, but the *relative difference* between this minimum value and all other values. Burnham and Anderson (2002; p. 72) explain the value of this with an analogy to racing: here, it is important to know the overall winner out of all competitors, but it is also useful to know how close the competition is. If the competition is very close, then on this particular occasion there may be a role of extraneous factors (e.g. weather) in which competitor edged out overall. Consider this example now in relation to the RT data here. It may be that one model (e.g. Eta) has the minimum AIC for an individual comparison. However, if the competition between models is close (e.g. the Full model was close to the minimum AIC) then it is important to be aware of this, as it indicates the models are both good explanations of this participant’s data. Further, if this participant was re-tested, it may be that changes in extraneous factors (e.g. fluctuations in attention, or fatigue) would cause the Full model to edge out over the Eta model. Using the relative difference scores, therefore, gives an indication of how likely it is that the result could change with a different sample of RTs. If two (or more) models provide AIC values within a certain range, then each provides a compelling account of the data.

To evaluate the relative difference between models with AIC, I computed the *AIC difference* score. Following Burnham and Anderson (2002, p.71) this is calculated as

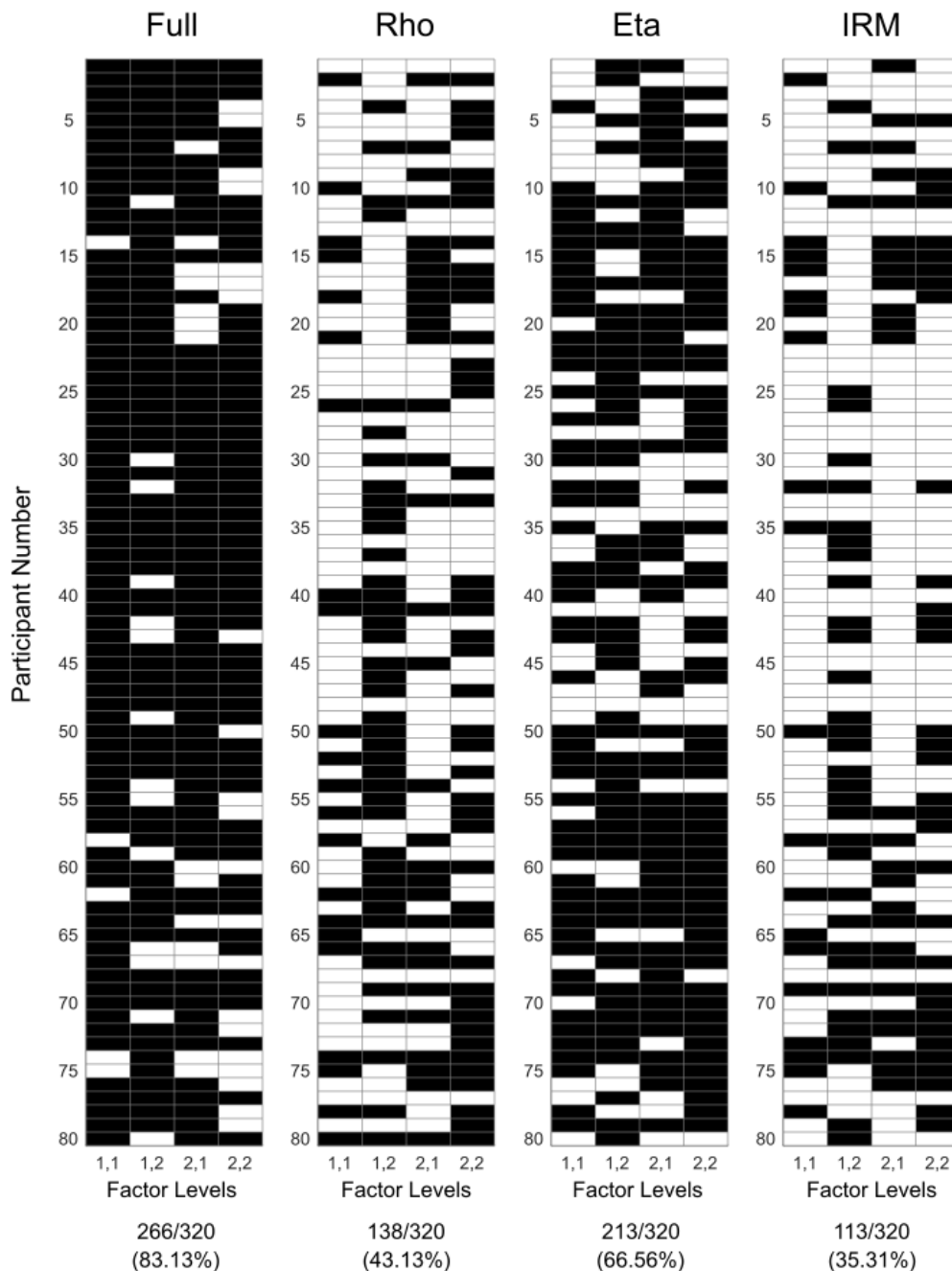
$$AIC\ Difference_i = AIC_i - \min(AIC), \quad (31)$$

where *AIC* is the set of AIC values being compared and *i* is the index for the model. I then followed Burnham & Anderson's (2002, p. 70) 'rules of thumb' for nested models. According to these criteria, any model with an AIC difference score between 0-2 is considered to have "substantial" empirical support. Instead of the overall percentage of times each model 'won', therefore, I computed the number of times each model received an AIC difference score between 0-2 (**Figure 6.8**).

Observing the results of the AIC comparison, the Full model was substantially supported across almost all participants in all conditions. This was followed by the Eta model, the Rho model, and finally the IRM. From this plot, a number of conclusions can be drawn. First, models which include interaction parameters received more support than the basic race model. This is in line with empirical observations of interactions across experiments. Second, as in previous comparisons, models including *eta* are generally more supported than models including *rho*. The fact that *rho* is quite weak (i.e. fairly close to 0) across the experiments presented here likely explains why this parameter is not often selected in comparisons. In contrast to the initial AIC comparison however, there is a substantial increase in the number of times each model was selected when computing the difference rather than the lowest AIC. This suggests that the competition between models was generally quite close, as in many cases different versions of the race model receive substantial support. Further, as it was selected in almost all cases, the AIC differences suggest that the context variant race model offers a parsimonious account of the RSE in the vast majority of cases. For future modelling work, this supports the idea that the context variant race model can be applied on the individual participant level, with little concern regarding overfitting.

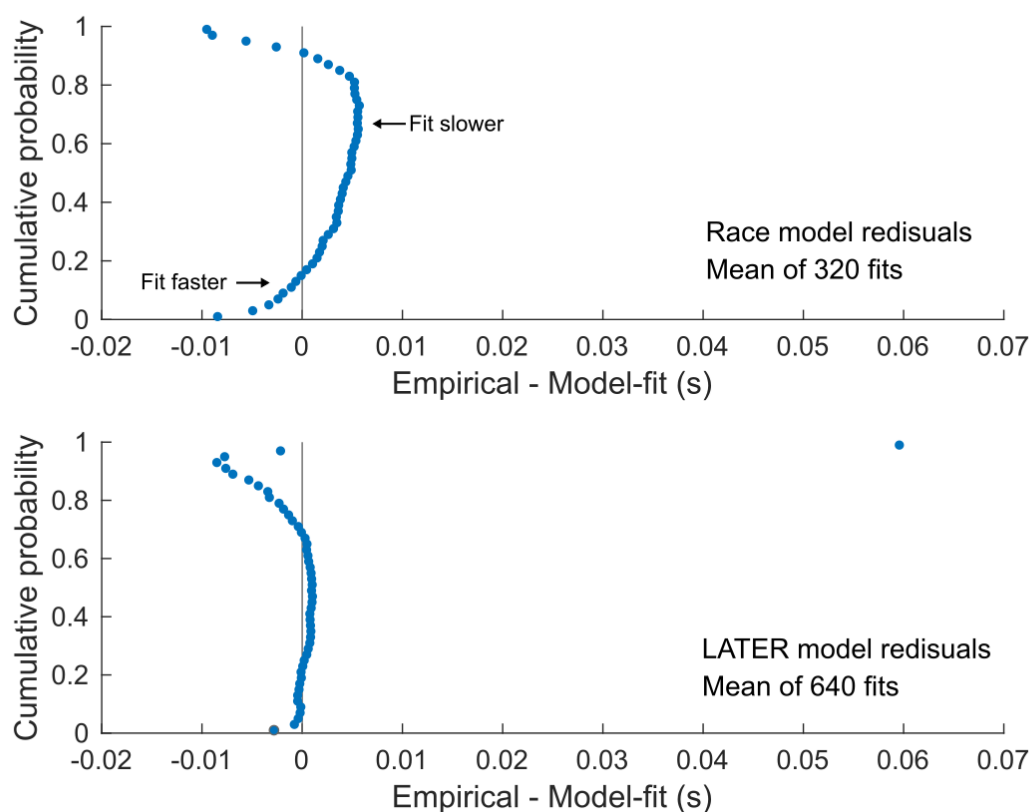
### 6.3.3. Limitations and Future Directions

Over all experiments, it has been shown that the context variant race model provides an excellent account of empirical benefit size on the individual subject level (**Figure 6.3**) It is important to note however that there are small, consistent deviations; specifically, empirical benefit is usually slightly larger than the benefit calculated on model-fit. In experimental chapters, this difference ranged from 0.001 s (**Chapter 5**) to 0.003 s (**Chapter 4**). Although this deviation is very small, it is interesting that it is so systematic. This can indicate potential biases in model fitting procedures. For this reason, it would be useful to evaluate the source of these model deviations.



**Figure 6.8 Nested model comparison using AIC difference score**

Each of the four plots represents a model included in the comparison (see **Table 6.1** for the parameters included in each model). Each cell represents a participant in one of the four experimental conditions (80 participants, for 320 model comparisons in total). A black cell indicates that the model had an AIC difference score between 0-2 (i.e. denoting “substantial empirical support”). The summary statistics below indicate the frequency (and overall percentage) that the model was supported according to the AIC difference score. Note that in this comparison, as there can be multiple models evaluated as equally supported, these do not sum to 100% as before. Factor levels (factor 1 level, factor 2 level) can be used in conjunction with the participant number to determine the experiment the data came from. Each experiment contained 20 participants, which were entered in order. For instance, Participant 61 (Condition Label 2, 1) is the first participant for Chapter 5 (in the Weak-Hand condition).



**Figure 6.9 Evaluating model residuals**

Plots show the difference between the empirical RT data and the model-fit RT data for 50 cumulative probability points. As can be seen for the race model (top panel), there are consistent deviations in which the best-fitting model is slower than the empirical redundant distribution. This is likely an error which propagates from LATER model fitting (bottom panel) which shows a similar deviation from the corresponding unisensory distribution. As 2 LATER model fits are required to fit the race model (one for each unisensory condition), double the amount of fits were included in the average shown.

To determine the source of this deviation, I computed the residual (i.e. the difference between the two-step model-fit distribution and the empirical redundant distribution) over 50 cumulative probability points (**Figure 6.9**). As shown averaged over all 320 participants, these residuals show a systematic misfit. At the extremes of the distribution, the best-fitting model *overestimates* the empirical benefit. Across the majority of cumulative probability points however, the best-fitting model *underestimates* the empirical redundant distribution. The average of these residual values (around 0.002 s) is roughly equivalent to the difference between model-fit and empirical benefit overall (between 0.001 to 0.003 s). Thus, this deviation in the model-fit is likely the source of the consistent difference between model-fit benefits and empirical benefits.

One explanation of this misfit is that it is a result of error propagation starting with the LATER model fitting. Comparing the LATER model fit to its corresponding unisensory distribution (**Figure 6.9**),

a similar trend in the average residual is observed. LATER model fits are generally useful as with just two parameters they quite effectively describe the entire RT distribution. However, as discussed in **Chapter 5**, LATER is a fairly simplistic model which equates RT with decision time (i.e. it does not explicitly model non-decision components of RT). Thus, while it may provide an effective fit to a particular RT distribution, a simplistic model such as LATER may not provide the most robust basis for additional modelling beyond this simple description. Consider here, for instance, that the context variant race model uses *two* LATER models, both of which will contain some degree of error (as shown in **Figure 6.9**). In future work, therefore, it may be useful to explore other models of unisensory RTs, such as the ex-Gaussian (e.g. Luce, 1986; Whelan, 2008) or ex-Wald (e.g. W. Schwarz, 2001). This however, would also require developing alternative implementations of the interaction parameters. As such differences are rather consistent and fairly small in relation to empirical benefit, however, they could possibly be neglected for most experimental purposes.

Overall, despite the minor limitations highlighted in this section, the context variant race model has been shown to offer a robust and effective fit to explain behaviour in the RSP. Depending on the modelling work and level of precision required, it may be useful to develop more complex and refined versions of the model, which even more precisely account for the benefit. However, given that the application of formal models is typically overlooked in most RSP studies (Gondan & Minakata, 2016), the abilities of the context variant race model are far beyond what is done in current practice.

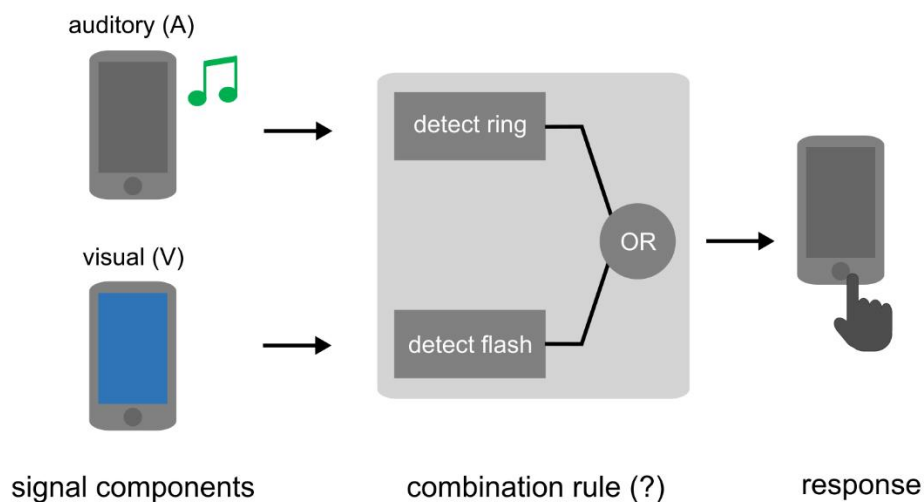
## 7. General Discussion

### 7.1. Race Models Offer a Common RSP Framework

Despite over 100 years of investigation (beginning with Todd, 1912), a common explanatory framework for multisensory RT benefits in the RSP has not emerged. This is because there has been a lack of clarity on the fundamental *combination rule* which explains how two redundant sensory signals are processed to produce a single response. Two main model classes, each suggesting a different combination rule, have so far been tested. The first candidate class, race models, suggests that two decision-units (one for each signal) accumulate evidence in parallel, and the faster of these units triggers the response. This model class, however, has not provided a common framework. The reason for this is that empirical benefits are larger than the upper limit to the basic race model's combination rule, which has been used as evidence to reject the entire model class (Miller, 1982). The second candidate class (pooling models), has since become the dominant candidate, as it can potentially explain larger benefits than race models. However, pooling has so far failed to account for benefits across experiments (Gondan & Minakata, 2016). Importantly, recent research has clearly shown that the previous rejection of the entire race model class can no longer be considered valid (Otto & Mamassian, 2017; C. T. Yang et al., 2018). As such, the potential for more race models to offer a comparative framework has not been adequately explored. The overall goal of this thesis was to explore the potential of the neglected race model class to offer a common framework to analyse redundant signal experiments.

To explore the potential of race models, a *comparative approach* to experimental design and analysis was developed. This was applied across 3 experimental chapters. The goal of this approach was to develop manipulations, based on the race model and perceptual decision-making frameworks, and compare changes in RT across these manipulations. The application of the approach's three analytical steps (see **Figure 1.19**) resulted in three corresponding conclusions. First, across all experiments, race models offered a powerful predictive and explanatory framework for multisensory benefits. The basic combination rule of the race model (**Figure 1.9**), therefore, offers a convincing general explanation of benefits. Second, empirical effects in RT previously thought to be beyond the explanation of race models can actually be interpreted effectively as additional *processing interactions* between different senses. Rather than support for the alternative pooling architecture, therefore, these can be effectively incorporated into the race model architecture. Third, benefits were near-perfectly explained by using a formal race model allowing processing interactions (Otto & Mamassian, 2012). As a formal description of the underlying combination rules and processing interactions, such a model offers a necessary intermediate stage between behaviour and brain processes (e.g.





**Figure 7.1 The race model as the combination rule for multisensory response time benefits**

This thesis began by considering the combination rule the brain employs to determine simple responses to redundant signals. The way the brain combines two signals (e.g. a flash and beep from a smartphone) to make a single response (e.g. answer a call) cannot be directly observed. One approach, as adopted in the experimental work presented here, is to model particular combination rules for similar tasks, and observe how the model's performance matches empirical behaviour. Overall, the results of this thesis suggest that for redundant signal tasks, the brain employs a combination rule similar to a race model, in which the faster of two decision processes ultimately triggers the response.

Forstmann, Wagenmakers, Eichele, Brown, & Serences, 2011; Mulder et al., 2014), which is a fundamental step for the development of further research. Overall, in stark contrast to the current rejection, the race model here has been shown to be a strong candidate for a common RSP framework.

#### 7.1.1. Variability is the Source of Multisensory Benefits

In the absence of a common framework, previous attempts to understand the sources of multisensory benefits have had limited success. Here, by considering benefits in relation to race model principles, and using this framework to develop manipulations, a clear common thread has emerged across experiments: factors which influence the variability of unisensory RTs influence the overall multisensory benefit. In **Chapter 3**, the variability of unisensory RTs was manipulated by changing *stimulus construction*: random (complex) signals elicited more variable RTs than non-random (simple) signals. Correspondingly, complex signals elicited larger multisensory benefits than simple signals. In **Chapter 5**, variability of unisensory RTs was manipulated by changing *signal strength*: weaker signals elicited more variable RTs than stronger signals. Correspondingly, weak signals also resulted in larger multisensory benefits than strong signals. Importantly, the race model framework is able to suggest *which sources* of variability will be important. Generally, as the benefit arises from statistical facilitation of decision times, post-decisional variability should not affect benefits (as one example of

an exception to this broad variability rule). In **Chapter 5**, post-decisional variability was introduced by changing the *response effector*. Foot responses produced more variable RTs than hand responses. However, in line with the suppositions of the race model architecture, this did not produce an overall change in multisensory benefit. Overall these experiments suggest that changes in benefits can be understood largely in relation to changes in the variability of RTs. As highlighted in **Chapter 6**, this appears to hold true across experiments, which suggests a general principle for understanding behaviour in the RSP.

#### 7.1.2. Race Model Principles as a Guiding Framework for Multisensory Research

The results of this thesis contribute to a wider understanding of how benefits can be elicited and manipulated. Previously, in accordance with the dominance of pooling models, explanatory frameworks for multisensory RT benefits have utilised principles derived from neuronal responses (for reviews, see Stein & Stanford, 2008; Stein et al., 2014). These principles, derived from the study of superior colliculus neurons in cats, suggest three guiding rules for multisensory responses; stronger multisensory responses should be observed when component unisensory signals are spatially aligned (the spatial rule), temporally aligned (the temporal rule), and when the individual unisensory signals elicit weak responses (the principle of inverse effectiveness). In the absence of a common framework, these principles have been assumed to apply to RT behaviour as well. This has led many authors to assume that larger multisensory RT benefits should occur with weak signals, aligned in space and in time.

The guiding principles of neuronal responses have been used to reflect on changes in multisensory benefits (e.g. Chandrasekaran et al., 2011; Juan et al., 2017; Minakata & Gondan, 2018). They are even discussed as motivating factors in the design of multisensory signals for behaviour (e.g. Ho et al., 2007; Pomper et al., 2014; van Erp et al., 2015). As noted by Otto et al. (2013), however, behavioural benefits do not necessarily follow these principles in the RSP literature. Consider the results of the experimental chapters of this thesis: no effort was made to align signals in space (in fact, in **Chapter 4 (Part 2)**, there is a very clear misalignment of tactile and visual signals). Despite this, clear benefits are observed throughout, suggesting spatial alignment is not strictly necessary. Further, temporal alignment of signals is not necessary; in fact, larger benefits occur with temporally-misaligned signals which elicit similar unisensory RTs (Harrar et al., 2017; Otto et al., 2013). Finally, the prediction of the inverse effectiveness principle, i.e. that weaker signals should elicit larger benefits, has been shown to fail. In fact, some experiment reporting larger benefits for strong signals compared to weak signals (Chandrasekaran et al., 2011; Juan et al., 2017). Using a neurophysiological framework for explaining behavioural benefits across conditions, therefore, is unsatisfactory, and in some cases predicts effects which are not observed in the behavioural data.

The experimental work in this thesis suggests that the race model's predictive principles (see **Section 1.6.1**) offer a clearer and more appropriate guiding framework for understanding behavioural benefits. These principles effectively make similar predictions to the three principles of multisensory integration, but with additional and important caveats linked to the race model mechanism. For instance, temporal correspondence is also important for race models. The distinction here, however, is that the correspondence of RTs (equal effectiveness) is most important, rather than precise synchrony of physical signals (e.g. Harrar et al., 2017; Otto et al., 2013). Further, the prediction of the inverse effectiveness principle (that weaker signals should elicit larger benefits) is also made by the variability rule of the race model principles; weaker signals should produce slower and more variable RTs, and thus elicit larger benefits. This was demonstrated clearly in **Chapter 5**. The caveat here, however, is that signals must be equally effective. If there are very large deviations in average RT which are introduced by making signals weaker, then there will be a reduced potential for a race mechanism to produce facilitation; this would actually lead to *smaller* benefits for weak signals. Indeed, in cases where the principle of inverse effectiveness was challenged (Chandrasekaran et al., 2011; Juan et al., 2017), weak stimuli demonstrated a large difference in average unisensory RTs (i.e. a lack of equal effectiveness). As shown by Otto et al. (2013), when both principles are taken into account in their group (Vincent averaged) RT data, the effect of weakening stimuli on benefit can be effectively predicted. Beyond this previous work, this thesis has demonstrated that this behavioural framework is also effective to understand benefits on the level of individual behaviour.

