Distributional constraints on cognitive architecture: A statistical evaluation

Inauguraldissertation

zur Erlangung des Doktorgrades der Philosophie

an der Ludwig-Maximilians-Universität München

vorgelegt von

Emil Ratko-Dehnert (geb. Ratko)

aus München

2013

Erstgutachter:	Dr. habil. Michael Zehetleitner
Zweitgutachter:	Prof. Dr. Hermann Müller
Drittgutachter:	Prof. Dr. Thomas Stoffer
Datum der mündlichen Prüfung:	24. Juni 2013

To Sonja and Tim.

Acknowledgements

The present thesis would not have seen the light of the day, if it were not for the help and support of various colleagues and friends in my life.

First and foremost, I would like to thank Dr. Michael Zehetleitner for his commitment, guidance and patience over the last almost four years. He has become more than a supervisor and mentor to me. He has accompanied me during my transition from formal mathematics to the field of empirical psychological research, while always taking into account my background. The intellectual debates on topics exceeding the boundaries of my scientific investigations have deepened my grasp of life and left me with a feeling of privilege and humility.

Also, I would like to express my deep gratitude to Prof. Hermann Müller for accepting an “immigrant” of the field of Experimental psychology to his working unit. Interdisciplinary research, while often demanded, is rarely fostered and seldom supported beyond lip service. This clearly is not the case with Hermann, as he has established a colorful, bright and invigorating scientific unit, with diverse research foci resting on miscellaneous professional backgrounds. His leadership and his vision for the big picture were always both intellectually challenging and inspiring.

I am greatly thankful to my former professor for mathematical logics, Prof. Dieter Donder, for spontaneously agreeing to join my disputation. Also, I would like to thank my brother, Christian Ratko, for proof-reading my thesis.

The people that probably have most constituted this episode in my life are the colleagues and friends I have got to know over the past few years. It is hard to limit oneself to a finite number of names. Nevertheless I want to explicitly thank my dear colleagues, Isabel Koch, Daniel Reutter, Elisabeth Hendrich, Andreas Hetmanek, Saurabh Dhawan and Thomas Fink for the countless discussions and debates (and sometimes laments) both concerning scientific and personal matters. As much as the scientific investigations have shaped my thinking, the interaction with these fine men and women has left a lasting impression on me.

Finally and most emphatically, I would like to thank my beloved wife Sonja and my precious son Tim, for their support, patience and understanding while always grounding me. I cannot imagine a world without them anymore and I am curious of the journey that still awaits us.

TABLE OF CONTENTS

Abstract

Mental chronometry is a classical paradigm in cognitive psychology that uses response time and accuracy data in perceptual-motor tasks to elucidate the architecture and mechanisms of the underlying cognitive processes of human decisions. The redundant signals paradigm investigates the response behavior in Experimental tasks, where an integration of signals is required for a successful performance. The common finding is that responses are speeded for the redundant signals condition compared to single signals conditions. On a mean level, this redundant signals effect can be accounted for by several cognitive architectures, exhibiting considerable model mimicry.

Jeff Miller formalized the maximum speed-up explainable by separate activations or race models in form of a distributional bound – the race model inequality. Whenever data violates this bound, it excludes race models as a viable account for the redundant signals effect. The common alternative is a coactivation account, where the signals integrate at some stage in the processing.

Coactivation models have mostly been inferred on and rarely explicated though. Where coactivation is explicitly modeled, it is assumed to have a decisional locus. However, in the literature there are indications that coactivation might have at least a partial locus (if not entirely) in the nondecisional or motor stage. There are no studies that have tried to compare the fit of these coactivation variants to empirical data to test different effect generating loci.

Ever since its formulation, the race model inequality has been used as a test to infer the cognitive architecture for observers' performance in redundant signals Experiments. Subsequent theoretical and empirical analyses of this RMI test revealed several challenges. On the one hand, it is considered to be a conservative test, as it compares data to the maximum speed-up possible by a race model account. Moreover, simulation studies could show that the base time component can further reduce the power of the test, as violations are filtered out when this component has a high variance.

On the other hand, another simulation study revealed that the common practice of RMI test can introduce an estimation bias, that effectively facilitates violations and increases the type I error of the test. Also, as the RMI bound is usually tested at multiple points of the same data, an inflation of type I errors can reach a substantial amount. Due to the lack of overlap in scope and the usage of atheoretic, descriptive reaction time models, the degree to which these results can be generalized is limited. State-of-the-art models of decision making provide a means to overcome these limitations and implement both race and coactivation models in order to perform large scale simulation studies.

By applying a state-of-the-art model of decision making (scilicet the Ratcliff diffusion model) to the investigation of the redundant signals effect, the present study addresses research questions at different levels. On a conceptual level, it raises the

question, at what stage coactivation occurs – at a decisional, a nondecisional or a combined decisional and nondecisional processing stage and to what extend? To that end, two bimodal detection tasks have been conducted. As the reaction time data exhibits violations of the RMI at multiple time points, it provides the basis for a comparative fitting analysis of coactivation model variants, representing different loci of the effect.

On a test theoretic level, the present study integrates and extends the scopes of previous studies within a coherent simulation framework. The effect of experimental and statistical parameters on the performance of the RMI test (in terms of type I errors, power rates and biases) is analyzed via Monte Carlo simulations. Specifically, the simulations treated the following questions: (i) what is the power of the RMI test, (ii) is there an estimation bias for coactivated data as well and if so, in what direction, (iii) what is the effect of a highly varying base time component on the estimation bias, type I errors and power rates, (iv) and are the results of previous simulation studies (at least qualitatively) replicable, when current models of decision making are used for the reaction time generation.

For this purpose, the Ratcliff diffusion model was used to implement race models with controllable amount of correlation and coactivation models with varying integration strength, and independently specifying the base time component.

The results of the fitting suggest that for the two bimodal detection tasks, coactivation has a shared decisional and nondecisional locus. For the focused attention experiment the decisional part prevails, whereas in the divided attention task the motor component is dominating the redundant signals effect. The simulation study could reaffirm the conservativeness of the RMI test as latent coactivation is frequently missed. An estimation bias was found also for coactivated data however, both biases become negligible once more than 10 samples per condition are taken to estimate the respective distribution functions. A highly varying base time component reduces both the type I errors and the power of the test, while not affecting the estimation biases.

The outcome of the present study has theoretical and practical implications for the investigations of decisions in a multisignal context. Theoretically, it contributes to the locus question of coactivation and offers evidence for a combined decisional and nondecisional coactivation account. On a practical level, the modular simulation approach developed in the present study enables researchers to further investigate the RMI test within a coherent and theoretically grounded framework. It effectively provides a means to optimally set up the RMI test and thus helps to solidify and substantiate its outcomes. On a conceptual level the present study advocates the application of current formal models of decision making to the mental chronometry paradigm and develops future research questions in the field of the redundant signals paradigm.

1 INTRODUCTION

1.1 Decision Making and Evidence Integration

The human perceptual system is composed of highly specialized sensory systems (e.g., vision, audition, olfaction, tactition, and gustation). In order to adequately respond to the demands of a dynamically changing and potentially threatening environment, the organism has to perform countless decisions based on the inputs of these systems, weighing evidence for and against different alternative behaviors. Although these decisions are embedded in specific contexts and employ different sources of evidence, they share common elements such as the accumulation of uncertain sensory evidence and the commitment to one alternative over another, once an internal deliberation process has finished. Investigations of simple decisions studied in the laboratory have revealed general principles that apply to the human system across these various domains and tasks. Current models of decision making have provided a mathematical formalism that enables the exploration and modeling of the human decision making on various levels of analysis (Gold & Shadlen, 2007). This approach allows for a broad applicability, as all the components of the decision process are directly mappable to experimentally controllable variables.

Whenever decisions necessitate the inputs of more than one signal or system (which, in ecologically valid settings marks the normal case), two central questions arise: the question of *integration mechanism* and the question of *cognitive architecture*. Signal stemming from e.g. different modalities need to be integrated, both temporally and spatially, to successfully form a decision that sufficiently represents the state of the environment the organism is operating in. In the visual domain for example, the visual search paradigm has been heavily employed to illuminate both the integration mechanisms and the potential cognitive architecture of categorical decision (detections, identifications, discriminations). A large class of models on visual attention assumes that attention itself is guided by a salience map. This map integrates inputs from feature maps to compute an overall representation of the visual field. It encodes for the local conspicuity in the visual scene, and controls where the focus of attention is currently deployed (Itti & Koch, 2000; Koch & Ullman, 1985; Wolfe, 1994). A core assumption is that maps from different dimensions are summed or integrated. That is, one location on the salience map should have higher activation if at the corresponding location on orientation and color than only on one of them. The activity of the master map at a given location in 2-D space is

$$M(x, y) = \sum_{i \in I} D_i(x, y) \quad (1)$$

where *M(x,y)* is the sum of the strength of feature contrast *D* in dimension *i* at location *(x,y)* (Zehetleitner, Krummenacher, & Müller, 2009b). Figure 1.1 illustrates the activity salience map representation of a singleton (pop-out) target amongst homogeneous distractors.

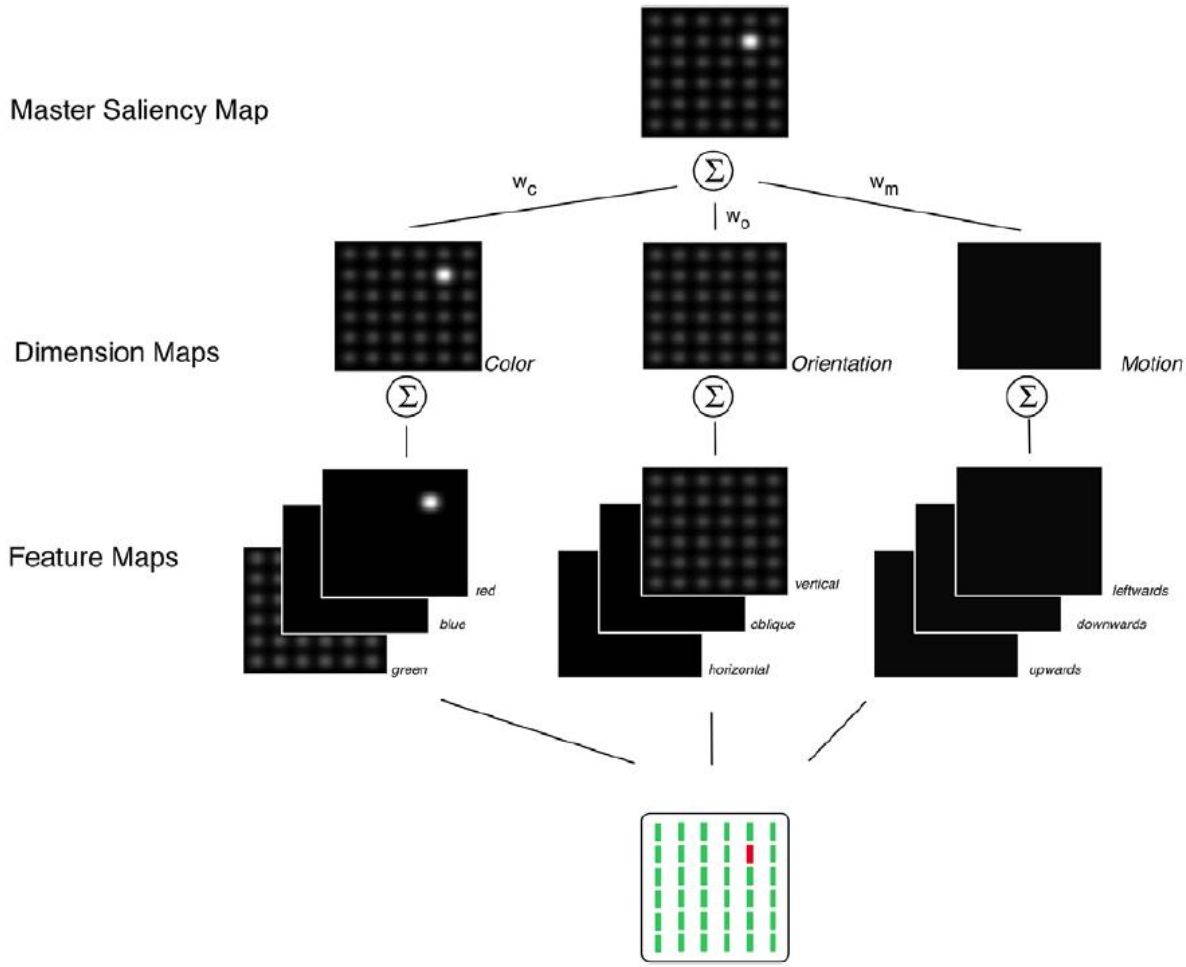

Figure 1.1 Example saliency summation model. The display is first analyzed by dimension specific feature analyzers (for color, orientation, and motion direction). The activity of each map is topographically organized; black areas represent no activity, whereas white areas represent strongest activity. The activity at each location is assumed to be summed up across the feature and dimension maps consequently. The summation is modulated by dimension-specific weights w_c, w_o, w_m. (Figure by Zehetleitner, Müller, & Krummenacher, 2008a , reprinted with permission)

1.2 Tests on Cognitive Architecture

In spite of the allure of its straight-forward design and operating mode, the neural underpinnings, locus and integration mechanisms of salience map models are still much debated questions. One paradigm which directly supports the idea of salience summation is the *redundant signals paradigm* (RSP). There, observers have to search for targets under distractors, which pop-out due to one or more differing features. Figure 1.2 shows display of a redundant target visual pop-out task. Panels A and B represent single target conditions (contrast and orientation pop-out), while panel C represents the combination of the single targets – a redundant target. The common finding is that redundantly defined targets (panel C), which differ in more than one dimension (for example contrast and orientation) on average lead to faster response times than singly defined targets (e.g. orientation or contrast only, panels A and B).

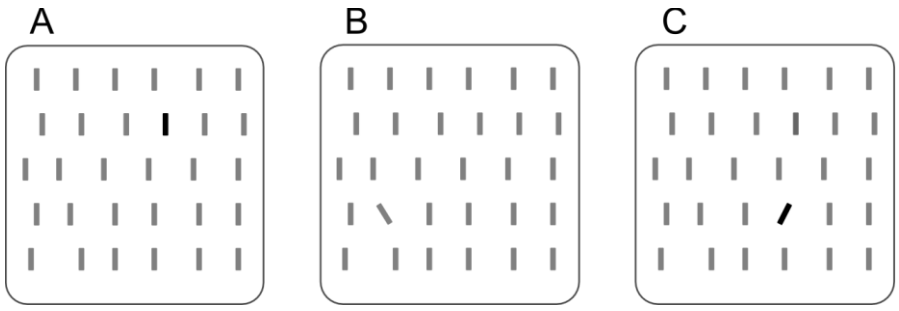

Figure 1.2 Redundant pop-out search displays. Panels A and B show single signals trials with only one dimension at variance with the surrounding distractors (brightness and orientation for panels A and B respectively). Panel C shows a redundant signals trial, where the target shares both dimensions of the single signals trials. On average, response times for redundant signals trials are faster than any single signals trial, termed the *redundant signals effect*.

The redundant signals paradigm taps into the ubiquitous question of serial versus parallel processing (Townsend & Wenger, 2004). It has been investigated with help of mental chronometry and several mechanisms and models have been proposed to account for the speed-up in reaction times. On a mean level, the redundant signals effect can be accounted for by statistical facilitation or a race model (Raab, 1962). This model assumes that features of the redundant target are processed independently and in parallel. The processing time of each feature channel is conceived as a random variable and as such exhibits varying

finishing times from trial to trial. When two channels race against each other, and the winner of the race determines the finishing time, this, on average, results in faster response times than one feature channel alone (see Figure 1.3).

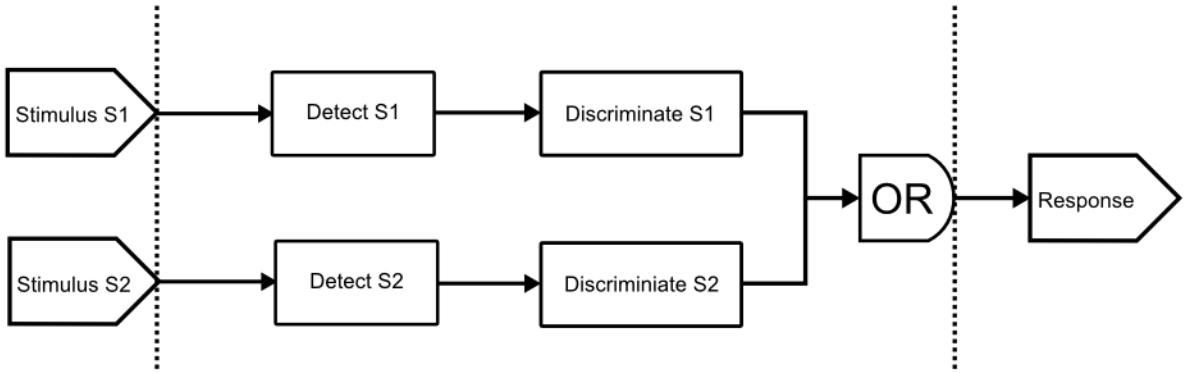

Figure 1.3 Illustration of a race model. Stimuli S1 and S2 are analyzed in separate and independent channels. The first channel to reach a critical activity determines the response time as it triggers the response behavior. The redundant signals effect in this model is a statistical facilitation: the response time variances of both channels are expedited by this minimum stopping rule.

Jeff Miller formalized a distributional bound for the entire class of race models – the *race model inequality* (1982):

$$F_{S_1S_2}(t) \leq F_{S_1}(t) + F_{S_2}(t), \qquad \forall t \tag{2}$$

where *F_* denotes the cumulative distribution function of the single signal conditions (S_1 and S_2) and the redundant signal condition (S_1S_2). It quantifies the maximum speed-up of response times accountable by a race model architecture. Whenever data violates this bound for any time point *t*, the entire class of race models must be eliminated as an architectural candidate for the redundant signals effect.

The alternative to a race model is a coactivation model: here it is assumed that the two feature signals integrate or coactivate into a common decision unit before reaching the response threshold. As the activity of two signals builds up faster, the threshold is reached earlier (see Figure 1.4).

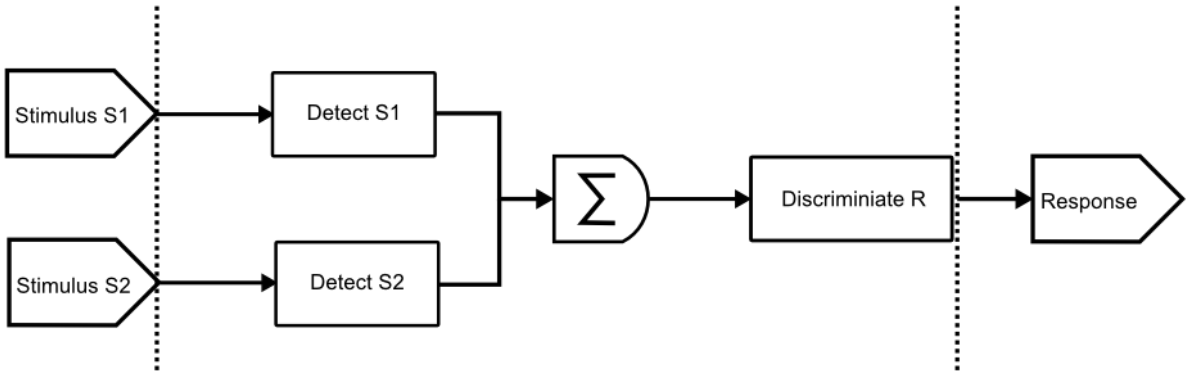

Figure 1.4 Illustration of a coactivation model. In a coactivation model, the processing of stimuli starts independently and then the activity is pooled across channels. This results in a higher activation than for one channel alone and shortens the response time.

Whenever the race model inequality is violated, this is seen as direct evidence against the class of race models. Coactivation accounts as the alternative are only inferred on as there is no direct test for them. In the literature, only few explicit coactivation models exist, all of which assume a coactivation at the decisional stage. On a conceptual level, the speed-up in the redundant signals paradigm must not be restricted to the decisional stage. There are studies that try to establish the RSE as a post-selective (i.e. nondecisional) effect. However, no comparison of competing coactivation accounts (with respect to the stage of coactivation) has been investigated in the literature so far. The RMI test itself is by design not able to discriminate between different coactivation accounts and can only rule out the null hypothesis (that is, a race model architecture).

From a test theoretic standpoint, the RMI test (as every other statistical classifier) can produce two types of errors: type I errors occur when race models are falsely rejected and type II errors occur when the test fails to reject race models although the data was generated by a coactivated mechanism. Because the RMI test is a non-parametric and compound test in nature, its properties have been investigated by means of numerical simulations. These simulation studies identified and confirmed several shortcomings of the test. First, one of the steps in the test' algorithm can introduce a bias towards violations. Second, because the test employs multiple *t*-tests on the same data, type I error or alpha accumulation can reach substantial values for specific experimental settings. Third, in presence of a moderately to highly varying base time the power of the RMI test is drastically reduced. But even without highly varying base times the test is deemed to be rather conservative as it pits the most extreme case of race models against experimental data of interest.

However, those studies have not employed the most current models of decision making (in particular sequential sampling models) for their response time generation. Instead the data was generated by distribution functions that share the general shape of response time distributions. They furthermore do not feature coactivated data or a plausible mechanism to generate it. And lastly, coactivation model variants (with coactivation occurring at different stages of the decision process) have not been tested against each other with implementations of formal models of decision making.

The race model inequality test has been employed in a multitude of experimental tasks and settings involving various modalities and response effectors. It is a pivotal instrument to infer the cognitive architecture of simple decisions. The validity and reliability of this test heavily relies on the fundamental understanding of its statistical properties.

The present study aims at applying state-of-the-art models of decision making to a classical experimental paradigm of mental chronometry – the redundant signals paradigm. It aims at improving the understanding of the race model inequality test and elucidating the mechanisms of coactivation within a formalized and computationally intensive framework.

1.3 Scope of the present Study

The present thesis pursues objectives at different levels. On a conceptual level, the thesis adds to the debate on the locus and mechanism of evidence integration in the redundant signals paradigm. Using a diffusion model analysis, the fitting performance of different coactivation model variants (reflecting different loci of integration) to data of two bimodal detection experiments is compared.

On a methodological and statistical level, the RMI test is evaluated by means of large scale simulations. At the core of these simulations, well established and state-of-the-art models of decision making are implemented as generators of reaction times. By that, an integration and extension of the previous numerical investigations of the RMI test is possible. Specifically, the interplay of different experimental and statistical variables is illuminated and provides a conclusive picture of the tests' statistical profile.

Overall the thesis advances the understanding of the RMI test and provides a principled way to improve its performance. Furthermore it investigates the locus of coactivation for bimodal simple reaction time tasks.

1.4 Summary of Findings

1.4.1 Chapter 2

Chapter 2 provides a literature review on the fundamental theoretical aspects of this thesis. It will cover the paradigm of mental chronometry as a means to investigate the cognitive architecture of decisions in the redundant signals paradigm. This is followed by an account of explanatory models, concerning architecture and locus of the effect. The race model inequality test is then formally derived and research on the tests statistical profile is reviewed. After that, current models of decision making are presented and applied to the redundant signals paradigms. The chapter is concluded by a motivation and formulation of the research questions of the present study.

1.4.2 Chapter 3

In chapter 3, the fitting performance of different coactivation accounts to empirical data is compared. The reaction times and accuracy data of two bimodal detection tasks were collected. In Experiment 1, a simple reaction time task with auditory and visual stimuli was performed and in Experiment 2, a two-alternative forced choice task with the identical bimodal stimuli was conducted. The quantile proportions of the response times for each condition (auditory, visual and audiovisual condition) was fit to the quantile proportions of coactivation models of different design: (i) a decisional model, where only the parameter for the rate of evidence accumulation could vary across conditions, (ii) a nondecisional or motor model, where only the parameters of the base time differed across conditions, and (iii) a combined model, where both decisional and nondecisional parameters were fit individually for the experimental conditions.

The empirical data displayed a large amount of coactivation as the RMI test was violated across nine and ten probability points for experiment 1 and 2 respectively. The numerical fitting of the data revealed, that the combined model possessed the best goodness-of-fit for both experiments. Based on these results, the conclusion can be made, that coactivation is not restricted to the decisional stage of the response times. Interestingly, for the data of Experiment 2 to larger proportion of the RSE can be explained by a coactivation in the motor regime. An evaluation of the fitting procedure itself revealed a differential performance across the models, as there was a high amount of failed fitting attempts for the nondecisional and combined model compared to the decisional model. This disparity is further treated in the discussion section.

1.4.3 Chapter 4

In chapter 4, a large scale Monte Carlo simulation study is performed to evaluate the statistical properties of the RMI test. The performance of the RMI test, in terms of type I errors, power rate and estimation bias is analyzed for several experimental and statistical manipulations. Experimental data is synthesized for race models and coactivation model using Ratcliff diffusion models as generating mechanisms. This marks an integration and validation of existing simulation studies in a coherent simulation framework, using state-of-the-art mechanistic model of decision making (here, the Ratcliff diffusion model). The amount of correlation between the racers is controlled by a parameter to investigate the type I error proneness in dependence of interchannel correlation. For the coactivation model, a weighted drift rate summation mechanism is utilized to produce both latently and manifestly coactivated data. By using the Ratcliff diffusion model, the decisional and nondecisional components of the decision times could be varied independently.

The simulation study looked into (i) quantitative estimates on the power of the RMI test, (ii) an investigation of the estimation bias both for race and coactivation models, (iii) the effect of base time variance on type I errors, power rates and estimation bias, and finally (iv) the validation of the previous simulation results by applying an explicit mechanistic model of decision making for the synthetization of all response time and accuracy data.

Overall, the simulation framework showed a strong degree of plausibility as all implemented manipulations affected the appropriate measures. The analysis of the power rates confirms that the RMI test is rather underpowered: only for high amounts of coactivation strength, low base time variance and sufficiently large sample and subject sizes an acceptable power rate of approx. 80% can be reached. The type I errors display a severe amount accumulation, but only in case of highly negative interchannel correlations. For moderately negatively correlated and uncorrelated conditions, type I error accumulation in fact does not occur. The estimation bias is negligible, once more than 40 samples are drawn per condition. Large base time variance reduces both type I errors and power rates, and thus should be taken into account by experimenters when choosing their response effectors and behavior.

The simulation framework and its results improve the understanding and use of the RMI test as a tool to infer on the cognitive architecture responsible for integrated decisions. The framework can effectively help researchers and experimenters to validate their results by providing quantitative estimates of type I errors and power rates for specific experimental and statistical parameter choices and is of high practical relevance.

1.4.4 General Discussion

In the last chapter, a critical evaluation of all the results of the present study is given. Theoretical and methodological limitations are detailed and discussed. Aspects like the underspecification of the RSE as a phenomenon, the critical relation of the context independence assumption for the RMI testing and potential adaptations of both the diffusion model and the simulation framework are treated.

Implications of the present study and its results are outlined and future research questions are formulated. These cover the expansion of the present scope and approach, such as combining this approach with other decision related measures, an application to the integration of more than two evidence sources and the modeling of decision chains as a logical next step. The impact and significance of the present study are emphasized and conclude the thesis.

2 TESTING COGNITIVE ARCHITECTURES

The aim of this chapter is to establish a profound theoretical groundwork for the ensuing empirical and computational investigations on the redundant signals effect and the race model inequality test. It consists of five parts: In the first part (subchapter 2.1), a concise literature review on the application of mental chronometry to the question of cognitive architecture is provided. The second part (subchapter 2.2.) introduces the redundant signals paradigm and provides a catalogue of explanatory models for the redundant signals effect. These are contrasted with respect to the proposed integration mechanism and the locus of integration. In the third part (subchapter 2.3), a formal account of race models is given together with the derivation of the race model inequality test on a level of means and distributions. This is relevant for the identification of the specific issues and shortcomings of RMI testing. The fourth part (subchapter 2.4) introduces the most current models of decision making, namely sequential sampling models. One prominent and well established exemplar - the Ratcliff diffusion model - will be detailed and applied to the redundant signals paradigm. This application enables the principled investigation of all the problems brought forward in the preceding subchapters. On a broader level it constitutes a theoretical and empirical advancement as it combines the logic and rationale of mental chronometry with the most recent models of decision making. It is also the core of the empirical (chapter 3) and simulation parts (chapter 4) of the present study. In the final part of this chapter (subchapter 2.5), the research questions of the present study are motivated on the basis of the theoretical groundwork and the worked out issues of the RMI test so far. The significance of the present study is carved out and concludes the chapter.

The content of chapter 2 is essential for understanding the ensuing theoretical and empirical work of this study. Where necessary, explicit references to parts of this chapter will be given in the empirical investigations that follow.

2.1 The Question of Cognitive Architecture

Over the past several decades, there has been remarkable progress in cognitive psychology by linking formal mathematical (rather than solely qualitative) models to experimental designs. This approach allowed for discriminating between competing theoretical accounts of mental processes with help of elaborated quantitative analyses.

One of the pivotal and most central questions in the investigation of psychological mechanisms is what architectures have afforded them. In which way to humans perceive,

remember, or cognize items in their mind - simultaneously (that is, in parallel) or sequentially (that is, serially)? The serial-parallel distinction has been treated on different levels of empirical and theoretical investigations.

2.1.1 Neurophysiology

On a neurophysiological level, Livingstone and Hubel could show that visual processing works in separate and parallel channels from the retinal level onwards (Hubel & Livingstone, 1985, 1987; Livingstone & Hubel, 1984, 1987, 1988). The separation of retinal cells specialized for high temporal− and, respectively, high spatial−frequency information is maintained in the laminar network of the lateral geniculate nucleus (LGN) and further in cortical areas. One pathway specialized for extracting motion information runs through distinct layers of the LGN, V1, and V2 on to the medial temporal area (MT), whereas the other pathway coding color and form information runs though distinct layers and sections (blobs, inter−blobs of V1, and thin−, inter−stripes of V2) on to V4 and higher−level areas in infero−temporal cortex. It is now commonly accepted that the neuronal visual processing stream operates in functionally separate and parallel pathways or channels. The interplay and exact nature of these pathways are still under debate (for a review, see Sincich & Horton, 2005). Similar to the visual domain, other modalities have been attested a parallel initial processing of external stimuli.

2.1.2 Mental Chronometry

On a behavioral level, the question of architecture and mode of cognitive processing has been investigated with the help of *mental chronometry*. There, response time and accuracy data are used to infer the content, duration and temporal sequencing of cognitive processes (Townsend & Ashby, 1983). The use of response times as the main dependent variable is old and has its origin more than a century ago (Posner, 2005). One of the earliest instances of mental chronometry to uncover the durations required by various mental processes was Donders' *method of subtraction* (Donders, 1969). It was based on the idea that complicated mental activities were compounded in a simple sequential fashion from less complex parts, with mean response times being used to estimate the duration of each of the component durations. Following that idea, the duration of a complex cognitive task can be inferred on by first measuring the duration for each of the supposed components of the task. Then the mean response time for the simpler parts will be subtracted from the overall response time to obtain an estimate of the complex part of the task.

2.1.3 Response Time Decomposition

Even the simplest mental activities - such as e.g. the detection of a stimulus - involves a cascade of processing stages. Response times reflect this composite nature, as they are canonically modeled as a compound of decision time and nondecision or base time (Luce, 1991). The decision time, *D*, is defined as the (latent) time necessary to form a decision. Its duration is determined (amongst other factors like stimulus properties) by the complexity of the experimental task. The base time component, *B*, on the other hand is an amalgam of all the processes not involved in the decision making. These include stimulus encoding and transmission to higher processing centers and the time for response selection, preparation and execution. The observable response time, *RT*, then is modeled as the sum of the decision time and the nondecision time: $RT = D + B$.

2.1.4 Complexity of Decisions

The complexity of the decision itself depends on the depth of processing necessary to form the decision. There are four basic types of decisions in a mental chronometry task. In *simple reaction times*, the observers are required to respond to the onset of any stimulus. For example, in a bimodal an observer has to press a button as soon as a visual or auditory stimulus appears. In *recognition* or *go/ no-go* reaction times, the observer also has to respond upon the detection of a specific stimulus type. In contrast to a simple reaction time however, the observer has to withdraw its response when another stimulus type appears. For example, the observer might be instructed to press a button for words of a prelearned list and withhold the button press for words that are not on that list. *Choice reaction times* share the basic protocol of go/no-go tasks, except that the observer does not withhold its response for the non-target stimulus set, but actively responds with a different response behavior (here, a different button press). *Discrimination reaction times* represent the most complex decisions in mental chronometry. There, observers must compare pairs of simultaneously presented stimuli according to predefined categories. For example, observers must discriminate which of two visual stimuli appear to be brighter, bigger, or the like.

Response times are not a constant entity but are conceptualized as the realizations of a random variable. Due to momentary attentional lapses and system inherent noise (e.g. Laughlin & Sejnowski, 2003), there is a considerable amount of variability in the observer's response time. To control for this, the observers are required to perform the experimental task for multiple trials to obtain reliable estimates of the response times. Also the amount of variability can be used as a source of information, as variances accumulate for more processing stages.

Donder's method of subtraction then rests on the assumption, that the specific mental processes act serially and independently of each other. As further research however would show, this assumption does not hold in general for mental processes. One striking reason why this was only uncovered later, is due an issue that is prevalent for the question of parallel and serial processing – the problem of model mimicking.

2.1.5 Model Mimicry

Model mimicry refers to the ability of one class of models of a psychological process to make the same predictions as other classes of models on the basis of strikingly different psychological assumptions. Some of the early attempts to settle the question tested certain types of parallel and serial models (e.g. Egeth, 1966). As it turned out, the methods used to test certain parallel and serial models, failed to test other models that might have explained the same data (for example, Atkinson, Holmgren, & Juola, 1969; Murdock, 1971; Townsend, 1971). Further analysis by means of mathematical modeling showed, that representations of diametrically opposed psychological principles could sometimes mimic one another and not be distinguished even on a level of their defining probability law (for an example, see Townsend & Wenger, 2004).

2.1.6 Levels of Analysis

Unfortunately, a great amount of information that is provided by response times is lost, when the data is compressed to the level of means. This appears all the more remarkable, when considering, that response times are almost always right skewed, so a compression to the central moment is concealing an informative aspect of the distribution. Effects in means however, can be produced by shifts of response time distributions, or stretching the slow tails of response time distributions or some combination of the aforementioned (Balota & Yap, 2011). The inclusion of the whole response time distribution for the analysis and investigation of mental processes marks an important advancement in mental chronometry. It can effectively help avoid model mimicry.

One paradigm that uses a distributional bound to increase the discriminative power and avoids model mimicry is the *redundant signals paradigm*. It investigates the performance of observers in perceptual-motor decisions, which necessitates an integration of evidence stemming from one or more signal sources. It epitomizes both the theoretical dilemma and the methodological advancements in mental chronometry. By use of a distributional bound it enables researchers to test the observed performance against an entire class of models –

the separate activations or race model accounts. It has been applied in a myriad of experimental settings and was subject to a multitude of theoretical and empirical studies. In the following, the redundant signals paradigm will be introduced as a means to infer the cognitive architecture for decisions based on multi-signal evidence. A formal derivation of the race model inequality test will be provided before turning to the explanation models of the redundant signals effect.

2.2 The Redundant Signals Paradigm

The *redundant signals paradigm* (RSP; Todd, 1912; Kinchla, 1974) forms the core of the present study. It is an established and well studied experimental paradigm that applies mental chronometry to the question of cognitive architecture of perceptual-motor decisions for multisensory and/ or multimodal stimulation. In an RSP task, observers have to respond as soon as a critical stimulus – an element from a predefined set of target stimuli – appears. The response is identical for each target stimulus and each trial can consist of either a single target (single signals trial, SST) or two (or more) targets (redundant signals trial, RST). The redundantly defined target is then the simultaneous presentation of either two identical target element at one or two locations or two different target elements at one or two locations. For the RST at one location, the target can be a combination of two or more features, dimension or modalities. The general finding for all redundant signals experiments is that the mean response time for the redundant signals trials is faster than the mean response time of either single signal condition. This speed-up has been termed *redundant signals effect* (RSE; Hershenson, 1962).

2.2.1 Classification of RSP Experiments

A general way to classify RSP experiments is (Töllner, Zehetleitner, Krummenacher, & Müller, 2011) to discriminate them by the spatial origin of the signals (same versus different), the modality of the involved signals (e.g. vision, audition, tactition), the number of relevant modalities (one, two or more) and the depth of signal processing (cf. complexity of the decision). This classification provides a means to expose vital differences in the experimental setup when comparing potentially conflicting results. Two examples for redundant signals experiments will be provided and classified according to this system, before turning to the catalogue of explanatory models of the RSE.

As a first example, a bimodal simple response time task is delineated. Here, observers are required to press a button with both their index fingers, once they detect the

onset of any stimulus. The stimulus can be a Gabor patch, an acoustic beep or the combination of the two with a varying interstimulus interval. As on every trial a stimulus will appear, observers have to respond on every trial. However, to prevent anticipatory effects, the intertrial interval is varied. In the aforementioned classification system, this experiment would feature one spatial origin of signals, two modalities (namely vision and audition) and require a simple decision performance (i.e. a simple reaction time task).

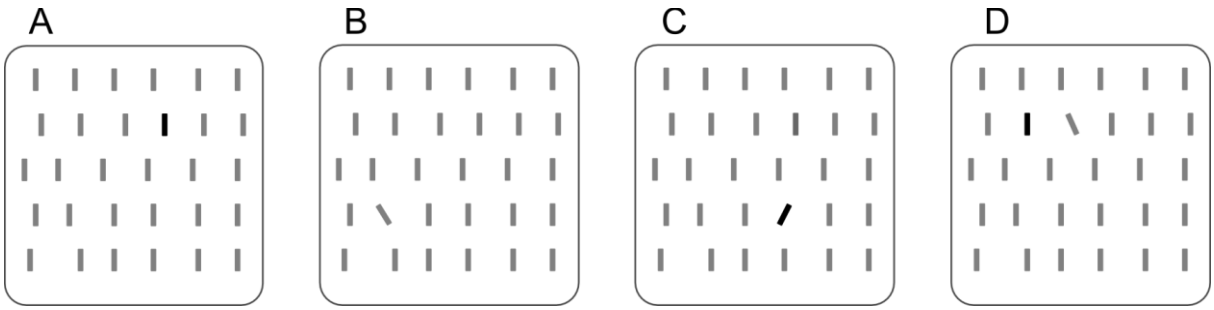

Figure 2.1 Redundant pop-out search displays. Observers are required to respond upon the detection of any pop-out target. Single signals trials (panels A and B) yield on average slower reaction times than redundant signals trials (panels C and B). Redundant signals trials are the combination of single signals, either at one location (panel C) or at neighboring locations (panel D).

In the second example, a visual pop-out search task is performed. There observers have to search for targets amongst distractors, which pop-out from their surrounding by one or more discrepant dimensions within the visual domain (see Figure 2.1). The target set is composed of black untilted bars (Panel A), gray tilted bars (Panel B), black tilted bars (Panels C) or the combination of a black untilted and a gray tilted bar amongst gray untilted distractors (Panel D). Here, the first two targets are exemplar for single signals, whereas the latter two are redundant signals. In the classification system, this experiment would be characterized by two spatial origins, one modality, namely vision and a moderately complex decision task, as stimuli do not appear in solitude but have to be discriminated against other stimuli.

2.2.2 Model to explain the RSE

The redundant signals effect was first reported by Todd (Todd, 1912), and has been thoroughly investigated, empirically and theoretically (Colonius, 1988, 1990a; Diederich, 1995; Giray & Ulrich, 1993; Miller, 1982; Mordkoff & Yantis, 1991; Schwarz, 1989, 1994;

Townsend & Colonius, 1997; Townsend & Nozawa, 1997; Townsend & Wenger, 2004) for go/no-go tasks (Egeth & Mordkoff, 1991; Heijden, Heij, & Boer, 1983; Gondan, Götze, & Greenlee, 2010), search RT tasks (Krummenacher, Müller, & Heller, 2001), for elderly subjects (Bucur, Allen, Sanders, Ruthruff, & Murphy, 2005), different modality combinations (Schröter, Frei, Ulrich, & Miller, 2009; Veldhuizen, Shepard, Wang, & Marks, 2010), and different neurological deficiencies (Pollmann & Zaidel, 1999; Marzi et al., 1996; Reuter-Lorenz, Nozawa, Gazzaniga, & Hughes, 1995; Miller, 2004).

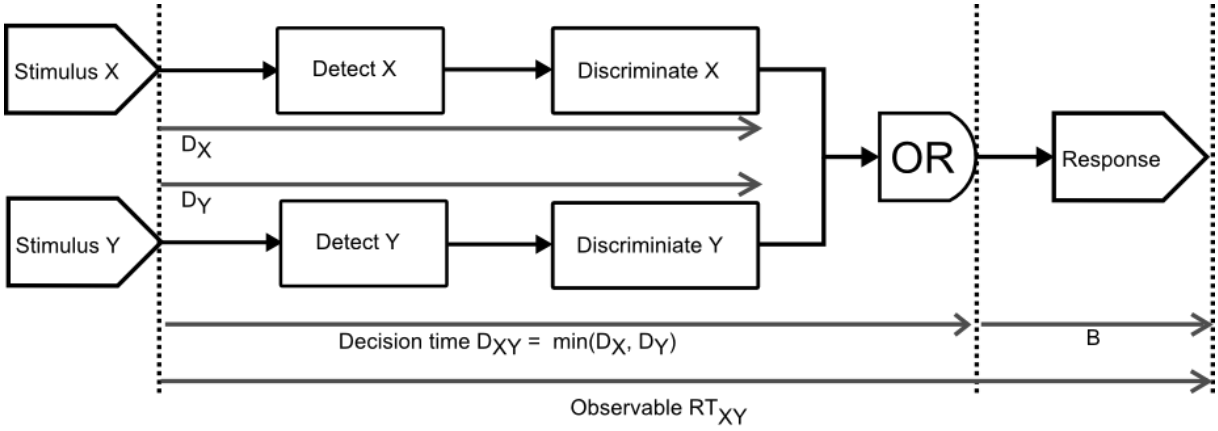

Figure 2.2 Statistical facilitation or race model account. Reaction times can be decomposed into decision (D_X, D_Y) and nondecision times (B). Decision and nondecision times are modeled as random variables with a channel inherent variance. The processing times of the channels are independent of each other and the faster channel determines the decision time. On average, the minimum of the decision times is lesser than each channel decision time alone.

It exemplifies the question of cognitive architecture as competing models postulate drastically different psychological assumption to account for it. Historically, the redundant signals effect has been initially explained by a separate activations account (Raab, 1962). Raab proposed that the processing of a redundant signal can be conceptualized as a horse race of the single signals. Sensory components of the redundant stimulus are processed independently and in parallel (Miller, 1982) and *race* against each other. The faster channel then decides the “race”, as it reaches a specific decision criterion earlier. Both channels process the response-relevant signals in a stochastic way, thus the minimum finishing time for two or more channels is on average faster than of one alone (see Figure 2.2).

In its first formulation, the redundant signals effect generated by a race model was bounded on a mean level:

$$E[\min(X, Y)] \leq min[E(X), E(Y)], \tag{3}$$

where *E( )* denotes the expectation of a random variable, and *X* and *Y* the random variables for the detection time of each target channel. (This is a special case of Jensen's inequality, see for example Rudin, 2006). What is remarkable about this first account of the RSE is that de facto no integration of signals is assumed. It rather states that the RSE is the result of a *statistical facilitation*: As both channels act independently of each other and vary from trial to trial, on average one of the channels will be faster than the other. The variability in response times is expedited and forms the actual cause for the observed speed-up.

2.2.3 Race Model Inequality

Miller (1982) formalized a distributional upper limit for the maximum amount of facilitation that race architectures can account for – the *race model inequality* (*RMI*). It relates the distribution function of the redundant signal to the distribution functions of the single signals:

$$F_{XY}(t) \leq F_X(t) + F_Y(t), \qquad \forall t. \tag{4}$$

Here, $F_X(t)$, $F_Y(t)$ and $F_{XY}(t)$ are the cumulative distribution functions (CDF) of the reaction times for the single signals trials of conditions *X* and *Y* and for the redundant signals trials *XY*. Equation 4 marks a critical test for all race models: Any response time data violating this inequality (for any time point *t*) by definition leads to a falsification of the whole class of separate activation models. It marks the best possible performance race model architecture. When both racers have a perfect negative correlation the variance of the channels is expedited in the strongest way (Ulrich & Giray, 1986; Colonius & Diederich, 2006). This then is not a stochastically independent processing of signals, but implies that fast processing times of one signal co-occur with slow processing times of the other signal.

As further empirical studies would show, the RMI is frequently violated excluding race models as a viable architecture for the RSE in many experimental situations (Diederich, 1992; Egeth & Mordkoff, 1991; Feintuch & Cohen, 2002; Grice, Canham, & Boroughs, 1984; Krummenacher et al., 2001; Krummenacher, Müller, & Heller, 2002; Miller, 1982; Mordkoff, Miller, & Roch, 1996). The common alternative accounts are coactivation models, proposed by Miller (1982; 1986).

Coactivation models assume that single signals originating from two (or more) afferent channels do interact and integrate at some stage (see Figure 2.3). Thus decision related activity builds up faster and to a higher level on redundant signals trials compared to single signals trials, resulting in redundancy gains that cannot be accounted for by race models.

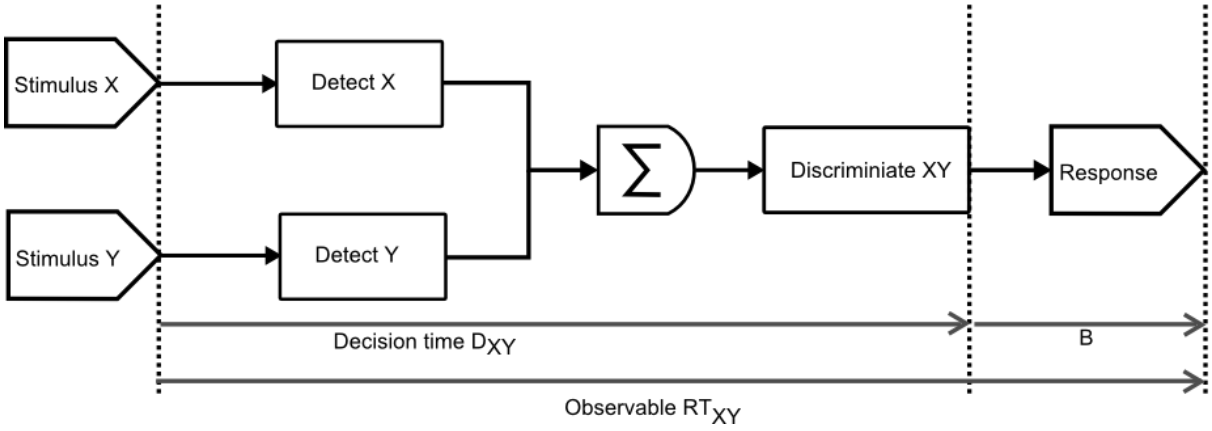

Figure 2.3 Coactivation model. The processing of the stimuli interacts and activity across channels is pooled to a higher and faster value. This leads to a speed-up in the decision time.

Coactivation models can produce such violations of the RMI and are the commonly assumed alternative hypothesis. It is important to remark, that violations of the RMI only falsify race models and provide no direct evidence for concrete coactivation mechanisms. From a theoretical standpoint however, there are other architectures that can produce RSE and violations of the RMI. (Further alternative accounts such as the interactive race model (Mordkoff & Yantis, 1991) and serial architectures with an exhaustive stopping rule (Townsend & Nozawa, 1997) have been proposed. At least in the visual domain these could be excluded empirically as potential models of the RSE and RMI violations, for a review see Zehetleitner, Müller, & Krummenacher, 2008b).

Mostly, researchers have implied coactivation models upon detecting violations of the RMI and have not explicated mechanisms that could generate the redundancy gains. Two coactivation models that have provided an explicit mechanism are the multichannel diffusion model by Diederich (1995) and the diffusion superposition model by Schwarz (1994) (Both models will be described in detail in the subchapter 2.4, where explicit mechanisms of the RSE are discussed.) For the moment, it suffices to note that both these models assume a

decisional stage for the integration of signals. From the stance of mental chronometry however, this is only one potential locus of the effect.

2.2.4 Locus of the Redundant Signals Effect

In cases where the RMI test has robustly attested violations and a race model account can be excluded, still the question remains, how exactly this data pattern comes about. Thus, orthogonal to the architectural question (independent parallel versus coactive parallel), researchers have investigated at what processing stage the RSE can occur – with different results. Consistent to the distinction of decision and nondecision times in mental chronometry (Luce, 1991), most cognitive models on decision making assume, that to perform a perceptual-motor task various distinct subprocesses have to be completed.

A fictitious audiovisual bimodal redundant signals experiment will be decomposed in this sense to illustrate all the involved processing stages. The observers' task is to detect items of a target set amongst distracting items. First, all the stimuli (auditory and visual) have to be encoded in parallel by the various modality systems. The activation of these systems is then transmitted to higher processing units. There, decision units accumulate the sensory evidence upon a threshold. Once this threshold is reached, a response signal is triggered. This leads to the next-to-last stage, where the response is prepared (e.g. by selecting and preparing a motor plan) and finally to the measurable response execution.

Owing to the fact that apart from the last stage (where the response is executed) all other stages are latent and unobservable, these multiple stages collapse into two reaction time components. Canonically the response time is decomposed into a decision time that is consumed by the process of research interest (e.g. detection, identification, discrimination) and a nondecision or base time, where all the stages not involved in the decision are merged. This aggregation of stages is not deemed to be problematic, because basic processes such as encoding and executing motor programs are thought to be relatively constant over time and display only little amount of variance (Luce, 1991).

Explicit accounts of coactivation model the RSE as a decisional effect. In theory however, any of the stages (or combinations of them) can be responsible for the speed-up in the overall response time. In the redundant signals paradigm, this has been addressed for the visual modality, as it is generally agreed on, that the visual RSE is a result of coactivation rather than parallel race models (Feintuch & Cohen, 2002; Krummenacher et al., 2001; Mordkoff & Miller, 1993). There is however an ongoing debate on the specific locus of the visual RSEs, which can be viewed as prototypical for other domains. Accounts favoring a pre-attentive perceptual locus (Zehetleitner, Krummenacher, & Müller, 2009a; Koene &

Zhaoping, 2007; Krummenacher et al., 2001, 2002; for a review Zehetleitner et al., 2008b) assume that at least part of the RSE is generated at the level of early stimulus analysis, where the relevant target attributes are coded and compared with those of non-target elements, without necessarily involving stages following focal-attentional target selection.

2.2.5 Decisional Accounts

Results supporting a pre-attentive view have been presented by Zehetleitner et al. (2009b) and Krummenacher et al. (2001, 2002). They conducted a visual pop-out search task where the redundant target was composed of one or two targets. Observers had to discriminate the presence (vs. the absence) of any target. Targets could be singly (orientation or color) or redundantly (orientation and color) defined in the search display. Crucially, on redundant-target trials, the distance between the color and the orientation singleton could vary from zero (one bar contains both target features) to three (two target bars were separated by two distractor bars). Krummenacher et al. (2002) found violations of the RMI, only with small spatial separations (one to two) but not with larger separations. This hints at an integration of signals before response selection or response execution, because these stages should be unaffected by the distance between targets.

More recently, Zehetleitner et al. (2009b) examined the size of the RSE and the amount of coactivation (using the geometrical measure by Colonius & Diederich, 2006) under pop-out search conditions of high- or low-feature contrast of the target relative to the distracters. The authors found both the size of the RSE (RSEs of 50 vs. 15 ms) and the amount of coactivation to be larger for low-contrast than for high-contrast redundant targets. This pattern argues against a post-selective origin of coactivation effects: given that the manipulation of feature contrast affects the speed of focal-attentional selection (which is determined by pre-attentive coding processes) a post-selective origin of coactivation would have predicted the RSE to be statistically comparable in both conditions rather than to be dependent on target feature contrast.

By contrast, other authors have proposed that the RSE arises exclusively at stages subsequent to attentional selection (Miller, Beutinger, & Ulrich, 2009; Feintuch & Cohen, 2002; for a review see Miller & Reynolds, 2003). Results that affirm this post-selective view have been reported by Miller et al. (Miller et al., 2009) who presented one or two target stimuli either at expected or unexpected locations. They found the RSE to be comparable in both conditions, from which they inferred that the redundancy gains arise after target elements were selected by focal attention. Using a similar rationale, Feintuch and Cohen (Feintuch & Cohen, 2002) presented two bars that had one of three possible colors and one

of three possible orientations, with one specific color and one orientation being defined as response-relevant target features. Observers had to produce a go-response if they discerned either the target color or the target orientation or both the target color and the orientation in the display. The target features could belong either to one bar or to two separate bars. Feintuch and Cohen found RMI violations only if the target features belonged to the same bar rather than to separate bars. However, when focal attention was guided to both bars simultaneously by presenting them within an outline ellipse (i.e., within a surrounding object cue), there was evidence for coactivation by target features belonging to separate bars. Feintuch and Cohen concluded that the coactivation mechanism is located in the processing stream after focal-attentional selection, and attentional selection is a prerequisite for the coactivation mechanism to come into play. This finding is at variance to Krummenacher et al. (2002) who found evidence of coactivation in a similar situation. The authors revisited this question and argued that coactivation takes place at a level before attentional selection. In their Experiment 3, they pre-cued the spatial location of an upcoming search target symbolically and examined the RMI at uncued and cued locations. As they found evidence for coactivation for both non-preselected and preselected locations they took this as evidence in favor of a pre-attentive rather than a post-selective locus of coactivation in pop-out search.

A recent study by Töllner, Zehetleitner, Krummenacher and Müller (2011) combining mental chronometry with electrophysiological markers could show that in visual pop-out search, redundancy gains originate entirely from processes that operate at the processing stages of pre-attentive perceptual encoding. There did not appear to be any (additional) contribution from post-selective stages of stimulus analysis and stimulus-response mapping and of response execution. This finding however, has to be taken with a grain of salt due to the asymmetrical epistemological value of absence of evidence compared to evidence of absence.

2.2.6 Motor Accounts

Proponents of an even later stage of coactivation localize the source of the speed up in the motor regime. In fact Miller himself pointed out that although coactivation appears to be most plausibly a decisional effect, further research is necessary to rule out the possibility that response execution is accelerated (Miller, 1982). Empirical evidence in favor of a motor account of coactivation has been reported by several authors.

Ulrich and Stapf (1984) present evidence for a speeding up of the motor processes, but not exclusively at that stage. They applied a double-response paradigm to an auditory

and a visual simple reaction time task with varying stimulus intensity. Observers had to perform a button press with both hands simultaneously upon detection of any stimulus. They analyzed the difference of both manual response times by isolating the motoric component. Let *S* be the time needed to detect the stimulus and prepare the motor command for movement execution and M_L, M_R be the motor time for the left hand and right hand response respectively. *S* is the same for both RT_L and RT_R as both movements are triggered upon the detection of the stimulus. The RT difference *D*

$$D = RT_L - RT_R = (S + M_L) - (S + M_R) = M_L - M_R, \quad (5)$$

then depends only on M_L and M_R, since the common term *S* is dropped. If stimulus intensity affects only the earliest stages of stimulus encoding, then the distribution of the RT difference should be invariant to a change of stimulus intensity. This logic is independent of the distributions M_L and M_R might have and whether the mean motor delays are equal for both hands. To determine whether stimulus intensity speeds up the motor delay, one has to make one additional assumption. As motor delays are not directly observable, one assumes that if the variance of the motor delay decreases with stimulus intensity, this gives evidence that mean motor delay is decreased too. The authors did find that the motor delay was decreased by an increase in stimulus intensity.

Diederich and Colonius (1987) performed a bimodal detection task with redundant stimuli using the double response paradigm by Ulrich and Stapf. They analyzed the RT differences between hands to obtain an estimate of the motor delay as a function of stimulus onset asynchrony (SOA). The SOA manipulation is assumed to give an insight into the time course of the RSE (see Miller, 1986). The authors reported a slight but significant effect on the variance of *D* in redundant-signals trials. Specifically they found a U-shaped functional dependence between the variance of *D* and stimulus onset asynchrony of the visual and auditory conditions. The function dependence was parallel to the way in which the RSE depended on the SOA. This finding can be seen as (indirect) positive evidence for the motor coactivation hypothesis.

Giray and Ulrich (1993) employed a psychophysical measure that is not commonly used in mental chronometry. They recorded response force of manual responses in a bimodal divided attention task. Based on the previous literature on motor coactivation, they hypothesized that the response forces for bimodal signals trials will be greater and that the facilitation mechanism is effective at a motoric level. They could show that observers produces more forceful responses under bimodal than unimodal stimulation. However, this effect depended on the type of task: a relatively large force effect was found for a divided-attention task and a weaker effect for the focused-attention task. They interpreted that the

motor system is more active in bimodal than unimodal stimulation and that the RSE is at least partially located in the motor system. The authors argue in favor of utilizing response force as an additional measure for the investigation of coactivation, as a correlation analysis between suggests that response force is not inherently redundant to reaction times.

At least for visual discrimination tasks, the study by Mordkoff, Miller and Roch (1996) could disconfirm the motor-coactivation hypothesis. They could show that response force is in their experiments was determined by the number of stimuli rather than the number of targets in the visual field, when target-distractor discriminations were required. Using three motor-related psychophysical measures (i.e. response force, event-related potentials and electromyograms), and despite large amounts of RMI violations, the authors could not find evidence favoring a motor-coactivation. In fact, all three measures indicated that the motor process started earlier in RST than in SST, implying that all of the effect of redundant targets lies before the onset of motor processing.

2.2.7 Intermediate Conclusion

In a nutshell, there are several accounts that try to explain the RSE (and violations of the RMI). They can be classified according to their proposed architecture (serial, independent parallel or coactive parallel) and in their locus of the effect. Moreover, there are several points that become evident from the literature review. First, the RMI test can be a valuable tool to infer the architecture of multisignal decision tasks, namely by falsifying one class of explanatory models – the race models. If the RMI holds, one must agree with the most parsimonious and empirically supported null hypothesis of a race model. If it is violated, that whole class (i.e. race models) must be considered as falsified. Second and beyond that, the RMI alone can not give insight into the specifics of the generally assumed alternative coactivation model. It can neither help to specify the concrete mechanism of the effect nor can it elucidate the source (or sources) of the effect with respect to the different processing stages of the decision. Lastly, all the explicit models of coactivation assume that it occurs at a decisional stage, but have been tested only on a mean level.

To aggravate matters, research focusing on the statistical properties and performance could reveal several intrinsic flaws of the RMI test itself. In theory, these ambiguous results might partially be due to a potentially premature elimination of race models as the source of the RSE. As every statistical test can be characterized by its false alarm and hit rate, an evaluation of the RMI test is of paramount importance for the weight of the conclusions drawn from it. In order to assess the statistical side of the RMI test, the next chapter will first provide a rigorous formalization of race models, then derive the race mode inequality for

decision times only and overt response times (i.e. decision and base times). After a description of the standard practices of RMI testing, the chapter will discuss already known deficiencies as well as knowledge gaps of the test. These will ultimately lead to the motivation and research questions of the present study.

2.3 The Race Model Inequality Test

The previous elaborations attest that the race model inequality marks a pivotal progress within the mental chronometry paradigm. With help of the race model inequality it is now possible to distinguish between two radically different processing architectures (parallel race and coactivation models), that on a mean level can mimic each other. In the following, a rigorous formal description of race models and a concise derivation of the race model inequality will be provided. Furthermore the explicit computational algorithm by Ulrich, Miller and Schröter (2007) as the standard practice of RMI testing will be detailed. A critical assessment of classical and novel issues of RMI testing will conclude the chapter.

2.3.1 Assumptions of Race Models

In the redundant signals paradigm, reaction times of single signal trials (SST) are compared with reaction times of redundant signal trials (RST), where both stimuli are presented in combination. Observers are instructed to respond as soon as they detect any target (that is, singly or redundantly defined). The redundant signals effect (RSE) is defined as the speed-up in mean reaction times between SSTs and RSTs.

The first model to account for this effect is the separate activations or race model account (Raab, 1962). It assumes that the different modalities (or feature dimensions within a modality) are processed (a) in *separate channels*, along which evidence is accumulated towards a response eliciting threshold. Let s_X and s_Y be the two single signal channels of abstract modalities or dimensions *X* and *Y*. Further it assumes (b) that these signals are processed in *parallel*, that is, one channel does not have to wait for the other channel to finish before it can begin processing its input. It is also assumed, (c) that the rate of processing along each channel is invariant across SS and RS conditions. This assumption is known as *context independence* (Ashby & Townsend, 1986; Luce, 1991). It forms the crucial theoretical link to justify the comparison of data gathered in SST and RST. The race property (or statistical facilitation) is formalized in the stopping rule of the model. The finishing time in RST is determined by the winner of the race, ergo the faster of the two channels. This assumption is called (d) *minimum-time stopping rule*, since the completion of the entire

system is the minimum of the completion time of the individual channels. Sometimes it is also assumed, (e) that the signals of each channel are processed *independently* that is the rate of evidence accumulation of one channel is not affected by the rate of evidence accumulation of the other. A model that satisfies the assumptions (a) to (d) is commonly referred to as a race model. A model that additionally suffices (e) is known as an unlimited capacity, independent parallel (UCIP) model with a minimum-time stopping rule (cf. the taxonomy of information-processing models by Townsend & Nozawa, 1995).

2.3.2 Formalization of Race Models

The decision latencies for each channel are modeled using non-negative random variables, say *X* and *Y*. These variables are theoretical quantities – the variables governing the empirical (measurable) reaction times in each condition are written as RT_X and RT_Y. For the moment, the reaction times are assumed to be identical to the decision times for each channel respectively. Thus, each measured reaction time is a sample realization of the corresponding theoretical quantity: $RT_X = X$, $RT_Y = Y$ and $RT_{XY} = min\,(X', Y')$, where the random variable $X' =^d X$ and $Y' =^d Y$. The notation $A =^d B$ means that *A* and *B* have identical distribution functions (that is, they are "equal in distribution"). The distinction between *X* and *X'* is made to in order to emphasize that although the variables across conditions are equal in distribution (usually due to context independence) their realizations are entirely independent.

For a derivation of the RSE on a mean level, it follows from a special case of Jensen's inequality (e.g. Billingsley, 2008) that

$$E[\min(X, Y)] \leq \min[E(X), E(Y)] \tag{6}$$

for any distribution of (*X, Y*). Random variables *X* and *Y* are not observable (only their minimum is, in the redundant signal condition), but from the equal-in-distribution assumptions a testable analogon of Inequality 6 follows:

$$E(RT_{XY}) \leq min[E(RT_X), E(RT_Y)] \tag{7}$$

This mean level version of the race model inequality has been used to test whether data from an RSP experiment can be accounted for by a stochastically independent race between the single signal detection times. Whenever the redundant condition generates smaller (that is faster) average RTs than this model, all race model can be excluded. Jeff Miller developed a more general test on a distributional level that proved to be more sensitive than the mean variant (Miller, 1978, 1982). He showed, that the race model inequality

$$P(RT_{XY} \leq t) \leq P(RT_X \leq t) + P(RT_Y \leq t) \tag{8}$$

must hold for all $t \geq 0$. This inequality follows directly from

$$P[\min(X, Y) \leq t] \leq P(X \leq t) + P(Y \leq t) \quad (9)$$

a special case of Boole's inequality (Billingsley, 2008). Figure 2.4 depicts the race model inequality bound together with the single stimulus CDFs for fictitious RSP experiment.