In addition to *understanding* benefits, these principles also suggest how to *manipulate* benefits, by calibrating unisensory performance. This will prove useful for designing stimuli for experimental purposes e.g. investigating the effects of signal strength. For example, it is difficult to say whether two signals in different modalities (a noise sound and coherent rotation of dots, for instance) are equally 'weak' according to inverse effectiveness, because signal strength for each stimulus is measured in different units. The strength of the noise, for instance, is measured in dB SPL, whereas the motion coherence is measured by the percentage of dots which rotate. In comparison, it is much more straightforward to say whether a participant's *performance* to the signals is similar (following equal effectiveness) and more variable (following the variability rule), as these require only a simple evaluation of RTs. Effectively therefore, maximally beneficial stimuli could be determined on the individual participant level by properly calibrating unisensory stimuli to elicit similar performance. In addition to experimental uses, such application of these principles potentially also has wider-ranging implications for designing maximally effective signalling procedures, such as those used by smartphones (**Figure 7.1**) or by multisensory warning and guidance systems (e.g. Biondi et al., 2017; Ho et al., 2007; Spence & Ho, 2008; van Erp et al., 2015).

## 7.2. Understanding Sources of Processing Interactions

In addition to the basic combination rule posed by the race model, there are at least two empirical processing interactions observed in RSP experiments: trial history effects and violations of Miller's bound. In all of the experiments presented here, these processing interactions were significantly present. Importantly, these processing interactions contribute to benefits beyond what the basic combination rule of the race model can account for, so a complete account of the RSE relies on understanding them. However, it is more difficult to generate a clear overall impression of the factors which give rise to interactions. This is because, unlike the race model framework which exists for benefits, there is no clear candidate framework for interactions. This may be because quantification and modelling of processing interactions is still at a relatively early stage, and further work may eventually provide a clear framework for interactions. At this early juncture, however, significant insights can still be made into common sources of each interaction. Here I review common sources of interactions based on the results of the present experiments.

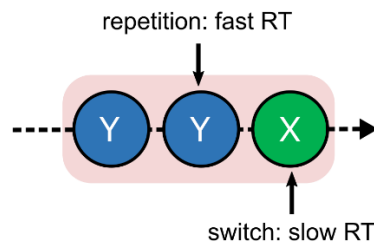
### 7.2.1. Higher-Level Processing is Important for Trial History Effects

The experimental work presented in this thesis adds to a small body of research which has quantified trial history effects in the RSP previously (e.g. Gondan et al., 2004; Harrar et al., 2014; Juan et al., 2017; Miller, 1982, 1986; Otto & D'Souza, 2015; Otto & Mamassian, 2012). As this body of evidence is still quite limited, much more work is needed before a comprehensive understanding will emerge on a level equivalent to our understanding of benefits. However, the experiments here suggest that certain interpretations can be ruled out. One hypothesis, following repetition priming experiments in visual search (Kristjansson & Campana, 2010), was that history effects arise from repetition of low-level features. Following **Chapter 3**, however, it is clear that history effects are not strongly rooted in *signal features*. Indeed, by introducing alternating signal features within a block, there was no reduction in the history effect compared to when signal features were consistent. This suggests that repetition of low-level signal features (and the underlying activity of the corresponding neurons) is not causally related to history effects. Further, across very different examples of *stimulus construction*, the history effect also remained relatively constant. Consider also in **Chapter 3** that non-random, sudden onset stimuli (simple signals) elicited a comparable history effect as randomised stimuli with background noise (complex signals). One possible exception was observed in **Chapter 5**, as history effects were larger when *signal strength* was weak compared to strong. However, the overall contribution of history effects to benefit was constant across signal strength conditions. This would suggest that history effects in this case were in proportion to unisensory RT variability generally. Broadly, therefore, studies attempting to target history effects in the RSP are unlikely to find causal mechanisms in low-level signal properties.

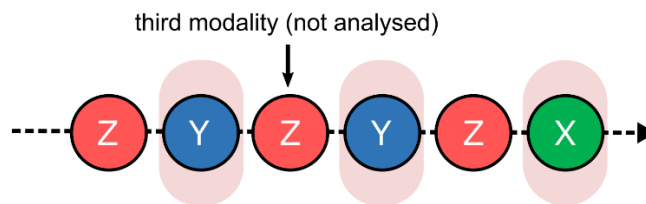
In light of the above, an alternative approach considered that history effects were linked to higher-level mechanisms more similar to attention (Spence et al., 2001) and task-switching (Kiesel et al., 2010; Monsell, 2003). These experiments show that higher-level explanations of history effects are more promising. For instance, in **Chapter 4**, it was expected that, if history effects arise from a similar mechanism to attention-switching between modalities, the history effect should be modulated by *signal duration*. Indeed, it was shown that for AV signal pairs, a larger history effect occurred for short signals than for long signals. Though this effect was not observed in a replication with TV signals, this null result may actually be linked to the low intensity of the short tactile signal in this particular version of the experiment. More reliably (across both AV and TV signal pairs), history effects were manipulated by *task-irrelevant stimulation*; history effects were reduced when an irrelevant third modality stimulus was present on all trials, compared to absent. This result also supports a higher-level explanation. Under an attentional framework, for example, this effect might be explained by the task-irrelevant stimulus causing the observer to divide attention between modalities, ‘resetting’ attentional biases that would otherwise be caused by unisensory stimuli. This would be in line with similar results observed in visual attentional with non-relevant auditory stimuli (Doyle & Snowden, 2001). Broadly, these factors are in line with the classic conceptualisation of the RSP as a form of divided attention task (e.g. Miller, 1982). As a template for further research, therefore, these higher-level manipulations are most likely to reveal the mechanisms of history effects.

It should be noted that the history effects observed in the present work were rather small in comparison to previous work. The overall mean history effects, ranging from 0.023 s in **Chapter 4 (Part 1)** to 0.040 s in **Chapter 4 (Part 2)**, are smaller than the 0.074 s history effect observed by Miller (1982), for instance. Similarly, the corresponding mean  $\rho$  parameters are only weakly negative, from -0.155 in **Chapter 4 (Part 1)** to -0.295 in **Chapter 4 (Part 2)**; this is much smaller than the  $\rho$  of -0.70 reported by Otto and Mamassian (2012) on their group data. This may be due to experimental differences (e.g. inclusion of catch trials in the present work). More direct manipulations to trial sequences, therefore, will undoubtedly be useful to understand these differences across studies. Alternatively, as these effects are smaller than previously reported, further work may try to remove history effects altogether (and thus the corresponding  $\rho$  parameter could be removed from the model). One manipulation which may achieve this would be to build upon the result of task-irrelevant stimulation observed here. The difference would be to employ the third modality stimulus as an ‘intermediate’ unisensory signal between all trials, which is not analysed (see **Figure 7.2**). This would help to remove history effects by effectively engaging the same ‘neutral’ modality before each experimental trial; thus, all signal trials analysed would be affected equally by the preceding trial. A similar task was originally trialled by Otto and D’Souza (2015), who did not find an influence of intermediate signals. Crucially, however, these

### Typical RSP



### Adding Intermediate Trials



**Figure 7.2 Adding third modality intermediate signals to remove history effects**

In the typical RSP history effects arise between unisensory trials as a consequence of the randomisation of trials. All signal trials are analysed, and each individual trial is affected differently by the previous trial. One possible method to remove these history effects would be to add an additional third modality signal as an intermediate between all trials. These intermediate trials would be discarded from analyses. Adding this intermediate third modality signal (Z) would mean that all unisensory trials analysed (X, Y) were preceded by the same signal, and thus the effect of the previous trial would be more similar for all RTs.

researchers used AV intermediate signals within an experiment with AV signal pairs. Given the promising role of third-modality stimuli here, it would be expected that such modifications would be more effective.

#### 7.2.2. Context is Crucial for Violations of Miller's Bound

Following the currently-dominant approach to the RSE, violations of Miller's bound have been interpreted as benefit which exceeds the limits of the race model architecture, and thus justifies an alternative pooling model approach. Importantly however, Miller's bound assumes context invariance (i.e. assumes that the processing of one unisensory decision-unit unaffected by the processing of the other). Therefore, pooling is just one interpretation of violations of Miller's bound. An equally valid alternative, following the previously-rejected race model approach, is that the violations of Miller's bound actually represent context variance (i.e. the processing of one unisensory decision-unit makes the accumulation of the other faster or more variable). The latter approach has not fully been explored given the previous rejection of race models.

In the experiments presented in this thesis, I have provided empirical support for the context variance interpretation. Under this interpretation, context variance (and thus violation of Miller's

bound) arises because of the difference between stimulation on redundant trials compared to unisensory trials. Thus, any factor which increases the similarity of stimulation across trials should reduce context variance, and correspondingly the violations observed. In **Chapter 3**, stimuli were made more similar by changing *stimulus construction*. By introducing continuous AV background stimulation (complex signals), the overall level of stimulation was more similar across trials than when signals were sudden onset (simple signals). Correspondingly, there was a reduction in violation for complex signals compared to simple signals. In **Chapter 4**, stimuli were made more similar across trials by introducing *task-irrelevant stimulation*. By having a third modality task-irrelevant stimulus present on all trials, the onset of task-relevant signals became more similar for redundant and unisensory trials than when the third modality stimulus was absent (i.e. the classic RSP). Correspondingly, there was a reduction in violation when the task-irrelevant stimulus was present compared to absent. This effect was consistent across AV and TV signal pairings, suggesting that such mechanisms apply generally across the different senses. In these experiments, there appears to be a role of strong signal onset transients in creating context variance, and thus violations of Miller's bound. As changes in violations generally followed changes in the additional noise parameter (*eta*), one explanation might be that sudden onsets in one modality disrupt accumulation in the other modality (i.e. make the accumulation process more variable). Overall, by following manipulations suggested by the context variance assumption, violations of Miller's bound were reliably manipulated according to predictions. This suggests a useful framework to understand violations of Miller's bound.

These results support a novel, parsimonious interpretation of the literature on violation. Previously, the race model has been implemented as a "nullmodel" (W. Schwarz, 1989), with the dominant pooling models employed in most cases following violations. However, this approach implicitly assumes that two different combination rules are actually used for the RSP (race, pooling) in different contexts. Further, it does not offer much in the way of a guiding framework for suggesting why a race (statistical facilitation) would be used in one context and pooling (integration) in another when the task demands of the RSP are always the same. Under the race model framework, however, only one combination rule is employed in all contexts. However, different forms of signals can elicit different degrees of violation by changing the context variance between unisensory and redundant trials. By understanding violations as evidence of the context variance processing interaction, and not evidence of a change in combination rule, it is much clearer to understand how such effects arise, and how they can be manipulated.

### 7.3. Application and Development of Models in the RSP

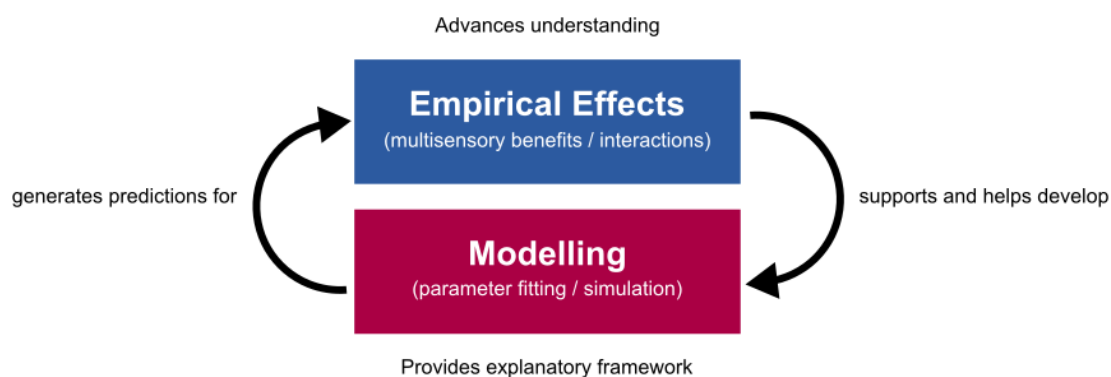
Despite the lack of effective models in multisensory RT research, there appears to be a resistance to acknowledging and applying context variance race models as a candidate to explain the RSE. One

reason for this, given the previous rejection of the basic race model, is that the entire model class is considered ineffective. It is worth, therefore, considering how candidate models are defined. McClelland (2009), for instance, states the following: “Models become candidates for further exploration (and criticism) when they account for a newsworthy amount of data (i.e., *more data than it was previously thought they could account for, or data that another approach cannot account for*)...” (p. 23, emphasis added). As noted in the introduction (**Section 1.3.4**), the pooling model has so far failed to demonstrate a consistent account of benefits and interactions across different studies with different signals. Race models however, at least in this thesis, were able to near-perfectly explain multisensory benefits and account for empirical interactions across a wide range of experimental conditions. Therefore, this thesis contributes to establishing race models (and the specific implementation of the context variant race model) as one of the most promising candidates for a model of the RSP.

Another concern, however, is that context variant race models are sometimes thought to be unfalsifiable (e.g. Miller, 2016). This is certainly true in the sense that Miller’s bound cannot be used as a valid rejection of the entire race model class (Otto & Mamassian, 2017; C. T. Yang et al., 2018). However, this does not mean that there is no way to determine which model class (race or pooling) is able to explain the RSE most effectively. As explored in **Chapter 6**, tools are available which allow researchers to determine which model is most effective at offering a parsimonious explanation of RTs. Race models, therefore, can be directly compared with pooling models in future work to evaluate their relative explanatory power. Further, specific model implementations can be falsified if they fail to capture empirical behavioural effects (Palminteri et al., 2017). By following a formal modelling approach in RSP studies, the success or failure of models in different conditions should ultimately guide research towards a dominant model class. In the experiments presented here, the context variant race model has provided a clear account of the behaviour observed. Regarding interactions, for instance, trial history effects were captured via changes in the *rho* parameter and violations of Miller’s bound were captured via changes in the *eta* parameter (see **Figure 6.4**). By holding race or pooling models to such standards of evaluation as in the comparative approach presented here, progress towards a clear model understanding should be facilitated.

It is important to note, in McClelland’s definition above, the crucial role of criticism in candidate model development. In this thesis, limitations to the context variant race model have been revealed, which can be addressed in further study. For instance, while changes in trial history effects broadly correspond to the *rho* parameter, this relationship is somewhat weak on the level of individual participants (and even on the level of the group average). This in part may represent a level of ‘uncertainty’ in the fitting of the *rho* parameter with the number of trials (~100) collected here (see





**Figure 7.3 Interaction between behavioural and formal modelling approaches**

Empirical effects observed in the RSP (multisensory benefits and interactions) can be explained by modelling approaches. In turn, models can generate novel predictions for future behavioural studies. Applying and refining a modelling approach based on behavioural results advances our understanding much more rapidly than studying behaviour in isolation.

the confidence intervals estimated in **Section 9.7**, for instance). In addition, a clear limitation of the model was shown when comparing interactions across different effectors (**Chapter 5**, see also **Section 9.7** of the appendices). This shows that by not including explicit post-decisional components in models of RT, the *eta* parameter can reveal effects which are (most likely) not genuine. These are two clear areas for model improvement beyond this thesis.

Overall, this thesis highlights that the application of formal models is always a fruitful exercise. Ultimately, models are judged by their ability to predict and explain data, and the success or failure of specific models will always inform future directions for multisensory research. To most effectively facilitate RSP research, therefore, the application of models (race or pooling) should rather become the rule. By applying models in a consistent comparative framework, it should be possible to achieve a “desired constant interaction” (Lewandowsky, 1993, p. 236), whereby clear model theory suggests predictions for experimental data, and in turn experimental data suggests methods for refining formal models (**Figure 7.3**).

#### 7.4. Experimental Applications of the Comparative Approach

The novel comparative approach developed here has provided an effective explanation of benefits and interactions across a range of different stimuli and contexts. For instance, while the majority of signals investigated were AV, the comparative approach generalised to TV (**Chapter 4, Part 2**) without issue. It should be highlighted, however, that the stimuli and contexts explored in this thesis are just a small selection of those tested in the broader literature (see **Section 1.3.1** in the introduction for an

overview). In this section, I hope to demonstrate that the comparative approach is sufficiently broad to be readily applied to a number of additional research contexts.

An immediate and simple extension to the present work would be to further explore alternative sensory pairings. Indeed, the RSP has also been observed with AT signals (e.g. Gondan et al., 2004; Marinovic et al., 2015; Nava et al., 2014), in addition to the AV and TV signal pairings evaluated here. A comparative analysis may help to highlight similarities and differences between sensory systems. In particular, the comparative approach may be especially beneficial in situations where the existing literature is less substantial. One example of this is visual-olfactory pairings (e.g. Amsellem et al., 2018; Hochenberger, Busch, & Ohla, 2015). In such investigations, researchers have found little evidence of interactions (particularly, a lack of violations), and thus have concluded that the basic race model alone is sufficient to explain the data. Applying the comparative race model approach here would likely help elaborate on *why* interactions are not observed, and suggest manipulations which might elicit them. For instance, in both visual-olfactory studies cited, RTs for vision were on average much faster than those for olfaction. According to the race model principles, therefore, these stimuli did not demonstrate equal effectiveness. As such, the signals simply may not have been maximised for proper investigation of benefits and interactions. This could be addressed in further research by calibrating signals for equivalent performance in RT. A comparative approach would then be more informative and reveal novel sources of benefit and interactions for sensory pairings beyond the typical A, V and T pairings.

#### 7.4.1. Beyond Bisensory Pairings

In addition to the substantial literature on bisensory pairings, the RSP has also been applied in experiments where different numbers of modalities are activated. However, the comparative approach (and the race model framework it builds on) can easily be extended to accommodate such designs. In the following section, I consider some examples of how this may be done.

##### 7.4.1.1. Unisensory Designs

Though the RSP is often conducted with two signals in different modalities (multisensory designs), there is also a substantial literature in which redundant signals are presented in the same modality (unisensory designs). A large body of research has demonstrated the RSE occurs when two signals are presented in vision (e.g. Corballis, 2002; Mishler & Neider, 2018; Mittelstadt & Miller, 2018; Moradi et al., 2016; Mordkoff & Danek, 2011; Ridgway et al., 2008; Ritchie et al., 2014; Savazzi & Marzi, 2008; Vrancken et al., 2018). There have also been reports of the RSE in audition (e.g. Schröter, Ulrich, & Miller, 2007). The race model easily accommodates unisensory designs, as it does not specify that the two decision-units must be in different modalities. As such, it is already common practice to evaluate the race model in unisensory designs in the same way as for multisensory research. Though it has not

been explicitly reported in any published study, it would therefore also be expected that the predictive race model principles explored here would extend to unisensory designs. If this is the case, the framework would also provide a consistent account of the unisensory RSE. Beyond accounting for benefits, it would also be informative to explore how interactions change when signals are presented within the same modality. As violations of Miller's bound have been observed with unisensory stimulation, for instance, it would be interesting to see if the manipulating factors identified here (i.e. signal onset transients, task-irrelevant stimulation) have similar effects.

#### *7.4.1.2. Trisensory Designs*

In addition to bisensory signals, the RSE has been observed with trisensory stimulation (since Todd, 1912), for example AVT stimulation (e.g. Couth et al., 2018; Haggmann & Russo, 2016; Pomper et al., 2014). Typically, a further speedup (i.e. beyond bisensory RSE) is observed in these cases (Diederich & Colonius, 2004). Again, however, the basic race model framework is easily extended by simply assuming that the minimum decision time is taken from three parallel decision-units rather than two. This would predict faster and less variable RTs than both unisensory and bisensory components. In addition, the simple race model is able to make multiple predictions for benefits from unisensory and bisensory RT distributions. For example, individual RT distributions for A, V and T can be combined for an prediction of AVT benefit, but equivalently AV and T distributions can also provide an prediction of AVT benefit. As recently demonstrated by Otto (2018), these trisensory predictions offer similar predictive power to those shown in present experimental chapters. Further, recent efforts have shown how the limits of the race model (Miller, 1982) can be extended to trisensory stimulation, which would allow for a similar quantification of interactions (Colonius et al., 2017). One possible limitation here is that there is no race or pooling model which can also account for trisensory interactions on the level of distributions. This would therefore be a clear area for model development. However, quantification of benefits and interactions (following Steps 1 and 2 of the comparative approach) may already substantially improve our understanding of this more complex multisensory situation.

#### *7.4.2. Between-Subjects Comparisons*

Rather than testing different signal properties in the same participant group, the comparative approach can be extended to test the effects of the same signals on different participant groups. This would give a clear idea of how multisensory processing changes across different individuals. Here, I highlight a few examples in the current literature which may benefit from a detailed analysis via the comparative approach.

##### *7.4.2.1. Lifespan and Ageing*

The RSP has often been used in comparisons of age groups to understand how redundant sensory processing changes across the lifespan (e.g. Couth et al., 2018; Downing et al., 2014; Mahoney &

Verghese, 2018; Murray, Eardley, et al., 2018; Peiffer et al., 2007; Ren et al., 2017; Wang et al., 2018; W. Yang & Ren, 2018). A common interpretation of these studies is that multisensory processing is in some way enhanced in older adults, as they typically demonstrate larger benefits (see also Linnet & Roser, 2012 for a unisensory example). It is not clear why multisensory processing should improve with age, given that cognitive functions typically tend to decline in older age. One explanation of this effect, however, is that it is evidence of a “compensatory strategy” (Laurienti, Burdette, Maldjian, & Wallace, 2006). The idea here is that multisensory processing counteracts degraded unisensory abilities (e.g. poorer visual acuity) which come with old age (Dumas, Holtzer, & Mahoney, 2016).