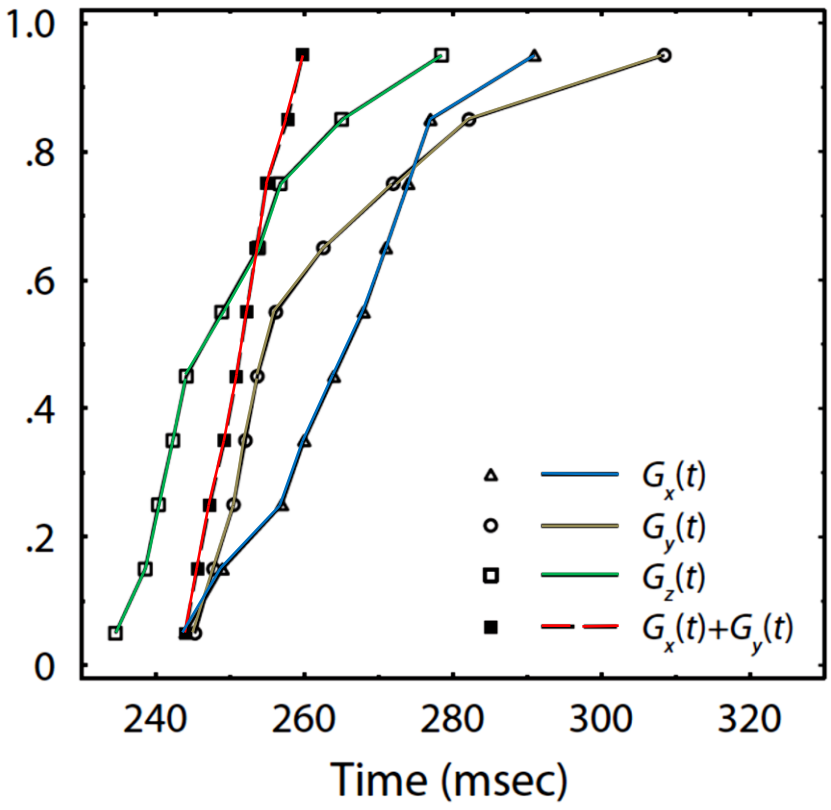

Figure 2.4 Race model inequality bound. G_X and G_Y signify the cumulative distribution functions of the single signal trials *X* and *Y* and G_Z of the redundant signals trial. The Race model inequality bound is the sum of the single signals distribution functions. Whenever G_Z is left of G_X+G_Y, this marks a speed-up that is not explainable by any model with a race architecture. (reprinted and adapted with permission, from Ulrich et al., 2007)

A common way to depict the amount of RMI violation is to subtract the single signals distributions from the redundant signals distribution and plot this as a function R^*_{XY} of *t* (Miller, 1986)

$$R^*_{XY}(t) = P(RT_{XY} \leq t) - P(RT_X \leq t) - P(RT_Y \leq t) \quad (10)$$

This function R^*_{XY} is called the *race model test function*. By the inequality shown in equation 8, positive values of $R^*_{XY}(t)$ indicate violations of the RMI. Given that the left-hand side of the inequality shown in equation 10 is always bounded by 1, the inequality can be rewritten as

$$P[\min(X,Y) \leq t] \leq min[P(X \leq t) + P(Y \leq t), 1] \quad (11)$$

resulting in the modified RMI test function

$$R_{XY}(t) = P(RT_{XY} \leq t) - min[P(RT_X \leq t) - P(RT_Y \leq t), 1] \quad (12)$$

Violations of the RMI will again result in positive values of $RT_{XY}(t)$, whereas negative or zero values of $RT_{XY}(t)$ are compatible with the race model.

It is important to emphasize, that the inequality holds for any race model, not only for the independent case. This is relevant because dependent processing does affect the predictions of the race model. Indeed, assuming negative dependence — that is, relatively fast detection latencies for signal s_X co-occurring with relatively slow detection latencies for signal s_Y and vice versa – it is obvious that the smaller of the two random latencies RT_X and RT_Y tends to be small as compared to the smaller of two independent latencies, as long as the individual latencies' means do not vary. Note that if the inequality is violated not only if the independent model is falsified but in fact any race model is (as the RMI is invariant to the dependence between *X* and *Y*).

2.3.3 Base Time Component as a Filter

In the explication above it was assumed, that the measured reaction times RT_X, RT_Y and RT_{XY} were samples of the process variables *X*, *Y*, and *min(X',Y')*. However, reaction times are canonically modeled as a sum of two random variables, the processing time and the base time (here, B_X, B_Y). The decision component (*X, Y*) involves the (latent/ unobservable) time necessary to accumulate evidence until the threshold of a channel is reached. The base time component is an amalgam of all of all processes not involved in the decision making that is the stimulus encoding and transmission to higher processing centers and the time for response selection, preparation and execution. It is commonly assumed to be uniformly of normally distributed. Following Ulrich and Giray (1986) the race model incorporating base time is

$$RT_X = X + B_X,\ RT_Y = Y + B_Y,\ RT_{XY} = \min(X', Y') + B_{XY}, \quad (13)$$

where $X =^d X'$, $Y =^d Y'$ and $B_X =^d B_Y =^d B_{XY}$. This added base time component requires an additional assumption, analogous to the context independence – a *motor context*

independence. It states that the base time is invariant across all experimental conditions, SST and RST (Luce, 1991 Ulrich & Giray, 1986). When applying this base time component to the RMI, this gives

$$P[\min(X,Y) + B_{XY} \leq t] \leq P(X + B_X \leq t) + P(Y + B_Y \leq t) \quad (14)$$

It is evident, that this inequality (eq. 14) is identical to the RMI (eq. 9); the only difference is that the quantities in equation 14 incorporate base time while those in equation 9 do not. Clearly, then violations of the RMI falsify race models whether or not the race models include base time (see Ulrich & Giray, 1986). What is not evident at first sight is that the presence of a base time (with a specific variance) does affect the sensitivity of the RMI test.

On a theoretical level the attachment of a random variable to the distribution function is equivalent to the convolution of two independent random variables (Townsend & Honey, 2007). Let $F_X(t) = P(X \leq t)$, the distribution function of the decision time. When a function is convolved with another function (especially a unimodal function with smaller support), then it is usual to refer to the simpler function as a *kernel* and to view the result of the convolution transformation as a filtered version of the *input*. In the context of the race model inequality, F_X the distribution function of the decision time represents the input and the base time is conceived as the filtering kernel f_B:

$$F_{X+B}(t) = \int_{-\infty}^{t} F_X(s) f_B(t-s) ds = F_X * f_B(t). \quad (15)$$

Kernels of this type can be understood to act as local averaging mechanisms: each point of the output function F_{X+B}, the measurable response time distribution, is a kernel-weighted average of corresponding points in the input function.

Townsend and Honey (2007) applied this reasoning and derived the empirical equivalent of the race model test function featuring base times.

$$R_B^*(t) = R^*(t) * f_B(t), \quad (16)$$

where R^*_B is the empirical race model test function and R* is the theoretical/ decisional test function. By definition, $R^*(0) = 0$ and asymptotically runs to $R^*(\infty) = -1$. Any t^* where the race model test function is positive $R^*(t^*) > 0$ constitutes a violation of the race model inequality. A standard result from theory of function convolution (Kecs & Giurgiuţiu, 1982) gives that the maximum positive value of $R^*(t)$ is always greater than the maximum positive value R^*_B. More generally, applying Hölder's inequality, it can be shown that when an input function is filtered with a probability density function, the output (here the distribution

function of the decision times) will always have a smaller absolute magnitude at its extrema than the input function.

2.3.4 Effects of Base Time Variance on Statistical Power

The results of Townsend and Honey are of high importance when evaluating the statistical performance of the RMI test. In essence, the authors could prove analytically that the presence of a base time can filter out areas of positivity of the race model function. Furthermore, if the breadth of the filter (here, the variance of the base time) is large enough, any positivity of the race model function can be eliminated and thus violations prevented. This poses a serious problem for RMI testing: whenever researchers fail to find violations of the RMI, this might also be due to the filtering effect of the base time component.

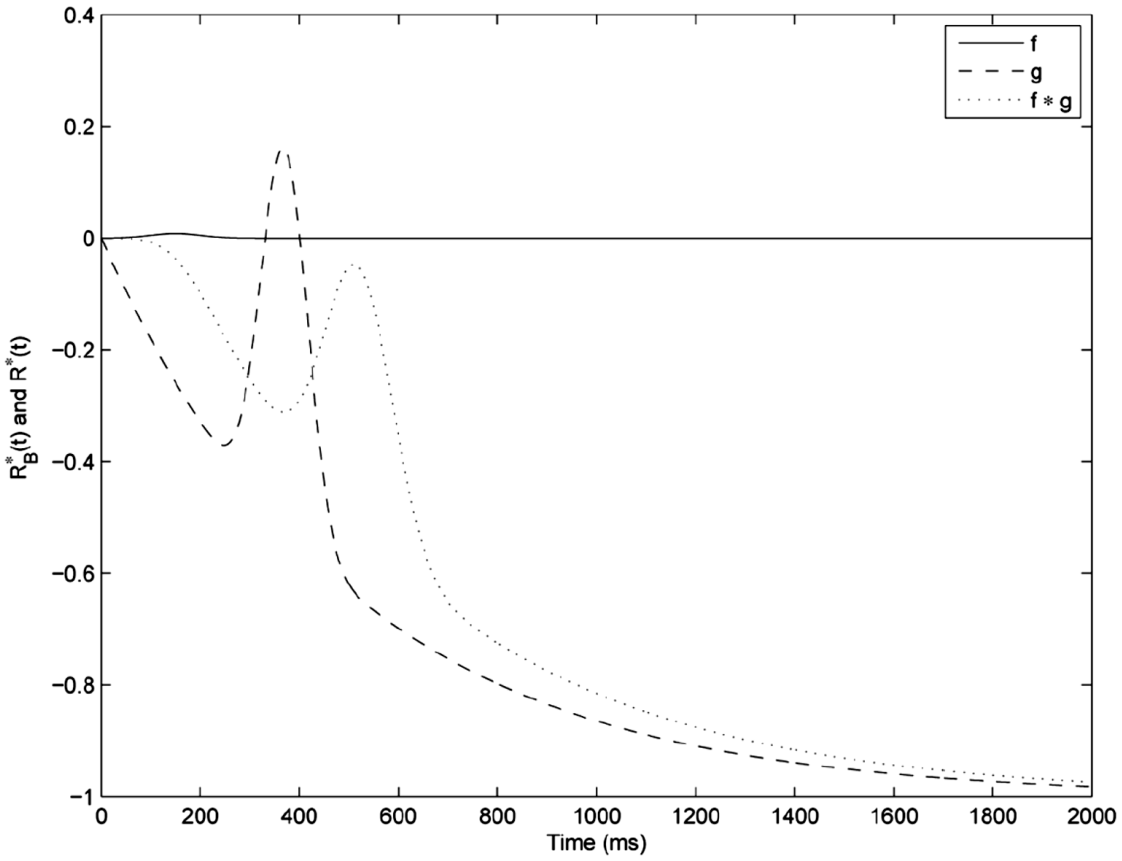

Figure 2.5 Filtering effect of base times on the race model test function. The race model test function f is convolved with a base time component, g. This results in a filtering out of areas of positivity (i.e. violations of the race model inequality; reprinted with permission from Townsend & Honey, 2007)

As different response effectors have inherently different motor variances, some response types might preclude violations and lead to false conclusions, here retaining race models as the explanatory model of choice (see Figure 2.5). This caveat in RMI testing is one research question of the present study. It will be treated in a novel way using mechanistic models of reaction times to investigate the interplay of decisional and nondecisional components in the simulation study (chapter 4).

2.3.5 Explicit Algorithm for RMI Testing

After deriving formal accounts of race models and the race model inequality test, next the question of how to explicitly perform the RMI test will be addressed. In spite of the ubiquity of its usage, it took more than two decades to formulate a standardized algorithm for its conductance. This was important and necessary, as the usage of some commercial software (e.g. Excel) provides certain percentile estimators which are inappropriate for the race model inequality testing. The following algorithm by Ulrich, Miller and Schröter (2007) was used in the present study as the RMI test version of interest. It is based on the estimation of cumulative distribution functions and consists of four algorithmic steps.

Let X, Y, denote the two single signal conditions and Z the redundant signal condition. Then F'_X is the empirical cumulative distribution function of the response time for condition X. The first step in the algorithm is to interpolate the empirical cumulative distribution functions for all three conditions of a redundant signals paradigm. Let G_X, G_Y and G_Z denote the estimates of the empirical CDF's F_X, F_Y and F_Z for the single and redundant signals trials. Second, the RMI bound $S(t)$ is computed as the sum of the CDFs G_X and G_Y, $S(t) = G_X(t) + G_Y(t)$ for each participant. Third, at certain prespecified probabilities, p, (usually, equally spaced probabilities like p = 0.05, 0.10, 0.15, … , 0.95 are chosen) percentile values s_p and z_p for S and for G_Z are estimated according to the percentile definition proposed by Hazen (1914). (For alternative quantile estimation functions see Hyndman & Fan, 1996). And fourth, percentile values s_p and z_p are aggregated over participants and for each percentile value a paired t-test is computed to evaluate whether G_Z is larger than S. The race model is rejected if G_Z is larger than S at any percentile. Figure 2.6 shows the progression of the test: in panels A to C the CDFs for each condition are estimated. Panel D depicts the computation of the bound $S(t)$, whereas panel E shows how the percentiles to the prespecified probabilities are obtained. In the last panel E, the computed bound and the estimated redundant CDF are compared via multiple t-tests.

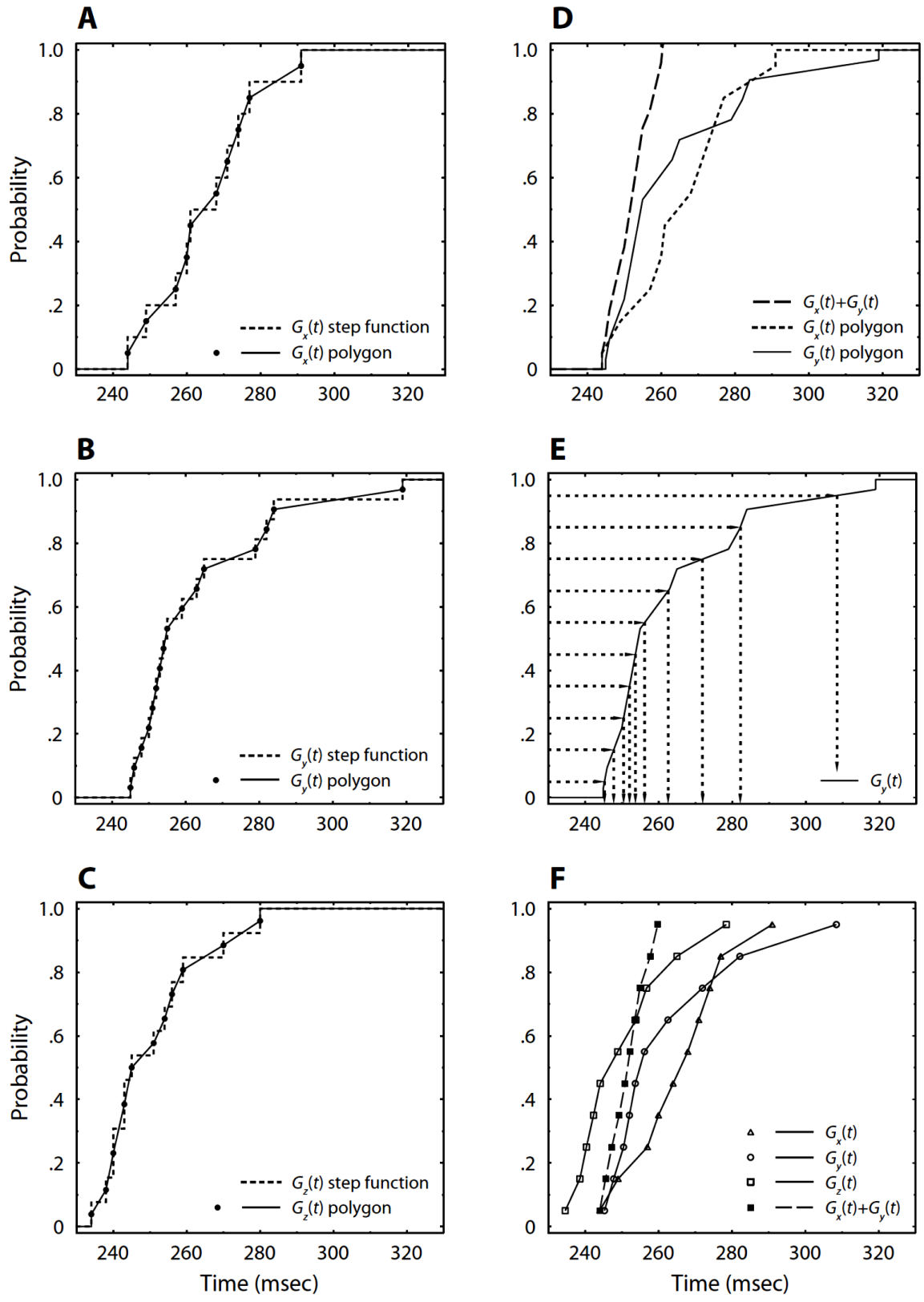

Figure 2.6 The four steps of the explicit algorithm by Ulrich, Miller and Schröter (2007). First the cumulative distribution functions of the single signals (G_X, G_Y) and the redundant signals trials (G_Z) are estimated via cumulative frequency polygons (Panels A to C). Then, the race model inequality bound is

computed by adding the cumulative frequency polygons of the single signals trials (G_X+G_Y, Panel D). Panel E shows how to obtain the estimated percentile points corresponding to the percentiles .05, .15, … , .95 for the function G_Y. Finally, the estimate of the redundant signals trials distribution function (G_Z) is *t*-tested against the estimated race model inequality bound (G_X+G_Y) at the prespecified percentiles. Significant deviations are seen as evidence against race models (reprinted with permission).

This procedure is thought to be rather conservative (Miller, 1982; Patching & Quinlan, 2002; Schwarz & Ischebeck, 1994) in the sense that it favors the race model. However, there are no reliable quantitative estimates of the power of the test are available. This conservativeness stems from the fact, that the data is tested against the most extreme exemplar of race models – a maximally negatively correlated race. Nevertheless many studies applying the redundant signals paradigms using this procedure have found violations and thus rejected the race models.

2.3.6 Issues of RMI Testing

Contrary to this commonly accepted notion of conservativeness, Kiesel, Miller and Ulrich (2007) have presented simulation data which suggests, that the RMI test is prone to false alarms, that is the inadequate rejection of race models based on flaws and biases in the tests' procedure. The authors identified two weighty sources for such erroneous test outcomes – an *estimation bias* working against race models and an *accumulation of type I errors*, when testing the RMI at multiple time points.

In the estimation literature (e.g. Gilchrist, 2000) it is a general finding that quantile estimators tend to overestimate the lowest percentiles and underestimate the highest percentiles. In the standard RMI test proposed by Ulrich et al. (2007) the estimation of the cumulative distribution functions for all experimental conditions plays a central role for computing the RMI bound and testing potential violations thereof (see algorithm step 1). Kiesel et al. (2007) defined a bias measure that quantifies the amount of deviation from the true cumulative distribution functions per probability point. Whenever this bias is negative, the RMI test is biased towards violations and thus incorrectly rejecting race models. When it is positive, it further pushes the test towards conservatism as violations are hindered. The authors used synthesized reaction times (using the ex-Wald distribution, Schwarz, 2001) to emulate racers with samples sizes of 10, 20 or 40 per condition. The results show that (depending on the sample sizes of the three conditions) substantial systematic bias pattern

can emerge. These are mostly negative in value and thus tend to produce violations of the RMI. Unfortunately these biases were not restricted to lower sample sizes but would also emerge for the largest sample size in their study (i.e. 40 samples per condition). The authors advise researchers to remain cautious, when they find violations of the RMI for studies with sample sizes lower than 20.

Another aspect this study addressed was the amount of type I error or α accumulation of the test. In the last step of the algorithm the empirical CDF of the redundant condition is compared to the estimated RMI bound. This comparison is realized by applying *t*-tests at typically multiple probability points. As the RMI test is generally considered to be conservative, this multiple testing should increase the power of the test, as it is given more data points to become significant. Each *t*-test is set up with a significance level α to control for the amount of false alarms. Usually α is set to be 0.05, meaning that 1 out of 20 tests is prone to falsely find a statistically significant violation of the RMI bound. When conducting more than one *t*-test, this error probability inflates with each additional *t*-test. A numeric example shall illustrate this fact: Let $\alpha = P(\text{“rejecting } H_0\text{”} \mid \text{“}H_0 \text{ is true”}) = 0.05$, where H_0 is *“the data was generated by a race model”*. Then $P(\text{“retaining } H_0\text{”} \mid \text{“}H_0 \text{ is true”}) = 1 - \alpha$ for one testing the RMI at one probability point of the quantile function. When doing this k times on the same data however, the probability then is $P^*(k) = P(\text{“retaining } H_0 \text{ for k probabilities”} \mid \text{“H0 is true”}) = P(\text{“retaining } H_0\text{”} \mid \text{“}H_0 \text{ is true”})^k = (1 - \alpha)^k$, for $k = 5$ this yields $P^*(k = 5) = 0.774$, and the complementary event of $P^*(k = 5)$, that is, at least on *t*-test is committing a false alarm is $1 - P^*(5) = 0.226$. This effective α error of 22.6% is then more than four times higher than the inner significance level, originally set to control the false alarm rate. Kiesel et al. (2007) simulated race models where the number of participants (20 or 40), RMI test points (5 or 10) and the inner significance level ($\alpha = 0.01$ and 0.05) were varied systematically. The simulations revealed that type I errors (i.e. false alarms) are accumulated to a remarkable degree. And that despite the fact that the *t*-tests are highly correlated across percentiles). The authors provide practical advice how to combat α accumulation, e.g. by applying a stricter significance level (cf. Bonferroni correction, see Holm, 1979), restricting the range of the *t*-test to only early (and theoretically more violation prone) percentile points or replication via independent experiments.

The investigations by both Townsend and Honey (2007), and Kiesel et al. (2007) have increased the knowledge about the RMI test. However, their results simultaneously also increase the uncertainty of what to conclude from applying the RMI test. In cases where violations are missing, a potentially high base time variance might have filtered them out of the race model test function. And in cases where violations are found, this might be due to an estimation artifact of the test algorithm. Furthermore, both simulation studies have several

shortcomings that prevent a direct generalization of their results to other experimental settings or conditions. The most severe deficiency is the lack of a plausible model for the generation of reaction times in their simulations.

In the next chapter, state-of-the-art models of decision making (and in consequence, models of reaction times) will be presented in the form of sequential sampling models. The Ratcliff diffusion model will be detailed, as it forms the theoretical and computational basis for the implementation of both race and coactivation models in the present studies' simulations. The chapter will be concluded by an application of Ratcliff diffusion models to the redundant signals paradigm and the RMI testing. Both known and novel issues of RMI testing will be presented that motivate the simulation and empirical work of the present study.

2.4 Sequential Sampling Models

Experimental psychology has tried to unravel the complex structure of human cognition by using relatively simple tasks, e.g. decisions which requires the observers to choose between two responses. Mental chronometry, in particular, is one standard method to infer the architecture and structure of cognitive processes by analyzing observers' performance. Traditionally, this implied comparing only mean reaction times (e.g. stage theory models; Townsend & Ashby, 1983 Sternberg, 1969) or only the proportion of correct responses (Green & Swets, 1966), across experimental conditions. An experimental manipulation that increases mean reaction times or decreases the accuracy of responses is thought to lower the efficiency of stimulus processing.

This approach, although straight forward and general, suffers from several important limitations. First, it neglects the shape of the response time distributions for correct responses and discards the response times for incorrect responses altogether. This implies ignoring of a large amount of data that can be informative for discerning competing models or architectures of processing. Second, the standard method does not factor in the strong inverse relationship between response time and response accuracy, i.e. the *speed-accuracy trade-off,* (Pachella, 1973 Schouten & Bekker, 1967 Wickelgren, 1977). It is unable to combine the response times and the accuracy in one single index. Third, it is not grounded on any explicated theoretical framework. Thus inferences on their results do not speak directly to the details of the underlying cognitive processes (cf. model mimicry, chapter 2.1).

A conceptual framework that overcomes these limitations and has had a remarkable success in cognitive psychology is the class of *sequential sampling models* (SSM). Sequential sampling models are mathematical models of decision making for simple

decisions, e.g. two-alternative forced choice tasks, discrimination tasks or the like. They have been successfully applied to many response time paradigms, such as lexical decision, short-term and long-term recognition memory tasks, same/ different letter-string matching, numerosity judgments, visual-scanning tasks, brightness discrimination, letter discrimination, and visual search tasks (e.g. Purcell et al., 2010 Ratcliff, 1978, 1981, 2002; Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Rouder, 1998, 2000; Ratcliff, Van Zandt, & McKoon, 1999; Thornton & Gilden, 2007; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008; Ward & McClelland, 1989; Zehetleitner, Koch, Goschy, & Müller, 2013; Zehetleitner & Müller, 2010). More recently, evidence accumulation models have been applied more widely, for example, as general tools to measure cognition in the manner of psychometrics (Schmiedek, Oberauer, Wilhelm, Suss, & Wittmann, 2007; Wagenmakers, Van Der Maas, & Grasman, 2007 Vandekerckhove, Tuerlinckx, & Lee, 2011), and as models for the neurophysiology of simple decisions (e.g. Forstmann et al., 2008; Ho, Brown, & Serences, 2009; Smith & Ratcliff, 2004). They provide an explicit mechanism for decision formation and rest on neurologically plausible assumptions while allowing for direct psychological interpretation of the featuring model parameters.

They can be classified according to four dimensions: their *stopping rule* (relative versus absolute), their *time scale* (discrete vs. continuous), the *granularity of evidence* (discrete vs. continuous), and in their *integration mechanism* (leaky versus perfect) (for a review see Ratcliff & Smith, 2004). Sequential sampling models that have a relative stopping rule assume that evidence in favor of one response is evidence against the other response alternative. They are also called *random walk models* (Laming, 1968; Link & Heath, 1975; Ratcliff, 1988; Stone, 1960). Random walk models can be further partitioned into pure *random walk* and *diffusion models*. Random walk models accrue amounts of evidence in discrete time steps, while diffusion models assume a continuous time scale. Finally, diffusion models can have a constant drift rate (so called *Wiener diffusion model*) or a drift that decays with time (so called *Ornstein Uhlenbeck model*). In recruitment or absolute stopping rule models, evidence for each response is accumulated independently. Amongst recruitment models, the subclass with discrete time steps are the *accumulator models* (e.g. Smith & Vickers, 1988; Vickers, 1970, 1978, 1979; Vickers, Caudrey, & Willson, 1971) and the variant with a continuous time scale the *Poisson counter models* (LaBerge, 1994; Pike, 1966, 1973; Townsend & Ashby, 1983).

In the following, the *Ratcliff diffusion model*, as a distinguished exemplar of SSM and the generating mechanism for the response times in the present study, will be described. (Implications of selecting this model above other sequential sampling models will be discussed at the end of the present study). Then, the advantages and disadvantages of a

diffusion model analysis will be presented, before applications of SSM to the redundant signals paradigm will be presented. A motivation of the research question of the present study will conclude this chapter.

2.4.1 The Ratcliff Diffusion Model

Historically, the Wiener diffusion process, which is at the basis of the Ratcliff diffusion model, was first studied in physics (e.g. Einstein, 1905). Later, Roger Ratcliff implemented and modified it to decision making in simple two-alternatives forced choice (2AFC) tasks (Ratcliff, 1978). The RDM provides a detailed account of observers' performance, that is, response time distributions for both correct and incorrect responses. For concreteness sake, let an observer perform a simple discrimination task, of discriminating left versus right oriented Gabor patches. The RDM assumes the observer initiates a process of noisy information accumulation. That is, the observer samples information via its sensors (here, the photoreceptors in the retina) to determine the extent to which each new sample of information supports one decision versus the alternative decision (here, left versus right orientation of the Gabor patch). The process is illustrated in Figure 2.7.

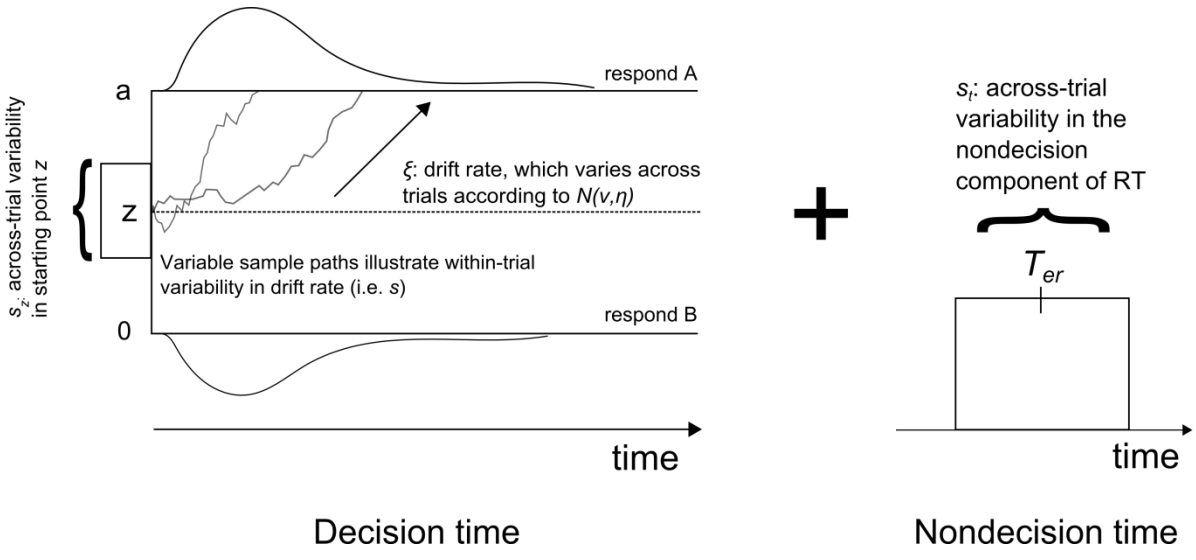

Figure 2.7 Ratcliff diffusion model with parameters. The decision process is modeled as an accumulation of noisy evidence over time. Once an absorbing barrier is reached, the respective response behavior is triggered. The overall response time is the sum of the decision time and a uniformly distributed nondecision or base time. See text for a detailed account of the parameters.

An important difference to signal-detection theory (Green & Swets, 1966) is that the observer does not base its decision on a single noisy sample, but on an entire sequence of noisy samples. In that sense, the RDM (and SSM's in general) can be viewed as a natural extension of signal detection theory over time. The diffusion model assumes that the signal-to-noise ratio of the evidence accumulation process is higher for stimuli that are easy to classify than it is for stimuli that are difficult to classify. The "ease of processing" is quantified in diffusion models by a parameter called drift rate. When the drift rate is high, decisions are fast and accurate, when it is low processing is determined to a stronger extent by noisy fluctuations and as a consequence decisions are slow and inaccurate. In the previous example, Gabor patches with a low feature contrast would have a lower drift rate than Gabor patches that are visually clear with a high feature contrast. The drift rate reflects an inherent property of stimuli and participants, and is generally not believed to be under subjective control. The evidence accumulation process is formally described by the following stochastic differential equations (e.g. Gardiner, 2004):

$$dX(t) = vdt + sdW(t) \tag{17}$$

where *dX(t)* is the change in the accumulated evidence *X* for a small time interval *dt*, *v* is the drift rate and *sdW(t)* are zero-mean random increments with infinitesimal variances s^2dt. The factor *W(t)* represents the Wiener noise process (i.e. idealized Brownian motion). The drift rate then is the deterministic and the Wiener process is the stochastic component of the accumulation process. The amplitude of the noise in the process is governed by the parameter *s*: It is a scaling parameter, as a doubling of *s* causes a doubling of the other parameters in the model. Therefore the choice of a specific *s* is arbitrary; in the literature s is usually fixed at 0.1.

Two parameters of the diffusion models that are under subjective control is the boundary separation *a* and the starting point *z* of the accumulation. Both of them are assumed to be set before the start of each new trial. The boundary separation quantifies the response conservatism and modulates the speed-accuracy trade-off: when an observer is focused not to make a mistake, the boundaries are set wide apart. As a result noisy fluctuations are less likely to provoke an incorrect response. The downside of this decrease in error rates is an increase in the response times. When the boundary separation is large, the evidence must accumulate to a higher value in order to reach the response-eliciting threshold. This results in slower response times and lower error rates.

The starting point *z* reflects the a priori bias of an observer favoring one response over the other. This parameter is usually manipulated via payoff or proportion manipulations (Diederich & Busemeyer, 2006; Edwards, 1965). In the thought Gabor experiment, a high

prevalence of left oriented Gabor patches would shift the starting point towards the left oriented boundary. In consequence, this would lead to relatively fast and accurate responses for left oriented patches, compared to relatively slow and inaccurate responses for right patches. The fourth parameter of the diffusion model, *Ter,* quantifies the nondecision component of response times. In line with the canonical decomposition of response times, it encompasses all the time taken up by processes not affecting the decision process. Therefore, this nondecision time shifts the entire response time distribution by a constant amount.

Three additional parameters were subsequently introduced, as the model so far could not capture relevant empirical findings in the response time literature. First, uniformly distributed trial-to-trial variability in nondecision time, denoted by s_t was added to account for the relatively gradual rise in the leading edge of the response time distribution (Ratcliff & Tuerlinckx, 2002). Second, uniformly distributed trial-to-trial variability in starting point, called s_z, was introduced to account for error responses that are systematically faster than correct responses (Laming, 1968). Third, normally distributed trial-to-trial variability in drift rate, called η, was introduced to capture error responses that are systematically slower than correct responses (e.g. Ratcliff, 1978; Ratcliff & Rouder, 1998).

Taken altogether, this yields the following seven parameters of the Ratcliff diffusion model:

1. mean drift rate v,
2. across-trial variability in drift rate η,
3. boundary separation a,
4. mean starting point (or bias) z,
5. across-trial range in starting point s_z,
6. mean of the nondecision component of processing *Ter*,
7. across-trial range in the nondecision component of processing s_t

2.4.2 Evaluation of the Ratcliff Diffusion Model

The Ratcliff diffusion model embodies the most recent advancements in mental chronometry. It provides a principled and explicit account of how decisions are formed as a function of properties of the observer and the environment the observers is exposed to. Its advantages can be ascribed across different levels. On a behavioral level, the model can account for changes in both accuracy and response time data on a coherent theoretical framework (i.e. noisy evidence accumulation) that is compatible to the speed-accuracy trade-off. Its parameters can be interpreted in a psychologically direct way, as they unambiguously

map to psychological and experimental properties (for an empirical validation of the parameters, see Voss, Rothermund, & Voss, 2004). Thus it provides a means to estimate the unobserved psychological processes (decision and nondecision) that together determine the observed performance. This further distinguishes the model from other response time models (e.g. ex-Gaussian, ex-Wald, Gamma and Weibull, see Palmer, Horowitz, Torralba, & Wolfe, 2011) that merely allow for an atheoretic description and compression of the data. While these response time models have their value in statistically describing and compressing data, inferences, and psychological interpretations on their parameter changes should be avoided (see Matzke & Wagenmakers, 2009). Also on a statistical level, the RDM extracts a far greater amount of information provided by the performance of the observers than standard methodology. It acknowledges both the response distribution for correct and incorrect responses and has displayed excellent fits to the data.

More importantly, it is a falsifiable model, as there is data it can not account for. Ratcliff performed a simulation study for a brightness discrimination experiment and could show that all the experimental manipulations were fit plausibly to the model but not fake data that only appeared plausible (Ratcliff, 1978). The basic properties of the model are also consistent with neurophysiological data on the build-up of evidence (Brown, Steyvers, & Wagenmakers, 2009; Usher & McClelland, 2001).

Arguments against the use of diffusion model analyses are few. From a model fitting aspect, a diffusion model is not the most parsimonious model. It features seven parameters and - depending on the concrete implementation - at least one parameter per experimental condition. This however is not an instance of overfitting, as the amount of data to account for is increased compared to the standard approach (two distributions compared to one mean RT and accuracy rate, see Matzke & Wagenmakers, 2009). From a computational aspect, the model requires a decent amount of data to accurately estimate the parameters (at least 200 per condition), which can be troublesome for some experimental settings and research questions. Moreover, the fitting procedure is both computationally and theoretically demanding as it involves numerical optimization in a multidimensional parameter space. By now several program tools exist, that enable researcher to quickly and reliably fit their data to the model (see Vandekerckhove & Tuerlinckx, 2008; Voss & Voss, 2007; Wagenmakers et al., 2007).

2.4.3 Applications of Sequential Sampling Models to the paradigm

Sequential sampling models have been already applied to the redundant signals effect. Initially Schwarz (1989) proposed a *superposition model* based on Poisson counters.

There, for each stimulus channel an ordinary renewal process (with a Poisson distribution; Smith & Vickers, 1988) accumulates counts until a response triggering critical value of counts has been registered. The main assumption of the model is, that for the redundant conditions the two single renewal processes simply superposed – that is, the counts of both channels are summed up. This leads to a shorter time to reach the critical amount of counts. Further analysis revealed that this model imposed severe restrictions on the resulting response time distributions (the mean and the standard deviation of the intercount intervals must necessarily be equal). Also, the stochastic independence between channel latencies may often prove to be too restrictive. Subsequently, Schwarz presented a modified superposition model which allows for a more flexible type of evidence accumulation than simple Poisson counting (Schwarz, 1994), which also permits arbitrary amounts of channel correlation. The modification included that the channel activity is modeled by a Wiener diffusion process. The critical assumption was that the drift rates of both stimulus diffusors sum up for the redundant diffusor. Schwarz illustrated the performance of this model using data from Miller (1986)

Diederich on the other hand, proposed a *multichannel diffusion model* based on the Ornstein-Uhlenbeck process for the RSE. The author compared the fit of this model to the fit of Schwarz' linear superposition model for a trimodal intersensory facilitation experiment (Diederich, 1995). She came to the conclusion that the fit (with respect to the means) was good for the linear superposition model and even better for the multichannel diffusion model. Also both models agreed on the ordering of the experimental conditions. However, both models failed to capture the standard deviations of both the double and triple stimulus conditions. Her conclusion was that the reason for the poor fit of the standard deviation might be due to the fact that the base time is not independent of the decision time. It is important to point out that neither model featured variability in base times.

While these studies did apply current models of decision making to the RSE, both assumed a decisional stage for the locus of the effect (although Diederich briefly discusses the possibility of a partial base time locus). To justify this assumption, it is necessary to demonstrate that this locus model is the best fitting model variant. Moreover, the fit has to be assessed beyond the level of means.

In the final part of this chapter, an integration of all the open questions is provided before presenting the research questions of the present study.

2. 5 What is missing?

The previous parts of this chapter highlighted several shortcomings and open questions concerning the analysis of cognitive architecture for decisions based on multiple signals via mental chronometry. Next, a recapitulatory catalogue of question is compiled that motivated the research questions of the present study.

2.5.1 Catalogue of Knowledge Gaps

First, both simulation studies available use descriptive and atheoretic distribution functions to synthesize reaction time data. This prevents researchers from inferring on the model parameters in a psychologically valid manner. Also, the functional form of the synthesized data is not validated and thus reduces the generalizability and confidence in the acquired results. Current models of decision making (such as sequential sampling models) offer a potent alternative for simulations, as the present study will demonstrate.

Second, no study has implemented non-extremal cases of either coactivation or race models. This would require controlling for the strength of coactivation and the amount of correlation between the racers respectively. This in turn requires the implementation of channel correlation and an explicit model of coactivation, with help of parameters controlling these properties of the data. While the extreme cases of race models and full integration models can be informative for the border area of the RMI test, it is quite probable, that empirical data is not of that sort (see Ulrich & Giray, 1986 for the challenges of estimating channel correlation). All the more, it is important to ask how the RMI test performs (in terms of type I error or α accumulation, and power rates), when the data is only moderately correlated or coactivated.

Third, due to lack of a coactivation account in the Kiesel, Miller and Ulrich (2007) study, and the missing analysis of estimation bias in the Townsend and Honey (2007) study, it is unknown whether there is an analogous estimation bias for coactivation models, and in what direction it might work. In case there is no estimation bias for the coactivated case or there is one working systematically against violations, the recommendations by Kiesel et al. would render the test even more conservative.

Fourth, it is unclear what effect a highly varying base time component will have on the estimation bias, the type I errors and the power rates for an empirically plausible reaction time model (here, the Ratcliff diffusion model). While this has been addressed theoretically and computationally for implausible reaction time models (Townsend & Honey, 2007), it is still unexplored for mechanistic models of decision making (i.e. sequential sampling models).

Fifth, there have been attempts to compare different sequential sampling models in the redundant signals context. Those were however restricted to decisional accounts only and assessed the goodness-of-fit just at the level of means (Diederich, 1995; Schwarz, 1994). In order to justify this locus assumption of the RSE, a fitting study is necessary that compares goodness-of-fit of all the coactivation model variants in. This implies testing explicit coactivation accounts, where the effect is generated either by decisional, nondecisional or a combination of both components of the reaction times. Again, sequential sampling models are the model of choice, as all of these variants can be naturally adopted within their framework, as will be shown in the present study.

Finally, to guarantee the explanation of the data to a full(er) extend, the fitting must be expanded to the whole spread of the reaction time data and not the mean level only. This way the pervasive issue of model mimicry in mental chronometry can be alleviated. Specifically, this comprise implementing fitting procedures that quantify the difference between the cumulative distributions functions (Ratcliff, Thapar, & McKoon, 2003).

2.5.2 Research Questions

Motivated by these shortcomings, the present study pursues theoretical and methodological goals. On the theoretical side, the locus of coactivation is assessed. For this purpose, a comparative fitting of highly coactivated empirical RSP data to the different coactivation model type (decisional, nondecisional or a combination of both) is performed and analyzed. This diffusion model analysis will provide an answer to the question of which coactivation account is in best agreement with experimental data. Apart from using sequential sampling models to test the fit of different loci models of the RSE (and violations of the RMI), they can also help in evaluating the RMI test itself. Due to their formal nature and explicit formulation, sequential sampling models allow for a numerical implementation and enable researchers to conduct large scale simulation studies to scrutinize the already presented shortcomings of the RMI test. Methodologically, the present study validates and generalizes the findings by Kiesel et al. (2007) and Townsend and Honey (2007), using the Ratcliff diffusion model to synthesize both race and coactivation data. Being able to control crucial aspects of both these models (in form of the correlation and strength of integration between the channels respectively) as well as the base time component will help to assess their impact on the performance of the RMI test. This approach also signifies an integration of the previous disjunct scopes into a common framework that will help researchers to optimally set up the RMI test for different experimental conditions.

Specifically, the present study aims to answer the following explicit research questions: (i) what is the locus of coactivation for two bimodal detection tasks (chapter 3), (ii) what is the power of the RMI test, (iii) is there an estimation bias in the coactivated case and if so, in what direction is it working, (iv) what is the effect of a high base time variance on type I errors, power rates, and estimation bias, (v) and particularly when using a sequential sampling model for the reaction times (both chapter 4).

2.5.3 Significance of the present Study

The present study marks a natural and necessary next step in probing and understanding the structure of cognitive processes. It combines the logic and rationale of mental chronometry with the most recent and auspicious models of decision making to the questions of architecture for multisignal integration and test quality. Answers to the questions will directly improve the theoretical understanding of both coactivation and the RMI test. In addition, the practical value and efficiency of the RMI test will be improved, by solidifying results that have already been or will be obtained in the future.

3 MODELING COACTIVATION IN BIMODAL DETECTION TASKS

The redundant signals effect is a shortening in mean reaction times (RT), when observers process a redundantly defined stimulus compared to processing a target, defined in only one quality (i.e. feature, dimension or modality; see chapter 2.2 for a detailed description of the redundant signals paradigm). The race model inequality test allows inferring the cognitive architecture by putting a distributional constraint on this speed-up. Whenever this RMI bound is violated, this is taken as evidence against the whole class of race models and in favor of integrative or coactivation models (see chapter 2.3 for a formal derivation of the RMI test and its statistical profile).

Even when researchers find violations of the RMI, it is still unclear at what stage(s) of the decision process an integration occurred. Decisional accounts assume that the RSE is at least to some degree due to a speed-up at the level of early visual stimulus analysis. This is the stage, where target attributes are coded and compared with those of non-target elements, without necessarily involving stages following attentional selection (Koene & Zhaoping, 2007; Krummenacher et al., 2001, 2002; Töllner et al., 2011; Zehetleitner et al., 2009b; for a review see Zehetleitner et al., 2008b). There already exist two explicit models of a decisional account of the RSE (Diederich, 1995; Schwarz, 1994), however both only show a good fit on the level of means, while being unable to account for the spread of the data.

Contrary decisional accounts, there are researchers who believe that the RSE is a nondecisional effect and post-selective in nature (Corballis, 1998; Feintuch & Cohen, 2002; Iacoboni & Zaidel, 2003; Miller, 2007; Miller et al., 2009; for a review see Miller & Reynolds, 2003). From mean level only, those coactivation accounts cannot be discriminated as only the overall reaction time is measurable and not its latent subcomponents (see however, motor related measures in Ulrich & Giray, 1986; Ulrich & Stapf, 1984). Part of the reason is that the used reaction time models were of a descriptive and atheoretic nature.

Sequential sampling models in contrast, by design allow for an implementation of both decisional and nondecisional accounts to the RSE (see chapter 2.4 for an elaborate description of this model class). However, so far only decisional accounts of the RSE have been empirically tested, as both Diederich's (1995) and Schwarz's (1994) models assume a summation in the rate of evidence accumulation for redundant signals over single signals. There are no studies which have tried to fit nondecisional or combined coactivation accounts (where both decision and nondecision parameters may vary) in a comparative fashion.

Apart from this conceptual lack, all of the models so far only look at fits to the means. As reaction times are usually positively skewed, there are distinct patterns, which can be missed, when focusing only on the central moment. Balota and Yap (2011) showed that effects in means can be produced by either shifts of RT distributions, or stretching of slow tails of RT distributions, or a combination of the two. In the context of sequential sampling models, Ratcliff, Thapar, Gomez and McKoon (2004) have argued, that a complete explanation of processing, requires accounting for all aspects of the experimental data. This encompasses the distributions for correct and incorrect responses as well as the proportion of correct and incorrect responses.

As the redundant reaction times are not only mean shifted but also differ in their skewness (mean-variance relation, Wagenmakers, Grasman, & Molenaar, 2005) an analysis on the mean level is in principle not able to distinguish subtle parameter shifts, which do not exclusively manifest in a change in means (for the problem of model mimicry, see chapter 2.1). Also, response times are known to show a positive correlation between mean and variance: higher mean response times are afflicted with a higher variance and vice versa. Diffusion accounts intrinsically produce this relation of means and variances for response times. Due to this correlation, a nondecisional model alone is unlikely to fit the data adequately, as there a change in the base time parameter is only capable to shift the whole distributions horizontally. It is unclear whether a combined (decision and nondecision) model can outperform a decision-alone measured in its goodness-of-fit.

This study contributes to the debate on the source of RMI violations and coactivation both conceptually and methodically. A diffusion model analysis was performed to fit quantile proportions of the response times in two bimodal audiovisual RSP experiments to three model variants reflecting different sources of coactivation: (a) a decisional model (drift rate may vary), (b) a nondecisional model (base time may vary), and (c) a combined model (both drift rates and base times may vary). This way, the question of the source(s) of RMI violations (and the RSE in consequence) can be addressed: Does coactivation occur at a decisional stage, a nondecisional or at both stages and to what degree?

On a methodical level, the present study promotes the application of sequential sampling models (the Ratcliff diffusion model in particular) to the redundant signals paradigm, as it enables researchers to utilize all of their empirical data (RT distributions for both correct and incorrect response, and error rates) to plausibly map it to latent psychological variables and generate theory-driven hypothesis to further shed light on the redundant signals paradigm.

To answer this question satisfactorily however, diffusion model analyses of a large class of redundant signals experiments with varying tasks and stimuli have to be performed. In this study, the endeavor is commenced by analyzing to bimodal RSP tasks: Experiment 1 utilizes a simple reaction time task and Experiment 2 a two-alternatives forced choice task.

3.1 Experiment 1 (audiovisual SRT)

In Experiment 1, a simple RT task was performed, where participants had to respond to the central onset of either visual, auditory or audio-visual stimuli by pressing both mouse button simultaneously. A variable intertrial interval was used to prevent anticipatory responses and contaminate response times.

3.1.1 Method

Subjects

In Experiment 1, 15 subjects (11 of them female) participated in a single 45-minutes session in return for 6€ or a course participation. Their average age was 25.7 years (range, 20 – 34) and they were all right-handed and had normal or corrected-to-normal vision.

Apparatus

The Experiment was conducted in a sound-insulated booth, and was controlled by a program using MATLAB (R2009bSP1, Natick, Massachusetts: The MathWorks Inc., 2010) and the PsychToolbox (Brainard, 1997; Pelli, 1997) running on an Apple Mac mini (Cupertino, California: Apple Inc.) machine (with Mac OS X). The stimuli were presented on a 20” Mitsubishi Diamond Pro 2070SB monitor at a resolution of 1,280 × 1,024 pixels and a refresh rate of 100 Hz with a viewing distance of approximately 73 cm.

Stimuli

Visual stimuli were grey discs (CIE Yxy 10.9, .286, .333; 1° in diameter), auditory stimuli were 400 Hz beeps (with a duration of 150 ms) delivered by headphones, and redundant stimuli were both visual and auditory stimuli presented simultaneously (stimulus onset asynchrony of zero). The participants responded to the onset of the stimuli by simultaneously pressing both mouse buttons with both their index fingers.

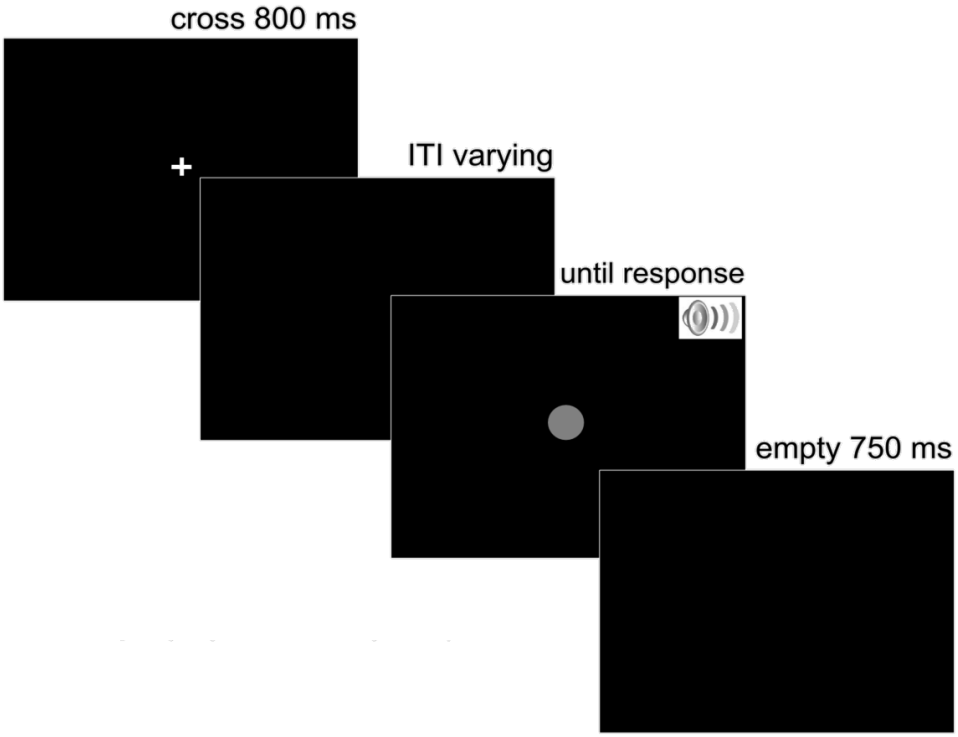

Figure 3.1 Display progression of the SRT Experiment. First, a fixation cross is presented centrally for 800 ms. After a variable intertrial interval the target (single or redundant) appears and remains until the observer responds bimanually. A blank screen follows for 750 ms before the next trial begins.

Procedure

Every trial was structured in the following way: First, a white fixation cross (0.5° x 0.5° of visual angle) was presented centrally on a black screen for 800 ms. Then, after an intertrial interval (uniformly varying between 500 and 1500 ms), the stimulus or stimuli appeared and remained on the screen until a response was elicited. After the observers' response, a 750 ms waiting period followed, before the next trial started again with the fixation cross (see Figure 3.1 for the display progression).

The whole experiment was divided into 17 blocks à 45 trials, where unimodal and bimodal trials were interchanging randomly. Overall, this amounted to 765 trials (255 trials per condition). Participants could make a break in between blocks and feedback on their mean RT and error rate was given. They were instructed to respond as fast as possible while remaining below 5% error rate.

3.1.2 Data Analysis

All analyses and the numerical parameter fitting were done with GNU R (version 2.14.0). For the fitting procedures the R-package "*optim"* was used.

Implementation of the Models

The diffusion model variants were implemented in the following way. The *free model* allowed all of the RDM parameters (that is, drift rate, criterion, base time and base time variance) to vary across the three conditions (auditory, visual and redundant trials). It was implemented to assess the performance of a completely unconstrained RDM. The free model is not a viable candidate, as it is theoretically implausible: it allows a conditions-specific adaptation of the criterion. This would imply an instantaneous detection of the target condition, which in itself necessitates a successful target detection and identification.

Theoretically plausible candidates are the *decisional model*, the *nondecisional* and the *combined model*. In contrast to the free model, some parameters are fixed and cannot vary. For the decisional model, all the parameters except the drift rates v_c are held fixed for the three target conditions c = auditory, visual and redundant (i.e. audiovisual). The nondecisional model on the other hand, allowed only the base time parameters (Ter_c and s_{tc}) to vary across the conditions, whereas the *combined model* allowed both drift rates v_c and base time Ter_c to be fit individually for each condition c.

The parameter *s,* controlling the starting point of the evidence accumulation process was set to $a/2$, resulting in an unbiased decision process. The parameters η and s_z of the RDM were both set to zero (for example, the EZ-diffusion model Wagenmakers et al., 2007 makes the same simplifying assumptions). As these represent only auxiliary parameters to increase certain aspects of the fit, these were eliminated here to improve the fitting performance (see chapter 2.4.1 for an account of all RDM model parameters). Table 3.1 gives an overview of the free and constrained parameters for each coactivation model variant.

Table 3.1

Coactivation Models with fixed and shared Parameters, and Degrees of Freedom.

Model	Free Parameters	Constrained Parameters	Degrees of Freedom
Decisional	v_a, v_v, v_{av},	a, Ter, s_t	3 + 3 = 6
Nondecisional	Ter_a, Ter_v, Ter_{av}	v, a, s_t	6 + 2 = 8
Combined	v_a, v_v, v_{av}, Ter_a, Ter_v, Ter_{av}	a, s_t	9 + 1 = 10
Free	v_a, a_a, Ter_a, s_{ta}, v_v, a_v, Ter_v, s_{tv}, v_{av}, a_{av}, Ter_{av}, s_{tav}	(none)	3 x 4 = 12

Notes. In the decisional model, only the drift rates *v* could vary across the Experimental conditions (auditory targets *a*, visual targets *v* and redundant targets *av*). In the nondecisional model only the motor parameter *Ter* was allowed to vary. In the combined model both the drift rates and the motor parameters could be fit individually for each condition. The free model allowed every parameter to be fit for each condition and is used to obtain a base line of how good diffusion models can fit the data.

Quantile Distribution Functions

In order to find the model (and the corresponding parameters) that can best explain the data of Experiment 1, a fitting of quantile proportions was performed. These were computed by use of quantile probability functions. Quantile probability functions plot response probabilities against quantile response times. The probability of a response for a particular stimulus type determines the position of a point on the X-axis, and the quantile RTs for that stimulus type determine position on the Y-axis (Ratcliff, Thapar, et al., 2004). Figure 3.2 displays the empirical quantile proportions of Experiment 1. Consistent to the mean-variance relation, the fastest condition (here, the bimodal redundant targets) also displays the narrowest response time range.

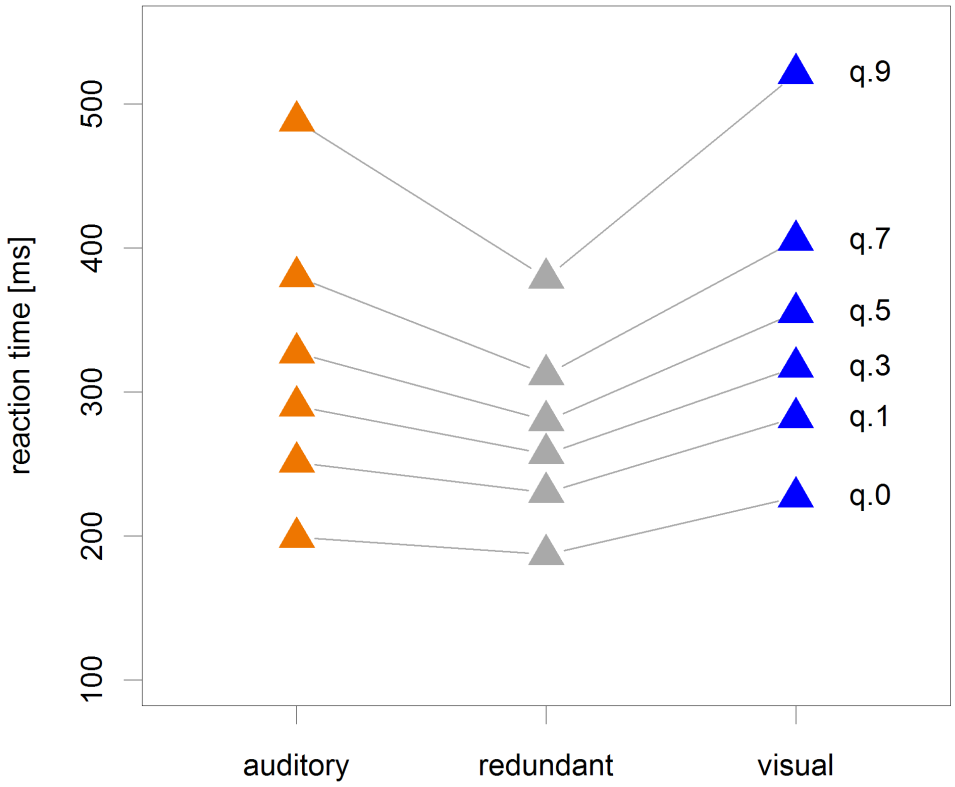

Figure 3.2 Quantile function plot of the response times for the three Experimental conditions in Experiment 1. Quantile functions give a fuller description of the reaction time data than mean and standard deviation values only, as the proportion in each quantile bin is visible as well as the spread of the entire distribution.

Cost Function

In order to quantify the goodness-of-fit of the Experimental data to a model, a cost function is necessary. In this study, the *BIC statistic* (Raftery, 1986) summated with a weighted error rate term was implemented. The original BIC statistic is defined as follows:

$$BIC = 2[\sum N_i p_i \ln(\pi_i)] + Mln(N) \quad (18)$$

where p_i and π_i are the proportion of observations in the *i*-th bin for the empirical data and the prediction of the model, respectively, and $MIn(N)$ is the penalizing term related to the number of free parameters *M* and the sample size (*N*; the number of observations) (Gomez, Ratcliff, & Perea, 2007). As all of the models under consideration have different numbers of free parameters, this term is necessary to penalize overfitting as a result of added parameters to the model. Because of the low error rates, the quantile proportions of the incorrect responses were not used for the fit. Instead, the error rates of the empirical data

and the model data were subtracted and then added to the BIC score as the weighted absolute value of this subtraction. This third summand was added to equation 18. The overall cost function then was,

$$BIC' = 2[\sum N_i p_i \ln(\pi_i)] + Mln(N) + w \cdot \sum |err_c^e - err_c^m| \quad (19)$$

where, err_c^e denotes the empirical error for target type c = auditory, visual, audiovisual and err_c^m the model error for target type. The weight *w*, was chosen to be 600 for each condition. This amounts to roughly one sixth of the minimal BIC score (that is, when all model and empirical proportions are equal) for sample size 1000.

Fitting Procedure

The generic fitting procedure for each model variant consisted of four computation steps. First, a starting parameter vector was randomly sampled. By design, it consisted of the RDM parameters for every of the three target types (that is, auditory, visual, and audiovisual). The exact composition of this vector varied depending on the model that was being tested. For example, the decisional model only allowed the drift rates to vary; all other RDM parameters were fixed across the target types.

Second, a sufficiently large number of reaction times was synthesized, instantiated with this initial parameter vector. Here 10‘000 response times were generated, in order to get an unbiased estimate of the (empirical) cumulative distribution function of each condition. The CDF was used to extract the seven quantiles, 0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0.

Third, the quantile response times of experiment and diffusion model were used to generate the predicted cumulative probability of a response by that quantile response time. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between each quantile (ideally this yields 0.1, 0.2, 0.2, 0.2, 0.2, 0.1). The observed and expected proportions were multiplied by the number of observations to produce expected frequencies.

Fourth, the cost for this parameter value was computed (equation 19) and minimized, using a general SIMPLEX minimization routine (Nelder & Mead, 1965; implemented in the “*optim*” package for R), that adjusts the parameters to find the parameters that give the minimum score for each model. The model with the lowest BIC can be considered the model that jointly maximizes descriptive accuracy (goodness of fit) and parsimony (small number of free parameters).

Model Selection

The fitting procedure was performed in a Monte Carlo fashion (using 1000 iteration runs) with the four computation steps outline before. The minimum cost value for every condition was used to assess which model was in best agreement to the data and with what specific parameter vector. In addition, the raw BIC values were transformed to a probability scale, enabling a more intuitive comparison of the probabilities of each model being the best model (Wagenmakers & Farrell, 2004).

The transformation of BIC values to probability values consists of three steps: First, for each model *i*, the difference in BIC with respect to the model with the lowest BIC value is computed, i.e. $\Delta_i(BIC)$. Second, the relative likelihood *L* of each model *i* is estimated by means of the following transformation:

$$L(M_i|data) \propto \exp[-0.5 \cdot \Delta_i(BIC)], \quad (20)$$

where $\propto$ stands for "is proportional to". Last, the model probabilities are computed by normalizing the relative model likelihoods, which is done by dividing each model likelihood by the sum of the likelihoods of all models.

3.1.3 Results

Errors

Errors were defined as anticipatory responses (*RT* ≤ 150 ms) or misses (*RT* > 1600 ms). Observer errors were 3.10%. One observer produced more than 10% errors; this data was discarded.

Mean reaction times

The mean RTs for the unimodal and bimodal conditions are listed in Table 3.2. The participants were generally faster in the auditory condition than in the visual condition (360 ms compared to 390 ms). There was a pronounced RSE of 55 ms, as the bimodal condition produced a mean RT of 305 ms. Table 3.2 summarizes the mean reaction times and standard deviations aggregated across all participants.

Table 3.2

Mean Response Times for Experiment 1 (SRT)

Condition	Mean RT [ms]	Std. Dev.
Auditory	345	84
Visual	379	73
Audiovisual	290	57
RSE	55	30

Notes. Mean response times (in ms) of the unimodal, single signals trials (auditory, visual) and the bimodal, redundant signals trials (audiovisual) with corresponding standard deviations of Experiment 1. The data exhibits a pronounced redundant signals effect of 55 ms across all observers.

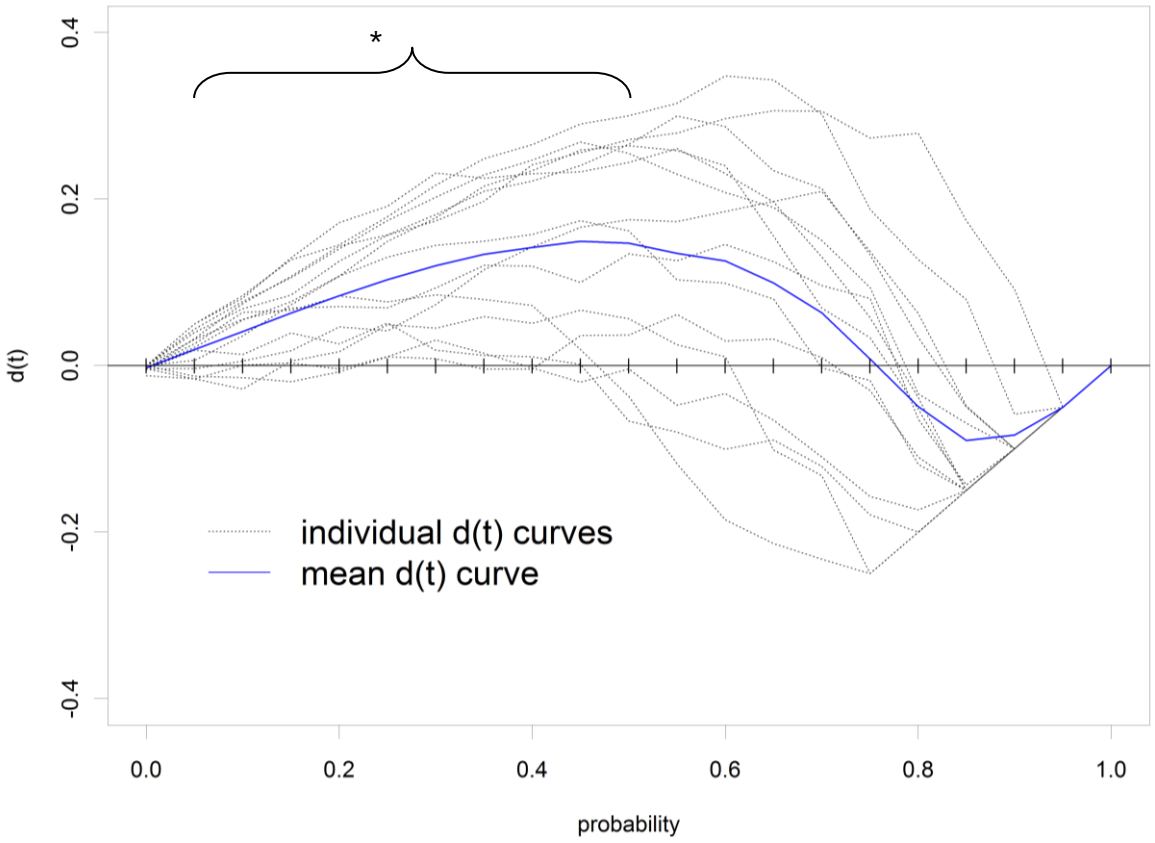

Figure 3.3 Race model test function aggregated across individual observers (blue full line) and for each individual observer (grey dotted line). Violations of the Race model inequality were found for probability points 0.05 to 0.50 using multiple *t*-tests with a Bonferroni-corrected significance level of 0.0026.