Interestingly, these results might be directly understood by the race model framework and by applying the comparative approach. This framework would state that, rather than counteracting degrading unisensory performance, it is precisely *because* unisensory performance degrades that benefits increase with age. If older participants’ unisensory performance is less reliable (i.e. responses are more variable), then according to the variability rule, larger multisensory benefits are clearly expected. It is difficult to properly assess race model principles based on previous results because intra-individual RT variability (SD or MAD, as assessed here) is rarely reported. However, there has been previous indication that most measures of RT variability increase with age (e.g. Hultsch, MacDonald, & Dixon, 2002). This may also help to disentangle conflicts within the existing literature. For instance, some research on ageing has noted a lack of benefits for participants in which one RT distribution is faster overall (Mahoney, Holtzer, & Verghese, 2014). This, as established, would be directly understood by the equal effectiveness principle. Further, in some studies violations are also observed to change with age (Couth et al., 2018). Quantifying violation on the level of individual subjects (and fitting the corresponding *eta*) may allow us to disentangle these conflicting findings.

#### 7.4.2.2. Clinical Investigations

The RSP is a simple, easily-administered task to assess multisensory processing. As such, it has a number of clinical applications, allowing researchers to investigate how multisensory processing is affected by clinical conditions. The RSP has already been applied in the study of schizophrenia (Williams et al., 2010; Wynn et al., 2014) and Parkinson’s disease (Plat et al., 2000; Ren et al., 2018). Interestingly, as an illustration of the wide-ranging applications of the task, it has also been employed in comparatively less common clinical conditions, including Prader-Willi syndrome (Salles et al., 2016), Niemann-Pick Type C Disease (Andrade et al., 2014), and Cerebellar Agenesis (Ronconi et al., 2017). While such studies alone are clearly valuable, they could also benefit from a comparative approach building on a clear common framework. This would provide insight into the sources of any processing differences between clinical groups.

One promising area of study, building on the race model principles, would be clinical conditions in which increased variability of RTs is established, such as Attention Deficit Hyperactivity Disorder (Castellanos & Tannock, 2002) and bipolar disorder (Adleman et al., 2014). Assuming this comes from variability in pre-decisional processes, the race model may uncover larger multisensory benefits in these populations. Further, as both disorders are associated with attentional differences, application of the comparative approach may reveal differences in processing interactions (and corresponding model parameters). Interestingly, the context variant race model has previously been applied in such clinical investigations to reveal differences in processing interactions. For instance Harrar et al. (2014) find evidence of weaker negative correlations for individuals with dyslexia, interpreted in relation to attentional shifting. A recent investigation by Crosse, Foxe, and Molholm (2019) also applied the context variant race model approach in an investigation of individuals with autism. More consistent application of these analytical steps, therefore, could provide a more solid platform for understanding individual differences.

#### 7.4.3. Involving Cognitive Neuroscience Techniques

One of the advantages of model-based approaches is that they provide a link between brain processes and behaviour (Mulder et al., 2014). Given that benefits and interactions can be successfully targeted with a comparative approach, therefore, an important further step will be to attempt to relate these interactions (and corresponding model parameters) back to brain processes. The RSP has already been applied in neuroimaging (e.g. Martuzzi et al., 2007) and neuroelectrophysiological (e.g. Molholm et al., 2002; Murray, Thelen, et al., 2018; Wang et al., 2018; Wynn et al., 2014) contexts. Such studies could benefit from a clear behavioural analysis offered by the comparative approach. Further, a formal modelling approach, which is rarely applied in RSP research (Gondan & Minakata, 2016), is also useful to link this behaviour to specific brain processes. One potentially fruitful line of enquiry, for example, could be to explore the link demonstrated here between increased context variance (demonstrated by violations of Miller's bound) and increased noise in accumulation (demonstrated by the *eta* parameter). A clear avenue here would be to use neurostimulation techniques, such as transcranial magnetic stimulation (e.g. Bolognini et al., 2009; Romei et al., 2007 for examples in the RSP). This technique has the capability to causally manipulate neural noise in targeted brain regions (Pascual-Leone, Walsh, & Rothwell, 2000; Walsh & Cowey, 2000). By explicitly fitting the context variant race model to different stimulation conditions, this could help provide further conceptual links between noise in decision-making models and noise in the brain.

## 7.5. Beyond the RSE: Implications for Multisensory Research

### 7.5.1. Is The Race Model Part of a Wider Set of Combination Rules?

The overarching message of this thesis is the race model offers an effective explanation for an example of RT behaviour which is part of everyday life (**Figure 1.1**). However, it should be considered what the results of this thesis tell us about behaviour for different tasks. As noted by Colonius and Diederich (2017), one of the restrictive factors of most multisensory models is that they only really explain behaviour in one paradigm, and may not be equipped to offer a general account of multisensory behaviour. Certainly, the race model alone does not provide a complete account of all multisensory behaviour. What I hope to have highlighted in this thesis, however, is that by applying similarly thorough approaches to different tasks, it will be possible to develop a clear understanding of different behaviours, and eventually identify commonalities between them.

Useful guidance for this is provided in the wider computational modelling literature. Marr (2010), for instance, classically distinguished between different *levels of understanding* for behaviours of interest. The first key level is to understand the demands of the behavioural task (i.e. the *computation* necessary to perform the behaviour). This then informs how we construct models (or *algorithms*) of the behaviour. These models then allow us to understand the link between behaviour and the underlying brain processes (or the *implementation*). Each stage is important, as developing a clear understanding of one stage informs the others (Krakauer, Ghazanfar, Gomez-Marin, Maclver, & Poeppel, 2017). Such stages can also be thought of as corresponding to the levels of investigation, with a modelling approach forming a link between behaviour and brain processes (e.g. Forstmann et al., 2011; Mulder et al., 2014). In this thesis, the race model was explored as the potential combination rule for redundant signal tasks. One of the main reasons this combination rule is appealing, and possibly why it has been observed to be successful here, is that it corresponds directly to the task demands (see **Section 1.3.1** in the General Introduction). A clear approach for further study, following this guidance, would be to first clearly identify combination rules which suit task demands.

The idea that combination rules may follow task demands was explored by Otto and Mamassian (2012). These authors examined participants' RTs to the same signals used in an RSP; however, in an additional experiment, they gave their participants a different task instruction, requiring a different combination rule. Specifically, participants had to respond only if *both* unisensory signals were present (i.e. only on multisensory trials). In contrast to the task demands of a classic RSP (**Figure 1.8b** in the introduction) which correspond to a logical disjunction (OR-gate), the alternative task-demands followed a logical *conjunction* (AND-gate). Similarly, instead of being determined by the *minimum* decision time (as in the RSP), RTs here would have been determined by the *maximum* unisensory decision time. The corresponding model, therefore, predicted *slower* RTs in multisensory

conditions. In much the same way that a basic race model accounts well for multisensory benefits, the corresponding combination rule for this alternative task effectively reproduced the empirical distribution of multisensory RTs. This suggests that combination rules may be flexibly implemented by the brain to suit different demands on decision-making. More so, these authors noted that the model was further improved by fitting the same interaction parameters (*rho* and *eta*) as used in context variant race model. If combination rules are employed to match the behavioural task, then this latter result may suggest that processing interactions (e.g. trial history effects, noise exchanged between decision-units) are more general effects across different multisensory tasks. Further research, therefore, may wish to evaluate the extent to which combination rules and processing interactions are distinct.

### 7.5.2. Linking Frameworks for Response Times and Accuracy

In the experiments presented here, my goal was to evaluate multisensory benefits in RTs. As noted in the introduction to the unisensory framework (**Section 1.2**), however, RT is known to trade-off with accuracy. Typically, as RTs are made faster, there is a corresponding decline in accuracy (e.g. Luce, 1986; Smith & Ratcliff, 2004). To study RTs in isolation, therefore, a common approach adopted in many studies (and in this thesis) is to ensure ceiling performance (i.e. close to 100% accuracy in all conditions; see **Section 2.2.1**). The rationale is that if accuracy is always at ceiling, it is not changing across conditions, and thus does not need to be considered in any detail. While this is a useful simplification for experimental and modelling purposes, it is unfortunately not able to provide a complete understanding of multisensory behaviour in everyday life. In many cases, individuals are required to detect unclear signals, in noisy environments, at a performance level far from ceiling. A complete understanding of multisensory behaviour (following the unisensory framework) should therefore account for accuracy as well as RT.

As noted in the introduction (**Section 1.1**), paradigms for accuracy already exist in multisensory research. There are also successful model accounts for how multiple estimates of stimulus properties from different modalities can be combined; for instance, the successful maximum likelihood estimation framework (Ernst & Banks, 2002) suggests that estimates are combined by weighting each estimate according to its reliability. It is possible that there are commonalities between such accuracy frameworks and RT frameworks (provided here by race models). Otto et al. (2013) have already noted similarities between race model principles (**Section 1.6.1**) and maximum likelihood estimation. Maximum likelihood estimation, for instance, also finds that benefits are related to the similarity of unisensory components (relating to equal effectiveness) and to the degree of uncertainty (relating to the variability rule). While these are promising commonalities, however, an explicit approach which links RT and accuracy frameworks for multisensory behaviour is still to be developed.

This is unsurprising because a common framework for multisensory RTs in isolation is not yet fully established. As the race model here provides a very successful candidate, however, further research towards a complete approach should now be explored. This should allow for similar advances in multisensory decision-making as seen in the established unisensory decision-making framework (e.g. Forstmann et al., 2016; Gold & Shadlen, 2007; Smith & Ratcliff, 2004).

### 7.5.3. Towards Conceptual Clarity in Multisensory Research

The ultimate goal of multisensory research will be to work towards a complete understanding of multisensory behaviour. One of the key steps towards this goal will be to establish clear definitions of empirical effects, which are consistent regardless of paradigm or level of investigation. One attempt to achieve this in recent years, which can be considered in light of present results, is in the definition of “multisensory integration”. Recognising such a need for clarity in research, a large cohort of authors (Stein et al., 2010) attempted to provide a broad definition which would apply to both neuronal responses (where the term originated) and behavioural effects. Specifically, the authors defined integration as “a response (neural or behavioural) that is significantly different from the responses evoked by the modality component stimuli” (p. 1719).

Such attempts at clear definitions are important as they improve communications between different fields and levels of investigation. The issue here is that even with such statements, definitions can remain so broad that many different behavioural measures qualify as evidence for the same general term. In the RSP, the multisensory benefit (i.e. RSE) immediately seems to fit such a definition of multisensory integration, as the redundant RT distribution is significantly faster than the component unisensory RT distributions. Alternatively, however, the violation area would also meet such a definition, as it represents an area of facilitation which is significantly larger than the prediction based only on unisensory responses. Indeed, following the dominant pooling approach, parallels have been drawn between the summation of evidence in the model and the integration of unisensory information by multisensory neurons (see also the dominant use of neuronal principles as a guiding framework, as discussed in **Section 7.1.2**). This is perhaps why violation is sometimes considered a “benchmark for integrative processing” (Martuzzi et al., 2007, p. 1674). It is not immediately clear, therefore, which of these behavioural measures (benefit or violation) best fits the definition put forward by Stein et al. (2010).

Unfortunately, despite attempts towards clear definitions, there remains a lack of clarity in the RSP. Implicitly or otherwise, there appears to be an assumption that benefits and violations are equivalent measures of integration. For instance, Stevenson et al. (2014) list both the RSE (i.e. benefit) and violations of Miller’s bound as RT measures of integration, amongst other measures in alternative areas of investigation (e.g. neuroimaging). To further highlight this, consider also a few examples of



definitions from papers in the 2018 literature review earlier (**Section 1.3.4.2**, see also **Section 9.2**). Amsellem et al. (2018) state that a violation of Miller's bound "provides evidence for information pooling across channels, or multimodal integration" (p. 330). W. Yang and Ren (2018), however, describe "audiovisual integration" as the effect by which "bimodal stimuli can be discriminated or detected more rapidly and accurately than unimodal visual or auditory stimuli presented alone" (p. 41). The former rather highlights the link between violation and integration, while the latter is more consistent with benefits as a measure of integration. Note however that under the suggested definition of multisensory integration, both interpretations appear to be valid.

The assumed equivalence between these measures, however, is in direct contrast to empirical observations in the experiments of this thesis. This is because benefits and violations often produced different main effects across experimental factors. In some cases, main effects are elicited in benefits where no changes occur in violations (**Chapter 5**). Conversely, main effects in violations were elicited where no corresponding changes in benefits occurred (**Chapter 4, Part 1**). Crucially, the main effects of benefits and violations were even observed to *dissociate* across factors (**Chapter 3**). In stark contrast to current assumptions of multisensory research, therefore, the choice between benefits and violations is not arbitrary. In fact, the pattern of behavioural results (and thus the interpretation for multisensory processing) may be very different depending on which is adopted as evidence of integration. Despite a movement towards clarity, therefore, this thesis has highlighted a conceptual confusion in the RSP literature at present. It is possible that such confusion, arising from assumed equivalence between measures, has contributed to the overall lack of development of a common approach to analysing the RSE. In contrast, under a race model framework, both benefits and violations have reason to be defined, quantified and examined in their own right. As I hope to have demonstrated, this offers a much-needed step towards conceptual clarity in multisensory research.

Having observed this conceptual confusion within a single paradigm, it is worrying to consider whether different measures of multisensory integration used *between* paradigms (such as those reviewed by Stevenson et al., 2014) show a similar lack of correspondence. Incorrect interpretations of results may easily propagate from one field of research to another if assumed equivalencies do not stand up to scrutiny. Ultimately, this will impede the ability to draw links between levels of understanding, rather than facilitate it. For multisensory research going forward, therefore, it will be important to work under clear definitions which apply to all levels of understanding, as crucially suggested by Stein et al. (2010). Further, proper scientific validation of relationships between variables should always be determined as a first step. By applying clear investigative frameworks, such as the novel comparative approach developed here, links between multiple levels will be validated more effectively.

## 8. References

- Adleman, N. E., Yi, J. Y., Deveney, C. M., Guyer, A. E., Leibenluft, E., & Brotman, M. A. (2014). Increased intrasubject variability in response time in unaffected preschoolers at familial risk for bipolar disorder. *Psychiatry Research*, *219*(3), 687-689. doi:10.1016/j.psychres.2014.06.047
- Akaike, H. (1974). A new look at statistical-model identification. *Ieee Transactions on Automatic Control*, *19*(6), 716-723. doi:10.1109/Tac.1974.1100705
- Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing and Perceiving*, *23*(1), 3-38. doi:10.1163/187847510X488603
- Amsellem, S., Hochenberger, R., & Ohla, K. (2018). Visual-Olfactory Interactions: Bimodal Facilitation and Impact on the Subjective Experience. *Chemical Senses*, *43*(5), 329-339. doi:10.1093/chemse/bjy018
- Andrade, G. N., Molholm, S., Butler, J. S., Brandwein, A. B., Walkley, S. U., & Foxe, J. J. (2014). Atypical multisensory integration in Niemann-Pick type C disease - towards potential biomarkers. *Orphanet Journal of Rare Diseases*, *9*, 149. doi:10.1186/s13023-014-0149-x
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154-179. doi:10.1037/0033-295X.93.2.154
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28. doi:10.21500/20112084.807
- Bailey, H. D., Mullaney, A. B., Gibney, K. D., & Kwakye, L. D. (2018). Audiovisual Integration Varies With Target and Environment Richness in Immersive Virtual Reality. *Multisensory Research*, *31*(7), 689-713. doi:10.1163/22134808-20181301
- Biondi, F., Strayer, D. L., Rossi, R., Gastaldi, M., & Mulatti, C. (2017). Advanced driver assistance systems: Using multimodal redundant warnings to enhance road safety. *Applied Ergonomics*, *58*, 238-244. doi:10.1016/j.apergo.2016.06.016
- Bolognini, N., Miniussi, C., Savazzi, S., Bricolo, E., & Maravita, A. (2009). TMS modulation of visual and auditory processing in the posterior parietal cortex. *Experimental Brain Research*, *195*(4), 509-517. doi:10.1007/s00221-009-1820-7
- Bolognini, N., Olgiati, E., Rossetti, A., & Maravita, A. (2010). Enhancing multisensory spatial orienting by brain polarization of the parietal cortex. *European Journal of Neuroscience*, *31*(10), 1800-1806. doi:10.1111/j.1460-9568.2010.07211.x
- Bompas, A., Hedge, C., & Sumner, P. (2017). Speeded saccadic and manual visuo-motor decisions: Distinct processes but same principles. *Cognitive Psychology*, *94*, 26-52. doi:10.1016/j.cogpsych.2017.02.002
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436. doi:10.1163/156856897x00357
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153-178. doi:10.1016/j.cogpsych.2007.12.002
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.). New York, NY: Springer.
- Carpenter, R. H., & Williams, M. L. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*, *377*(6544), 59-62. doi:10.1038/377059a0
- Castellanos, F. X., & Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nature Reviews Neuroscience*, *3*(8), 617-628. doi:10.1038/nrn896
- Cavina-Pratesi, C., Bricolo, E., Prior, M., & Marzi, C. A. (2001). Redundancy gain in the stop-signal paradigm: Implications for the locus of coactivation in simple reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 932-941. doi:10.1037//0096-1523.27.4.932

- Chandrasekaran, C., Lemus, L., Trubanova, A., Gondan, M., & Ghazanfar, A. A. (2011). Monkeys and Humans Share a Common Computation for Face/Voice Integration. *Plos Computational Biology*, 7(9). doi:10.1371/journal.pcbi.1002165
- Collignon, O., Girard, S., Gosselin, F., Saint-Amour, D., Lepore, F., & Lassonde, M. (2010). Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia*, 48(1), 220-225. doi:10.1016/j.neuropsychologia.2009.09.007
- Colonus, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology*, 34, 253-275. doi:10.1016/0022-2496(90)90032-5
- Colonus, H., & Diederich, A. (2006). The race model inequality: interpreting a geometric measure of the amount of violation. *Psychological Review*, 113(1), 148-154. doi:10.1037/0033-295X.113.1.148
- Colonus, H., & Diederich, A. (2017). Measuring multisensory integration: from reaction times to spike counts. *Scientific Reports*, 7(1), 3023. doi:10.1038/s41598-017-03219-5
- Colonus, H., Wolff, F. H., & Diederich, A. (2017). Trimodal Race Model Inequalities in Multisensory Integration: I. Basics. *Frontiers in Psychology*, 8, 1141. doi:10.3389/fpsyg.2017.01141
- Corballis, M. C. (2002). Hemispheric interactions in simple reaction time. *Neuropsychologia*, 40(4), 423-434. doi:10.1016/S0028-3932(01)00097-5
- Couth, S., Gowen, E., & Poliakoff, E. (2018). Using Race Model Violation to Explore Multisensory Responses in Older Adults: Enhanced Multisensory Integration or Slower Unisensory Processing? *Multisensory Research*, 31(3-4), 151-174. doi:10.1163/22134808-00002588
- Crosse, M. J., Foxe, J. J., & Molholm, S. (2019). Developmental recovery of impaired multisensory processing in autism and the cost of switching sensory modality (preprint available at bioRxiv). doi:10.1101/565333
- Diederich, A. (1995). Intersensory Facilitation of Reaction-Time - Evaluation of Counter and Diffusion Coactivation Models. *Journal of Mathematical Psychology*, 39(2), 197-215. doi:10.1006/jmps.1995.1020
- Diederich, A., & Colonius, H. (1987). Intersensory Facilitation in the Motor Component - a Reaction-Time Analysis. *Psychological Research*, 49(1), 23-29. doi:10.1007/Bf00309199
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, 66(8), 1388-1404. doi:10.3758/BF03195006
- Downing, H. C., Barutchu, A., & Crewther, S. G. (2014). Developmental trends in the facilitation of multisensory objects with distractors. *Frontiers in Psychology*, 5, 1559. doi:10.3389/fpsyg.2014.01559
- Doyle, M. C., & Snowden, R. J. (2001). Identification of visual stimuli is improved by accompanying auditory stimuli: the role of eye movements and sound location. *Perception*, 30(7), 795-810. doi:10.1068/p3126
- Dumas, K., Holtzer, R., & Mahoney, J. R. (2016). Visual-Somatosensory Integration in Older Adults: Links to Sensory Functioning. *Multisensory Research*, 29(4-5), 397-420.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433. doi:DOI 10.1038/415429a
- Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162-169. doi:10.1016/j.tics.2004.02.002
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, 17(5), 738-743. doi:10.1038/nn.3689
- Forster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research*, 143(4), 480-487. doi:10.1007/s00221-002-1017-9

- Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*, *67*, 641-666. doi:10.1146/annurev-psych-122414-033645
- Forstmann, B. U., Wagenmakers, E. J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract? *Trends in Cognitive Sciences*, *15*(6), 272-279. doi:10.1016/j.tics.2011.04.002
- Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332-343. doi:10.1007/s00221-002-1262-y
- Freeman, L. C. A., Wood, K. C., & Bizley, J. K. (2018). Multisensory stimuli improve relative localisation judgments compared to unisensory auditory or visual stimuli. *Journal of the Acoustical Society of America*, *143*(6), EL516. doi:10.1121/1.5042759
- Genest, W., Hammond, R., & Carpenter, R. H. S. (2016). The random dot tachistogram: a novel task that elucidates the functional architecture of decision. *Scientific Reports*, *6*, 30787. doi:10.1038/srep30787
- Girard, S., Pelland, M., Lepore, F., & Collignon, O. (2013). Impact of the spatial congruence of redundant targets on within-modal and cross-modal integration. *Experimental Brain Research*, *224*(2), 275-285. doi:10.1007/s00221-012-3308-0
- Giray, M., & Ulrich, R. (1993). Motor Coactivation Revealed by Response Force in Divided and Focused Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(6), 1278-1291. doi:10.1037/0096-1523.19.6.1278
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535-574. doi:10.1146/annurev.neuro.29.051605.113038
- Gondan, M., Lange, K., Rosler, F., & Roder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review*, *11*(2), 307-313. doi:10.3758/Bf03196575
- Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention Perception & Psychophysics*, *78*(3), 723-735. doi:10.3758/s13414-015-1018-y
- Gondan, M., Vorberg, D., & Greenlee, M. W. (2007). Modality shift effects mimic multisensory interactions: an event-related potential study. *Experimental Brain Research*, *182*, 199-214. doi:10.1007/s00221-007-0982-4
- 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*(6), 565-570. doi:10.3758/Bf03207517
- Hagmann, C. E., & Russo, N. (2016). Multisensory integration of redundant trisensory stimulation. *Attention, Perception, & Psychophysics*, *78*(8), 2558-2568. doi:10.3758/s13414-016-1192-6
- Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement. *Science*, *274*, 427-430. doi:10.1126/science.274.5286.427
- Harrar, V., Harris, L. R., & Spence, C. (2017). Multisensory integration is independent of perceived simultaneity. *Experimental Brain Research*, *235*(3), 763-775. doi:10.1007/s00221-016-4822-2
- Harrar, V., Tammam, J., Perez-Bellido, A., Pitt, A., Stein, J., & Spence, C. (2014). Multisensory integration and attention in developmental dyslexia. *Current Biology*, *24*(5), 531-535. doi:10.1016/j.cub.2014.01.029
- Hélie, S. (2006). An Introduction to Model Selection: Tools and Algorithms. *Tutorials in Quantitative Methods for Psychology*, *2*(1), 1-10. doi:10.20982/tqmp.02.1.p001
- Hershenson, M. (1962). Reaction-Time as a Measure of Intersensory Facilitation. *Journal of Experimental Psychology*, *63*(3), 289-&. doi:10.1037/h0039516
- Ho, C., Reed, N., & Spence, C. (2007). Multisensory in-car warning signals for collision avoidance. *Human Factors*, *49*(6), 1107-1114. doi:10.1518/001872007x249965
- Hochenberger, R., Busch, N. A., & Ohla, K. (2015). Nonlinear response speedup in bimodal visual-olfactory object identification. *Frontiers in Psychology*, *6*, 1477. doi:10.3389/fpsyg.2015.01477

- Houghton, G., & Grange, J. A. (2011). CDF-XL: computing cumulative distribution functions of reaction time data in Excel. *Behavior Research Methods*, *43*(4), 1023-1032. doi:10.3758/s13428-011-0119-3
- Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *57*(2), P101-P115. doi:10.1093/geronb/57.2.P101
- Innes, B. R., & Otto, T. U. (2019). A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. *Scientific Reports*, *9*(1), 2921. doi:10.1038/s41598-019-39924-6
- Jaekl, P., Perez-Bellido, A., & Soto-Faraco, S. (2014). On the 'visual' in 'audio-visual integration': a hypothesis concerning visual pathways. *Experimental Brain Research*, *232*(6), 1631-1638. doi:10.1007/s00221-014-3927-8
- Juan, C., Cappe, C., Alric, B., Roby, B., Gilardeau, S., Barone, P., & Girard, P. (2017). The variability of multisensory processes of natural stimuli in human and non-human primates in a detection task. *Plos One*, *12*(2), e0172480. doi:10.1371/journal.pone.0172480
- Keyes, H., Whitmore, A., Naneva, S., & McDermott, D. (2018). The priming function of in-car audio instruction. *Quarterly Journal of Experimental Psychology*, *1747021818773293*. doi:10.1177/1747021818773293
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching--a review. *Psychological Bulletin*, *136*(5), 849-874. doi:10.1037/a0019842
- Kinchla, R. A. (1974). Detecting Target Elements in Multielement Arrays - Confusability Model. *Perception & Psychophysics*, *15*(1), 149-158. doi:10.3758/Bf03205843
- Klein, M. D., & Stolz, J. A. (2015). Looking and listening: A comparison of intertrial repetition effects in visual and auditory search tasks. *Attention Perception & Psychophysics*, *77*(6), 1986-1997. doi:10.3758/s13414-015-0908-3
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 14-14.
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., Maclver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480-490. doi:10.1016/j.neuron.2016.12.041
- Kristjansson, A., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention Perception & Psychophysics*, *72*(1), 5-18. doi:10.3758/App.72.1.5
- Kroese, D. P., Brereton, T., Taimre, T., & Botev, Z. I. (2014). Why the Monte Carlo method is so important today. *WIREs Computational Statistics*, *6*(6), 386-392. doi:10.1002/wics.1314
- Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology*, *4*(1), 35-45. doi:10.20982/tqmp.04.1.p035
- Lanz, F., Moret, V., Rouiller, E. M., & Loquet, G. (2013). Multisensory Integration in Non-Human Primates during a Sensory-Motor Task. *Frontiers in Human Neuroscience*, *7*, 799. doi:10.3389/fnhum.2013.00799
- Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging*, *27*(8), 1155-1163. doi:10.1016/j.neurobiolaging.2005.05.024
- Lewandowsky, S. (1993). The Rewards and Hazards of Computer-Simulations. *Psychological Science*, *4*(4), 236-243. doi:10.1111/j.1467-9280.1993.tb00267.x
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764-766. doi:10.1016/j.jesp.2013.03.013
- Li, X. R., Liang, Z., Kleiner, M., & Lu, Z. L. (2010). RTbox: A device for highly accurate response time measurements. *Behavior Research Methods*, *42*(1), 212-225. doi:10.3758/Brm.42.1.212

- Linnet, E., & Roser, M. E. (2012). Age-related differences in interhemispheric visuomotor integration measured by the redundant target effect. *Psychology and Aging, 27*(2), 399-409. doi:10.1037/a0024905
- Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research, 17*(2), 447-453. doi:10.1016/S0926-6410(03)00160-5
- Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York, NY: Oxford University Press
- Mahoney, J. R., Holtzer, R., & Verghese, J. (2014). Visual-somatosensory integration and balance: evidence for psychophysical integrative differences in aging. *Multisensory Research, 27*(1), 17-42.
- Mahoney, J. R., & Verghese, J. (2018). Visual-Somatosensory Integration and Quantitative Gait Performance in Aging. *Frontiers in Aging Neuroscience, 10*, 377. doi:10.3389/fnagi.2018.00377
- Manly, B. F. J. (2007). *Randomization, Bootstrap and Monte Carlo Methods in Biology* (3rd Ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Marinovic, W., Milford, M., Carroll, T., & Riek, S. (2015). The facilitation of motor actions by acoustic and electric stimulation. *Psychophysiology, 52*(12), 1698-1710. doi:10.1111/psyp.12540
- Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: MIT Press. (Original work published 1982).
- Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J. P., Maeder, P. P., Clarke, S., & Meuli, R. A. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral Cortex, 17*(7), 1672-1679. doi:10.1093/cercor/bhl077
- McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science, 1*(1), 11-38. doi:10.1111/j.1756-8765.2008.01003.x
- Meijers, L. M. M., & Eijkman, E. G. J. (1977). Distributions of Simple RT with Single and Double Stimuli. *Perception & Psychophysics, 22*(1), 41-48. doi:10.3758/Bf03206078
- Miller, J. (1982). Divided Attention - Evidence for Co-Activation with Redundant Signals. *Cognitive Psychology, 14*(2), 247-279. doi:10.1016/0010-0285(82)90010-X
- Miller, J. (1986). Timecourse of Coactivation in Bimodal Divided Attention. *Perception & Psychophysics, 40*(5), 331-343. doi:10.3758/Bf03203025
- Miller, J. (2016). Statistical facilitation and the redundant signals effect: What are race and coactivation models? *Attention, Perception, & Psychophysics, 78*(2), 516-519. doi:10.3758/s13414-015-1017-z
- Miller, J., & Lopes, A. (1991). Bias Produced by Fast Guessing in Distribution-Based Tests of Race Models. *Perception & Psychophysics, 50*(6), 584-590. doi:10.3758/Bf03207544
- Minakata, K., & Gondan, M. (2018). Differential coactivation in a redundant signals task with weak and strong go/no-go stimuli. *Quarterly Journal of Experimental Psychology, 174*7021818772033. doi:10.1177/1747021818772033
- Mishler, A. D., & Neider, M. B. (2018). Redundancy gain for categorical targets depends on display configuration and duration. *Visual Cognition, 26*(6), 393-404. doi:10.1080/13506285.2018.1470587
- Mittelstadt, V., & Miller, J. (2018). Redundancy gain in the Simon Task: Does increasing relevant activation reduce the effect of irrelevant activation? *Journal of Experimental Psychology: Human Perception and Performance, 44*(8), 1153-1167. doi:10.1037/xhp0000523
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research, 14*(1), 115-128. doi:10.1016/S0926-6410(02)00066-6
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140. doi:10.1016/s1364-6613(03)00028-7

- Moradi, Z., Yankouskaya, A., Duta, M., Hewstone, M., & Humphreys, G. W. (2016). Coactive processing of sensory signals for in-group but not out-group stimuli. *Visual Cognition*, 23(9-10), 1124-1149. doi:10.1080/13506285.2016.1147512
- 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, & Psychophysics*, 73(1), 103-112. doi:10.3758/s13414-010-0025-2
- Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, 17(2), 520-538. doi:10.1037/0096-1523.17.2.520
- Mulder, M. J., van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences - a model-based review. *Neuroscience*, 277, 872-884. doi:10.1016/j.neuroscience.2014.07.031
- Murray, M. M., Eardley, A. F., Edgington, T., Oyekan, R., Smyth, E., & Matusz, P. J. (2018). Sensory dominance and multisensory integration as screening tools in aging. *Scientific Reports*, 8(1), 8901. doi:10.1038/s41598-018-27288-2
- Murray, M. M., Thelen, A., Ionta, S., & Wallace, M. T. (2018). Contributions of Intraindividual and Interindividual Differences to Multisensory Processes. *Journal of Cognitive Neuroscience*, 31(3), 360-376. doi:10.1162/jocn\_a\_01246
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1), 90-100. doi:10.1016/S0022-2496(02)00028-7
- Nadarajah, S., & Kotz, S. (2008). Exact distribution of the max/min of two Gaussian random variables. *Ieee Transactions on Very Large Scale Integration (Vlsi) Systems*, 16(2), 210-212. doi:10.1109/Tvlsi.2007.912191
- Nava, E., Bottari, D., Villwock, A., Fengler, I., Buchner, A., Lenarz, T., & Roder, B. (2014). Audio-tactile integration in congenitally and late deaf cochlear implant users. *Plos One*, 9(6), e99606. doi:10.1371/journal.pone.0099606
- Nelder, J. A., & Mead, R. (1965). A Simplex-Method for Function Minimization. *Computer Journal*, 7(4), 308-313. doi:10.1093/comjnl/7.4.308
- Noel, J. P., Modi, K., Wallace, M. T., & Van der Stoep, N. (2018). Audiovisual integration in depth: multisensory binding and gain as a function of distance. *Experimental Brain Research*, 236(7), 1939-1951. doi:10.1007/s00221-018-5274-7
- Noorani, I., & Carpenter, R. H. S. (2011). Full reaction time distributions reveal the complexity of neural decision-making. *European Journal of Neuroscience*, 33(11), 1948-1951. doi:10.1111/j.1460-9568.2011.07727.x
- Noorani, I., & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience and Biobehavioral Reviews*, 64, 229-251. doi:10.1016/j.neubiorev.2016.02.018
- Oswal, A., Ogden, M., & Carpenter, R. H. S. (2007). The time course of stimulus expectation in a saccadic decision task. *Journal of Neurophysiology*, 97(4), 2722-2730. doi:10.1152/jn.01238.2006
- Otto, T. U. (2018). *Trimodal is faster than bimodal and race models can explain it all*. Presented at New Ideas in Hearing & Seeing: Cross-modal Processing (from Physiology to Behaviour). Paris, France.
- Otto, T. U. (2019). *RSE-box: An analysis and modelling package to study response times*. Manuscript submitted for publication.
- Otto, T. U., & D'Souza, R. (2015). *Tackling history effects in multisensory decisions*. Presented at the 16th Annual Meeting of the International Multisensory Research Forum (IMRF). Pisa, Italy
- Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of Multisensory Behavior. *Journal of Neuroscience*, 33(17), 7463-7474. doi:10.1523/Jneurosci.4678-12.2013
- Otto, T. U., & Mamassian, P. (2012). Noise and Correlations in Parallel Perceptual Decision Making. *Current Biology*, 22(15), 1391-1396. doi:10.1016/j.cub.2012.05.031
- Otto, T. U., & Mamassian, P. (2017). Multisensory Decisions: the Test of a Race Model, Its Logic, and Power. *Multisensory Research*, 30(1), 1-24. doi:10.1163/22134808-00002541

- Palminteri, S., Wyart, V., & Kochlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, 21(6), 425-433. doi:10.1016/j.tics.2017.03.011
- Pascual-Leone, A., Walsh, V., & Rothwell, J. (2000). Transcranial magnetic stimulation in cognitive neuroscience - virtual lesion, chronometry, and functional connectivity. *Current Opinion in Neurobiology*, 10(2), 232-237.
- Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E., & Laurienti, P. J. (2007). Age-related multisensory enhancement in a simple audiovisual detection task. *Neuroreport*, 18(10), 1077-1081. doi:10.1097/WNR.0b013e3281e72ae7
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437-442. doi:10.1163/156856897x00366
- Pfister, M., Lue, J. C., Stefanini, F. R., Falabella, P., Dustin, L., Koss, M. J., & Humayun, M. S. (2014). Comparison of reaction response time between hand and foot controlled devices in simulated microsurgical testing. *Biomed Research International*, 2014(1), 769296. doi:10.1155/2014/769296
- Plat, F. M., Praamstra, P., & Horstink, M. W. I. M. (2000). Redundant-signals effects on reaction time, response force, and movement-related potentials in Parkinson's disease. *Experimental Brain Research*, 130(4), 533-539. doi:10.1007/s002219900276
- Pomper, U., Brincker, J., Harwood, J., Prikhodko, I., & Senkowski, D. (2014). Taking a call is facilitated by the multisensory processing of smartphone vibrations, sounds, and flashes. *Plos One*, 9(8), e103238. doi:10.1371/journal.pone.0103238
- Posner, M. I. (2005). Timing the brain: mental chronometry as a tool in neuroscience. *Plos Biology*, 3(2), e51. doi:10.1371/journal.pbio.0030051
- Raab, D. H. (1962). Statistical Facilitation of Simple Reaction Times. *Transactions of the New York Academy of Sciences*, 24(5), 574-590. doi:10.1111/j.2164-0947.1962.tb01433.x
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59-108. doi:10.1037/0033-295X.85.2.59
- Ratcliff, R. (1979). Group Reaction-Time Distributions and an Analysis of Distribution Statistics. *Psychological Bulletin*, 86(3), 446-461. doi:10.1037//0033-2909.86.3.446
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873-922. doi:10.1162/neco.2008.12-06-420
- Reinmueller, K., Koehler, L., & Steinhauser, M. (2018). Adaptive warning signals adjusted to driver passenger conversation: Impact of system awareness on behavioral adaptations. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58, 242-252. doi:10.1016/j.trf.2018.06.013
- Ren, Y., Suzuki, K., Yang, W., Ren, Y., Wu, F., Yang, J., . . . Hirata, K. (2018). Absent Audiovisual Integration Elicited by Peripheral Stimuli in Parkinson's Disease. *Parkinson's Disease*, 2018(1), 1648017. doi:10.1155/2018/1648017
- Ren, Y., Yang, W., Nakahashi, K., Takahashi, S., & Wu, J. (2017). Audiovisual Integration Delayed by Stimulus Onset Asynchrony Between Auditory and Visual Stimuli in Older Adults. *Perception*, 46(2), 205-218. doi:10.1177/0301006616673850
- Ridgway, N., Milders, M., & Sahraie, A. (2008). Redundant target effect and the processing of colour and luminance. *Experimental Brain Research*, 187(1), 153-160. doi:10.1007/s00221-008-1293-0
- Ritchie, K. L., Bannerman, R. L., & Sahraie, A. (2014). Redundancy gain in binocular rivalry. *Perception*, 43(12), 1316-1328. doi:10.1068/p7808
- Romei, V., Murray, M. M., Merabet, L. B., & Thut, G. (2007). Occipital Transcranial Magnetic Stimulation has opposing effects on visual and auditory stimulus detection: Implications for multisensory interactions. *Journal of Neuroscience*, 27(43), 11465-11472. doi:10.1523/JNEUROSCI.2827-07.2007



- Ronconi, L., Casartelli, L., Carna, S., Molteni, M., Arrigoni, F., & Borgatti, R. (2017). When one is Enough: Impaired Multisensory Integration in Cerebellar Agenesis. *Cerebral Cortex*, *27*(3), 2041-2051. doi:10.1093/cercor/bhw049
- Saenz, M., & Langers, D. R. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research*, *307*, 42-52. doi:10.1016/j.heares.2013.07.016
- Salles, J., Strelnikov, K., Carine, M., Denise, T., Laurier, V., Molinas, C., . . . Barone, P. (2016). Deficits in voice and multisensory processing in patients with Prader-Willi syndrome. *Neuropsychologia*, *85*, 137-147. doi:10.1016/j.neuropsychologia.2016.03.015
- Savazzi, S., & Marzi, C. A. (2008). Does the redundant signal effect occur at an early visual stage? *Experimental Brain Research*, *184*(2), 275-281. doi:10.1007/s00221-007-1182-y
- Schröter, H., Ulrich, R., & Miller, J. (2007). Effects of redundant auditory stimuli on reaction time. *Psychonomic Bulletin & Review*, *14*(1), 39-44. doi:10.3758/BF03194025
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461-464. doi:10.1214/aos/1176344136
- Schwarz, W. (1989). A new model to explain the redundant-signas effect. *Perception & Psychophysics*, *46*(5), 498-500. doi:10.3758/BF03210867
- Schwarz, W. (1994). Diffusion, Superposition, and the Redundant-Targets Effect. *Journal of Mathematical Psychology*, *38*(4), 504-520. doi:10.1006/jmps.1994.1036
- Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods Instruments & Computers*, *33*(4), 457-469. doi:10.3758/Bf03195403
- Senkowski, D., Saint-Amour, D., Hofle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage*, *56*(4), 2200-2208. doi:10.1016/j.neuroimage.2011.03.075
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*(4), 1916-1936. doi:10.1152/jn.2001.86.4.1916
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161-168. doi:10.1016/j.tins.2004.01.006
- Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception & Psychophysics*, *65*(2), 318-328. doi:10.3758/Bf03194803
- Spence, C., & Ho, C. (2008). Multisensory interface design for drivers: past, present and future. *Ergonomics*, *51*(1), 65-70. doi:10.1080/00140130701802759
- Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*(2), 330-336. doi:10.3758/Bf03194473
- Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Alex Meredith, M., Perrault, T. J., Jr., . . . Lewkowicz, D. J. (2010). Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience*, *31*(10), 1713-1720. doi:10.1111/j.1460-9568.2010.07206.x
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255-266. doi:10.1038/nrn2331
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, *15*(8), 520-535. doi:10.1038/nrn3742
- Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., . . . Wallace, M. T. (2014). Identifying and quantifying multisensory integration: a tutorial review. *Brain Topography*, *27*(6), 707-730. doi:10.1007/s10548-014-0365-7
- Sürig, R., Bottari, D., & Röder, B. (2018). Transfer of Audio-Visual Temporal Training to Temporal and Spatial Audio-Visual Tasks. *Multisensory Research*, *31*(6), 556-578. doi:10.1163/22134808-00002611
- Todd, J. W. (1912). Reaction to multiple stimuli. *Archives of Psychology*(3), 1-65.




- 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*(4), 659-681. doi:10.3758/s13423-011-0106-9
- Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient Visual Search from Synchronized Auditory Signals Requires Transient Audiovisual Events. *Plos One*, *5*(5). doi:10.1371/journal.pone.0010664
- van Erp, J. B. F., Toet, A., & Janssen, J. B. (2015). Uni-, bi- and tri-modal warning signals: Effects of temporal parameters and sensory modality on perceived urgency. *Safety Science*, *72*, 1-8. doi:10.1016/j.ssci.2014.07.022
- Van Zandt, T., & Townsend, J. T. (2013). Designs for and Analyses of Response Time Experiments. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology, Vol. 1*. New York, NY: Oxford University Press.
- Veldhuizen, M. G., Shepard, T. G., Wang, M. F., & Marks, L. E. (2010). Coactivation of Gustatory and Olfactory Signals in Flavor Perception. *Chemical Senses*, *35*(2), 121-133. doi:10.1093/chemse/bjp089
- Vrancken, L., Vermeulen, E., Germeys, F., & Verfaillie, K. (2018). Measuring facial identity and emotion integration using the redundancy gain paradigm. *Attention, Perception, & Psychophysics*. doi:10.3758/s13414-018-1603-y
- Wagenmakers, E. J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830-841. doi:10.1037/0033-295X.114.3.830
- Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192-196. doi:10.3758/Bf03206482
- Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience*, *1*(1), 73-79. doi:10.1038/35036239
- Wang, B., Li, P., Li, D., Niu, Y., Yan, T., Li, T., . . . Xiang, J. (2018). Increased Functional Brain Network Efficiency During Audiovisual Temporal Asynchrony Integration Task in Aging. *Frontiers in Aging Neuroscience*, *10*, 316. doi:10.3389/fnagi.2018.00316
- Werner, S., & Noppeney, U. (2011). The contributions of transient and sustained response codes to audiovisual integration. *Cerebral Cortex*, *21*(4), 920-931. doi:10.1093/cercor/bhq161
- Whelan, R. (2008). Effective analysis of reaction time data. *Psychological Record*, *58*(3), 475-482. doi:10.1007/Bf03395630
- Wiggett, A. J., & Tipper, S. P. (2015). Priming of hand and foot response: is spatial attention to the body site enough? *Psychonomic Bulletin & Review*, *22*(6), 1678-1684. doi:10.3758/s13423-015-0823-6
- Williams, L. E., Light, G. A., Braff, D. L., & Ramachandran, V. S. (2010). Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia*, *48*(10), 3128-3136. doi:10.1016/j.neuropsychologia.2010.06.028
- Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research*, *50*(14), 1304-1311. doi:10.1016/j.visres.2009.11.002
- Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*, 131. doi:10.3389/fnhum.2015.00131
- Wynn, J. K., Jahshan, C., & Green, M. F. (2014). Multisensory integration in schizophrenia: a behavioural and event-related potential study. *Cognitive Neuropsychiatry*, *19*(4), 319-336. doi:10.1080/13546805.2013.866892
- Yang, C. T., Altieri, N., & Little, D. R. (2018). An examination of parallel versus coactive processing accounts of redundant-target audiovisual signal processing. *Journal of Mathematical Psychology*, *82*, 138-158. doi:10.1016/j.jmp.2017.09.003
- Yang, W., & Ren, Y. (2018). Attenuated audiovisual integration in middle-aged adults in a discrimination task. *Cognitive Processing*, *19*(1), 41-45. doi:10.1007/s10339-017-0838-1

- Yeung, N., Nystrom, L. E., Aronson, J. A., & Cohen, J. D. (2006). Between-task competition and cognitive control in task switching. *Journal of Neuroscience*, 26(5), 1429-1438. doi:10.1523/JNEUROSCI.3109-05.2006
- Zehetleitner, M., Ratko-Dehnert, E., & Muller, H. J. (2015). Modeling violations of the race model inequality in bimodal paradigms: co-activation from decision and non-decision components. *Frontiers in Human Neuroscience*, 9, 119. doi:10.3389/fnhum.2015.00119

## 9. Appendices

### 9.1. Illustrating the Basic Race Model Mechanism

As the race model is key to understanding much of the work done both in previous decades and the current thesis, it is worth highlighting how the basic mechanism works. A useful metaphor is to consider running times (e.g. Miller, 2016). Assume, as in **Figure 9.1**, that we have two runners, which we label X and Y. For each runner, we want to know how quickly they can run a particular circuit, and so measure them separately over 5 laps. We might find that both perform about equally well, with a similar mean lap time. Now, suppose that we want to evaluate, on laps 1 to 5, what the winning time would have been had the two runners been racing against each other. To do this, we would take the minimum time of both X and Y for each lap. In doing so, it would be found that these ‘winning’ times are faster and more reliable than either individual runner. By being able to take the smaller of two outcomes, therefore, a purely statistical outcome is that the resulting minimum array has a smaller average and is less variable overall. This is the core of the race model explanation of the RSE; by having two decision-units working in parallel and always letting the faster (minimum) decision time trigger the response, RTs are faster and less variable across multisensory trials.