RMI Violations

The violations of the RMI were tested according to the explicit algorithm by Ulrich, Miller and Schröter (2007). At the probabilities p = 0.05, 0.10, …, 0.95 (two-sided) t-tests were performed with a Bonferroni-corrected (see Holm, 1979) α of 0.0026 (= 0.05 / 19 probability points). Significant violations ($p < 0.0026$) were found across the ten probabilities 0.05 to 0.50. Figure 3.3 shows the individual and mean RMI test function curves for Experiment 1.

Fitting Results

The outcome of the fitting procedure for Experiment 1 were cost values of 10626, 10631, 10551, and 10578 for the decisional, nondecisional, combined and free model respectively. Thus, the best fitting model was the combined model. Even though it had comparably many free parameters, it outperformed the second best fitting model (here, the free model) by 27 units in a logarithmic scale. The probability transformation further emphasizes the clear advantage of the combined model over the others, as its probability is at over 99.99% (see Table 3.3 for an overview of the cost values and probability estimates). The fitting performance of the decisional and nondecisional model is numerically close, however on a logarithmic scale, their difference in BIC score amounts to a clear advantage for the decision model.

Table 3.3

Minimum BIC values for each model in the SRT Experiment.

Model	Degrees of Freedom	BIC	Probability
Decisional	6	10626	5.18×10^{-17}
Nondecisional	8	10631	4.25×10^{-18}
Combined	10	**10551**	9.99×10^{-01}
Free	12	10578	1.37×10^{-06}

Notes. The combined model exhibited the lowest BIC score and had the best fit to the data. (For the computation of the model probability, see equation 20 in 3.1.2.)

Table 3.4

Mean RTs (in ms) and Standard Deviations of empirical and model Data for Experiment 1

Condition	Empirical	Decisional	Nondecisional	Combined	Free
Auditory	345 (84)	345 (90)	339 (89)	346 (106)	345 (105)
Visual	379 (73)	385 (119)	368 (89)	379 (101)	375 (102)
Audiovisual	290 (57)	293 (56)	311 (89)	290 (66)	291 (64)
RSE	55 (30)	52	28	56	54

Notes. The mean reaction times of the models are in good agreement to the empirical data. The standard deviations (in braces) are at variance with the data. Except for the nondecisional model, all models capture the redundant signals effect of the empirical data.

On the level of means, the goodness-of-fit of the models was in good agreement to empirical data. Only the nondecisional model did not succeed in appropriately reproducing the mean reaction times and RSE (see Table 3.4).

Error Rates

Looking at the error rates that the models produced, reveals that only the free model could approximately mimic the empirical errors; the decisional model generated virtually no errors, while the nondecisional model (due to its design) could only generate identical error rates across the conditions. The combined model can qualitatively reproduce the empirical errors, except for the redundant condition, where it only generates around 0.48% compared to the empirical 2.79% (see Table 3.5).

Table 3.5

Error Rates of Experiment 1 and the Coactivation Models.

Condition	Empirical	Decisional	Nondecisional	Combined	Free
Auditory	2.81%	0.00%	1.92%	2.59%	2.62%
Visual	1.67%	0.00	1.96%	2.21%	1.51%
Audiovisual	2.79%	0.00	1.96%	0.48%	2.52%

Notes. The free model was the only model to reproduce the empirical error pattern. The decisional model produced no errors, the nondecisional model (due to its parameter constriction) only identical errors across conditions. The combined model produced plausible error rates for the single signals trials (auditory and visual).

Goodness-Of-fit

The quantile function plots of the respective model reveal a picture coherent to the BIC scores and mean results (see Figure 3.4). Apart from the lowest quantile, both the combined and free model display an exceptionally good fit to the distribution of the empirical data. The decisional model generally underestimates the lower quantiles (0.0, 0.1) for all conditions, and overestimates the highest quantiles (0.7, 0.8) for the visual condition. The nondecisional model fails to reproduce the spread of all the conditions, most pronounced for the redundant condition.

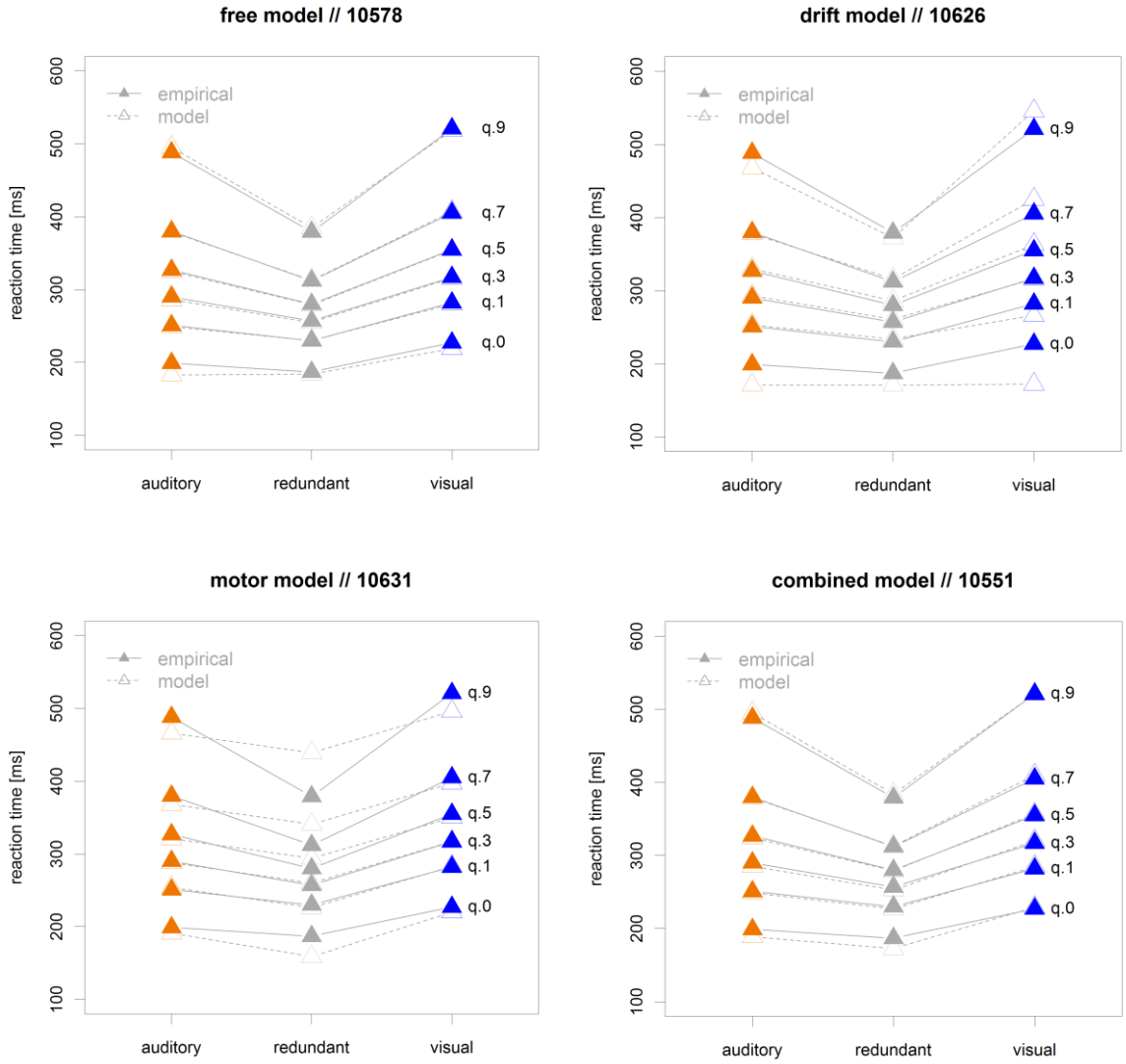

Figure 3.4 Quantile function plots for all the models. Full triangles denote empirical data points, while unfilled triangles show the data points of the respective models. The BIC score of each model is given in the title of each graph.

Parameter Analysis

The parameters, producing the best fit for each model are depicted in Figure 3.5. From a qualitative view, the free, motor and combined model agree in the range of the drift rates, criteria and base times for the three conditions arguably well. All models exhibit the highest drift rate for the redundant condition, while only the drift rates of the decisional model reflect the rank of the conditions, that is, the redundant condition having the highest drift rate and the visual condition having the lowest drift rate. The base times are lowest for the decisional model. In the other three models (where this parameter is allowed to vary across

the conditions), there is a systematic pattern for the base times: the redundant condition displays the lowest base time and the visual condition the highest base time. The base time variances were fixed for all models (except for the free model) and were different for all the models. For all the numerical values, see Table 3.6.

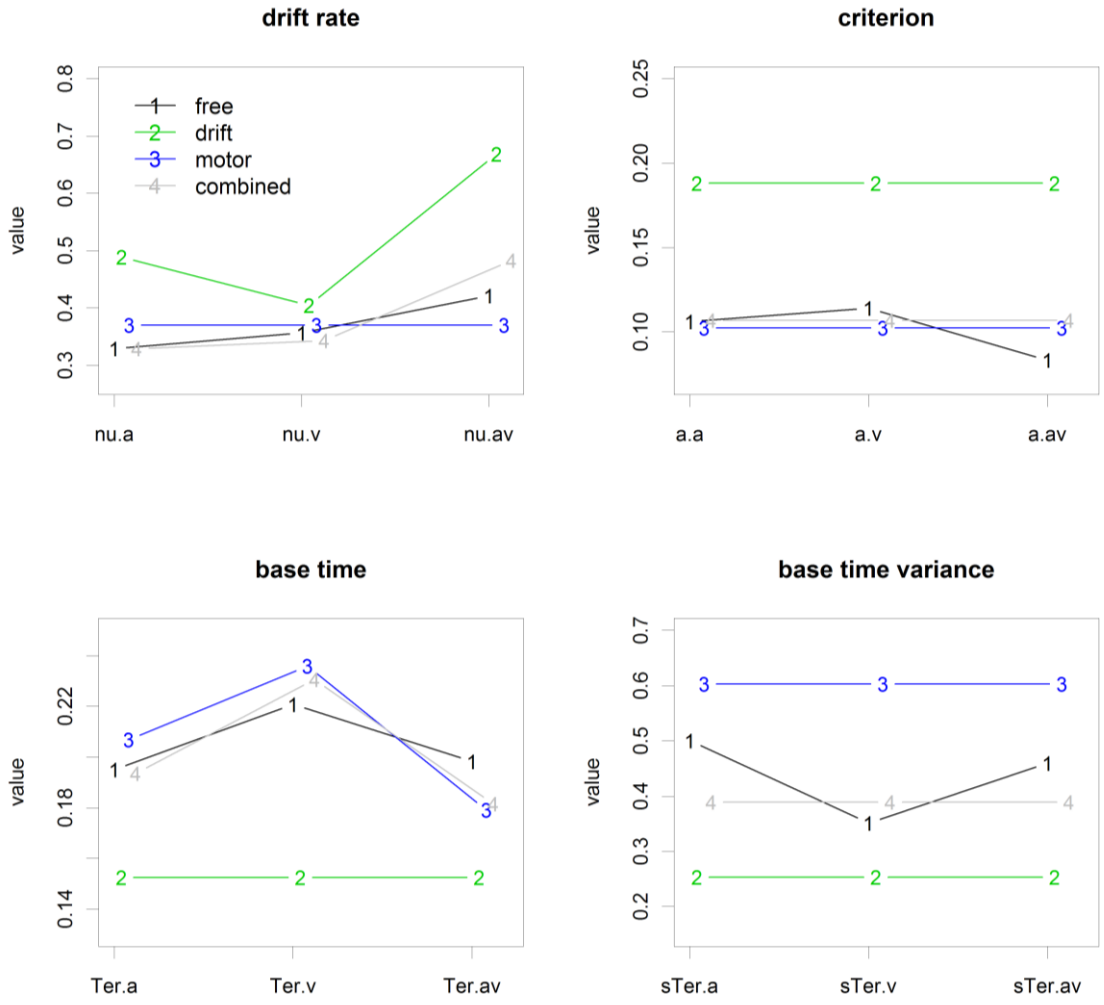

Figure 3.5 Parameter values for all coactivation models. The upper left panel shows the drift rates across conditions (*a* for auditory, *v* for visual and *av* for audiovisual) for each coactivation model (drift for decisional, motor for nondecisional model).

Table 3.6

Parameters for Best Fitting Exemplar per Model

Model	Mean Parameter Value		
	Auditory	Visual	Audiovisual
Decisional			
v	0.489	0.405	0.668
a*	0.188	0.188	0.188
Ter*	0.152	0.152	0.152
s_t*	0.253	0.253	0.253
Nondecisional			
v*	0.370	0.370	0.370
a*	0.102	0.102	0.102
Ter	0.206	0.236	0.179
s_t*	0.603	0.603	0.603
Combined			
v	0.329	0.343	0.483
a*	0.107	0.107	0.107
Ter	0.194	0.231	0181
s_t*	0.389	0.389	0.389
Free			
v	0.329	0.357	0.421
a	0.106	0.114	0.083
Ter	0.195	0.221	0.198
s_t	0.500	0.351	0.460

Notes. Values of the coactivation models across experimental conditions. Parameters marked with an asterisk were fixed across conditions for the respective model.

The analysis of BIC scores revealed that 12.0%, 91.5%, 92.3%, and 21.1% of the 11000 fitting attempts per model failed for the decisional, nondecisional, combined and free model respectively. This was due to either the fit being too poor (for at least one condition) or the parameters leaving the allowed range. But even for the successful fits, the BIC scores and parameter values varied considerably. In order to assess, whether the best fitting parameters are representative for the fitting capability of each model (and not statistical outliers), the parameters of the 1% best fits were used to compute the mean and standard deviations of the parameters for each model as well as the distribution of the BIC scores. This resulted in a subset of 110 data points per model (1% from the overall 11000 attempts to fit the data). Table 3.7 gives the mean, standard deviation, minimal and maximal BIC scores for the 1% best fits of each model.

Table 3.7

Mean BIC Score of the 1% Best Fitting Parameters.

Model	**Mean BIC (Std. Dev.)**	**Min BIC**	**Max BIC**
Decisional	10643 (3.96)	10626	10647
Nondecisional	10647 (4.17)	10630	10652
Combined	**10575** (15.57)	**10551**	10603
Free	10587 (**3.27**)	10578	**10591**

Notes. Mean, minimum and maximum BIC scores for the 1% best fitting parameters per model. Standard deviation in braces, minimum values in bold face.

Figure 3.6 shows the mean and standard deviations of the parameter values together with the best fitting values (as red diamonds). All parameter values (apart from the base time variance) stay well inside the two standard deviations of the mean value and can be considered as representative for the capability of each model to fit the data. Table 3.8 provides the numerical values of the means and standard deviations for all parameters and each model.

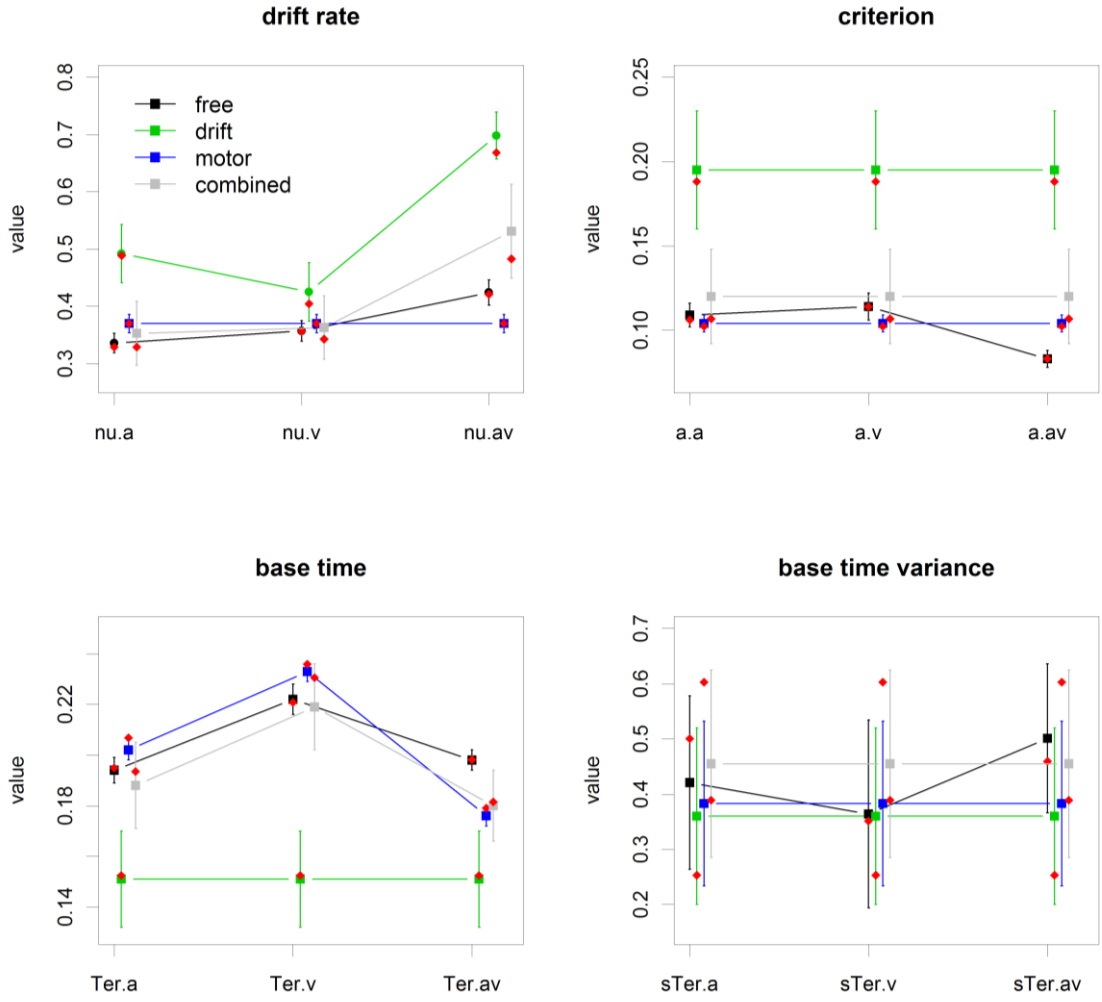

Figure 3.6 Means and standard deviations of parameter values are plotted for the fitted models. Only parameters of the 1% best fits are plotted, red diamonds denote the best fitting parameters for each model. Except for the base time variances, these values remain well within two standard deviations of the mean 1% best fits.

Table 3.8

Parameter Estimates for the 1% Best Fits for each Model

Model	Mean Parameter Values (Std. Dev.)		
	Auditory	Visual	Audiovisual
Decisional			
v	0.492 (0.051)	0.425 (0.051)	0.698 (0.041)
a*	0.195 (0.035)	0.195 (0.035)	0.195 (0.035)
Ter*	0.151 (0.019)	0.151 (0.019)	0.151 (0.019)
s_t*	0.360 (0.160)	0.360 (0.160)	0.360 (0.160)
Nondecisional			
v*	0.370 (0.016)	0.370 (0.016)	0.370 (0.016)
a*	0.104 (0.005)	0.104 (0.005)	0.104 (0.005)
Ter	0.202 (0.004)	0.233 (0.004)	0.176 (0.004)
s_t*	0.383 (0.149)	0.383 (0.149)	0.383 (0.149)
Combined			
v	0.353 (0.056)	0.363 (0.055)	0.531 (0.082)
v*	0.120 (0.028)	0.120 (0.028)	0.120 (0.028)
Ter	0.188 (0.017)	0.219 (0.017)	0180 (0.014)
s_t*	0.455 (0.170)	0.455 (0.170)	0.455 (0.170)
Free			
v	0.336 (0.017)	0.357 (0.018)	0.424 (0.022)
a	0.109 (0.007)	0.114 (0.008)	0.083 (0.005)
Ter	0.194 (0.005)	0.222 (0.006)	0.198 (0.004)
s_t	0.421 (0.157)	0.364 (0.170)	0.501 (0.135)

Notes. Parameter values of the 1% best fitting coactivation models across Experimental conditions. Parameters marked with an asterisk were fixed across conditions for the respective model. Standard deviations in braces.

3.2 Experiment 2

In this experiment observers had to respond to the onset of the stimuli from Experiment 1, which could now appear to the left or right of the fixation cross (a two-alternatives forced choice task). Apart from that, all aspects were identical to Experiment 1.

3.2.1 Method

Subjects

In Experiment 2, 21 new subjects (14 of them female) participated in a single 60-minutes session in return for 8€ or a course participation. Their average age was 27.2 years (range, 18 – 46) and they had normal or corrected-to-normal vision. One subject was left-handed.

Apparatus

The Experiment was conducted in a sound-insulated booth, and was controlled by a program using MATLAB (R2009bSP1, Natick, Massachusetts: The MathWorks Inc., 2010) and the PsychToolbox (Brainard, 1997; Pelli, 1997) running on an Apple Mac mini (Cupertino, California: Apple Inc.) machine (with Mac OS X). The stimuli were presented on a 20” Mitsubishi Diamond Pro 2070SB monitor at a resolution of 1,280 × 1,024 pixels and a refresh rate of 100 Hz with a viewing distance of approximately 73 cm.

Stimuli

Visual stimuli were grey discs (CIE Yxy = 10.9, .286, .333; 1° in diameter), auditory stimuli were 400 Hz beeps (of duration 150 ms) delivered by headphones, and redundant stimuli were both visual and auditory stimuli presented simultaneously (stimulus onset asynchrony of zero). On each trial the stimuli appeared either on the left or right side of the screen and/ or headphone respectively (no spatial conflicts, as only left or only right stimuli were presented). The participants responded to the onset of the stimuli by pressing the left mouse buttons with their left index finger for a left target and vice versa for a right target.

Procedure

Every trial was structured in the following way: First, a white fixation cross (0.5° x 0.5° of visual angle) was presented centrally on a black screen for 800 ms. Then, after an intertrial interval (uniformly varying between 500 and 1500 ms) the stimulus or stimuli

appeared on either the left or right side (1° of visual angle to the left or right) and remained on the screen until a response was elicited. After the observers' response (left or right mouse button), a 750 ms waiting period followed, before the next trial started again with the fixation cross (see Figure 3.7 for the display progression).

The whole experiment was divided into 20 blocks à 45 trials, where target positions (left and right), unimodal and bimodal trials were interchanging randomly. Overall, this amounted to 900 trials (150 trials per condition). Participants could make a break in between blocks and feedback on their mean RT and error rate was given, as they were instructed to respond as fast as possible while remaining below 5% error rate.

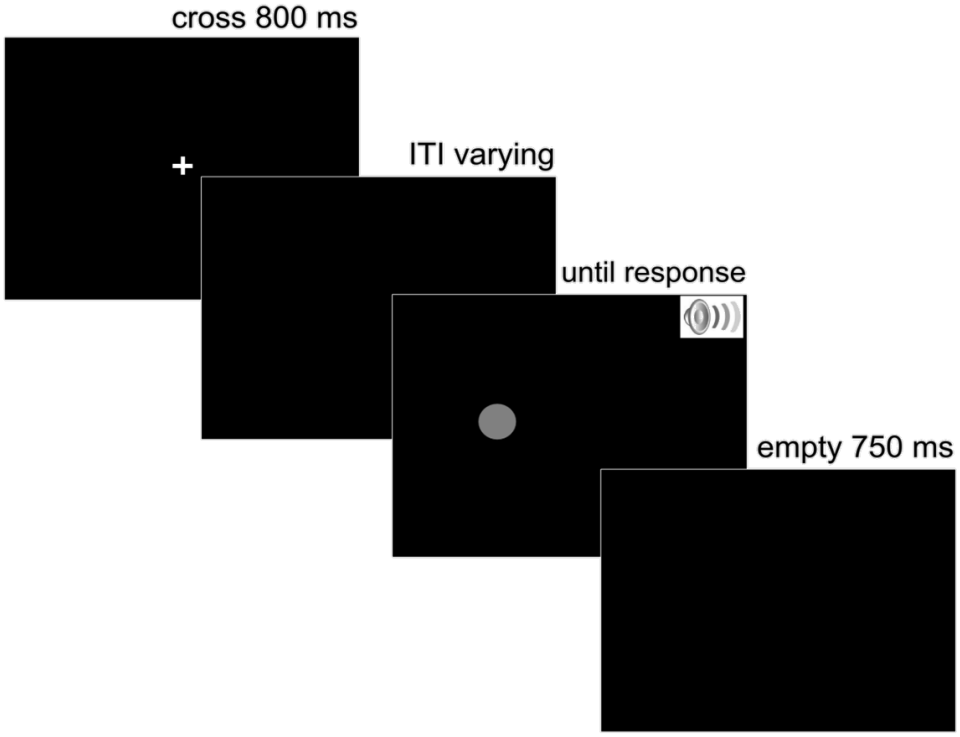

Figure 3.7 Display progression of the 2AFC Experiment. First, a fixation cross is presented centrally for 800 ms. After a variable intertrial interval the target appears left or right of the fixation cross and remains until the observer responds bimanually. A blank screen follows for 750 ms before the next trial begins.

3.2.2 Data Analysis

For the data analysis, calculation of the quantile proportions and the fitting procedures, please see the data analysis section of Experiment 1. Data for the left and right target positions were aggregated for the unimodal and bimodal conditions. Again, only correct responses and error rates were used for the fitting.

3.2.3 Results

Errors

Errors were defined as anticipatory responses (RT <= 150 ms), misses (RT > 1600 ms) or false responses (for example, pressing left when right target appeared). Observers' errors were on average at 3.35%. One subject produced more than 20% errors, this data was discarded.

Mean Reaction Times

The mean RTs for the unimodal and bimodal conditions are listed in Table 3.9. The participants were equally fast in the auditory condition and in the visual condition (403 ms). Reaction times did not differ for left or right targets. There was a pronounced RSE of 51 ms, as the bimodal condition produced a mean RT of 341 ms.

Table 3.9

Mean Response Times for Experiment 2 (2AFC)

Condition	Mean RT [ms]	Std. Dev.
Auditory	403	65
Visual	403	59
Audiovisual	341	53
RSE	50	16

Notes. Mean response times (in ms) of the unimodal, single signals trials (auditory, visual) and the bimodal, redundant signals trials (audiovisual) with corresponding standard deviations of Experiment 2. The data exhibits a pronounced redundant signals effect of 50 ms across all observers.

RMI Violations

The violations of the RMI were tested according to the explicit algorithm by Ulrich, Miller and Schröter (2007). At the probabilities p = 0.05, 0.10, …, 0.95 (two-sided) t-tests were performed with a Bonferroni-corrected (see Holm, 1979) α of 0.0026 (= 0.05 / 19 probability points). Significant violations (p < 0.0026) were found across the ten probabilities 0.05 to 0.45. Figure 3.8 shows the individual and mean RMI test function curves of Experiment 2. Significant positive values across subjects produce violations (Miller, 1982).

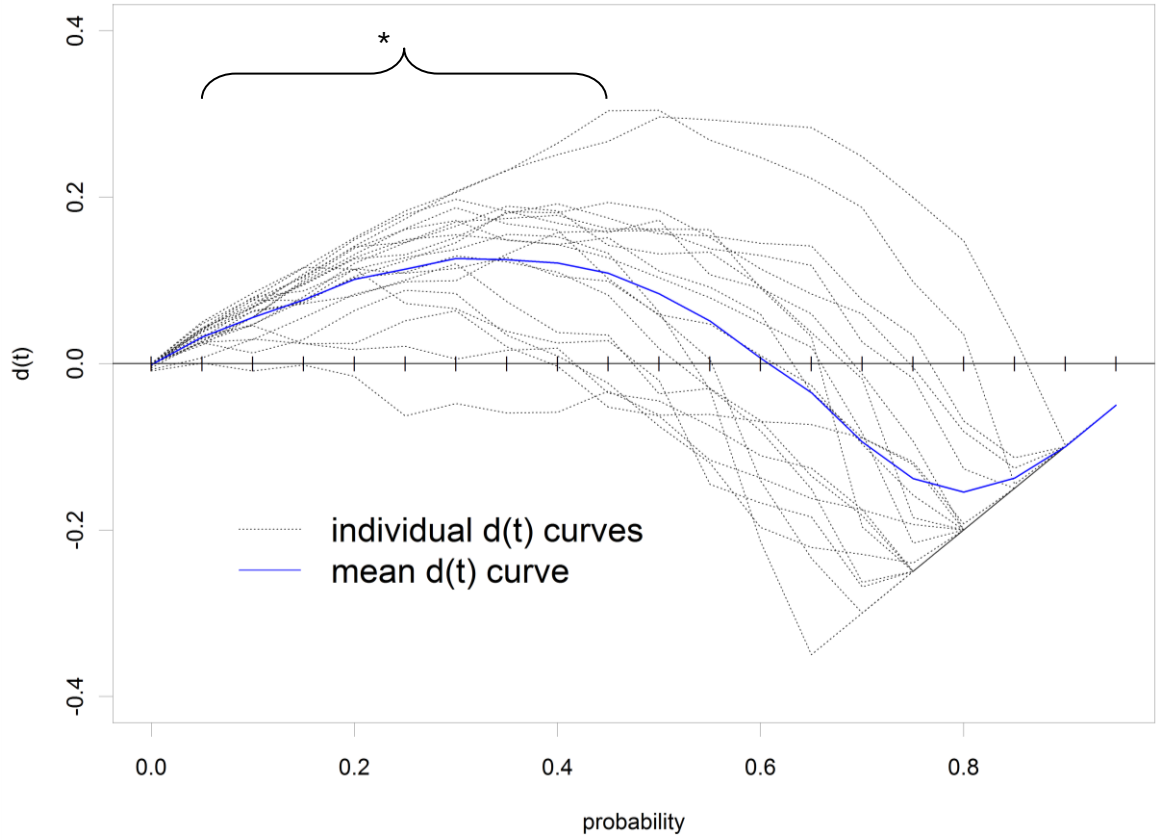

Figure 3.8 Race model test function aggregated across individual observers (blue full line) and for each individual observer (grey dotted line). Violations of the Race model inequality were found for probability points 0.05 to 0.45 using multiple t-tests with a Bonferroni-corrected significance level of 0.0026.

Fitting Results

The outcome of the fitting procedure for Experiment 2 was cost values of 10596, 10571, 10538, and 10578for the decisional, nondecisional, combined and free model respectively. Thus, the best fitting model was the combined model. It outperformed the second best model (the free model) by 40 units in a logarithmic scale. The probability

transformation yields a strong advantage of the combined model over the other models, as its probability is at over 99.99% (see Table 3.10 for an overview of the BIC values and probability estimates).

Table 3.10

Minimum BIC Values for each Model in the Experiment 2 (2AFC).

Model	Degrees of Freedom	BIC	Probability
Decisional	6	10596	2.54×10^{-13}
Nondecisional	8	10571	6.83×10^{-08}
Combined	10	**10538**	9.99×10^{-01}
Free	12	10578	2.06×10^{-09}

Notes. The combined model exhibited the lowest BIC score and had the best fit to the data. (For the computation of the model probability, see equation 20 in 3.1.2.)

On the level of means, the goodness-of-fit of the models was in good agreement to empirical data. The standard deviations of all models were considerably higher than in the empirical data. Only the nondecisional model did not succeed in reproducing the mean reaction times and RSE (see Table 3.11).

Error Rates

A look at the error rates the models produced, reveals that only the combined model could quantitatively reproduce the empirical errors; the decisional model generated virtually no errors for the redundant condition, while the nondecisional model (due to its design) could only generate identical error rates across the conditions. The free model produced to high error rates for all conditions (see Table 3.12).

Table 3.11

Mean RTs (in ms) and Standard Deviations of Empirical and Model Data for Experiment 2.