		Run time (s)					
		Lap 1	Lap 2	Lap 3	Lap 4	Lap 5	Mean (SD)
X		34	48	39	51	42	42.8 (6.8)
Y		38	45	41	33	59	43.2 (9.9)
min (		34	45	39	33	42	38.6 (5.1)

**Figure 9.1 Statistical facilitation with runners’ lap times**

The table shows hypothetical lap times for different runners (X and Y). Over five laps, these two runners produce similar average times. By taking the minimum time on each lap, however, the resulting ‘winning’ times are on average faster and less variable than either individual runner. This effect is known as statistical facilitation; by taking the smaller of two random numbers, the resulting distribution will be smaller and less variable than either component. The runner icons used are modified from a public domain vector image, which is free to use under the Creative Commons licence Public Domain Dedication (CC0 1.0). The images were downloaded from [www.svgsilh.com](http://www.svgsilh.com).

## 9.2. Literature Review of RSE Studies

To identify candidate articles, I followed the same search procedure as Gondan & Minakata (2016), i.e. papers citing Miller (1982) from Google Scholar and Web of Knowledge, which were published by a journal in the year 2018. This search returned 55 papers, to which following inclusion criterion were applied:

- The paper must be empirical (i.e. reviews and theoretical papers excluded)
- The paper must include analysis of human RTs
- The paper must evaluate the race model in some way (e.g. compute Miller's bound)

This left 22 papers which were analysed for their use of a formal RT model.

First Author (Journal)	Race Model Evaluation Used	Formal RT Model?
Amsellum, S.	Miller's bound / Capacity	No
Bailey, H. D.	Miller's bound	No
Cheng, X. J.	Capacity	Yes
Couth, S.	Miller's bound	No
Farid, B.	Independent race	No
Fitousi, D.	Capacity	No
Freeman, L. A. C.	Unclear (No equation / description)	No
Mahoney, J. R.	Miller's bound	No
Minakata, K.	Miller's bound	No
Mishler, A. D.	Miller's bound	No
Mittelstadt, V.	Miller's bound	No
Morey, S. A. (CR:PI)	Capacity	No
Morey, S. A. (PotHFaRSAM)	Capacity	No
Murray, M. M.	Independent race	No
Noel, J. -P.	Miller's Bound	No
Ren, Y.	Independent race	No
Surig, R.	Miller's bound	No
Vrancken, L.	Miller's bound	No
Wang, B.	Independent race	No
Yamani, Y.	Capacity	No
Yang, C. -T.	Capacity	Yes
Yang, W.	Independent race	No

### 9.3. Simulation Parameters for Validation of Down-Sampled Benefit Procedure

Note that the unit for each parameter is  $s^{-1}$  (except for correlation,  $\rho$ )

Parameters				
$\mu_X$	$\mu_Y$	$\sigma_X$	$\sigma_Y$	$\rho$
$N(2.5, 0.25)$	$N(2.5, 0.25)$	$N(0.4, 0.04)$	$N(0.4, 0.04)$	$U(-1, 1)$

### 9.4. Equations for the Context Variant Race Model

The context variant race equation is described by Otto and Mamassian (2012; Supplemental Information). Consider two unisensory signals (X and Y) and the redundant signal (XY). Drift rates elicited by the unisensory signals are modelled as two normal distributions,  $D_X$  and  $D_Y$ . Each are defined by their corresponding  $\mu$  ( $\mu$ ) and  $\sigma$  ( $\sigma$ ) parameters, such that

$$D_X \sim N(\mu_X, \sigma_X^2), \quad (\text{A1})$$

and

$$D_Y \sim N(\mu_Y, \sigma_Y^2). \quad (\text{A2})$$

The next step is to calculate the redundant drift rate distribution. According to the race model, the redundant decision time is determined by the minimum of the two unisensory decision time distributions (see **Section 1.3.3.3**). As smaller drift rates correspond to larger decision times, this corresponds to the *maximum* of the two drift rate distributions. The maximum function of two Gaussian distributions, including a correlation coefficient, is provided by Nadarajah and Kotz (2008). First, a bivariate Gaussian random vector ( $D_X, D_Y$ ) is defined using the  $\mu$  ( $\mu_X, \mu_Y$ ) and  $\sigma$  ( $\sigma_X, \sigma_Y$ ) parameters of the unisensory drift rate distribution, including a correlation ( $\rho$ ) parameter. The probability density function (PDF) of the maximum distribution  $D_{XY} = \max(D_X, D_Y)$  is given by the function  $f(x) = f_X(-x) + f_Y(-x)$ . The functions  $f_X(x)$  and  $f_Y(x)$  are defined as

$$f_X(x) = \frac{1}{\sigma_X} \varphi\left(\frac{x + \mu_X}{\sigma_X}\right) \times \Phi\left(\frac{p(x + \mu_X)}{\sigma_X \sqrt{1 - \rho^2}} - \frac{x + \mu_Y}{\sigma_Y \sqrt{1 - \rho^2}}\right) \quad (\text{A3})$$

and

$$f_Y(x) = \frac{1}{\sigma_Y} \varphi\left(\frac{x + \mu_Y}{\sigma_Y}\right) \times \Phi\left(\frac{p(x + \mu_Y)}{\sigma_Y \sqrt{1 - \rho^2}} - \frac{x + \mu_X}{\sigma_X \sqrt{1 - \rho^2}}\right) \quad (\text{A4})$$

respectively, where  $\varphi$  represents the standard normal distribution PDF and  $\Phi$  represents the standard normal distribution CDF.

Following the above equations allows for the implementation of a race model with a correlation parameter, *rho* ( $\rho$ ). To add the noise parameter, *eta* ( $\eta$ ), a constant is added to the *sigma* values of the unisensory drift rate distributions. Thus, in redundant conditions, the drift rates are defined as:

$$D_X \sim N(\mu_X, (\sigma_X + \eta)^2), \quad (\text{A5})$$

and

$$D_Y \sim N(\mu_Y, (\sigma_Y + \eta)^2). \quad (\text{A6})$$

In practice, unisensory drift rates were modelled by fitting a LATER model (see **Section 2.4.4.1**) to each unisensory 1/RT distribution. To find the best fitting *rho* and *eta* parameters, an optimisation algorithm (see **Section 2.4.1.2**) was used, and Maximum Likelihood Estimation (see **Section 2.4.2.2**) was used to evaluate the fit.

## 9.5. Analysis of Misses for Chapter 4 (Part 2)

In addition to the main effects reported in the main chapter (**Section 4.6.1**), the following interactions were observed.

There was a significant interaction between signal duration and task-irrelevant stimulation,  $F(1, 19)=20.374$ ,  $p=0.001$ ,  $\eta p^2=0.445$ . For short signals, the percentage of misses was 1.788% ( $\pm 0.228\%$ ) when task-irrelevant stimulation was absent and 0.748% ( $\pm 0.154\%$ ) when present. For long signals, the percentage of misses was 0.231% ( $\pm 0.094\%$ ) when task-irrelevant stimulation was absent and 0.067% ( $\pm 0.031\%$ ) when present.

There was a significant interaction between signal duration and signal modality,  $F(1.278, 24.285)=20.374$ ,  $p<0.001$ ,  $\eta p^2=0.517$ . In short conditions, tactile signals by far had the highest percentage of misses (2.871  $\pm 0.487\%$ ), followed by visual (0.886  $\pm 0.236\%$ ) and redundant (0.047  $\pm 0.047\%$ ). In long conditions, however, visual signals had the highest percentage of misses (0.273  $\pm 0.099\%$ ), followed by tactile (0.174  $\pm 0.090\%$ ) and redundant (0%).

There was also a significant interaction between task-irrelevant stimulation and signal modality,  $F(1.491, 28.324)=19.781$ ,  $p<0.001$ ,  $\eta p^2=0.510$ . In absent conditions, tactile signals by far had the highest percentage of misses (2.221  $\pm 0.354\%$ ), followed by visual (0.761  $\pm 0.161\%$ ) and redundant (0.047  $\pm 0.047\%$ ). In present conditions, tactile signals also had the highest percentage of misses (0.823

$\pm 0.196\%$ ), followed by visual ( $0.399 \pm 0.163\%$ ) and redundant (0%). However the difference between tactile and other signals was much smaller in present conditions.

There was also a significant three-way interaction between signal duration, task-irrelevant stimulation, and signal modality  $F(1.551, 29.466)=10.956, p=0.001, \eta p^2=0.366$ . Misses for each condition are shown in **Table 4.3**. In short-absent conditions, tactile signals by far had the highest percentage of misses, followed by visual and redundant. In short-present conditions, tactile signals also had the highest percentage of misses, followed by visual and redundant. In long-absent conditions, however, visual signals had the highest percentage of misses, followed by tactile and redundant. In long-present conditions, visual signals also had the highest percentage of misses, followed by tactile and redundant. Overall, as discussed in the main chapter, these interactions can be understood by the large percentage of misses for tactile signals in the short-absent condition.

## 9.6. History Effects for Each Modality

All values reported here are in seconds (s).

### Chapter 3

	Auditory		Visual	
	Consistent	Alternating	Consistent	Alternating
<b>Simple</b>	0.049 (0.010)	0.056 (0.008)	0.010 (0.006)	0.022 (0.006)
<b>Complex</b>	0.042 (0.009)	0.048 (0.011)	0.016 (0.010)	0.032 (0.012)

### Chapter 4 (Part 1)

	Auditory		Visual	
	Absent	Present	Absent	Present
<b>Short</b>	0.060 (0.011)	0.057 (0.009)	0.047 (0.006)	0.031 (0.005)
<b>Long</b>	0.042 (0.007)	0.032 (0.006)	0.027 (0.005)	0.024 (0.006)

### Chapter 4 (Part 2)

	Tactile		Visual	
	Absent	Present	Absent	Present
<b>Short</b>	0.019 (0.009)	0.022 (0.009)	0.036 (0.007)	0.022 (0.008)
<b>Long</b>	0.029 (0.009)	0.013 (0.009)	0.027 (0.007)	0.015 (0.006)

### Chapter 5

	Auditory		Visual	
	Hand	Foot	Hand	Foot
<b>Strong</b>	0.024 (0.006)	0.019 (0.007)	0.018 (0.006)	0.027 (0.005)
<b>Weak</b>	0.071 (0.011)	0.067 (0.016)	0.031 (0.011)	0.025 (0.012)



## 9.7. Simulating the Effect of Non-Decisional Components on Model Fitting

In **Chapter 5**, a main effect of response effector (hand, foot) was found for the *eta* parameter of the context variant race model. The additional noise added by *eta* was smaller for foot responses than for hand responses. This effect may have arisen from the simplicity of the LATER model, which is used to fit the unisensory RT distributions, rather than genuinely reflecting changes in decision-making. A large change in median RT which arises from *non-decisional* components (such as the change in response effector) can only be modelled by LATER as a change in *decisional* components (i.e. the *mu* and *sigma* parameters, which describe the evidence accumulation process across trials). As such, the fitting of the interaction parameters (which build on these LATER fits) may also be influenced by this issue when comparing across different effectors with different median RTs.

To demonstrate this, I simulated RTs by using the context variant race model (see **Section 2.5.3**). The parameter values used to initialise the model are shown in the top row of **Table 9.1**. 100 RTs for all three signal types (X, Y, XY) were sampled from the model over 1000 repetitions. As a first step, for an initial check, the context variant race model was fit back to these RTs. Overall, the average recovered parameter values shown in **Table 9.1 (RT)** match the parameters of the model very well, as expected. Second, I simulated an additional non-decisional component of RT, similar to the delay introduced by moving from responding by hand to by foot. As a simple procedure, this additional motor delay was simulated by adding a constant delay to all RTs (0.1 s). Importantly, this does not change the variability of RTs, but only shifts them in time. The context variant race model was then fit again to these RT data (*RT + Motor*). Overall, the additional motor delay was modelled by changes in the decision parameters, as seen in **Table 9.1 (RT + Motor)**. The *mu* parameters decreased, to account for the slower median RT following the addition of the motor component. Similarly, the *sigma* parameters decreased; this is because less variability in the rate values is needed to model the same variability in RTs (as measured by MAD, for instance) if the median becomes slower. This comes from the model geometry (see **Figure 1.6b**). Concerning *rho*, the correlation was not largely affected. Concerning *eta*, however, the additional noise added was smaller when the constant motor delay was added. In relation to the results of **Chapter 5**, therefore, this suggests that the main effect of response effector comes from the simplicity of the LATER model. As non-decisional components of RT can only

**Table 9.1 Model and average recovered parameter values ( $\pm$ SEM) for the motor component simulation (all values in  $s^{-1}$  except *rho*)**

	Parameters					
Model values	<i>mu<sub>X</sub></i>	<i>mu<sub>Y</sub></i>	<i>sigma<sub>X</sub></i>	<i>sigma<sub>Y</sub></i>	<i>rho</i>	<i>eta</i>
	2.5	2.5	0.4	0.4	-0.2	-0.1
Fit values	<i>mu<sub>X</sub></i>	<i>mu<sub>Y</sub></i>	<i>sigma<sub>X</sub></i>	<i>sigma<sub>Y</sub></i>	<i>rho</i>	<i>eta</i>
<i>RT</i>	2.501 (0.001)	2.499 (0.001)	0.399 (0.001)	0.399 (0.001)	-0.191 (0.007)	0.099 (0.001)
<i>RT + Motor</i>	1.992 (0.001)	1.991 (0.001)	0.256 (0.001)	0.256 (0.001)	-0.184 (0.007)	0.051 (0.001)

be modelled here by changes in decision components, erroneous effects can arise when comparing responses made by different effectors with different associated time-courses for the non-decisional response component.

Notably, on the straightforward model recovery ( $RT$ ), there was a larger variability in the value of  $\rho$  than all other parameters. As an additional assessment of the fitting of interaction parameters, therefore, I computed 95% confidence intervals for the recovery of both  $\rho$  and  $\eta$  in this simulation. For  $\rho$ , there was a large variability around the mean (-0.191), 95% CI [-0.582, 0.302]. For  $\eta$ , recovery was more reliable around the mean ( $0.099 \text{ s}^{-1}$ ), 95% CI [0.013, 0.184]. Overall, this indicates that with the number of trials used in experiments here ( $N=100$  in each condition), there is some degree of uncertainty in the recovery of interaction parameters, particularly  $\rho$ .

## 9.8. Average Model Parameters for Experimental Chapters

Note that the unit for each parameter is  $s^{-1}$  (except for correlation,  $\rho$ )

### 9.8.1. Two-Step Fitting Procedure

#### Chapter 3

	$\mu_A$	$\mu_V$	$\sigma_A$	$\sigma_V$	$\rho$	$\eta$
Sim / Con	3.278 (0.106)	3.163 (0.081)	0.736 (0.033)	0.649 (0.021)	-0.240 (0.085)	0.226 (0.032)
Sim / Alt	3.243 (0.108)	3.169 (0.085)	0.709 (0.035)	0.624 (0.025)	-0.215 (0.101)	0.219 (0.017)
Com / Con	2.420 (0.090)	2.321 (0.090)	0.576 (0.022)	0.493 (0.017)	-0.108 (0.081)	0.092 (0.014)
Com / Alt	2.382 (0.086)	2.285 (0.076)	0.532 (0.018)	0.506 (0.018)	-0.151 (0.080)	0.085 (0.013)

#### Chapter 4 (Part 1)

	$\mu_A$	$\mu_V$	$\sigma_A$	$\sigma_V$	$\rho$	$\eta$
Short / Abs	2.885 (0.134)	3.131 (0.116)	0.639 (0.033)	0.609 (0.028)	-0.286 (0.082)	0.206 (0.022)
Short / Pres	2.829 (0.140)	3.150 (0.133)	0.716 (0.042)	0.653 (0.036)	-0.154 (0.097)	0.141 (0.021)
Long / Abs	3.054 (0.129)	3.069 (0.114)	0.595 (0.027)	0.586 (0.028)	-0.159 (0.103)	0.194 (0.026)
Long / Pres	2.961 (0.137)	3.115 (0.121)	0.665 (0.028)	0.598 (0.035)	-0.020 (0.110)	0.172 (0.019)

#### Chapter 4 (Part 2)

	$\mu_T$	$\mu_V$	$\sigma_T$	$\sigma_V$	$\rho$	$\eta$
Short / Abs	2.897 (0.094)	3.215 (0.105)	0.733 (0.028)	0.676 (0.016)	-0.369 (0.079)	0.224 (0.026)
Short / Pres	3.105 (0.114)	3.378 (0.119)	0.772 (0.035)	0.771 (0.039)	-0.226 (0.072)	0.083 (0.019)
Long / Abs	2.954 (0.088)	3.227 (0.114)	0.622 (0.033)	0.645 (0.020)	-0.430 (0.060)	0.253 (0.025)
Long / Pres	3.143 (0.117)	3.388 (0.108)	0.708 (0.038)	0.734 (0.034)	-0.155 (0.080)	0.142 (0.021)

#### Chapter 5

	$\mu_A$	$\mu_V$	$\sigma_A$	$\sigma_V$	$\rho$	$\eta$
Strong / Hand	2.918 (0.093)	2.601 (0.075)	0.543 (0.024)	0.457 (0.019)	-0.189 (0.089)	0.123 (0.020)
Strong / Foot	2.384 (0.080)	2.154 (0.057)	0.400 (0.019)	0.363 (0.018)	-0.204 (0.092)	0.074 (0.009)
Weak / Hand	2.087 (0.060)	2.097 (0.074)	0.580 (0.027)	0.487 (0.019)	-0.145 (0.075)	0.078 (0.020)
Weak / Foot	1.769 (0.055)	1.779 (0.053)	0.459 (0.018)	0.375 (0.012)	-0.077 (0.076)	0.045 (0.013)

## 9.8.2. One-Step Fitting Procedure

### Chapter 3

	$\mu_{u_A}$	$\mu_{u_V}$	$\sigma_{\tau_A}$	$\sigma_{\tau_V}$	$\rho$	$\eta$
Short / Abs	3.291 (0.106)	3.176 (0.082)	0.727 (0.031)	0.649 (0.020)	-0.218 (0.084)	0.220 (0.032)
Short / Pres	3.256 (0.109)	3.184 (0.087)	0.699 (0.034)	0.626 (0.025)	-0.191 (0.099)	0.211 (0.017)
Long / Abs	2.427 (0.091)	2.330 (0.091)	0.571 (0.021)	0.493 (0.017)	-0.081 (0.081)	0.088 (0.015)
Long / Pres	2.392 (0.086)	2.295 (0.077)	0.527 (0.017)	0.506 (0.018)	-0.120 (0.078)	0.080 (0.013)

### Chapter 4 (Part 1)

	$\mu_{u_A}$	$\mu_{u_V}$	$\sigma_{\tau_A}$	$\sigma_{\tau_V}$	$\rho$	$\eta$
Short / Abs	2.901 (0.136)	3.151 (0.117)	0.638 (0.033)	0.605 (0.028)	-0.229 (0.096)	0.192 (0.022)
Short / Pres	2.849 (0.142)	3.171 (0.135)	0.715 (0.042)	0.648 (0.036)	-0.129 (0.094)	0.128 (0.019)
Long / Abs	3.068 (0.131)	3.084 (0.115)	0.592 (0.027)	0.584 (0.028)	-0.134 (0.101)	0.186 (0.025)
Long / Pres	2.971 (0.138)	3.126 (0.122)	0.660 (0.028)	0.596 (0.034)	0.005 (0.109)	0.167 (0.018)

### Chapter 4 (Part 2)

	$\mu_{u_T}$	$\mu_{u_V}$	$\sigma_{\tau_T}$	$\sigma_{\tau_V}$	$\rho$	$\eta$
Short / Abs	2.922 (0.096)	3.242 (0.105)	0.733 (0.027)	0.671 (0.017)	-0.334 (0.076)	0.206 (0.026)
Short / Pres	3.128 (0.116)	3.409 (0.121)	0.772 (0.034)	0.763 (0.038)	-0.153 (0.085)	0.060 (0.023)
Long / Abs	2.975 (0.088)	3.255 (0.118)	0.625 (0.032)	0.636 (0.020)	-0.386 (0.055)	0.233 (0.028)
Long / Pres	3.158 (0.118)	3.406 (0.108)	0.707 (0.037)	0.727 (0.033)	-0.121 (0.079)	0.135 (0.021)

### Chapter 5

	$\mu_{u_A}$	$\mu_{u_V}$	$\sigma_{\tau_A}$	$\sigma_{\tau_V}$	$\rho$	$\eta$
Strong / Hand	2.931 (0.094)	2.611 (0.076)	0.534 (0.023)	0.459 (0.019)	-0.155 (0.086)	0.117 (0.020)
Strong / Foot	2.391 (0.080)	2.159 (0.057)	0.395 (0.018)	0.364 (0.018)	-0.180 (0.090)	0.072 (0.009)
Weak / Hand	2.091 (0.060)	2.101 (0.075)	0.579 (0.027)	0.483 (0.019)	-0.127 (0.078)	0.075 (0.020)
Weak / Foot	1.776 (0.054)	1.786 (0.053)	0.456 (0.017)	0.374 (0.012)	-0.015 (0.091)	0.039 (0.014)

## 10. Abbreviations

Abbreviation	Term
AIC	Akaike's Information Criterion ( <b>Section 2.4.3.2</b> )
AT	Auditory-Tactile
AV	Auditory-Visual
BIC	Bayesian Information Criterion ( <b>Section 2.4.3.3</b> )
CDF	Cumulative Distribution Function ( <b>Section 1.2.2</b> )
CI	Confidence Interval
CP	Cumulative Probability ( <b>Section 1.2.2</b> )
DDM	Diffusion Decision Model
IRM	Independent Race Model ( <b>Section 1.4.1</b> )
LATER	Linear Approach to Threshold with Ergodic Rate ( <b>Section 1.2.3.1</b> )
LBA	Linear Ballistic Accumulator
LRT	Likelihood Ratio Test ( <b>Section 2.4.3.1</b> )
MAD	Median Absolute Deviation
MLE	Maximum Likelihood Estimation ( <b>Section 2.4.2.2</b> )
MVUE	Minimum Variance Unbiased Estimator
PDF	Probability Density Function
RT	Response Time
RMSE	Root Mean Squared Error ( <b>Section 2.4.2.1</b> )
RSE	Redundant Signal Effect ( <b>Section 1.3.2</b> )
RSP	Redundant Signal Paradigm ( <b>Section 1.3.1</b> )
TV	Tactile-Visual

## General acknowledgements

I am foremost grateful to my parents, Vanessa and Nigel, who have never hesitated to offer every possible support. I would not have made it here without them.

First, I would like to sincerely thank my supervisor, Dr. Thomas Otto. Tom was kind enough to take me on as his student part-way through a difficult period in my PhD journey, a gesture I will not soon forget. In the 3 years I have worked as his student, he has been nothing but patient in his teaching and supportive in his role as supervisor. It was with his encouragement that I was able to travel to my first conference abroad, and through his helpful comments and attention to detail that I was able to make substantial progress in my writing and my presentation skills. Under his guidance, I have learned so much more than I ever thought I was capable of.

Second, I would like to thank the other members of the Logical Calculus Lab for their insight and input over the years: in particular, I would like to thank Yue (Serena) Liu, Cleo Pike, & Harry Coulson. To have been part of this group has been the highlight of my studies.

Third, I would like to thank my second supervisor Dr. Dhanraj Vishwanath and the other members of the Vision Labs Journal Club, who have provided me with useful feedback on my presentations. I would also like to thank my fellow PhD students at the University of St. Andrews. I am particularly grateful to Andrew Mackenzie, David Walker, Abigail Lee, and Alexandra Mitchell.

A PhD is its own unique and peculiar weight; a strange state of being which reveals as much about one's self as it does about the subject matter under investigation. For this reason, I am also exceptionally grateful for the shared understanding and moral support of two of my fellow PhD students outside the University of St. Andrews: Jamal Kinsella and Ben Beswick.

No task is complete without its distractions. It is for this reason that I thank the St. Andrews Comedy Society, and the close friends I have made as part of the committee and through performing at Sandy's Sundown Standup. All of you have made my time in St. Andrews more pleasant, and allowed me to finally experience a different side of university life. I am particularly grateful to Joe Irvine, who, in addition to being an outstanding comedian, is also a talented friend in his downtime.

Beyond words, I am grateful to Amy Callaghan, who supported me through the final, darkest days of this work. Though I'm sorry I couldn't be more present while working on this, I will always be thankful for the time we got to spend together. You made it so much easier, and I'm not sure I would have made it through without your encouragement.

Finally, at the end of it all, I can also say that I am truly grateful for that lost first year, and for every day I was told I wouldn't get this far. Whatever happens next, here I am.

## **Funding**

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/J01446X/1].

### **Research Data/Digital Outputs access statement**

All research data underpinning this thesis are available on PURE, the University of St. Andrews' Research Information System.

The data underpinning Chapter 3 (published as Innes & Otto, 2019) are available at:  
<https://doi.org/10.17630/c8cbd7b7-e2b3-4e62-bb3d-66fce081ff59>

The data underpinning Chapter 4 are available at:  
<https://doi.org/10.17630/f7d830a6-aea0-452f-8efc-54fb68f9184c>

The data underpinning Chapter 5 are available at:  
<https://doi.org/10.17630/a21df9d8-0298-4360-9c62-9425870a1b28>

### **Candidate's declaration**

I, Bobby Richard Innes, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 61,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2014.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date

Signature of candidate

### **Supervisor's declaration**

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date

Signature of supervisor

### **Permission for publication**

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Bobby Richard Innes, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:



**Printed copy**

No embargo on print copy.

**Electronic copy**

Embargo on all of electronic copy for a period of 1 year on the following ground(s):

- Publication would preclude future publication

**Supporting statement for electronic embargo request**

I am hoping to use material within this thesis as a basis for published journal articles. As such, I would request that the thesis is not published online until I have had time to write and submit these articles (to agree with the requirements of any journals I may submit to).

**Title and Abstract**

- I agree to the title and abstract being published.

Date

Signature of candidate

Date

Signature of supervisor

## **Underpinning Research Data or Digital Outputs**

### **Candidate's declaration**

I, Bobby Richard Innes, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date

Signature of candidate

### **Permission for publication of underpinning research data or digital outputs**

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

Embargo on all of electronic files for a period of 1 year on the following ground(s):

- Publication would preclude future publication

### **Supporting statement for embargo request**

As previously stated, I am hoping to use material in this thesis as a basis for published articles. I would like therefore to make the data available publicly only after I have had the chance to write and submit these articles.

Date

Signature of candidate

Date

Signature of supervisor



University Teaching and Research Ethics Committee

24 June 2016

Dear Bobby

Thank you for submitting your ethical application which was considered at the School of Psychology & Neuroscience Ethics Committee meeting on 9<sup>th</sup> June 2016; the following documents have been reviewed:

1. Ethical Application Form
2. Advertisements
3. Participant Information Sheet
4. Participant Consent Form
5. Mailing List Consent Form
6. Data Management Plan

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

<b>Approval Code:</b>	PS12181	<b>Approved on:</b>	23/06/2016	<b>Approval Expiry:</b>	23/06/2021
<b>Project Title:</b>	Investigating the stochastic nature of multisensory response times				
<b>Researcher:</b>	Bobby Innes				
<b>Supervisor:</b>	Dr Thomas Otto				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (<http://www.st-andrews.ac.uk/utrec/guidelinespolicies/>). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee

cc Dr Thomas Otto (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP  
Email: [psyethics@st-andrews.ac.uk](mailto:psyethics@st-andrews.ac.uk) Tel: 01334 462071

The University of St Andrews is a charity registered in Scotland: No SC013532



University Teaching and Research Ethics Committee

16 May 2018

Dear Bobby

Thank you for submitting your amendment application which comprised the following documents:

1. Ethical Amendment Application Form
2. Advertisement: Poster, SONA, Email
3. Equipment Details: C-2 Tactor

The School of Psychology & Neuroscience Ethics Committee is delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has approved this ethical amendment application. The particulars of this approval are as follows –

<b>Original Approval Code:</b>	PS12181	<b>Approved on:</b>	23/06/2016
<b>Amendment Approval Date:</b>	11/05/2018	<b>Approval Expiry Date:</b>	23/06/2021
<b>Project Title:</b>	Investigating the stochastic nature of multisensory response times		
<b>Researcher:</b>	Bobby Innes		
<b>Supervisor:</b>	Dr Thomas Otto		

Ethical amendment approval does not extend the originally granted approval period of five years, rather it validates the changes you have made to the originally approved ethical application. If you are unable to complete your research within the original five year validation period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the understanding that you adhere to the 'Guidelines for Ethical Research Practice' (<http://www.st-andrews.ac.uk/media/UTRECguidelines%20Feb%2008.pdf>).

Yours sincerely

Convener of the School Ethics Committee

cc Dr Thomas Otto (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP  
Email: [psyethics@st-andrews.ac.uk](mailto:psyethics@st-andrews.ac.uk) Tel: 01334 462071

The University of St Andrews is a charity registered in Scotland: No SC013532


## Published Work

The following article is reproduced from:

Innes, B.R., & Otto, T.U. (2019). A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. *Scientific Reports*, 9(1), 2921. doi: 10.1038/s41598-019-39924-6

The article is a published version of the work I have presented in this thesis (Chapter 3), which is included here for reference purposes. The article is available for sharing, adaptation, distribution and reproduction in any medium or format under the Creative Commons licence Attribution 4.0 International (CC BY 4.0). No changes have been made to the published version.

# SCIENTIFIC REPORTS



OPEN

## A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent

Bobby R. Innes  & Thomas U. Otto 

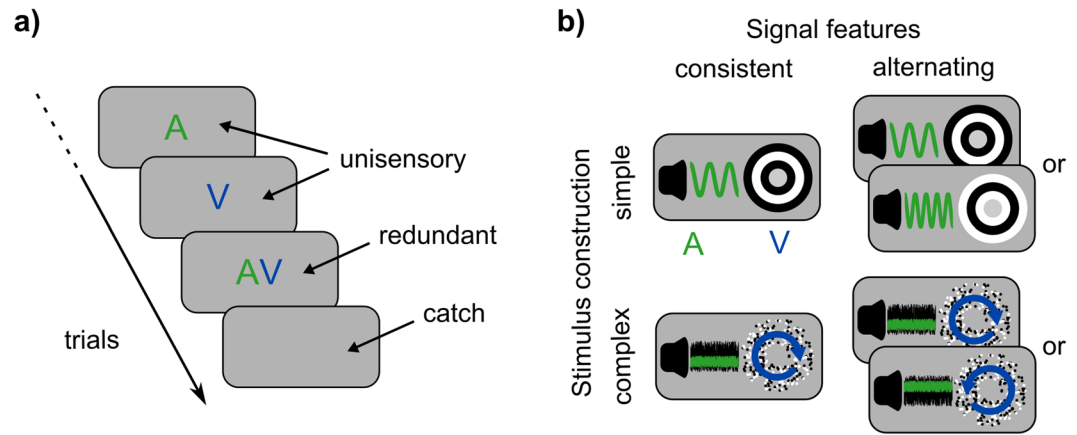
Multisensory signals allow faster responses than the unisensory components. While this redundant signals effect (RSE) has been studied widely with diverse signals, no modelling approach explored the RSE systematically across studies. For a comparative analysis, here, we propose three steps: The first quantifies the RSE compared to a simple, parameter-free race model. The second quantifies processing interactions beyond the race mechanism: history effects and so-called violations of Miller's bound. The third models the RSE on the level of response time distributions using a context-variant race model with two free parameters that account for the interactions. Mimicking the diversity of studies, we tested different audio-visual signals that target the interactions using a  $2 \times 2$  design. We show that the simple race model provides overall a strong prediction of the RSE. Regarding interactions, we found that history effects do not depend on low-level feature repetition. Furthermore, violations of Miller's bound seem linked to transient signal onsets. Critically, the latter dissociates from the RSE, demonstrating that multisensory interactions and multisensory benefits are not equivalent. Overall, we argue that our approach, as a blueprint, provides both a general framework and the precision needed to understand the RSE when studied across diverse signals and participant groups.

Everyday life requires processing of multiple signals coming from different sensory modalities like audition and vision. An incoming phone call, for example, can be signalled either in one modality only (e.g. a ringtone as a unisensory signal) or in multiple modalities simultaneously (e.g. a ringtone and a flashing screen as redundant signals). As observed in the redundant signal paradigm (Fig. 1a), the benefit of redundant signals is that individuals respond faster compared to the unisensory components, which is the *redundant signals effect* (RSE) as described in classic experiments<sup>1-3</sup>.

Given the broad scope of multisensory benefits, the RSE has been investigated for over 100 years with different research questions and diverse signal sets. Basically all sensory modalities have been tested, including audio-tactile<sup>4</sup>, visual-tactile<sup>5</sup>, and gustatory-olfactory pairings<sup>6</sup>. The largest body of research has focused on audio-visual pairings, but even within this domain, stimuli have varied considerably across studies. Some have used abstract signals, such as pure tones or noise sounds that were paired with flashes, letters, or simple geometric shapes<sup>7-9</sup>. Others have used real-world signals, such as animal sounds and images<sup>10</sup> or signals with emotional content<sup>11</sup>. In most experiments, signals have had sudden-onsets and were presented briefly with little variation across trials, but the RSE was also observed when signals were randomly generated on each trial and presented within a continuous audio-visual background<sup>12,13</sup>. While the basic RSE has reliably replicated, the diverse studies differ in several aspects including overall mean RTs and the exact size of the RSE. Due to these differences, a recent review argued that a common modelling framework for the RSE is unlikely to be found<sup>14</sup>. In this view, a comparative approach between studies is difficult and a complete understanding of the RSE is lacking.

To develop a common framework, it is a good starting point to build upon models of unisensory decision-making, which typically assume some key processing steps e.g.<sup>15-17</sup>. Sensory evidence for a signal is available from the environment, and is sampled by neurons with corresponding receptive fields. Due to fluctuations in neuronal activity, the evidence is subject to *noise*. To minimise the influence of noise on signal detection,

School of Psychology & Neuroscience, St. Mary's Quad, South Street, St. Andrews, KY16 9JP, United Kingdom. Correspondence and requests for materials should be addressed to B.R.I. (email: [bobby.richard.innes@gmail.com](mailto:bobby.richard.innes@gmail.com)) or T.U.O. (email: [to7@st-andrews.ac.uk](mailto:to7@st-andrews.ac.uk))



**Figure 1.** Experimental design. **(a)** Redundant signals paradigm. Participants respond to trials with auditory (A), visual (V), and combined signals (AV), but not on catch trials. Signals were presented in a random order within a block of trials. **(b)** Following a  $2 \times 2$  design, we tested four signal sets in separate blocks. Stimulus construction was either simple (e.g. a pure tone, in audition) or complex (e.g. a noise tone). The sequence of signal features was either consistent (e.g. one frequency) or alternating (e.g. randomly one of two frequencies).

the evidence is accumulated over time until a threshold level is reached and a response is triggered. To extend the framework to the RSE, some sort of *combination rule* is needed that specifies how evidence from two senses is processed to trigger a single response. Broadly, there are two main model architectures with fundamentally different combination rules. Firstly, so-called *race models* suggest that evidence for each signal is accumulated separately by parallel decision units (e.g. one for audition and one for vision) and that the first unit to reach its threshold triggers a response. Secondly, *pooling models* propose that evidence for both signals is summed in a single decision unit and that a response is triggered when the combined evidence reaches a threshold level. Both architectures could in principle explain the RSE.

The race model architecture was first proposed by Raab<sup>18</sup>. The idea is that selecting the faster of two stochastic decision processes results on average in faster and less variable RTs with redundant compared to unisensory signals. Thus, race models explain the RSE mainly by statistical facilitation. Miller<sup>8</sup>, however, argued that if race models were correct, an upper bound for the RSE can be computed on the level of RT distributions, which is named Miller's bound. As this bound is typically violated in experiments, Miller<sup>8</sup> and many follow-up studies concluded that race models cannot explain the RSE, which has led to the dominance of pooling models. However, to the best of our knowledge and in agreement with the recent review<sup>14</sup>, no pooling model has been shown to account for the RSE across studies. Hence, the lack of a common framework is in essence a failure of the dominating class of pooling models. The possibility that race models provide a common framework remained unexplored due to the general rejection.

This omission becomes critical as the general rejection of race models is not a valid argument<sup>19</sup>. The issue is the untested and mostly neglected *context invariance* assumption, which states that the processing of one signal is not affected by another<sup>20,21</sup>. As Miller's bound implicitly assumes context invariance, a violation of the bound does not necessarily show that the basic race model architecture is wrong. Alternatively, the context invariance assumption may be wrong, which would show that some sort of interaction has taken place between the parallel decision units. If the context invariance assumption is dropped, we demonstrated that context variant race models can explain the RSE, including violations of Miller's bound<sup>12,13</sup>. These considerations show that the parallel architecture of race models could in fact provide a common framework for RSE studies.

To study the RSE with diverse signal sets, here, we develop a comparative approach based on race models. In a first of three steps, we test the ability of the framework to predict multisensory benefits. Following Raab's model<sup>18</sup>, directional predictions of the RSE for different sets of signals can be readily made according to the *principles of multisensory behaviour*<sup>12</sup>. Firstly, the *principle of equal effectiveness* states that benefits should increase when unisensory RTs become more similar. Secondly, the *variability rule* states that the variability of RTs is the key driving force of multisensory benefits, which should increase when unisensory RTs become more variable. In addition, following Raab's model<sup>18</sup>, a parameter-free prediction can be made of the exact size of the RSE based on the unisensory RT distributions and *probability summation*. The first step thus provides a simple tool to analyse the RSE if diverse signal sets yield different RTs.

The second step quantifies two processing interactions with multisensory signals that are not covered by Raab's model. The first interaction concerns *history effects* that describe the influence of previous trials on current-trial processing<sup>4,8,13</sup>. Typically, RTs to unisensory signals are faster if a modality is repeated (e.g. audition following audition) compared to a switch (e.g. audition following vision). This interaction challenges the *statistical independence* assumption that is made by the parameter-free prediction using probability summation<sup>19</sup>. The second interaction is observed by violations of Miller's bound, which shows that processing of one signal must affect processing of the other signal in some way if the parallel processing architecture is correct. This second interaction challenges the *context invariance* assumption made by the parameter-free prediction based on probability

summation<sup>19</sup>. By quantifying the two interactions, it becomes clear that Raab's basic model needs to be extended. Moreover, the interactions need to be critically considered to understand how multisensory interactions are eventually changed with diverse signals.

The third step is to account for the RSE on the level of entire RT distributions with a modelling approach. Specifically, we use here the context variant race model that includes two free parameters to account for the interactions<sup>12,13,19</sup>. Regarding history effects, the model includes the correlation  $Rho$ , which is used in the probability summation rule given that the assumption of statistical independence does not hold. Regarding violations of Miller's bound, the model includes the noise  $Eta$ , which models increased noise within the accumulation process for redundant compared to unisensory conditions. This increased noise may come, for instance, from a disruptive interaction between parallel decision units (i.e. having both accumulate evidence at the same time decreases their reliability). This second parameter is motivated by findings showing that empirical RT distributions with redundant signals are not faster but more variable than the best-fitting race model with only  $Rho$  as a free parameter<sup>13</sup>. A key advantage of considering entire RT distributions is that these provide a more detailed view compared to an approach that is based only on a subset of particularly fast RTs as in Miller's test. Such a detailed view is critically needed given that present understanding of multisensory interactions is still far from complete. For example, the source of additional noise is not yet known and subject to speculation<sup>19</sup>. Thus, it will be particularly important to identify experimental manipulations that specifically target one or the other interaction so that the underlying sources can be determined. In the end, we are convinced that such a comparative approach will lead to a better understanding of multisensory processing.

Following the comparative approach, we developed stimuli to mimic some of the diversity across RSE studies. Using a  $2 \times 2$  design, we manipulated *stimulus construction* and *signal features* in each modality (Fig. 1b). Regarding the first factor, the construction of auditory and visual stimuli was either simple (identical across trials; without background stimulation) or complex (randomly generated; with background stimulation). One aspect of this manipulation is that simple stimuli had strong onset transients, which were masked in complex conditions by the background. As onset transients have been considered relevant to multisensory processing<sup>22,23</sup>, we aim to evaluate whether processing interactions, as revealed by the modelling approach, change according to this factor. Regarding the second factor, signal features were either consistent (one variant per modality) or alternating (one of two variants per modality). With this factor, we aimed to understand the origins of history effects. One possibility is that history effects arise simply because of the repetition of low-level features (e.g. the specific frequency of a tone). If true, introducing alternating signals within each modality would reduce history effects. Alternatively, if no changes in history effects occur, this would suggest a basis in higher-level processes.

## Methods

**Participants.** 20 healthy adults (18–29 years; 14 females) were recruited via the University of St Andrews. All were naïve regarding the experiment's purposes and reported normal hearing and normal/corrected-to-normal vision. Informed consent was obtained from all participants. Participants were reimbursed with £9. All procedures were approved by the University of St. Andrews' *University Teaching and Research Ethics Committee* (UTREC, approval code: PS12181) and were performed in accordance with the Code of Human Research Ethics (British Psychological Society, 2014).