Condition	Empirical	Decisional	Nondecisional	Combined	Free
Auditory	403 (65)	393 (131)	395 (115)	393 (116)	398 (120)
Visual	403 (59)	394 (132)	384 (115)	392 (130)	391 (134)
Audiovisual	341 (53)	334 (78)	348 (115)	337 (103)	337 (103)
RSE	50 (16)	59	36	55	54

Notes. The mean reaction times of the models are in good agreement to the empirical data. The standard deviations (in braces) are at variance with the data. Except for the nondecisional model, all models capture the redundant signals effect of the empirical data.

Table 3.12

Error Rates of Experiment 2 and the Coactivation Models.

Condition	Empirical	Decisional	Nondecisional	Combined	Free
Auditory	3.12%	2.89%	3.17%	3.03%	3.96%
Visual	4.37%	2.95%	3.17%	4.29%	6.73%
Audiovisual	2.69%	0.45%	3.17%	1.98%	3.91%

Notes. The combined and the free model were the only models to approximately reproduce the empirical error pattern. The decisional model produces no errors in the audiovisual condition, the nondecisional model (due to its parameter constriction) identical errors across conditions.

Goodness-of-Fit

The quantile function plots of the respective models, revealed that all models (apart from the decisional model) overestimate the lowest quantile for the unimodal conditions. Apart from the lowest quantile, both the free and the combined model display an exceptionally good fit to the quantile proportions of the empirical data. The decisional model

generally underestimates the lower quantiles (0.0, 0.1) for all conditions, and overestimates the highest quantiles (0.7, 0.8) for the visual condition. The nondecisional model fails to reproduce the spread of all the conditions, most pronounced for the redundant and visual condition.

Parameter Analysis

The parameters, producing the best fitting for each model are depicted in Figure 3.10. From a qualitative view, the free, motor and combined model agree in the range of all parameters. The decisional model features the highest drift rate overall in the redundant condition and the highest base time variance

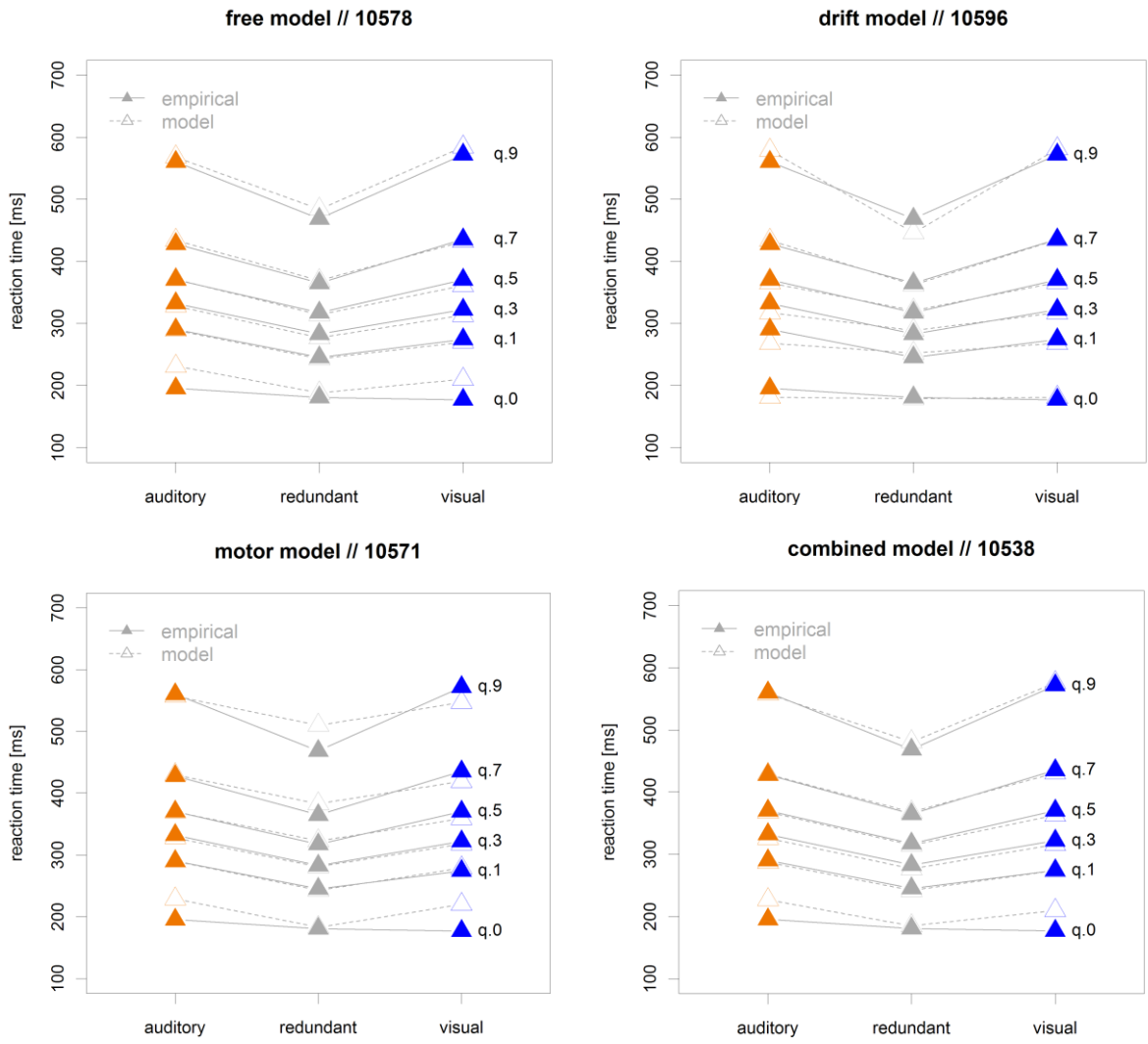

Figure 3.9 Quantile function plots for all the models. Full triangles denote empirical data points, while unfilled triangles show the data points of the respective models. The BIC score of each model is given in the title of each graph.

All models exhibit the highest drift rate for the redundant condition and reflect the rank of the conditions, that is the redundant condition having the highest drift rate and the visual condition having the lowest drift rate (the nondecisional model has the drift parameter fixed, so it cannot reproduce this pattern by design reasons). The base times are lowest for the redundant channel across all models (where this parameter can vary). In contrast to Experiment 1, the base time variances agree in their value for the free, nondecisional and combined model. For the numerical values, see Table 3.13.

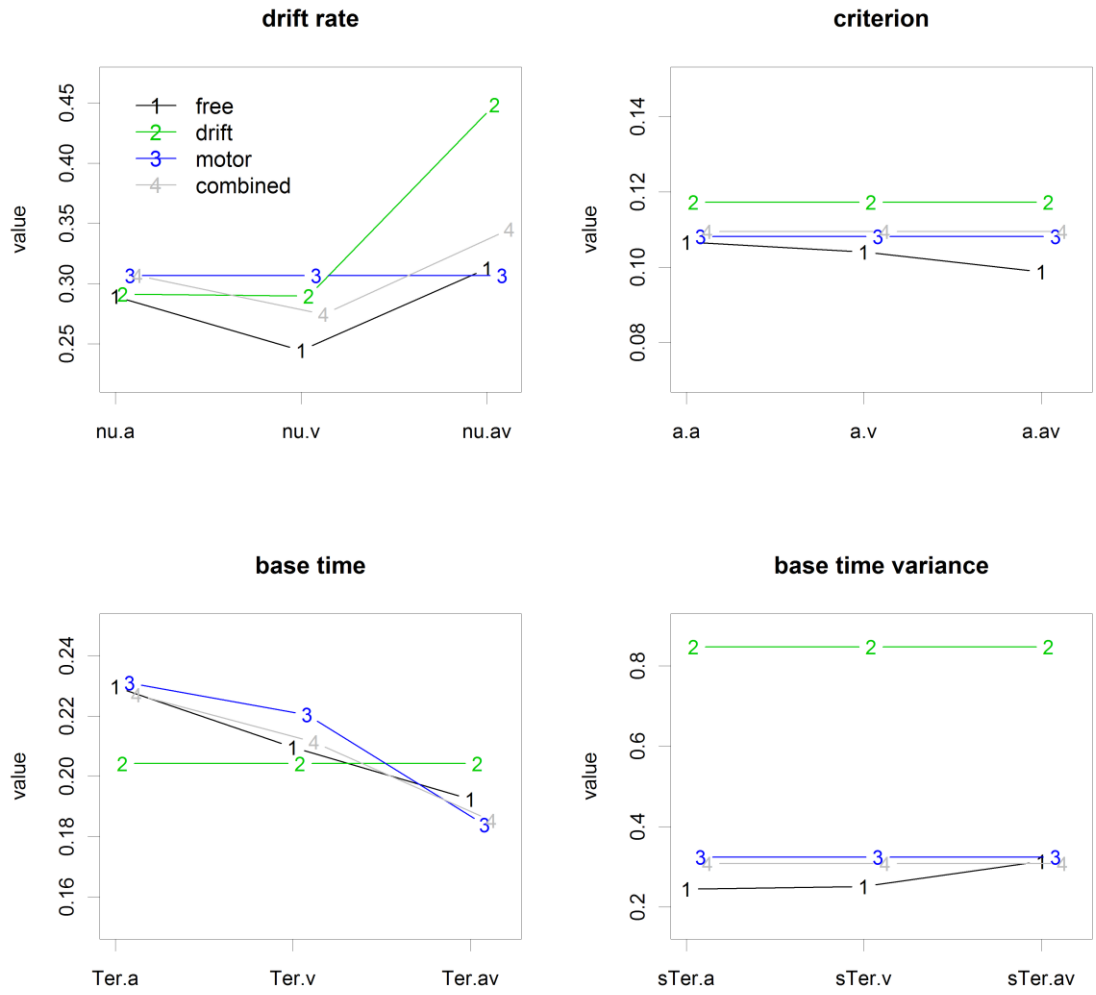

Figure 3.10 Parameter values for all coactivation models. The upper left panel shows the drift rates across conditions (*a* for auditory, *v* for visual and *av* for audiovisual) for each coactivation model (drift for decisional, motor for nondecisional model).

Table 3.13

Parameters for Best Fitting Exemplar per Model

Model	Mean Parameter Value		
	Auditory	Visual	Audiovisual
Decisional			
v	0.291	0.290	0.449
a*	0.117	0.117	0.117
Ter*	0.204	0.204	0.204
s_t*	0.848	0.848	0.848
Nondecisional			
v*	0.307	0.307	0.307
a*	0.108	0.108	0.108
Ter	0.231	0.220	0.184
s_t*	0.324	0.324	0.324
Combined			
v	0.307	0.275	0.345
a*	0.109	0.109	0.109
Ter	0.227	0.211	0.185
s_t*	0.308	0.308	0.308
Free			
v	0.289	0.245	0.313
a	0.106	0.104	0.099
Ter	0.230	0.210	0.192
s_t	0.244	0.251	0.313

Notes. Values of the coactivation models across experimental conditions. Parameters marked with an asterisk were fixed across conditions for the respective model.

Parameter Estimation

The analysis of BIC scores revealed that 22.7%, 93.1%, 92.5%, and 26.9% of the 11000 fitting attempts failed for the decisional, nondecisional, combined and free model respectively. Analogously to the analyses for Experiment 1, the range of BIC scores was computed. For that, the parameters of the 1% best fits were used to compute the mean and standard deviations of the parameters for each model as well as the distribution of the BIC scores. This resulted in a subset of 110 data points per model (1% from the overall 11000 attempts to fit the data). Table 3.14 gives the mean, standard deviation, minimal and maximal BIC scores for the 1% best fits of each model.

Table 3.14

Mean BIC Score of the 1% Best Fitting Parameters.

Model	Mean BIC (Std. Dev.)	Min BIC	Max BIC
Decisional	10613 (4.90)	10596	10619
Nondecisional	10589 (6.32)	10571	10601
Combined	**10559** (10.41)	**10538**	**10577**
Free	10590 (**3.00**)	10578	10593

Notes. Mean, minimum and maximum BIC scores for the 1% best fitting parameters per model. Standard deviation in braces, minimum values in bold face.

Figure 3.11 shows the mean and standard deviations of the parameter values together with the best fitting values (as red diamonds). In contrast to the results of Experiment 1, not all parameter values stay well inside the two standard deviations of the mean value. Most notably the best fitting parameters of the free model leave the confidence range for both the drift rates and the criterion. The other models stay well within their respective ranges. Table 3.15 provides the numerical values of the means and standard deviations for all parameters and each model.

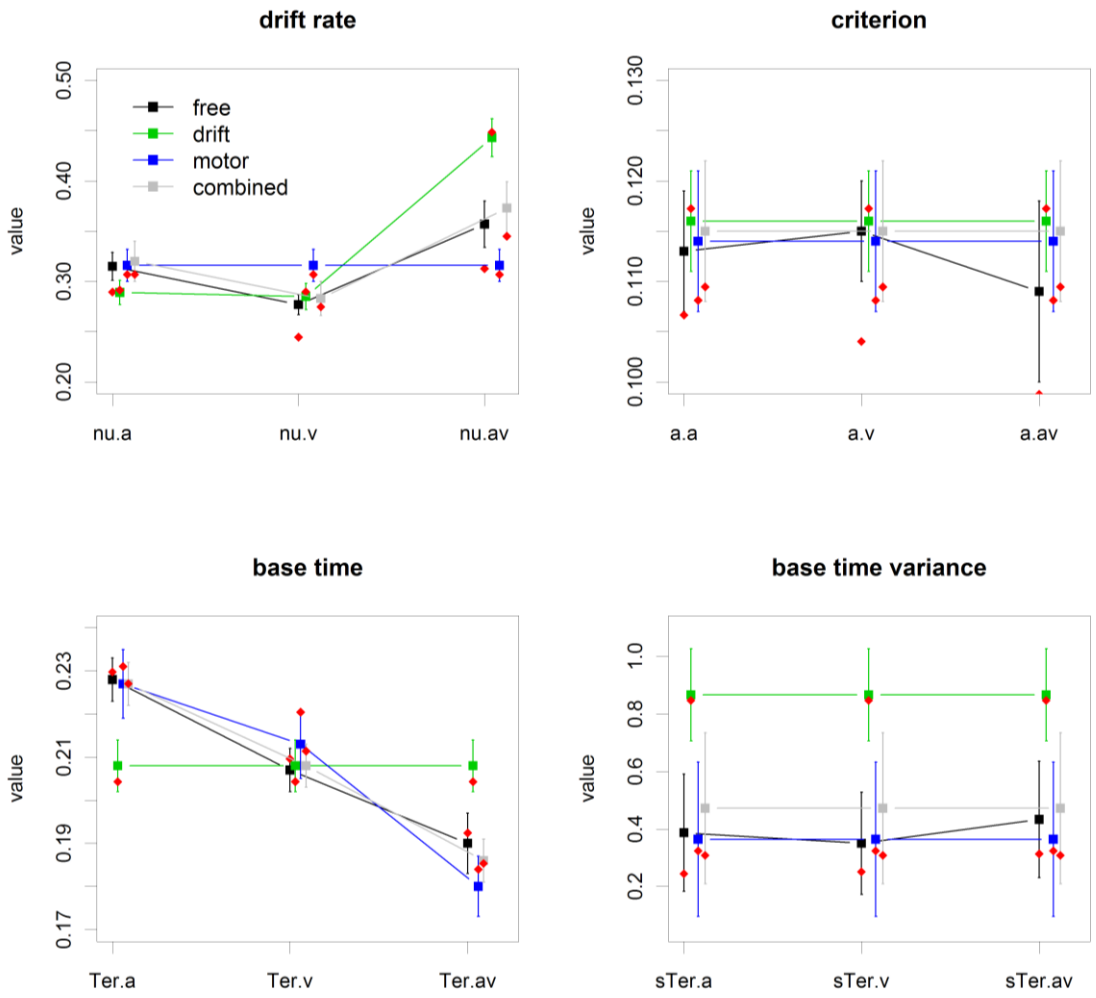

Figure 3.11 Means and standard deviations of parameter values are plotted for the fitted models. Only parameters of the 1% best fits are plotted, red diamonds denote the best fitting parameters for each model. Except for the drift rate and the criterion of the free model, all values remain well within two standard deviations of the mean 1% best fits. The standard deviation of the base time variance is relatively high.

Table 3.15

Parameter Estimates for the 1% Best Fits for each Model

Model	Mean Parameter Value (Std. Dev.)		
	Auditory	Visual	Audiovisual
Decisional			
v	0.289 (0.012)	0.285 (0.013)	0.443 (0.019)
a*	0.116 (0.005)	0.116 (0.005)	0.116 (0.005)
Ter*	0.208 (0.006)	0.208 (0.006)	0.208 (0.006)
s_t*	0.867 (0.160)	0.867 (0.160)	0.867 (0.160)
Nondecisional			
v*	0.316 (0.016)	0.316 (0.016)	0.316 (0.016)
a*	0.114 (0.007)	0.114 (0.007)	0.114 (0.007)
Ter	0.227 (0.008)	0.213 (0.008)	0.180 (0.007)
s_t*	0.364 (0.269)	0.364 (0.269)	0.364 (0.269)
Combined			
v	0.320 (0.020)	0.283 (0.017)	0.373 (0.026)
a*	0.115 (0.007)	0.115 (0.007)	0.115 (0.007)
Ter	0.227 (0.005)	0.208 (0.005)	0.186 (0.005)
s_t*	0.472 (0.263)	0.472 (0.263)	0.472 (0.263)
Free			
v	0.315 (0.014)	0.277 (0.010)	0.357 (0.023)
a	0.113 (0.006)	0.115 (0.005)	0.109 (0.009)
Ter	0.228 (0.005)	0.207 (0.005)	0.190 (0.007)
s_t	0.387 (0.204)	0.350 (0.178)	0.433 (0.203)

Notes. Parameter values of the 1% best fitting coactivation models across Experimental conditions. Parameters marked with an asterisk were fixed across conditions for the respective model. Standard deviations in braces.

3.3 Discussion and Conclusion

This study marks the first of a kind, as it employs a diffusion model analysis to the redundant signals paradigm to uncover the source of coactivation. The results suggest that there is coactivation both in the decision and nondecision component of the reaction times. This confirms previous conjectures on a hybrid source of RMI violations (see e.g. Diederich, Colonius). Surprisingly, the contribution of nondecision processes to the effect varied with the task. For the focal attention task (Experiment 1), the decision component was dominant, whereas in the divided attention task (Experiment 2), the nondecision component outweighs the decision component clearly. This interpretation has to be further validated by follow-up studies.

Methodologically, this study marks an important improvement in the use of mental chronometry. The pitfall of model mimicry is evaded as the whole distribution of each experimental condition is expedited for the fitting procedure. Using only the mean reaction time and Occam's razor, would have attested choosing the decisional model to be in best agreement with both the data of Experiment 1 and 2. And this even though in the second experiment the decisional model fared worst, when looking at the quantile proportion fits. Also, the detailed analysis of the fitting procedure is an important step to justify results from such model analyses, as it can give estimates on the reliability of each fit.

In this section, the experimental and fitting results will be discussed and embedded into the existing literature. Concrete future experimental ideas and improvements of the diffusion model analysis and fitting procedure are discussed together with a concluding summary of this study.

3.3.1 Observers Performance

The low error rates across the two experiments indicate the general simplicity of the task and the observers' ability to follow the experimental instructions. On a mean level analysis, the experiment showed very pronounced RSE's of 55 and 50 ms, which are not uncommon for simple detection tasks. Comparing the two single target conditions in Experiment 1, the auditory trials were processed faster than the visual trials. However not statistically significant, this is in good accordance to basic findings (Todd, 1912), where response times of auditory stimuli are faster than of visual stimuli (for medium intensity levels). In Experiment 2, both unimodal conditions do not differ numerically.

The large amount of violations (ten in Experiment 1, nine in Experiment 2) very strongly falsifies the class of race models as an explanatory model for these experiments.

This result is further reinforced, as both a conservative alpha-correction was used and response contingencies were avoided (Mordkoff & Miller, 1993). Moreover, the violations happen in the lower range of probability points, which is plausible by the makeup of the race model inequality. These results serve as justification that the empirical data indeed is of a coactivated manner and thus the subsequent fitting results can be linked to coactivation mechanisms in the redundant signals paradigm.

3.3.2 Decisional and Non-Decisional Processes contribute to Coactivation

The fitting results, that is, the BIC scores and model parameters, were in a comparable range for both experiments. Due to the high amount of similarity in Experiment 1 and 2, this can be taken as sanity check for the fitting. The results suggest that the best fitting model for both the SRT and 2AFC experiments is the combined account, where drift rates and base times are allowed to vary across all conditions. In Experiment 1, the free model had the second best fit, while in Experiment 2, the nondecisional model performed second best. The separation from the combined model to the other models is even more distinct, as the cost function is defined in a logarithmic scale (see equation 18).

Looking at the parameters (for the best fitting exemplars per model) revealed similar patterns across the two experiments. There was a ranking of drift rates, as the bimodal drift rate was higher than both the unimodal drift rates for all models. Together with the BIC scores, this can be seen as evidence for a drift rate coactivation in these bimodal RSP experiments. Interestingly, where they could vary, the base times also follow this ranking, as the bimodal base time is lowest (and thus response times are fastest) across all models.

Under the assumption that the fitting results reveal the true mechanism for the data in these two experiments, the decisional and nondecisional components seem to be differentially responsible for the observed RSE. In Experiment 1, out of the RSE of 56 ms, 13 ms can be accounted to the base time difference between the faster unimodal condition and the bimodal condition alone. In contrast to that, half the RSE (26 of 51 ms) in Experiment 2 can be attributed to the base time difference between the faster unimodal and the bimodal condition.

Studies that have tried to fit data to explicit coactivation are rare. One of the explicit models, that assumes a coactivation at the decisional stage is Schwarz' superposition model (1994). The basic assumption of the model is that on redundant-target trials the separate activation of the two stimulated channels superpose to form the overall-diffusion process $X_r(t) = X_a(t) + X_b(t)$. Activity in both channels can be adequately described by independent diffusion processes of the Wiener type and can have variable channel dependency. It has

been applied to data from Miller (1986) and predicted the data well on a level of mean reaction times. As in the present study, a large proportion of the response time in Schwarz' data is consumed by the nondecisional stage.

Diederich (1995) performed a trimodal SRT with visual, auditory and tactile stimuli with varying interstimulus intervals. The author performed a mean level fit to a race model and two coactivation models. As both coactivation models outperformed the separate activations model and resulted in excellent fits of the mean reaction time, the author states that the spread of the response times was not captured adequately well. The author suggested that this inability to account for the standard deviation in the data is possibly due toe dependencies of motor times. The present study is in good accordance to that assumption, as the combined model features a condition-dependent motor coactivation.

In accordance to the finding of the diffusion model analysis, that there is coactivation occurring at the motor stage, the study by Diederich & Colonius (1987) found positive evidence for this claim. An analysis of the distributions of RT differences between left- and right-handed responses showed, that there is a U-shaped functional dependence of the amount of facilitation in the motor component on the interstimulus interval. The method of analysis based on RT differences rests upon the disputable assumption that the motor delay constitutes and additive component of the entire observable RT (see, e.g. McClelland, 1979).

A comparison to the literature above is problematic for at least two reasons. Both have analysed the goodness-of-fit only to decisional models and only at the level of means. Relying only on the fit to the means alone in this study, would not help distinguish between the decisional and combined model in Experiment 1. In Experiment 2, such a limited analysis would not be able to rule out any of the models. As the decisional model features the lowest amount of parameters (that is 6) compared to the other models, the principle of parsimony would prefer decisional models, although in Experiment 2 the decisional model exhibited the weakest fit. On a methodical level, these differential outcomes prove a strong argument for the of diffusion model and against classical mean level analyses.

3.3.3 Methodical Caveats

The experimental data was highly coactivated and yielded a large redundant signals effect. This was due to easy detection task and the highly salient target stimuli. Anticipation was minimized by introducing temporal uncertainty (varying intertrial intervals), however there were no no-go trials and very little uncertainty about location of the targets (either centrally in Experiment 1 or at the same left/ right positions in Experiment 2).

Not unexpectedly, due to the high saliency of the targets and the simplicity of the experimental tasks, the error rates were rather low. This prevented the inclusion of the error response time distributions in the quantile fitting algorithm, however the error rates where accounted for by the modified cost function (see equation 19). In other Experimental contexts (with different tasks and/ or stimuli) the proportion on errors might be higher and thus make it possible to take into account the RT distribution of incorrect responses and not only the error rates. Increasing the difficulty of the task, by reducing the feature contrast and/ or adding noise to the stimuli, will increase the proportion of erroneous responses and thus allow incorporating that data to the fitting procedure. From a diffusion model aspect, such an intervention should only affect the drift rates of the respective models. This would additionally validate the results of both experiments, as such interventions have been performed (for example by Voss et al., 2004) to validate the RDM empirically. Additional fitting inquiries of more difficult SRT and 2AFC tasks and other visual search tasks featuring redundantly defined targets, mark the next logical step to broaden the results obtained in this study.

Another way to improve the fitting performance, is to constrain the all the models by an estimate of the empirical motor variance. Ulrich and Stapf (1984) used the trial-to-trial variation of left and right-handed RT's to approximate the motor variance. If for each trial, the difference of left and right reaction times, $D = RT_L - RT_R = (S + M_L) - (S + M_R) = M_L - M_R$, depends only on M_L and M_R, since the common term S is dropped. This means, that D depends only on those stages that follow the central motor command. Therefore, if stimulus intensity affects only the earliest stages of stimulus encoding, then the distribution of reaction time difference should be unaffected by a change in stimulus intensity. As both the experiments featured only one level of stimulus intensity, this proxy was not used to constrain the model to be fit. Also, this requires that the paradigm of interest features a double-hand response (which is not possible for a left – right discrimination in Experiment 2).

3.3.4 Evaluating the Fitting Procedure

One non-experimental measure to improve the validity and performance of the fitting procedure is the parametric bootstrap approach of Wagenmakers et al. (Wagenmakers, Ratcliff, Gomez, & Iverson, 2004) to assess the model mimicry. In this study, a more pragmatic approach to evaluate the fitting procedure itself was chosen and revealed notable results. The fitting performance as measured by the rates of failed and successful fitting attempts, showed a stark disparity between models. Only the free and the decisional model were constantly able to fit the empirical data (approximately 80 and 88% of the time) while the nondecisional and combined model only managed to fit seldom (approximately 8% for both). As the details of fitting procedure are not explicated in the literature, it remains unclear

what and if anything to conclude from these performance differences. Is the (relative) inability to successfully fit the data for a large proportion of attempts in itself an indication of the "unfittingness" of the model? Or is this an artefact of the optimization process and not intrinsically related to the goodness of the respective model and the data? For example, some models might feature more than one parameter range can mimic the data, while for other models only one parameter combination is able to fit the data. There are no papers reporting or discussing these aspects of fitting. (Vandekerckhove & Tuerlinckx, 2007 give four rules-of-thumb for the optimization part of the fitting procedure, which are not analytically derived.) Further theoretical and simulation investigations are necessary to solve this question.

To solidify the fitting results, the parameter values of the (1%) best fitting exemplars, were subject to closer scrutiny. They show a good amount of stability, as almost all of the parameters are constrained by one standard deviation around the mean for every model. An exception to this fact, are the base time variances which could vary rather drastically. This, however is of no practical consequence, as this noise fluctuations were present for all models and do not constitute the decisive reason for the respective models fit. Moreover, the amount of standard deviation added to the response time by the s_t parameter is marginal (Gomez et al., 2007). Another way to ensure the robustness and validity of the fits is to use the best fitting parameters for each model and compute confidence intervals of their BICs in a Monte Carlo fashion. This way the possibility of the best fit being an outlier is minimized.

3.3.5 Conclusion

Overall, the present fitting analyses pose a strong challenge to the view, that coactivation in the redundant signals paradigm is a purely decisional effect. This pattern is even more pronounced in the data of Experiment 2 where the decisional model even fared worst and the purely nondecisional was the second best model in terms of goodness-of-fit. Although two experiments are not sufficient to decisively rule out a decisional only model, they make a point to emphasize the role of the (late) motor stage for the coactivation. The results furthermore encourage a diffusion model analysis and avoid the implicit null hypothesis of a decisional coactivation in case of violations of the RMI.

In order to gain a converging picture of the question, what the sources of the RSE are, further experiments and fittings are necessary to look at the influence of different stimuli, modalities, response effectors and experimental tasks in the RSP on coactivation and its generating (diffusion model) parameters.

4 STATISTICAL EVALUATION OF THE RMI TEST

The race model inequality test has been a central instrument to discriminate between separate activations and coactivation architectures in the redundant signals paradigm. Since its conception by Miller (1982), it has been applied in a myriad of experimental settings (see chapter 2.3 for a rigorous derivation and evaluation of the test). Violations of the RMI bound exclude race models as a viable architecture for the RSE (Diederich, 1992; Egeth & Mordkoff, 1991; Feintuch & Cohen, 2002; Grice et al., 1984; Krummenacher et al., 2001, 2002; Miller, 1982; Mordkoff et al., 1996, see chapter 2.2 for models of the RSE).

The RMI bound itself formalizes the maximum amount of speed-up possible by race models. It enables researchers to test this bound as the null hypothesis for the class of race models at different percentiles. The RMI test and some of its variations and generalizations have been the subject of numerous theoretical and methodological studies (Ashby & Townsend, 1986; Colonius, 1986, 1990b, 1999; Colonius & Ellermeier, 1997; Colonius & Vorberg, 1994; Diederich, 1992; Miller, 1986, 2004; Miller & Lopes, 1991; Miller & Ulrich, 2003; Mordkoff & Yantis, 1991; Townsend & Colonius, 1997; Townsend & Nozawa, 1997; Townsend & Wenger, 2004; Ulrich & Giray, 1986; Ulrich & Miller, 1997).

As simulation studies on the RMI would show the test itself is affected by several issues challenging its outcome and value for researcher investigating the redundant signals effect. In order to deepen the knowledge and consequently avoid these caveats, it is essential to investigate the statistical properties of the RMI test in an exhaustive framework: In this simulation study, the RMI test is investigated in a Monte Carlo fashion using the Ratcliff diffusion model to synthesize reaction time data.

First, this investigation is motivated by detailing the known statistical properties of the test. Then the simulation framework and its modules, that will help integrate previous study scopes and increase the utility of the RMI test, are described.

4.1 Statistical Properties of the RMI Test

Generally, the RMI test (like every statistical classifier) can produce two types of errors. Type I errors (α-errors or “false alarms”) occur, when the null hypothesis is incorrectly rejected. In the case of the RMI test, this implies a scenario where the data was de facto produced by a race model, but the RMI test wrongly suggests a rejection of race models. The false alarm rate is controlled by the significance level (“α level”) of the *t*-test. Common

practice is to set α = 5%, as researchers want to avoid rejecting the null too promptly. This means that on average one out of 20 experiments will produce a false alarm.

Failing to reject the null hypothesis (here: race models), when the data was in fact generated by a non-race model mechanism, is termed a type II error (β error or "miss"). The power of the RMI test ($1 - \beta$) is generally believed to be considerably below the desired 80%, as the data is compared to the maximum amount of facilitation possible by race models. Yet this conservativeness (see Patching & Quinlan, 2002; Schwarz & Ischebeck, 1994) has been rarely quantitatively estimated. (There are experimental circumstances which can produce violations of the RMI apart from integration or coactivation models (see Mordkoff & Miller, 1993). These can be avoided by a balanced sampling of single and redundant trials.)

4.1.1 Alpha Accumulation

Despite its pivotal role for inferring on cognitive architectures, the RMI test has only lightly been investigated with help of simulation studies. Standard practices in RMI testing require researchers to *t*-test the RMI bound at several probability points (see Ulrich et al., 2007 for an explicit algorithm), where any probability point alone can attest a violation. With this multiple testing on the same set of data, the chances of falsely rejecting race models accumulate. Although, *t*-tests are highly correlated across percentiles (as neighboring points of distribution functions are close to each other), the effective overall type I error can be much higher than the inner significance level. Kiesel, Miller and Ulrich (2007) could show, that the amount of amount of type I error accumulation can be rather large - reaching up to 13% (Table 4, Kiesel et al., 2007). They recommend different strategies to control for type I error accumulation: restricting the range of percentiles to evaluate the race model, independent replication of experiments and/or lowering the inner significance level. This however, is bought dearly, as each of the measures goes along with a reduction of power.