**Apparatus.** The experiment was controlled by a Dell computer (Optiplex XE2) equipped with Matlab and the Psychophysics Toolbox extensions<sup>24–26</sup>. Visual stimuli were presented on a Dell UltraSharp U2713HM monitor (resolution:  $1,920 \times 1,080$  pixel; refresh: 60 Hz). Viewing distance was 57 cm. Auditory stimuli were presented via over-ear headphones (Sennheiser HD 280 Pro) at a 44.1 kHz sample frequency. The volume was calibrated using a Brüel and Kjær sound level meter (Type 2250) attached to an artificial ear (Type 4153). Participants responded by pressing a custom-built handheld button connected to an RTbox V5<sup>27</sup>. Prior to data collection, the RTbox was also used to calibrate audio-visual timing. As a test, audio-visual signal onsets were presented in 1000 trials. Calibrated stimuli were synchronous and onset jitter was reliably below 1 ms.

**Task.** We used the redundant signals paradigm (Fig. 1a). On signal trials (75%), we presented either an auditory (A), a visual (V), or redundant signals (both auditory and visual together, AV). Participants pressed a button as quickly as possible after detecting any signal. On catch trials (25%), no signal was presented and participants were asked to withhold a response.

**Stimuli.** *Auditory Stimuli.* We manipulated *stimulus construction* and *signal features* following a  $2 \times 2$  design (Fig. 1b). In simple conditions, signals were pure tones. In the consistent condition, the signal was always a 440 Hz tone. In the alternating condition, the signal was either a 440 or 660 Hz tone. Tones were presented at 45 dB SPL. In complex conditions, signals were noise sounds. To generate a new sound on each trial, Gaussian noise (i.e., a sequence of normally-distributed random numbers) was filtered with a 2<sup>nd</sup> order Butterworth bandpass filter. In the consistent condition, we always used edge frequencies of 1.0/1.1 kHz. In the alternating condition, we used edge frequencies of either 1.0/1.1 kHz or 1.1/1.2 kHz. Noise sounds were presented at 45 dB SPL. Noise sounds were presented within background noise (1<sup>st</sup> order Butterworth bandpass filter, edge frequencies: 0.5/2.4 kHz, 50 dB SPL). Auditory signals had a ramp onset of 10 ms.

*Visual Stimuli.* Signals covered the area of a notional annulus with an inner/outer radius of  $1^\circ/4^\circ$  (visual angle) around central fixation. In simple conditions, signals were composed of 3 concentric rings ( $1^\circ$  increments, alternating black-to-white; Fig. 1b). In the consistent condition, the ring pattern always started on black. In the alternating condition, the ring pattern started either on black or white. In complex conditions, 1000 dots (1 pixel; black or white) were uniformly-distributed within the annulus as visual background stimulation. Each dot moved



linearly in a random direction and speed (mean: 1°/s, *SD*: 0.2°/s). Dots were randomly replaced within the annulus when moving outside the annulus, or when the lifetime of 0.1 s had passed. For the signal, 50% of the dots started coherent rotation (mean: 0.67 rad/s, *SD*: 0.067 rad/s). In the consistent condition, the direction was always clockwise. In the alternating condition, the direction was either clockwise or anticlockwise. All visual stimuli were presented on a grey screen.

**Procedures.** Each trial started with a central, green fixation point (0.27°). In complex conditions, both the auditory and visual background were presented as well. The foreperiod duration consisted of a fixed component (1 s) and an exponentially-distributed random component (mean: 0.75 s). Then, a signal was presented for a max. duration of 1.5 s. On catch trials, no signal was presented and stimulation as during the foreperiod continued for 1.5 s. After a response, or at the end of the max. duration, the fixation point changed from green to red to indicate that the trial ended. Following false alarms and misses, a feedback screen indicated the mistake for 2 s. Stimulation, including the background in complex conditions, stopped for 0.25 s before the next trial started.

Stimuli were presented in blocks of 104 trials (26 trials per modality, 26 catch trials). The trial sequence was randomised, but always started with an additional dummy trial (presenting redundant signals). After a false alarm or miss, a corresponding trial was repeated, the trial sequence was reshuffled, and stimulation continued with another dummy trial. Dummy trials were not analysed. The experiment included 16 blocks (2 × 2 within-subjects design with 4 blocks per condition). Conditions were intermixed in the randomisation of block order, which was done for each participant using a Latin square. A block lasted approx. 4 min (a small pointer on fixation denoted progress). The entire experiment lasted around 105 min with breaks.

**Data analysis.** Any response within 1.5 s after signal onset was recorded as valid. Responses were false alarms when given during the foreperiod or at any time during catch trials. Signal trials with no response were misses. Before the main analysis, we performed an outlier correction. For this, we transformed the RTs of each condition into rates (1/RT). We excluded data points that deviated by more than 3 × 1.4826 × median absolute deviations (MADs) from the median<sup>28</sup>, corresponding to 3 SDs if data are normally distributed. Fast outliers occurred on 0.4% (±0.1) of trials and slow outliers on 0.5% (±0.1) of trials. A total of 24,720 valid RTs (approx. 103 per condition and participant) remained.

The main analysis focused on the speed-up of RTs with redundant signals compared to the unisensory components, which is the RSE. As defined previously, this multisensory benefit is best measured by the area between the cumulative distribution function (CDF) in the multisensory condition and the faster of the unisensory CDFs<sup>12</sup>. As a simple procedure, if the number of measured RTs is the same in all three conditions, the size of the area can be estimated:

$$benefit = \frac{\sum \min(A_i, V_i) - AV_i}{N} \quad (1)$$

$A_i$ ,  $V_i$ , and  $AV_i$  are quantiles of the empirical CDFs, where the index  $i$  indicates the rank ranging from 1 (the fastest) to  $N$  (the slowest RT).

This simple computation requires equal numbers of RTs in each condition to allow for equal quantile points. However, following outlier correction, the numbers of RTs are often unequal. As a solution, we used linear interpolation<sup>29</sup> to down-sample the collected RTs to a common sample size (we used 50 quantile points; the procedure was shown to be unbiased using Monte Carlo simulations). We computed empirical benefits for each participant and each condition of the 2 × 2 design.

To obtain quantitative predictions of the RSE, we used Raab's model<sup>18</sup>. Based on the CDFs in the unisensory conditions, a parameter-free prediction of the CDF in the multisensory condition can be obtained using probability summation:

$$P_{AV}(t) = P_A(t) + P_V(t) - P_A(t) \times P_V(t) \quad (2)$$

$P_A$  and  $P_V$  are the empirical CDFs in the unisensory conditions (we used linear interpolation to obtain continuous CDFs).  $P_{AV}$  is here the predicted CDF for multisensory condition. We then extracted 50 quantile RTs from the predicted CDF and calculated expected benefits analogous to the calculation of empirical benefits (Equation (1)).

Next, we quantified two processing interactions that go beyond Raab's model. First, RTs on a given trial can depend on previously presented signals. To quantify this history effect, we computed the RT difference between trials in which modalities were repeated, and trials in which modalities were switched:

$$history\ effect = \overline{RT}_{Switch} - \overline{RT}_{Repetition} \quad (3)$$

We computed the history effect for auditory and visual RTs separately and averaged these values to obtain a single measure. Second, we computed violations of Miller's bound<sup>8</sup>, which is given by the sum of the CDFs in the unisensory conditions:

$$P_{Miller}(t) = \min[P_A(t) + P_V(t), 1] \quad (4)$$

The CDF in the redundant signals condition has to fall below Miller's bound if both the race model and context invariance hold<sup>19</sup>. To quantify violations, we first used linear interpolation to obtain 50 quantiles for Miller's bound ( $Miller_t$ ). Then, similar to a previously detailed method<sup>30</sup>, we computed the size of the area between the CDF in the redundant condition and Miller's bound, where the redundant condition exceeded the bound:

$$violation = \frac{\sum \max(Miller_i - AV_i, 0)}{N} \quad (5)$$

$AV_i$  are the 50 quantile RTs from the empirical CDF in the redundant condition as used in the computation of benefits.

Finally, we applied the context variant race model<sup>12,13</sup>. All modelling was performed in the rate space (1/RT). As a first step, we fitted a LATER model<sup>31,32</sup> to each of the unisensory conditions. As the LATER model assumes that rates are normally distributed, we used the Matlab function *normfit* to obtain best-fitting values of the mean (*Mu*) and the standard deviation (*Sigma*) using the Minimum-Variance Unbiased Estimator (MVUE). As a second step, we fitted the context variant race model in the redundant condition. Using a race mechanism, the model assumes that a response on a given trial with redundant signals is triggered by the unisensory signal with the higher rate. Hence, the resulting rate distribution can be computed using the maximum distribution of two Gaussian random numbers<sup>33</sup>, which are given by the LATER fits in the unisensory conditions. The exact distribution depends on the correlation (*Rho*), which we kept as a free parameter to account for history effects. As a second free parameter, the additional noise (*Eta*) increases the variability of rates (i.e. the *Sigma* parameter) in both LATER units additively. This free parameter manifests a violation of the context invariance assumption, which allows for violations of Miller's bound. We used the Matlab function *mle* to obtain maximum likelihood estimates (MLEs) of *Rho* and *Eta* individually for each participant and each condition of the  $2 \times 2$  design.

Statistical analyses were performed in IBM SPSS Statistics 22. For main analyses, we used  $2 \times 2$  repeated-measures ANOVAs, which always tested the factors stimulus construction (simple, complex) and signal features (consistent, alternating). For initial analysis of median RTs, we used  $2 \times 2 \times 3$  repeated-measures ANOVAs, including the additional factor signal modality (A, V, and AV). We used a Greenhouse-Geisser correction where sphericity was violated. The alpha level for statistical testing was 0.05.

## Results

**General Performance.** To assess data quality, we first checked performance (false alarms, misses). In signal trials, the false alarm rate during the foreperiod was 1.04% ( $\pm 0.19\%$ , SEM). In catch trials, the false alarm rate was 1.53% ( $\pm 0.25\%$ ). The miss rate was 0.46% ( $\pm 0.19\%$ ). As performance was close to perfect, it is not further considered in the RT analysis (for a detailed analysis of false alarm and miss rates, see Supplementary Analysis 1).

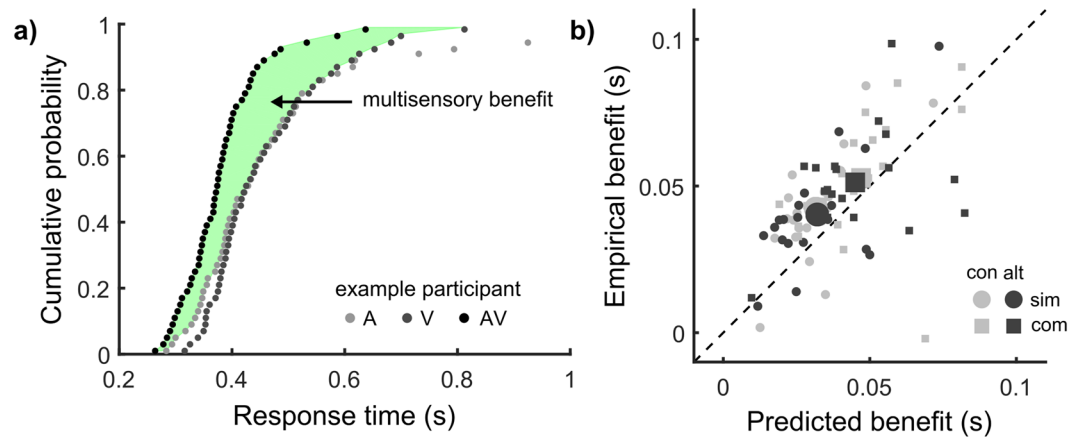
To assess whether the experimental manipulations mimicked the diversity of RSE studies, we next tested median RTs (Table S1). A  $2 \times 2 \times 3$  repeated-measures ANOVA showed a significant main effect of stimulus construction,  $F(1, 19) = 154.59, p < 0.001, \eta^2 = 0.89$ . RTs were faster with simple ( $0.300 \pm 0.010$  s) than with complex stimuli ( $0.416 \pm 0.015$  s). Hence, the experimental factor stimulus construction successfully manipulated RTs. There was neither a main effect of signal features nor any interaction (all  $F \leq 2.04, p \geq 0.165, \eta^2 \leq 0.10$ ). Expectedly, there was a main effect of signal modality,  $F(1.39, 26.37) = 53.32, p < 0.001, \eta^2 = 0.74$ . Pairwise comparisons revealed that redundant RTs ( $0.321 \pm 0.011$  s) were faster than both auditory ( $0.371 \pm 0.014$  s) and visual RTs ( $0.382 \pm 0.012$  s), both  $p < 0.001$ . Auditory and visual RTs were not significantly different,  $p = 0.626$ . This multisensory benefit is the RSE, which we investigate in full detail using our comparative approach.

Additionally, we also examined effects on variability of RTs, as measured by the Median Absolute Deviation (MAD) of RTs. A  $2 \times 2 \times 3$  repeated-measures ANOVA showed a main effect of stimulus construction,  $F(1, 19) = 58.851, p < 0.001, \eta^2 = 0.76$ . MAD of RTs was smaller with simple ( $0.040 \pm 0.003$  s) compared to complex ( $0.060 \pm 0.004$  s) stimuli. There was also a main effect of signal modality,  $F(1, 19) = 35.306, p < 0.001, \eta^2 = 0.65$ . As expected from probability summation, redundant RTs were less variable ( $0.040 \pm 0.003$  s) than auditory ( $0.057 \pm 0.004$  s) and visual ( $0.054 \pm 0.004$  s) RTs, both  $p < 0.001$ . MAD of auditory and visual RTs was not significantly different,  $p = 0.822$ . Finally, there was a significant interaction between stimulus construction and signal modality,  $F(2, 38) = 5.230, p < 0.010, \eta^2 = 0.22$ . Paired-samples were conducted for MAD (pooled across signal features) for each modality. This analysis found larger MAD values for complex conditions compared to simple conditions for all three modalities, all  $p < 0.001$ . The interaction is explained by the difference between simple and complex MAD values being larger for auditory (0.021 s) and visual (0.024 s) signals than it is for redundant signals (0.013 s).

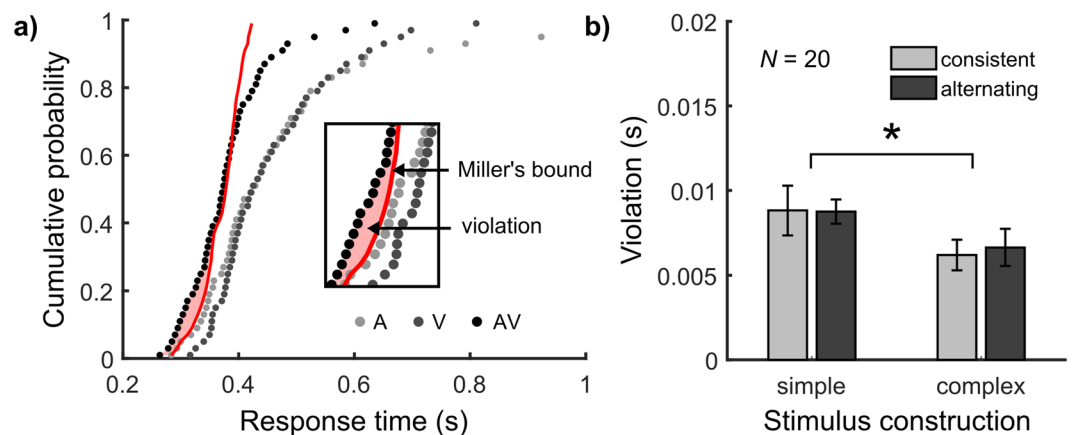
**Step 1: Multisensory Benefits.** As first step, we tested the ability of the race model framework to predict multisensory benefits, which are best measured based on RT distributions (Equation (1); for an illustration, see Fig. 2a). According to Raab's model, a parameter-free prediction of the RT distribution with redundant signals can be obtained using probability summation (Equation (2)), which allows calculation of predicted benefits analogous to empirical benefits. For predicted benefits (Supplementary Fig. S1), a  $2 \times 2$  repeated-measures ANOVA showed a main effect of stimulus construction,  $F(1, 19) = 41.54, p < 0.001, \eta^2 = 0.69$ . Predicted benefits were larger for complex ( $0.046 \pm 0.004$  s) compared to simple stimuli ( $0.032 \pm 0.003$  s). No other effects were significant (all  $F \leq 0.80, p \geq 0.383, \eta^2 \leq 0.04$ ). Hence, this simple analysis based on the unisensory RTs predicts larger benefits for complex compared to simple stimuli (which is here basically driven by the variability rule, see Supplementary Analysis 2).

We tested next if the RSE followed the prediction. For empirical benefits (Supplementary Fig. S1), a  $2 \times 2$  repeated-measures ANOVA showed a main effect of stimulus construction,  $F(1, 19) = 5.87, p < 0.026, \eta^2 = 0.24$ . Empirical benefits were larger for complex ( $0.052 \pm 0.003$  s) compared to simple stimuli ( $0.041 \pm 0.004$  s). No other effects were significant (all  $F \leq 0.41, p \geq 0.528, \eta^2 \leq 0.02$ ). Hence, empirical benefits were consistent with the prediction.

To assess the explanatory power, we correlated predicted and empirical benefits (Fig. 2b). All correlation coefficients were positive, and significant in 3 of 4 conditions (Table S2). One notable observation is that mean



**Figure 2.** Analysing multisensory benefits. (a) Empirical benefits are measured by the area between the cumulative RT distributions in the multisensory (AV) condition and the faster of the unisensory conditions (A, V; Equation (1)). Data from an example participant in the complex-alternating condition. (b) Empirical benefits as a function of benefits predicted by Raab's model (Equation (2)). Each point represents a participant in one of the four conditions. Large symbols represent the group mean.

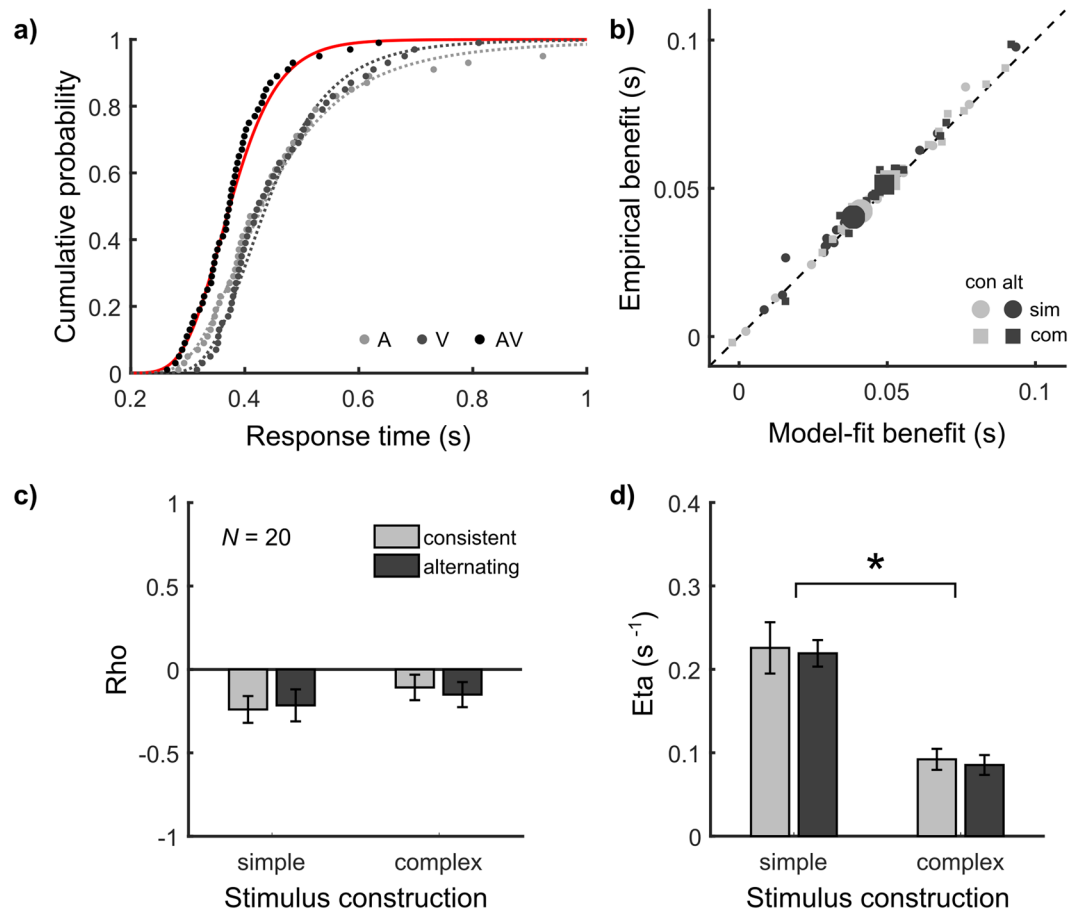


**Figure 3.** Measuring empirical interactions. (a) Violations of Miller's bound are measured by the area between the cumulative RT distributions in the multisensory (AV) and the sum of the unisensory conditions (Equations (4) and (5)). Example data as in Fig. 2a. (b) Violations of Miller's bound across conditions. Mean and SEM of 20 participants.

empirical benefits for each participant over all conditions ( $0.047 \pm 0.003$  s) was larger than mean predicted benefits ( $0.039 \pm 0.003$  s). This was confirmed by a paired-samples *t*-test,  $t(19) = 3.07$ ,  $p = 0.006$ . Thus, while Raab's parameter-free model broadly predicts the RSE, it does not provide a complete account.

**Step 2: Empirical Interactions.** As second step, we focus on two processing interactions that are not covered by Raab's model. Firstly, we measured history effects by analysing RTs as a function of previously presented signals (Equation (3)). Expectedly, a  $2 \times 2$  repeated-measures ANOVA showed a significant intercept of  $0.034$  s ( $\pm 0.004$  s),  $F(1, 19) = 75.62$ ,  $p < 0.001$ ,  $\eta^2 = 0.80$  (Fig. S3). Hence, unisensory RTs following a modality switch were slower compared to a repetition. If history effects arise simply due to the repetition of low-level signal features (e.g. the specific frequency of a tone), introducing alternating features within each modality should reduce history effects compared to consistent features. However, no effects were significant (all  $F \leq 4.13$ ,  $p \geq 0.056$ ,  $\eta^2 \leq 0.179$ ). Hence, while clearly present and thus to be considered by any model of the RSE, the history effect did not change across conditions here (for additional tests, see Supplementary Analysis 3).