4.1.2 Estimation Bias

Apart from type I error accumulation, Kiesel et al. (2007) present an unexpected caveat of RMI testing. The first step of the RMI algorithm, which is the estimation of the cumulative distribution functions (see the algorithm by Ulrich, Miller and Schröter, 2007 in chapter 2.3.5) can itself be biased favoring violations – and thus coactivation accounts. It is known that quantile estimators are in general problematic, as they tend to overestimate lower percentiles and underestimate higher percentiles (Gilchrist, 2000). Kiesel et al. simulated races, while manipulating the similarity of the two single target channel distributions and the sample sizes of each channel. Independent of the distribution relations, a sparse and

unbalanced sampling across channels (rather different distributions, see Figure 4.1), systematically biases the RMI test to work against race models (and thus promoting type I errors).

Contrary to classical power considerations (Cohen, 1988), an increase in sample size in that case still could not eliminate the bias, as it remained between 1 and 3 ms per probability point for the highest trials-per-condition case (40 samples for each target type).

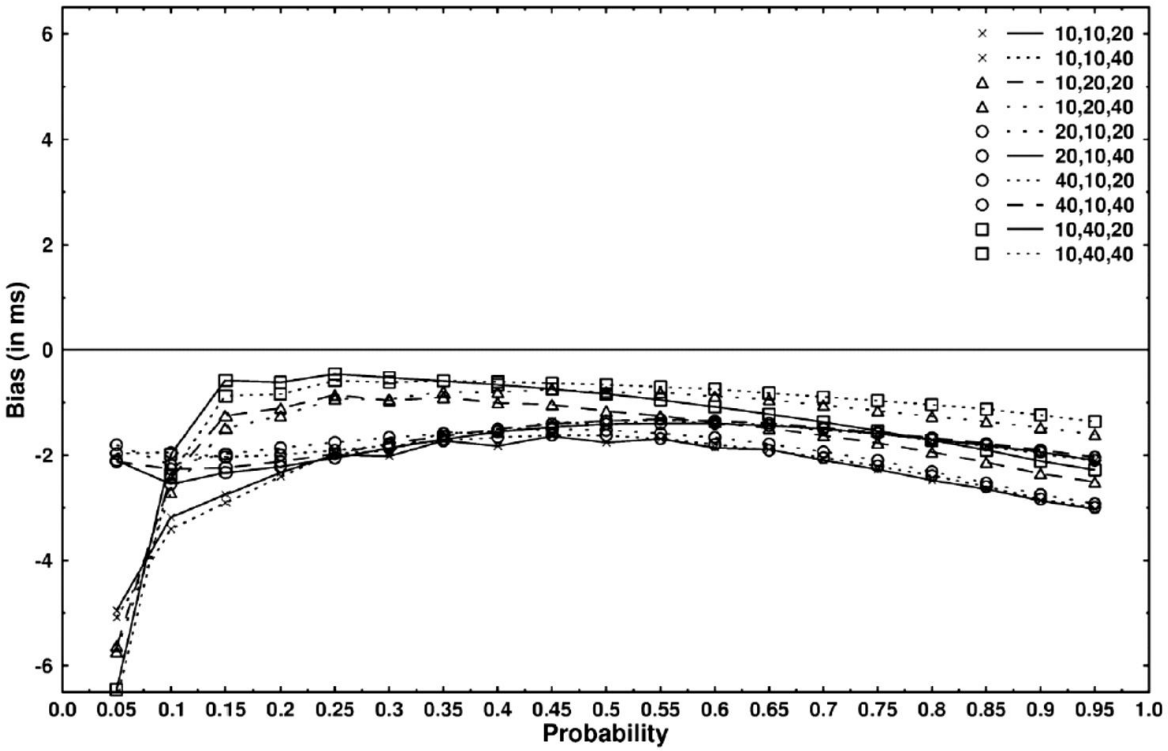

Figure 4.1 Estimation bias for race models with different sample sizes per condition. The first two numbers refer to the sample sizes of the single signals trials, the last number to the redundant signals trials. Values less than zero line favor violations, values greater than zero hinder the detection of violations. (middle panel of Figure 4 from Kiesel et al., 2007; reprinted with permission)

4.1.3 Effects of high Base Time Variance

In contrast to these findings, Townsend and Honey could demonstrate that the RMI test might still be prone to conservatism, when the response time is afflicted with a highly varying base time component (Townsend & Honey, 2007). Classically, reaction times *RT* can be split into a decision and a nondecision or base time component (Luce, 1991): $RT = D + B$. Here, *B* is the base time or residual latency reflecting all the processes involved in encoding,

response selection and response execution, whereas *D*, the decision latency, is the psychological process of interest (e.g. the detection or discrimination process). Townsend and Honey could show (analytically and via simulations), that the RMI test is rather underpowered, that is, in a range between 20 to 80% for reasonable RSE's, when the base time variance is high (200 ms). And this holds even though the reaction time data was in fact generated by a non-race data. They derived the effect of highly varying base times and likened them to a filter, smoothing away potential violations of the RMI. Here again, an increase in sample size (from 100 to 250 and to 500) was not recommended, as it reduced the power to detect actual coactivation even further (Figure 7, for low base time variance and Figure 8, for high base time variance in Townsend & Honey, 2007).

4.1.4 Limitations of previous Studies

Overall, the picture these two studies provide is inconclusive, as disjunct aspects of the RMI were probed. The Kiesel et al. (2007) study for example has no coactivation implementation and thus cannot investigate power or a potential estimation bias of the RMI test in the coactivated case. Furthermore, as it employs a descriptive (instead of a mechanistic) model of response times, it cannot illuminate the effects of base time on type I errors, power and the estimation bias. To generate their reaction time data, Kiesel et al. used the ex-Wald distribution, which is the convolution of an inverse Gaussian distribution with an exponential distribution. It is commonly used in experimental psychology to model reaction time data, and its parameters were believed to reflect the reaction time decomposition (with the inverse Gaussian or Wald-part being a model for the decision time and the exponential part signifying the base time component). The use of the ex-Wald and other distribution functions beyond the level of data description is problematic, as Matzke and Wagenmakers (2009) could show that the parameters of these models do not uniquely correspond to parameter changes of the theoretical and explicit model of reaction times. Townsend and Honey (2007) also used simplified, atheoretic response time models with empirically implausible distributions. Townsend and Honey use normally distributed data for their processing times data, which in contrast to right-skewed empirical reaction times is symmetrical around its modus. This data is convolved with a normally distributed base time.

Apart from these numerical issues, the studies are improvable in several other respects. Neither Townsend nor Kiesel implemented differentially correlated race models, but instead only use the extremal case of race models (i.e. maximally negatively correlated racers). Neither study investigated both low and high sample sizes, different significance levels of the inner *t*-test, subject sizes or correlation and coactivation strengths (respectively). This is necessary, as the knowledge of these parameters can effectively help setting up the

RMI test in an optimal sense (e.g. minimizing type I errors or maximizing power rates or a combination of the two). Most importantly, the Townsend and Honey study did not look into type I error accumulation or the estimation bias, while the Kiesel et al. study neglects power and base time manipulations altogether.

4.1.5 Research Questions

Both these studies make valid and important points, every researcher applying the RMI tests should incorporate in their design and analysis. Nevertheless, the validity of these results relies on their replication and extension within an integrated and fully crossed study. This simulation study aims at integrating and extending the work and scope of the preceding studies in a coherent and adaptive framework. Specifically, this study addresses the following open research questions: (a) What is the power of the RMI test, (b) is there an estimation bias in the RMI test when the data is actually coactivated (and if so, in what direction)? (c) what is the effect of base time variance on estimation bias, type I error accumulation, and power, and (d) are the results by Kiesel et al. (2007) and Townsend & Honey (2007) quantitatively or at least qualitatively replicable, when using mechanistic models instead of descriptive models of decision making? Table 4.1 gives an overview of the properties of previous simulation studies in comparison to the present study.

The RMI test has been introduced as a tool to elucidate the architecture question of multisensory integration. The results and inferences of RMI studies then are only as credible and solid as is the understanding of the tool that brings them about. In case of the RMI test, there are several known shortcomings, as were outlined so far. This simulation study aims at overcoming these issues and providing a means to assess the properties and performance of the RMI test in a principled way. It features the latest and most established models of decision making and will effectively improve both the knowledge on the RMI test and its application in an experimental context.

Table 4.1

Comparison of Scopes and Properties of Simulation Studies on the RMI Test.

Properties	**Study**		
	Kiesel et al. (2007)	Townsend & Honey (2007)	The present Study
Scope	Estimation Bias for Race; Type I Error Accumulation	Effects of Base Time Variance on Type I Errors and Power	Estimation Bias for Race and Coactivation; Type I Errors for correlated Racers; Power of latent and manifest Coactivation
Alpha error/ accumulation	X / X	X / --	X / X
Race model	ex-Wald distributed	normally, exponentially distributed	Ratcliff Diffusion Models
Interchannel Correlation	-1	(not specified)	0, -0.5, and -1[a]
Power	--	X	X
Coactivation Model	--	normally distributed (mean shifted)	decisional model (drift rate summation)
Estimation Bias Race/ Coactivation	X / --	-- / --	X / X
Range of Sample Size	10, 20, and 40	100, 250, and 500	10, 20, 40, 100, 200, and 400
Subject Size	20, 40	1	8, 20, and 40

Notes. X denotes a present and -- denotes missing aspects in the respective study. [a] nominal values of interchannel correlation only, the effective correlation lower due to the skewness of reaction time distributions. The present study integrated the scopes and parameters of interest into a coherent simulation framework.

4.2 Method

4.2.1 Hardware

The simulation framework and all analyses are programmed with GNU R (R: A language and environment for statistical computing, Vienna, Austria). The simulation was run on an Apple (Cupertino, California: Apple Inc.) Power Mac machine (12 cores, 24 GB of RAM, 64-bit, operated by Mac OS X). Apart from the Ratcliff diffusion model implementation, which is a C++ library (Berkeley, New Jersey; Bell Labs) all code is written in R. The whole simulation job is executed in a distributed fashion, using the *foreach*-package. There the conditions are processed in parallel on 24 virtual cores (via hyper threading).

Table 4.2

Simulation Modules and their respective Data Output

Module	**Output**
1) Synthesizing RT Data	Raw Reaction Times for Single Target Channels *X, Y*, Race and Coactivated Channel XY_R, XY_C
2) Testing the RMI	Number of Violations Percentile Point of Violations p Values and Estimates of Statistic Mean and Std. Dev. of Reaction Times Mean Error Rate
3) Aggregating and Analyzing	Mean Correlation between *X, Y* Mean and Standard Dev. of RSE Overall Type I Error and Power Estimation Bias

Notes. *X* and *Y* denote the synthesized single signals data, XY_R the race and XY_C the coactivated redundant signals data. For the different measures and analyses, see the 4.3.

4.2.2 Modules of the Simulation Framework

The simulation framework can be conceptually divided into three modules. In the first module (the experiment simulation), reaction times for both the race and coactivation model are synthesized based on Ratcliff diffusors. In the second module, the RMI is tested according to the explicit algorithm by Ulrich (Ulrich et al., 2007). In the final module, both the statistical and RMI results are aggregated and analyzed to give the resulting overall type I errors, estimation bias and power. This modular approach allows to adapt and extend the analysis of the RMI test to, for example, different response time generating models, alternative RMI tests (for example, the non-parametric test by Maris & Maris, 2003 or the geometric measure by Colonius & Diederich, 2006), different violation rules and more than two racers.

Next, the three simulation modules, the simulation parameters and the analyses will be described, before turning to the simulation results. Table 4.2 gives an overview of the simulation modules and their respective outcomes.

4.2.3 Generating Reaction Times (Module 1)

For both the race and coactivation models, the Ratcliff diffusion model (Ratcliff, 1988) was implemented - a well-established and widely applied sequential sampling model for decision times (Ratcliff & Smith, 2004; see chapter 2.4 for a detailed description of the Ratcliff diffusion model). It quantitatively describes the decision process as a noisy accumulation of evidence towards a response-eliciting threshold. It is characterized by seven parameters. The rate of evidence accumulation (or *drift rate*) varies across trials according to $N(v,\eta)$ and reflects the clarity of the signal (for example, how easy or hard the target is perceivable). The parameter s governs the within-trial variability in drift rate and can be used as a scaling factor for the whole model (Donkin, Brown, & Heathcote, 2009). The distance between the response boundaries a quantifies the observers caution and controls the speed-accuracy tradeoff: when a is small, decisions are reached faster at the risk of terminating at the wrong threshold; when a is high, the decision making is more conservative at the cost of reaction times. The starting point z is used to model response bias, which can also vary between trials (controlled via the parameter s_z).

The last two parameters of the model concern the nondecision time component, which is usually conceived as a uniformly distributed variable. *Ter* is the mean base time added to the decision time and s_t is the across-trial variability of *Ter*. The overall response time then is the sum of the decision process and the nondecision time. Similar to the *EZ-diffusion* model (Wagenmakers et al., 2007), in the present study the simplifying assumptions

are made, that the starting point is unbiased (that is $z = a/2$) and the across-trial variabilities η are all zero. The base time variance parameter s_t was not set to zero, as it marks a vital aspect the research questions.

The Ratcliff diffusion model is characterized by a close fit to the response accuracy and the response time distributions for both correct and error responses. Furthermore, in contrast to descriptive models of decision making, its parameters do have direct psychological interpretations (Matzke & Wagenmakers, 2009). Its ecological validity was shown both on a neurological (Smith & Ratcliff, 2004; Ratcliff, Cherian, & Segraves, 2003), empirical (Voss et al., 2004) and statistical level (as it is unable to fit fake but plausible data, see Ratcliff, 2002). In addition to inherently producing the ubiquitous speed-accuracy trade-off in response time literature it also reproduces the mean-variance relation typical for response times (Wagenmakers et al., 2005). For a detailed description of the Ratcliff diffusion model, please see chapter 2.4.

By using the RDM (and unlike Townsend & Honey, 2007, or Kiesel et al. 2007), it is possible to manipulate the base time component independently of the architecture (race versus coactivation) and test its influence on the measures. The base time variance for all channels was set to 50, 100 or 200 (added to the standard deviation of the decision process alone).

Similar to Kiesel (Kiesel et al., 2007), different cumulative distribution function relations are simulated to reflect unequally fast sensory channels, *X* and *Y*. In the equal distributions condition, *X* and *Y* share the same RDM parameters, chosen in a way to reflects typical values for a simple detection task, e.g. a visual pop-out search. The targeted means are around 290 ms and the standard deviations were 65 ms. In the different distributions condition, *X* is identical to the equal distributions condition and *Y* has RDM parameters that yielded a mean of 330 ms and standard deviation were of 85 ms. As the RMI bound is calculated as the sum of the single target channel CDFs (or estimates thereof), the equal distributions condition (equal distribution) will have a different RMI bound, than the different distributions condition.

4.2.4 Race Model Implementation

For the race model, the two channels, *X* and *Y* were synthesized and the minimum response time of both (for each trial) was used to determine the redundant race channel, XY_R. The RMI formalizes the most extreme case of race models- where both racers are maximally negatively correlated. As the correlation determines the possible values of the RSE and violations of the RMI, it is controlled as a parameter of the race model in the

simulation. A perfect positive correlation of 1 would result in an extinction of the statistical facilitation (it would produce an RSE of zero, for equally distributed channels *X* and *Y*), while a perfect negative correlation of -1 would produce the maximum RSE possible for race architectures. To induce negative correlations in the data, a variant of the method of antithetic variates is used (Thompson, 2009). After randomly sampling reaction times for both single target channels *X* and *Y*, a proportion *corr* of all these reaction times each is randomly picked and reverse rank ordered. The correlation parameter *corr* is set to be either 0 (uncorrelated racers), -0.5 (moderately correlated racers) and -1 (strongly correlated racers). (Positive correlations are not implemented, as they do not constitute a critical test for the RMI). As reaction times distributions are generally right-skewed, the effective amount of correlation cannot reach the nominal values of -0.5 and -1, but is generally lower.

4.2.5 Coactivation Model Implementation

Coactivation is no intrinsic property of the RDM. In this study, the working hypothesis is that the RSE is a decisional effect (Zehetleitner et al., 2009a; Koene & Zhaoping, 2007; for a review see, Zehetleitner et al., 2008b). Its implementation burrows from theoretical considerations by Schwarz (Schwarz, 1989). In his account, Schwarz proposes a superposition of neural renewal counting processes. A theoretical equivalent in the RDM would be a full summation of drift rates of the single target channels. As this setting leads to potentially high and empirically implausible RSEs, a gauge or weighting factor λ is introduced to control for subadditive (latent) coactivation:

$$\nu_{XY} = \max(\nu_X, \nu_Y) + \lambda \cdot (\nu_X + \nu_Y) \tag{21}$$

The first term on the right hand side of equation 21 is equivalent to the fastest single target channel (in this case always *X*), whereas the second term introduces an increase in drift rate, based on a weighted sum of the two single drift rates. If the values of λ are restricted to the interval [0, 1[, plausible values of RSE (generated by comparable race models) can be found. Values for λ here are set to 0.1 to 0.8 in the equal distribution condition and to 0.1 to 0.6 in the different distributions condition (in steps of 0.1 each). This way, an exhaustive range of weakly (latently) to strongly (manifestly) coactivated models is guaranteed. The other parameters of the RDM (*a, s, z, Ter, η,* s_z) are set to be equal to the faster of the single target channels (here, always *X*).

4.2.6 Testing the RMI (Module 2)

In module two, an implementation of Ulrich's explicit algorithm is used to detect violations of the RMI (Ulrich et al., 2007; see chapter 2.3 for detailed description of the algorithm). This study's framework however, in principle allows the implementation of various and multiple RMI tests (Maris & Maris, 2003; Colonius & Diederich, 2006). If race models hold true, then the observed cumulative distribution function (CDF) of the RTs *X, Y* and *XY* should satisfy the equation for all time points *t*.

$$F_{XY}(t) \leq F_X(t) + F_Y(t), \quad \forall t \tag{22}$$

Ulrich's algorithm is composed of four computational steps: First, the CDFs for F_X, F_Y and F_{XY} are estimated as cumulative frequency polygons (CFP; see for example, Gilchrist, 2000) from the observed (correct) RTs for the three conditions. These estimated CFPs are denoted as G_X, G_Y, and G_{XY}. Second, the sum S of the G_X and G_Y is computed (the estimated RMI bound): $S(t) = G_X(t) + G_Y(t)$ for each participant. Third, at prespecified probabilities p = 0.05, 0.10, 0.15, …, 0.95, the percentile values s_p and z_p for S and G_{XY} are estimated according to the percentile definition by Hazen (Hazen, 1914). Finally, the percentile values s_p and z_p are aggregated over participants and for each percentile value a paired *t*-test (one-sided with the alternative "greater") is performed to assess, whether G_{XY} is larger than S. Race models are rejected, if G_{XY} is larger than S at any percentile.

As the RMI is typically tested at more than one probability, a percentile range criterion is used, although other criteria could be implemented easily. Common violation criteria are the "any violation" criterion, where one probability point violating the RMI suffices to falsify the race models. Another possibility is the "max" criterion, where only the largest difference between s_p and z_p need to be significantly non-zero.

From test theory, an increase of the significance level can help to increase the power of a test, but it also makes it prone to more false alarms (that is, rejecting race models wrongfully). An increase in the sample size of the different channels generally also has a beneficial effect on the power. In this study, the following sample sizes (that is, trials-per-condition) are looked into: 10, 40, 100, 200, and 400.

In case of low sample sizes, the method of *vincentising* allows for averaging of RT distributions across subjects (Ratcliff, 1979). This is useful when small sample sizes prevent accurate estimates of RT distributions for individual subjects) but it is unknown, how subject size exactly influences the RMI test in general. To average RT distributions across subjects, the vincentiles are computed for each subject's data and then averaged. Because each vincentile corresponds to a particular percentile rank, the resulting averages can be used to

construct empirical CDF. Here, subject sizes of 8, 20, and 40 are used for each channel, as those parameters can be frequently found in the literature.

4.2.7 Aggregation of Data and Analyses (Module 3)

In module three, the results of module two are aggregated and analyzed with respect to the overall type I error, the power and the estimation bias. The data from every processing stage is written to the hard drive, so that for each analysis, multiple questions can be addressed (for example, use of different violation criteria) and re-analysis is facilitated.

4.3 Data Analysis

4.3.1 The Bias Measure

The bias measure applied here, is defined in accordance to Kiesel et al. (2007) as

$$Bias = Bias(z_p) - Bias(s_p), \tag{23}$$

where the first term quantifies the difference between the estimated and the true distribution,

$$Bias(z_p) = G_{XY} - F_{XY}, \tag{24}$$

at specific probabilities *p*. The second term quantifies the difference between the estimated and the true RMI bound,

$$Bias(s_p) = (G_X + G_Y) - (F_X + F_Y), \tag{25}$$

at the same probabilities *p*. Thus, a positive $Bias(z_p)$ and a negative $Bias(s_p)$ work in favor of the race model, that is rendering it harder to violate the RMI. In contrast, a negative $Bias(z_p)$ and a positive $Bias(s_p)$ make it easier to obtain violations. Overall, *Bias* values less than zero indicate a negative discrimination of race models, and *Bias* values greater than zero a positive discrimination. Since there is no closed form for the distribution of Ratcliff diffusors, the “true” distribution of *X*, *Y* and *XY* in this study is obtained by sampling 100 000 reaction times for each and using the empirical cumulative distribution function.

4.3.2 Overall Type I Errors and Power

To obtain estimates of the overall type I errors and the power, the data is aggregated with respect to the percentile point and the simulation condition. The overall type I errors and

the power then are the relative frequencies of “false alarms” and “hits” respectively, averaged by the amount of iterations.

4.3.3 Simulation Design and Parameters

In all, 3600 conditions are synthesized and analyzed in a Monte Carlo fashion for 1000 iterations each. Increasing the number of iterations had no noticeable effect on the confidence intervals of the results (overall type I, power, RSE, bias, and so on), as a pilot simulation with a subset of the conditions could show. The correlation parameter is only manipulated for race models, whereas the λ parameter is only manipulated for coactivation models. The maximum coactivation value is set higher in the equal distribution conditions to reach ceiling power, so lambda 0.7 and 0.8 are only used in the equal distribution conditions. This resulted in 1080 race conditions and 2520 coactivation conditions (1440 for equal distribution and 1080 for different distributions). Overall 540’000 raw reaction times were generated and analyzed. Table 4.3 gives an account of the (fully crossed) factors that are manipulated in this simulation study:

Table 4.3

Overview of all Parameters and Manipulations of the Simulation Study

Factor	**Levels**	**# of Levels**
Distribution Relation	equal and different distributions	2
Coactivation Strengths λ	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7*, and 0.8*	6 (8)
Interchannel Correlation *corr(X,Y)*	0, -0.5, and -1	3
Base Time Variance s_t	50, 100, and 200	3
Significance Level for *t*-Tests	0.01, 0.03, 0.05, and 0.10	4
Sample Sizes for all Conditions	10, 40, 100, 200, and 400	5
Number of Subjects	8, 20, and 40	3
	overall conditions	3600

Notes. Table of all manipulated parameters in the simulation. See text for a detailed account.

4.4 Results

Due to the full-factorial design of the simulation (seven factors with two to eight levels), an exhaustive analysis would go beyond the scope of the study. Focus is put on plausible and often used parameter combinations, for example $\alpha = 0.05$ or sample size 100. Researchers who want to further analyze and explore the whole dataset can acquire it by the author. First, data affirming the validity of the simulations and the efficacy of the manipulations is presented (see Table 4.3 in methods section). Then, the results for the aspects overall type I errors and power, estimation bias for race and coactivation models and base time variance effects on the RMI test are shown.

Table 4.4

Mean RTs of Single and Redundant Target D for Race and Coactivation Models

Channel		Equal Distrib.	Diff. Distrib.
	X	296 ms	296 ms
	Y	296 ms	335 ms
corr = 0		263 ms	275 ms
corr = -0.5	**XY_R**	258 ms	268 ms
corr = -1		251 ms	262 ms
λ = 0.2		271 ms	
λ = 0.3		262 ms	
λ = 0.4	**XY_C**	256 ms	
λ = 0.5		250 ms	

Notes. *X, Y* denote the single target channels, XY_R the redundant race channel with varying correlation between *X* and *Y*, and XY_C the redundant coactivation channel with varying integration strength *λ*. In the equal distributions condition *X* and *Y* were equally distributed, in the different distributions conditions, *Y* was slower than X (see. 4.4.1). (Only plausible *λ*'s are listed, i.e. *λ*'s that produced comparable mean RT as the race models; see results on the redundant signals effect in 4.4.2).

4.4.1 Mean Reaction Times and Standard Deviations

The simulation has two conditions controlling the similarity of the single target channels. In the *equal distributions* condition, both *X* and *Y* had means of 296 ms; in the different distributions condition, *X* was also 296 ms, while *Y* had a mean of 335 ms. The mean reaction times of the redundant race and coactivation channels successfully varied with the amount of correlation and the coactivation weight *λ* respectively. The mean RT for the correlated racers is 263, 258 and 251 ms (equal distributions condition) and 275, 268 and 262 ms (different distributions condition). With increasing lambda, the means of the coactivated channels go down from 281 to 239 ms in the equal distribution condition and from a very similar 282 to 246 ms in the different distributions condition. Table 4.4 gives an overview of the mean reaction times for the single target (*X, Y*), race (XY_R) and coactivation channel (XY_C) across correlations and coactivation strengths.

The standard deviation of each channel is determined by the decision and the nondecision component of the RDM. While the nondecision or base time variance can be directly controlled in the s_t parameter, the decision component of the standard deviation also depends on the other RDM parameters (*v, a, s, z,* and *η*). The proportion of the standard deviation generated by the nondecision component for the overall standard deviation is marginal (Gomez et al., 2007). The standard deviations for the single target channels are in a range of 64 to 85 ms for the faster channel (*X*) and 74 to 93 ms for the slower channel (*Y*, only in the different distributions condition). It is a common finding, that standard deviations increase, as mean RTs increase (Wagenmakers & Brown, 2007).

For the race models, the standard deviations are in a range of 39 to 68, 34 to 65 and 24 to 60 ms for the uncorrelated, mildly and strongly correlated racers respectively. In the coactivated case, the standard deviations monotonically drop with increasing coactivation strength *λ* from 53 to 22 (s_t = 50), 59 to 33 (s_t = 100) and 77 to 60 ms (s_t = 200) for different base time variance conditions respectively. Table 4.5 gives an overview of the standard deviations for the single target, redundant race and redundant coactivation target channels.

The correlation manipulation was successful, as it resulted in effective correlations of 0, -0.35 and -0.75 for the correlation coefficients *corr* = 0, -0.5 and -1. (For a better readability, these conditions will still be referred to as correlation 0, -0.5 and -1.) Error rates are generally low across all channels and conditions. They vary between 0.5 and 2 per cent, which is common, e.g. for a simple visual search tasks.

Table 4.5

Standard Deviations of Single and Redundant Data for Race and Coactivation Models.

Channel		Base Time Variance					
		50		100		200	
		equal distr.	diff. distr.	equal distr.	diff. distr.	equal distr.	diff. distr.
	X	65		70		86	
	Y	65	83	70	86	86	100
corr = 0		37	42	45	49	67	70
corr = -0.5	$\mathbf{XY_R}$	31	36	40	43	64	66
corr = -1		22	25	33	36	60	61
λ = 0.2		44		51		71	
λ = 0.3	$\mathbf{XY_C}$	37		45		67	
λ = 0.4		33		41		65	
λ = 0.5		26		38		63	

Notes. *X, Y* denote the single target channels, XY_R the redundant race channel with varying correlation between *X* and *Y*, and XY_C the redundant coactivation channel with varying coactivation strength *λ*. In the equal distributions condition X and Y were equally distributed, in the different distributions conditions, Y was slower than X (see. 4.4.1). (Only plausible *λ*'s are listed, i.e. *λ*'s that produced comparable mean RT as the race models; see results on the redundant signals effect in 4.4.2).

4.4.2 Redundant Signals Effect

Increasing the correlation coefficient, *corr,* or the strength of integration, *λ*, should result in higher redundancy gains. The simulation behaves accordingly, as the mean redundant signals effect are in a range of 31 to 43, and 21 to 34 ms for (increasingly) correlated race models in the equal and different distributions condition. Analogously, an increase of *λ* for the coactivated models results in an increase of RSE's: in the equal distributions conditions, it is in a range of 13 to 56 ms and in the different distributions conditions in the range of 14 to 50 ms. Thus comparable RSE's can be found for *λ* = 0.3, 0.4

and 0.5 (in the equal distribution condition) and λ = 0.2 and 0.3 (in the different distributions condition). Table 4.6 gives an overview of the mean RSE for race models with varying correlation and coactivation models with varying coactivation strength.

Overall, the simulation framework (with the chosen parameters) produces empirically plausible reaction times and RSE's, while effectively manipulating the desired features of the data (be it channel correlation, amount of coactivation or base time variance).

Table 4.6

Mean Redundant Signals Effect (in ms) for Race and Coactivation Models.

Model	**Redundant Signals Effect**	
	equal distrib.	different distrib.
Race		
corr = 0	31	21
corr = -0.5	37	27
corr = -1	43	34
Coactivation		
λ = 0.2	24	25
λ = 0.3	32	36
λ = 0.4	39	40
λ = 0.5	44	46

Notes. Mean RSE's for race models with varying correlation and coactivation with varying integration strength (for plausible ranges only, see text for a detailed description).

4.4.3 Estimation Bias for Race Models

Kiesel et al.'s study (2007) devised a bias indicator to measure the amount of systematic disadvantage (or preference) of acquiring violations of the RMI, when estimating the respective cumulative distribution functions (see chapter 4.3.1). Positive values indicate that the RMI test will find a violation at the respective percentile point (in ms) harder, while negative values indicate a pull towards violations. For the most part, the race bias in this

simulation is confined to a range of -0.5 and 0.5 ms. Only for the lowest sample size-condition (10) at the probability p = 0.10, there is a robust bias of around -1.5 ms favoring violations (see Figure 4.2).

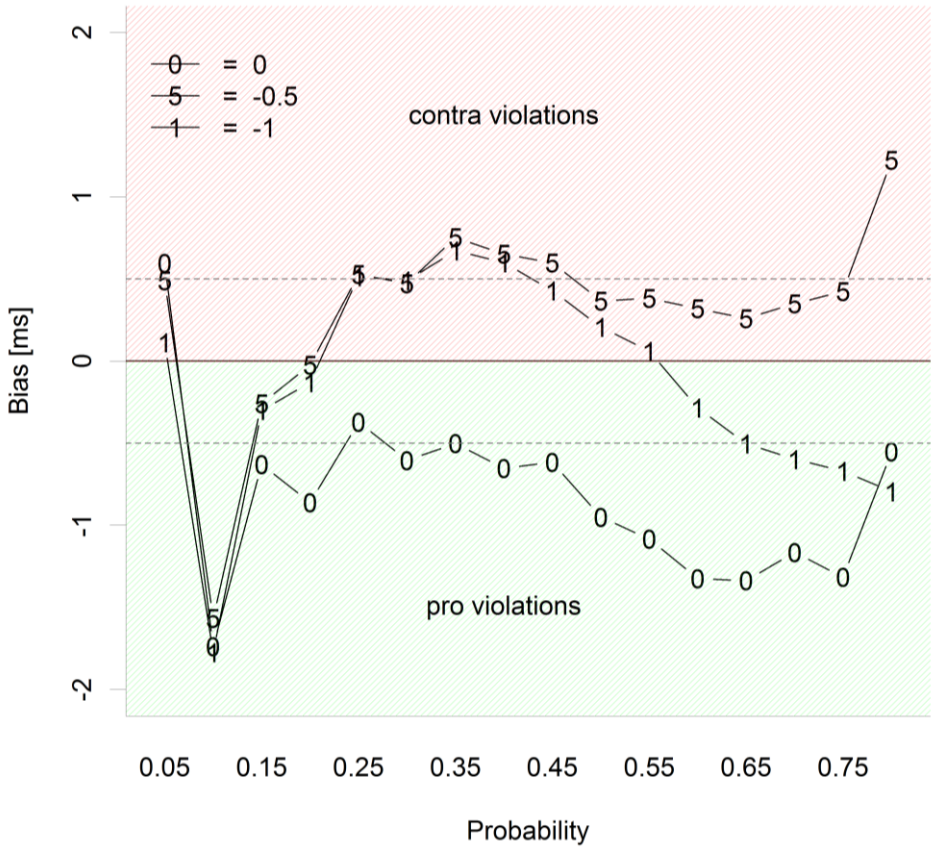

Figure 4.2 Estimation bias for race models with varying correlation for sample size 10 only. Bias values below zero indicate probability points, where violations are facilitated, and values above zero probability points, where violations are hindered.