Secondly, we measured violations of Miller's bound (Equations (4) and (5); for an illustration, see Fig. 3a). A  $2 \times 2$  repeated-measures ANOVA showed a significant intercept of  $0.008$  s ( $\pm 0.001$  s),  $F(1, 19) = 130.70$ ,  $p < 0.001$ ,  $\eta^2 = 0.87$  (Fig. 3b). Hence, in agreement with most RSE studies using audio-visual signals, Miller's bound was violated. In addition, there was a significant effect of stimulus construction  $F(1, 19) = 4.56$ ,  $p < 0.046$ ,  $\eta^2 = 0.19$ . The violation area for simple stimuli ( $0.009 \pm 0.001$  s) was larger than for complex stimuli ( $0.006 \pm 0.001$  s). No other effects were significant (all  $F \leq 0.05$ ,  $p \geq 0.819$ ,  $\eta^2 < 0.01$ ). Thus, violations are clearly present and therefore must be considered by race models.



**Figure 4.** Model fitting. (a) Best-fitting context variant race model (solid line). The model is constrained by the fits of the LATER model in the unisensory conditions (dashed lines) and has the correlation  $Rho$  and the additional noise  $Eta$  as free parameters. Example data as in Fig. 2a. (b) Empirical benefits as a function of benefits calculated from the best-fitting context variant race model. Each point represents a participant in one of the four conditions. Large symbols represent the group mean. (c,d) Best fitting values of  $Rho$  and  $Eta$  across conditions. Mean and SEM of 20 participants.

**Step 3: Modelling Approach.** As final step, we followed a modelling approach that accounts for the empirical interactions. We used the context variant race model<sup>12,13</sup>, which includes the correlation  $Rho$  and the noise  $Eta$  as free parameters to account for history effects and violations of Miller's bound, respectively. We fitted the model on the level of RT distributions (for an illustration, see Fig. 4a). To assess if the best-fitting model fully explained the RSE, we computed benefits according to the model-fit analogous to empirical benefits (Equation (1)). As with the parameter-free predictions, we correlated model-fit and empirical benefits (Fig. 4b). The model explained the empirical benefits almost perfectly as all correlation coefficients were positive and highly significant (see Supplementary Table S2). Hence, the context variant race model provides a viable explanation for the RSE across conditions.

We next inspected the best-fitting model parameters with respect to the empirical interactions. Regarding the first parameter  $Rho$ , a  $2 \times 2$  repeated-measures ANOVA showed a significant intercept of  $-0.178 (\pm 0.064)$ ,  $F(1, 19) = 7.72$ ,  $p = 0.012$ ,  $\eta^2 = 0.29$  (Fig. 4c). Hence, there was overall a negative correlation between unisensory RTs. There were no further significant effects (all  $F \leq 1.24$ ,  $p \geq 0.279$ ,  $\eta^2 \leq 0.06$ ). This finding is in agreement with the above finding that history effects did not differ across conditions. Regarding the second parameter  $Eta$ , a  $2 \times 2$  repeated-measures ANOVA showed a significant intercept of  $0.156 \text{ s}^{-1} (\pm 0.012 \text{ s}^{-1})$ ,  $F(1, 19) = 176.06$ ,  $p < 0.001$ ,  $\eta^2 = 0.90$  (Fig. 4d). The parameter manifests a violation of the context invariance assumption, which allows for violations of Miller's bound. We neither found a main effect of signal features nor an interaction (all  $F \leq 0.15$ ,  $p \geq 0.708$ ,  $\eta^2 \leq 0.01$ ). Interestingly, there was a main effect of stimulus construction,  $F(1, 19) = 44.62$ ,  $p < 0.001$ ,  $\eta^2 = 0.70$ .  $Eta$  was larger with simple ( $0.222 \pm 0.019 \text{ s}^{-1}$ ) compared to complex stimuli ( $0.089 \pm 0.010 \text{ s}^{-1}$ ). In agreement with the above finding that the violation of Miller's bound is larger for simple compared to complex stimuli, more noise was required to account for the redundant RT distributions with simple compared to complex stimuli according to the model. The analysis of  $Eta$  thus points to a striking dissociation between the processing interactions revealed here (Fig. 4d) and multisensory benefits, which in contrast were larger with complex compared to simple signals (Supplementary Fig. S1).

## Discussion

To mimic the diversity of RSE studies, we used a factorial design to create distinct audio-visual stimulus sets. With our sets, RTs were faster for simple compared to complex stimuli (with a rather substantial RT difference of 0.116 s). The comparative approach outlined in this paper, therefore, can be considered a blueprint to systematically analyse the RSE with distinct stimulus sets or different participant groups, which we think is critically needed to gain complete understanding of the RSE.

As a first step to this approach, we focused on multisensory benefits as given by the size of the RSE. Initially, we made a quantitative prediction of the RSE based on Raab's model<sup>18</sup>. At this point, we would like to highlight that this basic race model is fully constrained by the empirical RT distributions in the unisensory conditions, and does not contain free parameters. The quantitative prediction is simply computed based on probability summation (Equation (2)). Despite its simplicity, the model correctly predicted the increased RSE as observed for complex rather than simple stimuli (Fig. 2) and accounts considerably for the variation across participants (Table S2). To the best of our knowledge, no such straightforward predictions are provided by any published pooling model, which makes it difficult to understand why this alternative model class is favoured in RSE research e.g.<sup>34</sup>. In contrast, we argue that race models are a very promising candidate for a common framework for the diversity of RSE studies. At least, Raab's model provides a simple tool for analysing the RSE when distinct signal sets or different participant groups yield different RTs.

As second and third step, we analysed two empirical interactions not accounted for by Raab's model, and applied consequently the context variant race model, which is an extended race model including two free parameters to cover the interactions<sup>12,13</sup>. The first interaction concerns history effects, which challenges statistical independence (as assumed in Equation (2)). In agreement with previous studies<sup>4,8,13</sup>, we found significant history effects across conditions, which thus must be accounted for by any model of the RSE. In the race model, we dropped the statistical independence assumption and included instead the correlation *Rho* as a free parameter. As history effects are opposite for vision and audition (e.g. after presentation of a visual signal on the previous trial, fast visual and slow auditory RTs are expected), the best-fitting *Rho* value covered the effect by assuming a negative correlation (Fig. 4c).

One simple hypothesis, which we began with, is that history effects may arise due to the repetition of low-level signal features. For example, history effects may arise in auditory trials because tones with identical frequencies repeatedly stimulate the same population of sensory neurons. If this repeated stimulation is causally-linked, the history effect should be reduced when tones with different frequencies stimulate different populations of sensory neurons. Following this simple explanation, we expected to observe a main effect of signal features, whereby alternating signals would reduce history effects compared to consistent signals. However, such changes occurred neither for the measured history effect nor for the corresponding model parameter *Rho*. This suggests rather a basis in higher-level processes. For instance, in the field of task-switching<sup>35</sup>, costs are observed by switching between two tasks in the same modality. History effects may be the result of similar mechanisms (i.e. switching between auditory and visual detection tasks). Alternatively, in the attention literature, RT costs have been observed between targets in different modalities<sup>36</sup>. History effects may also arise from attentional mechanisms, whereby the previous unisensory trial directs attention to one modality at the expense of the other. Future investigations should develop manipulations to target such higher-level processes.

The second interaction concerns violations of Miller's Bound, which challenges the context invariance assumption<sup>19</sup>. In agreement with most audio-visual RSE studies e.g.<sup>4,5,8,9,13</sup>, we found significant violations of Miller's bound across conditions (Fig. 3b), which thus must be accounted for by race models. In the context variant race model<sup>12,13</sup>, we dropped the context invariance assumption and included instead the additional noise *Eta* as a second free parameter. Critically, for violations of Miller's bound we found a main effect of stimulus construction, with increased violations occurring with simple compared to complex stimuli. The best-fitting *Eta* value covered the effect by assuming further increased noise in the evidence accumulation process for simple compared to complex conditions (Fig. 4d).

Previously, we have only speculated about potential sources of this additional noise<sup>19</sup>. Here, however, one aspect of the experimental factor of stimulus construction is that simple stimuli are expected to trigger strong onset transients, which are masked by the background stimulation in complex conditions. As onset transients have been considered relevant for multisensory processing<sup>22,23</sup>, one possibility is that the additional noise interaction is linked to an interference caused by these transients. Although our behavioural data does not allow us to identify exact sources, one interpretation could be that strong neuronal activity linked to sudden signal onsets leaks as noise across sensory modalities and otherwise separate parallel processes.

As a more general outcome of our comparative approach, the systematic analysis of the RSE calls for a clarification of the rather vague term "multisensory integration". On the one hand, the term has been defined operationally as "as a multisensory response (neural or behavioural) that is significantly different from the responses evoked by the modality-specific component stimuli"<sup>37</sup>. Following this definition, we found that the multisensory benefit, and hence integration, was larger with complex compared to simple signals. On the other hand, in research on multisensory RTs, multisensory integration is often said to occur only if Miller's bound is violated (which is even considered "a psychometric benchmark of integrative processing"<sup>38</sup>). Regarding this definition, we found the opposite. Violations of Miller's bound (and consistently the best-fitting *Eta* value) were larger with simple compared to complex signals. Hence, the two variables clearly dissociate and should not be described using the same term or share the same concept.

The vagueness of definitions may have contributed to the lack of a common framework for the RSE across studies. In contrast, our comparative approach, which is based on Raab's original race model<sup>18</sup>, provides conceptual clarity. The basic model architecture, which consists of parallel decision units coupled by a logic OR operator, is convincing as it perfectly matches the task demands as imposed by the redundant signals paradigm<sup>19</sup>. Moreover, it provides a strong, parameter-free predictor of multisensory benefits<sup>12</sup>. As such, the combination rule

for redundantly defined signals (or the multisensory integration mechanism, under an alternative conception) is clearly defined. Beyond this basic mechanism, at least two very specific processing interactions emerge and can be precisely quantified by explicit parameters in a model-based approach.

The approach is general enough to be applied to various experimental setups and research questions. This includes not only experiments comparing different sensory modalities, such as audio-tactile<sup>4</sup> or visual-tactile stimulation<sup>5</sup>, but also studies that investigate unisensory RSE<sup>39,40</sup> or processing differences between uni- and multisensory versions of the redundant signals paradigms<sup>41</sup>. The approach is also suitable to shed light on processing differences between participant groups. For instance, studies on ageing typically show larger multisensory benefits for older compared to younger participants<sup>9</sup>. Such findings may be directly understood if older participants' RT distributions are more variable, as unisensory RT variability is the driving force behind multisensory benefits according to race models<sup>12</sup>. Finally, since the framework has successfully been used to study dyslexia<sup>42</sup>, it could of course also be applied to study other clinical populations, such as those with Schizophrenia<sup>43</sup> or Parkinson's Disease<sup>44</sup>. In the aggregate, we believe that our comparative approach provides both a general account and the precision needed to target and understand processing differences across diverse studies, which are key features for a common explicatory framework of the RSE.

Beyond the RSE, we believe this work contributes to a broader effort to achieve clarity of definitions in multisensory research<sup>37</sup>. In the wider literature, there are many instances of behavioural paradigms and associated measures assumed to reflect 'multisensory integration' equivalently. As shown here within one paradigm, however, this assumption can be incorrect. We hope therefore to demonstrate for multisensory researchers the importance of defining measures in detail and verifying their proposed relationships experimentally. Failing to do so may result in similar confusion when comparing results between paradigms.

## Data Availability

The research data supporting this publication can be accessed at <https://doi.org/10.17630/c8cbd7b7-e2b3-4e62-bb3d-66fce081ff59><sup>45</sup>. The data analysis and model-fitting tools are also freely available online via the RSE-box<sup>46</sup>: <https://github.com/tomotto/RSE-box>.

## References

- Hershenson, M. Reaction-Time as a Measure of Intersensory Facilitation. *J Exp Psychol* **63**, 289–&, <https://doi.org/10.1037/h0039516> (1962).
- Kinchla, R. A. Detecting Target Elements in Multielement Arrays - Confusability Model. *Percept Psychophys* **15**, 149–158, <https://doi.org/10.3758/Bf03205843> (1974).
- Todd, J. W. Reaction to multiple stimuli. *Arch. Psychol.*, 1–65 (1912).
- Gondan, M., Lange, K., Rosler, F. & Roder, B. The redundant target effect is affected by modality switch costs. *Psychon B Rev* **11**, 307–313, <https://doi.org/10.3758/Bf03196575> (2004).
- Forster, B., Cavina-Pratesi, C., Aglioti, S. M. & Berlucchi, G. Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp Brain Res* **143**, 480–487, <https://doi.org/10.1007/s00221-002-1017-9> (2002).
- Veldhuizen, M. G., Shepard, T. G., Wang, M. F. & Marks, L. E. Coactivation of Gustatory and Olfactory Signals in Flavor Perception. *Chem Senses* **35**, 121–133, <https://doi.org/10.1093/chemse/bjp089> (2010).
- Bolognini, N., Olgiati, E., Rossetti, A. & Maravita, A. Enhancing multisensory spatial orienting by brain polarization of the parietal cortex. *Eur J Neurosci* **31**, 1800–1806, <https://doi.org/10.1111/j.1460-9568.2010.07211.x> (2010).
- Miller, J. Divided Attention - Evidence for Co-Activation with Redundant Signals. *Cognitive Psychol* **14**, 247–279, [https://doi.org/10.1016/0010-0285\(82\)90010-X](https://doi.org/10.1016/0010-0285(82)90010-X) (1982).
- Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E. & Laurienti, P. J. Age-related multisensory enhancement in a simple audiovisual detection task. *Neuroreport* **18**, 1077–1081, <https://doi.org/10.1097/WNR.0b013e3281e72ae7> (2007).
- Molholm, S., Ritter, W., Javitt, D. C. & Foxe, J. J. Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cereb Cortex* **14**, 452–465, <https://doi.org/10.1093/cercor/bhh007> (2004).
- Collignon, O. *et al.* Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia* **48**, 220–225, <https://doi.org/10.1016/j.neuropsychologia.2009.09.007> (2010).
- Otto, T. U., Dassy, B. & Mamassian, P. Principles of Multisensory Behavior. *J Neurosci* **33**, 7463–7474, <https://doi.org/10.1523/Jneurosci.4678-12.2013> (2013).
- Otto, T. U. & Mamassian, P. Noise and Correlations in Parallel Perceptual Decision Making. *Curr Biol* **22**, 1391–1396, <https://doi.org/10.1016/j.cub.2012.05.031> (2012).
- Gondan, M. & Minakata, K. A tutorial on testing the race model inequality. *Atten Percept Psycho* **78**, 723–735, <https://doi.org/10.3758/s13414-015-1018-y> (2016).
- Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu Rev Neurosci* **30**, 535–574, <https://doi.org/10.1146/annurev.neuro.29.051605.113038> (2007).
- Bogacz, R. Optimal decision-making theories: linking neurobiology with behaviour. *Trends Cogn Sci* **11**, 118–125, <https://doi.org/10.1016/j.tics.2006.12.006> (2007).
- Shadlen, M. N. & Kiani, R. Decision Making as a Window on Cognition. *Neuron* **80**, 791–806, <https://doi.org/10.1016/j.neuron.2013.10.047> (2013).
- Raab, D. H. Statistical Facilitation of Simple ReactionTimes. *T New York Acad Sci* **24**, 574–590, <https://doi.org/10.1111/j.2164-0947.1962.tb01433.x> (1962).
- Otto, T. U. & Mamassian, P. Multisensory Decisions: the Test of a Race Model, Its Logic, and Power. *Multisens Res* **30**, 1–24, <https://doi.org/10.1163/22134808-00002541> (2017).
- Ashby, F. G. & Townsend, J. T. Varieties of Perceptual Independence. *Psychol Rev* **93**, 154–179, <https://doi.org/10.1037/0033-295x.93.2.154> (1986).
- Luce, R. D. *Response Times: Their Role in Inferring Elementary Mental Organization*. (Oxford University Press 1986).
- Jaekl, P., Perez-Bellido, A. & Soto-Faraco, S. On the 'visual' in 'Audio-visual integration': a hypothesis concerning visual pathways. *Exp Brain Res* **232**, 1631–1638, <https://doi.org/10.1007/s00221-014-3927-8> (2014).
- Werner, S. & Noppeney, U. The contributions of transient and sustained response codes to audiovisual integration. *Cereb Cortex* **21**, 920–931, <https://doi.org/10.1093/cercor/bhq161> (2011).
- Brainard, D. H. The psychophysics toolbox. *Spatial Vision* **10**, 433–436, <https://doi.org/10.1163/156856897x00357> (1997).
- Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? *Perception* **36**, 14–14 (2007).
- Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* **10**, 437–442, <https://doi.org/10.1163/156856897x00366> (1997).

27. Li, X. R., Liang, Z., Kleiner, M. & Lu, Z. L. RTbox: A device for highly accurate response time measurements. *Behav Res Methods* **42**, 212–225, <https://doi.org/10.3758/Brm.42.1.212> (2010).
28. Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J Exp Soc Psychol* **49**, 764–766, <https://doi.org/10.1016/j.jesp.2013.03.013> (2013).
29. Ulrich, R., Miller, J. & Schroter, H. Testing the race model inequality: An algorithm and computer programs. *Behav Res Methods* **39**, 291–302, <https://doi.org/10.3758/Bf03193160> (2007).
30. Colonius, H. & Diederich, A. The race model inequality: interpreting a geometric measure of the amount of violation. *Psychol Rev* **113**, 148–154, <https://doi.org/10.1037/0033-295X.113.1.148> (2006).
31. Carpenter, R. H. & Williams, M. L. Neural computation of log likelihood in control of saccadic eye movements. *Nature* **377**, 59–62, <https://doi.org/10.1038/377059a0> (1995).
32. Noorani, I. & Carpenter, R. H. S. The LATER model of reaction time and decision. *Neurosci Biobehav Rev* **64**, 229–251, <https://doi.org/10.1016/j.neubiorev.2016.02.018> (2016).
33. Nadarajah, S. & Kotz, S. Exact distribution of the max/min of two Gaussian random variables. *Ieee T Vlsi Syst* **16**, 210–212, <https://doi.org/10.1109/Tvlsi.2007.912191> (2008).
34. Chandrasekaran, C. Computational principles and models of multisensory integration. *Curr Opin Neurobiol* **43**, 25–34, <https://doi.org/10.1016/j.conb.2016.11.002> (2017).
35. Monsell, S. Task switching. *Trends Cogn Sci* **7**, 134–140, [https://doi.org/10.1016/s1364-6613\(03\)00028-7](https://doi.org/10.1016/s1364-6613(03)00028-7) (2003).
36. Spence, C., Nicholls, M. E. R. & Driver, J. The cost of expecting events in the wrong sensory modality. *Percept Psychophys* **63**, 330–336, <https://doi.org/10.3758/Bf03194473> (2001).
37. Stein, B. E. *et al.* Semantic confusion regarding the development of multisensory integration: a practical solution. *Eur J Neurosci* **31**, 1713–1720, <https://doi.org/10.1111/j.1460-9568.2010.07206.x> (2010).
38. Martuzzi, R. *et al.* Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb Cortex* **17**, 1672–1679, <https://doi.org/10.1093/cercor/bhl077> (2007).
39. Corballis, M. C. Hemispheric interactions in simple reaction time. *Neuropsychologia* **40**, 423–434, [https://doi.org/10.1016/S0028-3932\(01\)00097-5](https://doi.org/10.1016/S0028-3932(01)00097-5) (2002).
40. Schroter, H., Ulrich, R. & Miller, J. Effects of redundant auditory stimuli on reaction time. *Psychon B Rev* **14**, 39–44, <https://doi.org/10.3758/Bf03194025> (2007).
41. Girard, S., Pelland, M., Lepore, F. & Collignon, O. Impact of the spatial congruence of redundant targets on within-modal and cross-modal integration. *Exp Brain Res* **224**, 275–285, <https://doi.org/10.1007/s00221-012-3308-0> (2013).
42. Harrar, V. *et al.* Multisensory integration and attention in developmental dyslexia. *Curr Biol* **24**, 531–535, <https://doi.org/10.1016/j.cub.2014.01.029> (2014).
43. Williams, L. E., Light, G. A., Braff, D. L. & Ramachandran, V. S. Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia* **48**, 3128–3136, <https://doi.org/10.1016/j.neuropsychologia.2010.06.028> (2010).
44. Plat, F. M., Praamstra, P. & Horstink, M. W. Redundant-signals effects on reaction time, response force, and movement-related potentials in Parkinson's disease. *Exp Brain Res* **130**, 533–539, <https://doi.org/10.1007/s002219900276> (2000).
45. Innes, B. R. & Otto, T. U. Data underpinning - A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent, <https://doi.org/10.17630/c8cbd7b7-e2b3-4e62-bb3d-66fce081ff59>.
46. Otto, T. U. An analysis and modelling toolbox to study multisensory response times. Abstract from 19th International Multisensory Research Forum (IMRF), online at [http://imrf.info/wp\\_imrf/wp-content/uploads/2018/06/IMRF2018\\_Abtracts.pdf](http://imrf.info/wp_imrf/wp-content/uploads/2018/06/IMRF2018_Abtracts.pdf) (2018).

## Acknowledgements

We would like to thank Akira O'Connor, Cleo Pike, and Yue (Serena) Liu for their feedback on an earlier draft of the manuscript. This work was supported by the *Biotechnology and Biological Sciences Research Council* (BBSRC, grant number: BB/J01446X/1).

## Author Contributions

Both authors designed the experiment. B.R.I. implemented the research, collected and analysed the data, and wrote the initial draft. T.U.O. supervised the research. Both authors wrote the final article.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-39924-6>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019