For higher probabilities (p = 0.25, …, 0.55), both correlated racers (-0.5 and -1) have positive biases of around 1 ms (hindering violations), while the uncorrelated racer remains at negative values between -0.5 to -1.5 ms (favoring violations). A bias of this magnitude is of little to none consequence for the RMI test. Violations in this range are almost never observed and are unlikely to occur, as the RMI bound has an asymptote of 2 while a CDF, by definition, has an asymptote of 1.

The estimation bias for race models is also affected by the factor base time variance only in the lowest sample size condition (see Figure 4.3).

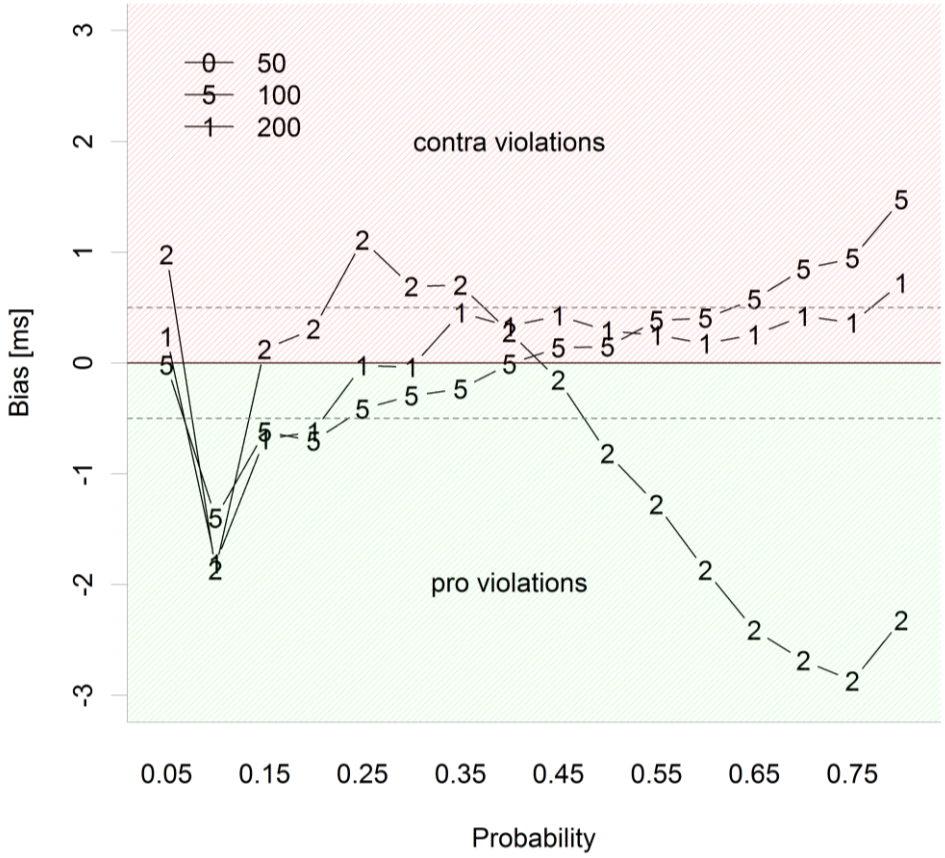

Figure 4.3 Estimation bias for race models with varying base time variance (50, 100, and 200) for sample size 10 only. Bias values below zero indicate probability points, where violations are facilitated, and values above zero probability points, where violations are hindered.

In addition to the already attested negative value at p = 0.1 of around -2 ms, there are positive biases of about +1 ms at p = 0.05 and around p = 0.25, but only for the race models with a high base time variance (s_t = 200). Also, only for the race models with highest base time variance, the bias curve falls into the negative range (p = 0.5 to 0.75 between -1 ms to -3 ms), while for the other base time variances it mostly remains positive and in a range between 0 and +1 ms. Again, those biases favoring violations should not be relevant because of the asymptotic behavior of the RMI.

4.4.4 Estimation Bias for Coactivation Models

As the Kiesel et al. (2007) study did not feature a coactivation model - it remains unclear whether the results hold for this class of models as well. Like in the race conditions, for sample sizes greater than 10, the bias in the simulation is generally low and confined to a

range of -0.5 and 0.5 ms. For the sample size-condition 10, there is a robust pattern favoring violations across plausible coactivation strengths. It amounts to -2 ms for equally distributed conditions and -1.5 ms for differently distributed conditions at p = 0.1 (see Figure 4.4). In the range of *p* = 0.50, …, 0.75 there again is a negative bias (favoring violations) of around -1.5 to -2 ms. There is also one positive bias value (hindering violations) for weakly coactivated models at *p* = 0.05 of around +0.7 ms. Subject sizes had no noticeable effect on the patterns of the bias curves.

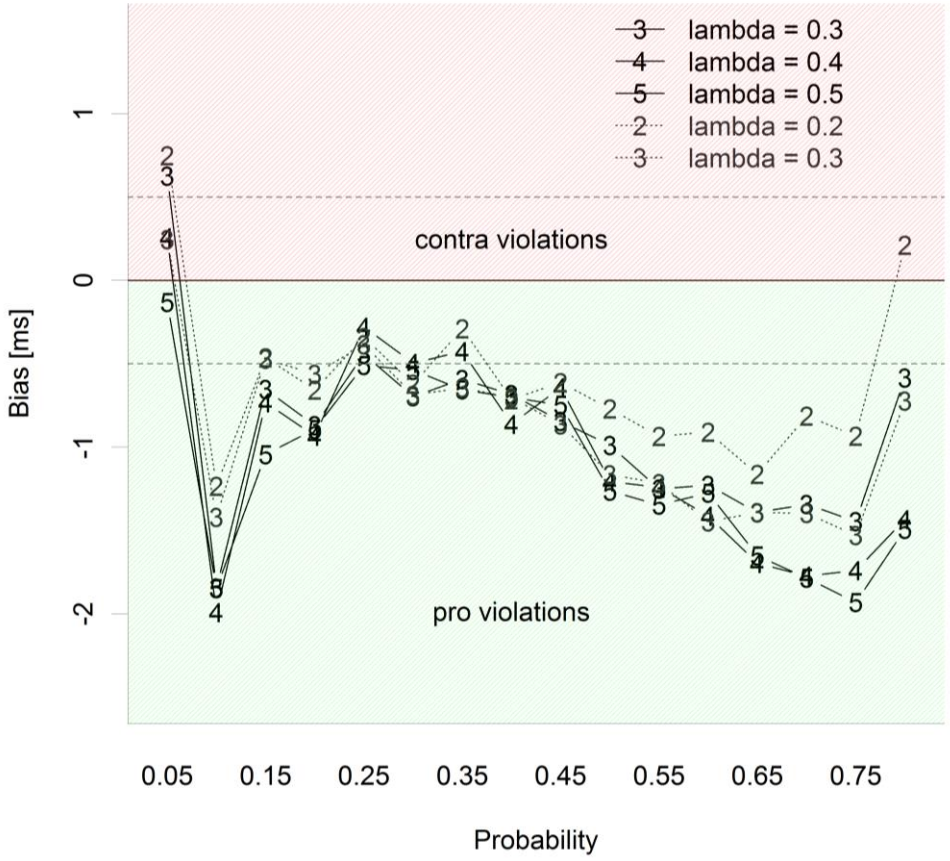

Figure 4.4 Estimation bias for coactivation models (for sample size 10) with different coactivation strengths. Full lines represent equal distributions conditions, and gray dotted lines represent different distributions conditions both for plausible *λ* values. Bias values below zero indicate probability points, where violations are facilitated, and values above zero probability points, where violations are hindered.

The coactivation bias responds in a similar way to different base time variances as the race bias. Figure 4.5 shows the aggregated data for plausible λ ranges in the equal distribution (λ = 0.3, 0.4, and 0.5) and different distributions (that is λ = 0.2, and 0.3) condition. Only for the lowest sample size condition 10 there are notable bias values (that is, beyond the range of -0.5 and 0.5 ms). There is a bias favoring violations at probability p = 0.1 for all base time variances, of around -2 ms to -1.5 ms for the equal and different distributions condition. Between the percentiles 0.45 and 0.75, the bias again drops into the negative range, but only for the highest base time variance conditions (s_t = 200). The other conditions are bounded by -1 and 1 ms.

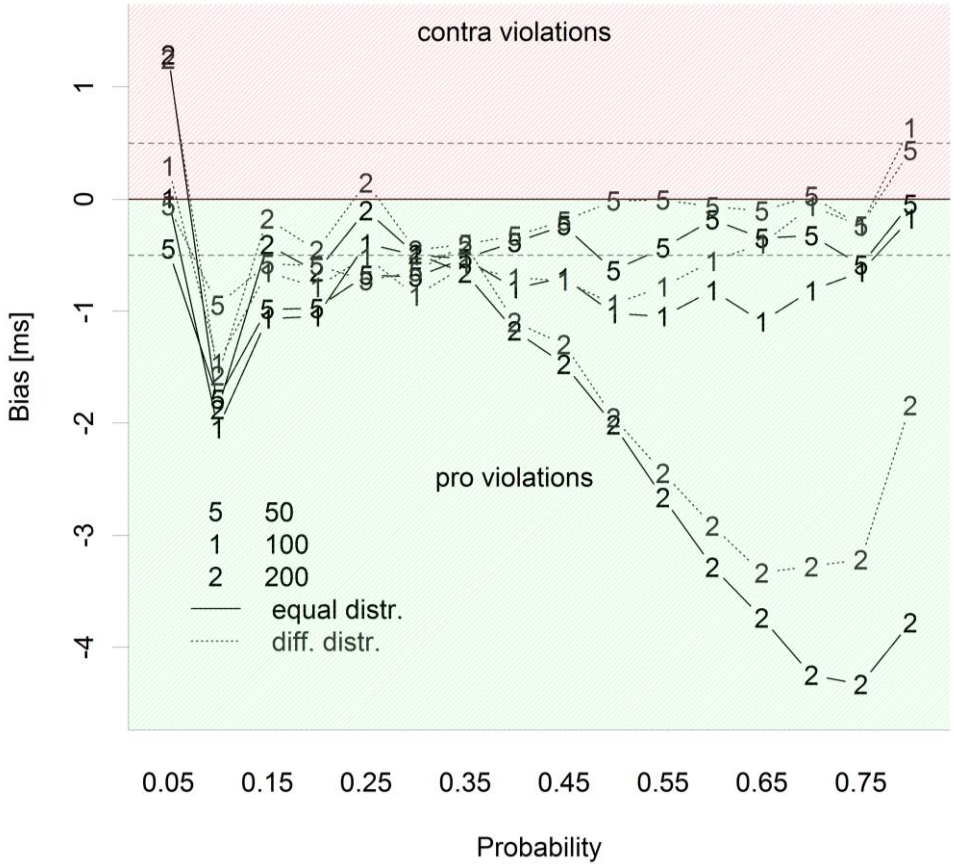

Figure 4.5 Estimation bias for coactivation models (for sample size 10) with base time variances. Full lines represent equal distributions conditions, and gray dotted lines represent different distributions conditions both for plausible λ values. Bias values below zero indicate probability points, where violations are facilitated, and values above zero probability points, where violations are hindered.

4.4.5 Overall Type I Errors

When turning to the performance of the RMI test with respect to “false alarms” or type I errors, it is plausible to expect them to increase as more percentiles are tested. The ranges of 5 - 10% (two percentiles), 5 - 20% (four percentiles), …, and 5 - 50% (ten percentiles) were looked into. Theoretically, the maximum amount of overall type I errors is then determined by the number of percentiles tested, multiplied by the significance level of the (inner or single) *t*-test. For example, testing in the range of 5 - 30% with a significance level $\alpha = 0.05$ can maximally render overall type I errors of $6 \cdot 5\% = 30\%$.

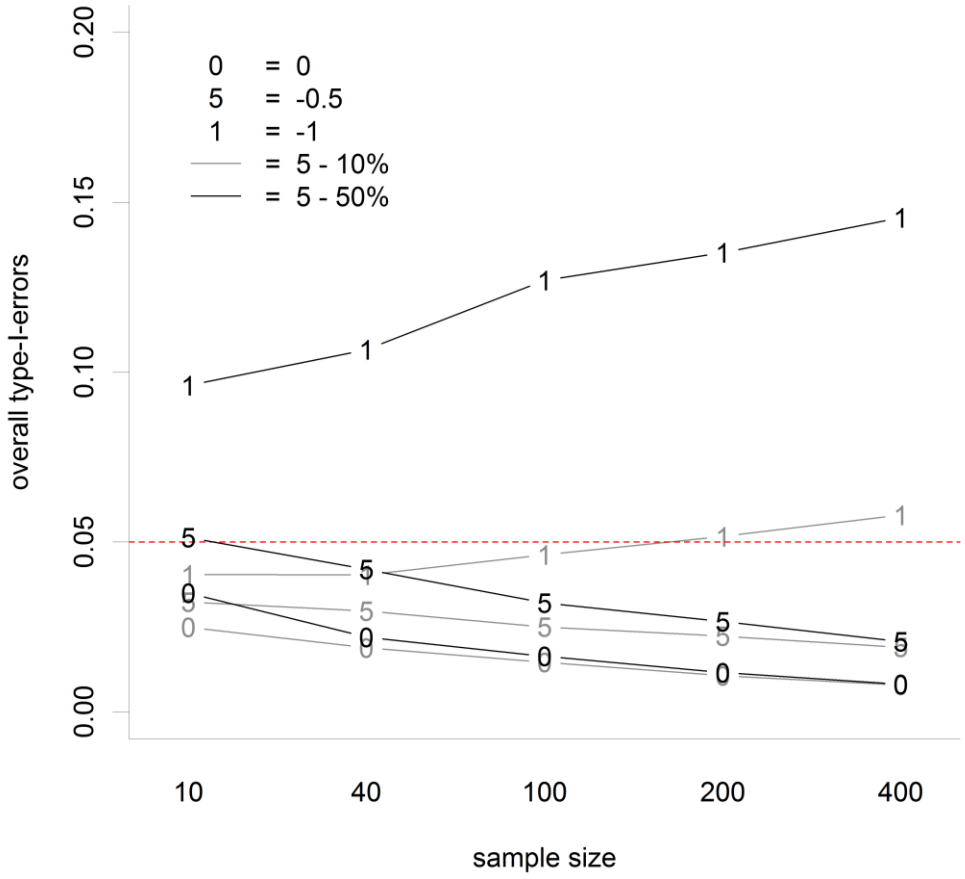

Figure 4.6 Overall type I errors for differently correlated race models across sample sizes tested at percentile ranges 5 – 10% (gray lines) and 5 - 50% (black lines). Correlation between channels *X* and *Y* is denoted by 0 for uncorrelated, 5 for moderately negatively correlated and 1 for maximally negatively correlated. (Significance level at 0.05).

From power theoretic reasons (Cohen, 1988), the (overall) type I error should decrease with increasing sample sizes. This is the case for uncorrelated and weakly correlated conditions (see Figure 4.6 for the ranges 5 – 10% and 5 – 50%). There, an increase in sample size can effectively push the overall type I error even below the inner *t*-test significance level, $\alpha = 0.05$. The highly negatively correlated racers constitute an exception: an increase in sample size results in even higher and more pronounced overall type I errors. The amount of overall type I error is similar to Kiesel et al.'s reported α-accumulation and results in a triplication of the (inner) significance level for the largest test range (when testing at 5 - 50%).

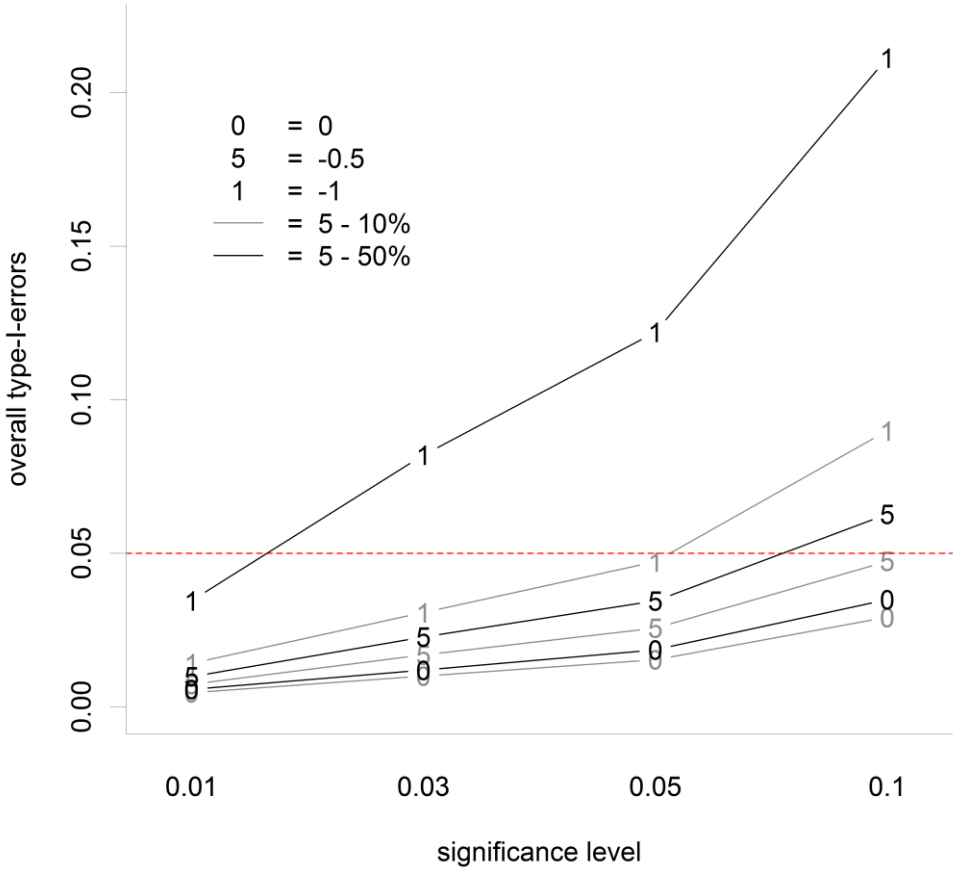

Figure 4.7 Overall type I errors for differently correlated race models across significance levels tested at percentile ranges 5 - 10% (gray lines) and 5 - 50% (black lines). Correlation between channels *X* and *Y* is denoted by 0 for uncorrelated, 5 for moderately negatively correlated and 1 for maximally negatively correlated.

By definition, an increase of the significance level *α* should lead to more frequent false alarms (for any test). The simulations provide consistent results, as the overall type I errors are directly proportional to the significance level *α* (see Figure 4.7). The rate of increase however, is moderated by the amount of negative correlation between the single target channels. For highly negatively correlated racers, an increase in the significance level is more than two-times higher than the inner significance level, when testing at six or more percentile points (see Figure 4.7, black line denoted by “1”).

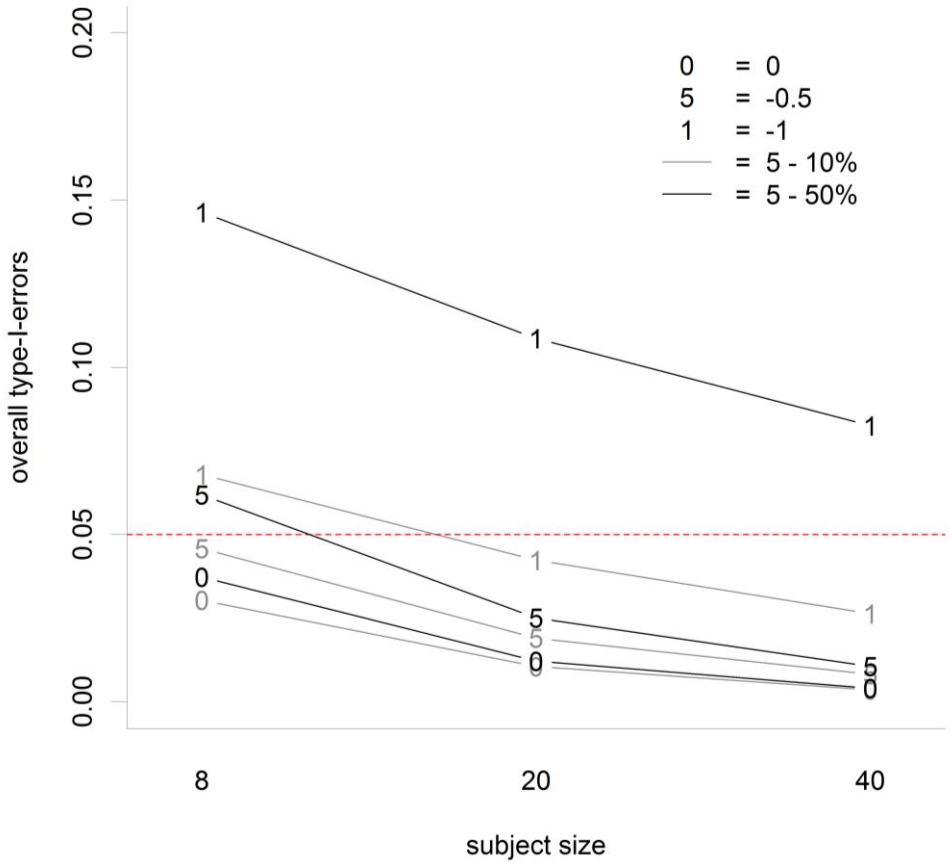

Figure 4.8 Overall type I errors for differently correlated race models across subject sizes tested at percentile ranges 5 - 10% (gray lines) and 5 - 50% (black lines). Correlation between channels *X* and *Y* is denoted by 0 for uncorrelated, 5 for moderately negatively correlated and 1 for maximally negatively correlated.

Specifically, for the RMI test applied in this simulation, there is another way to reduce type I errors, namely by vincentizing across subjects. Vincentizing is a method to average the percentile points of all subjects, to decrease the variance of the estimates of all cumulative

distribution functions (Ratcliff, 1979). Since this RMI test uses vincentizing, an increase of subjects should reduce overall type I errors. Indeed a beneficial effect of subject size on overall type I errors was found, as they can be halved, when using 40 instead of 8 participants (see Figure 4.8)

The different distributions condition in general produce slightly larger overall type I errors than the equal distributions condition (see Figure 4.9). This phenomenon will be treated in the discussion part.

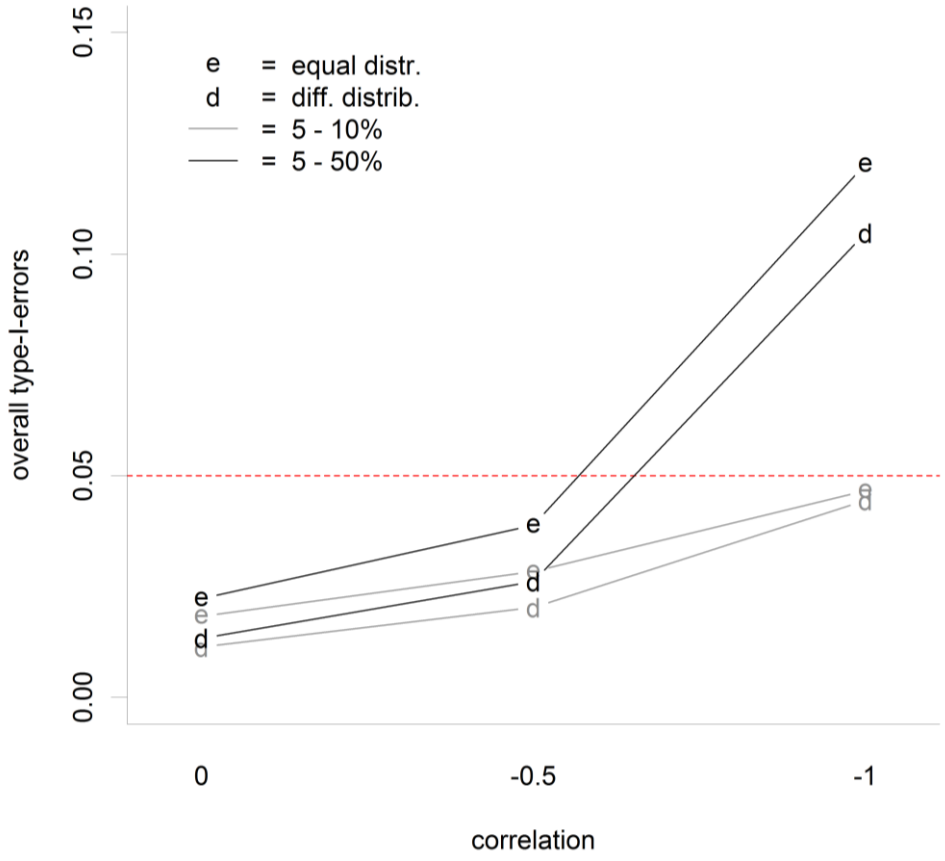

Figure 4.9 Overall type I errors for race models across correlations at percentile ranges 5 - 10% (gray lines) and 5 - 50% (black lines). Equal distributions condition is denoted by *e*, different distributions condition is denoted by *d*.

4.4.6 Effects of Base Time Variance

Specifically addressing the question of how the presence (of a potentially high) base time variance moderates the results, thus far the simulations integrate and extend the results

by previous simulation studies. The simulation by Townsend and Honey (2007) suggests, that the presence of a base time variance acts as a filter that can smear away violations (if the variance is high enough). As a consequence, an increase of base time variance should go along with a decrease in overall type I errors. The simulations, replicate this relation (see Figure 4.10).

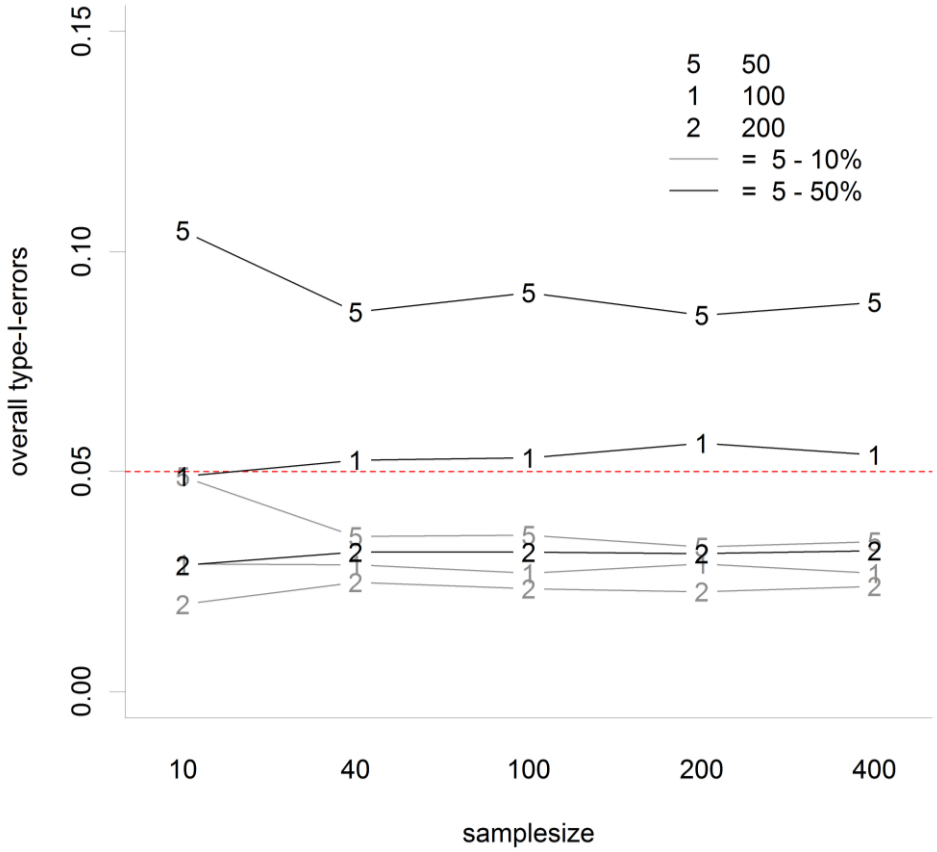

Figure 4.10 Overall type I errors for race models with different base time variances across sample sizes tested at percentile ranges 5 - 10% (gray lines) and 5 - 50% (black lines). Base time variances s_t of 50, 100, and 200 are denoted by 5, 1, and 2 respectively. (Significance level at 0.05).

A novel finding is that the amount of correlation mediates the effect of base time variance, as highly negatively correlated race models benefit more strongly from high base time variance (s_t = 200) with respect to the overall type I errors (see Figure 4.11).

Overall, the amount of negative correlation has the highest impact on overall type I errors, as these conditions alone show a substantial accumulation detriment. The uncorrelated or slightly correlated racers (*corr* = 0 or -0.5) produce overall type I errors well

within the inner significance level (provided, that the sample size is high enough, i.e. 40 or more). Table 4.7 summarizes the results for the overall type I errors in the simulation.

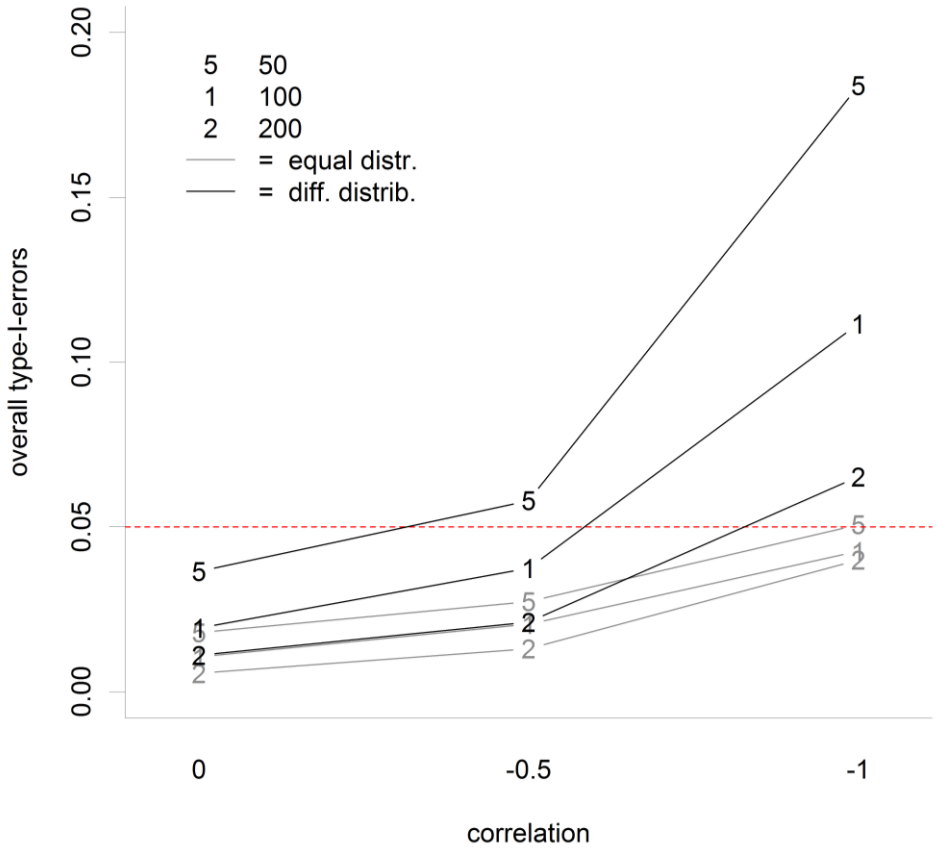

Figure 4.11 Overall type I errors for race models with different base time variances across correlations for equal (gray lines) and different distributions (black lines). Base time variances s_t of 50, 100, and 200 are denoted by 5, 1, and 2 respectively.

Table 4.7

Summary of Simulation Results for Overall Type I Errors

Factor	**Effect on Overall Type I Errors**	
	corr = 0 or -0.5	corr = -1
Sample Size	Increase of Sample Size can help to retain Inner Significance Level	Alpha Accumulation occurs (up to 2- or 3-fold)
Alpha Level	Sublinear Increase	Almost linear increase
Subject Size	Higher Subjects Sizes (linearly) lower Type I Errors	
Distribution Relation	Different Distrib. has higher Type I Errors than equal Distrib.	
Base Time Variance	Higher Base Time Variance results in lower Type I Errors	

Notes. See the respective elaborations in the text for more details.

4.4.7 Power Analysis

Avoiding false alarms is important to conduct reliable scientific results. Measures to ensure a low false alarms rate should however, not diminish the power of a test, in a way that actually existing effects are rendered undetectable. In the simulations, the power of the RMI test is generally poor (considerably below 80%) for most conditions. Only in the different distributions conditions and for (potentially) implausibly high amounts of coactivation and for high sample sizes, acceptable power rates can be achieved.

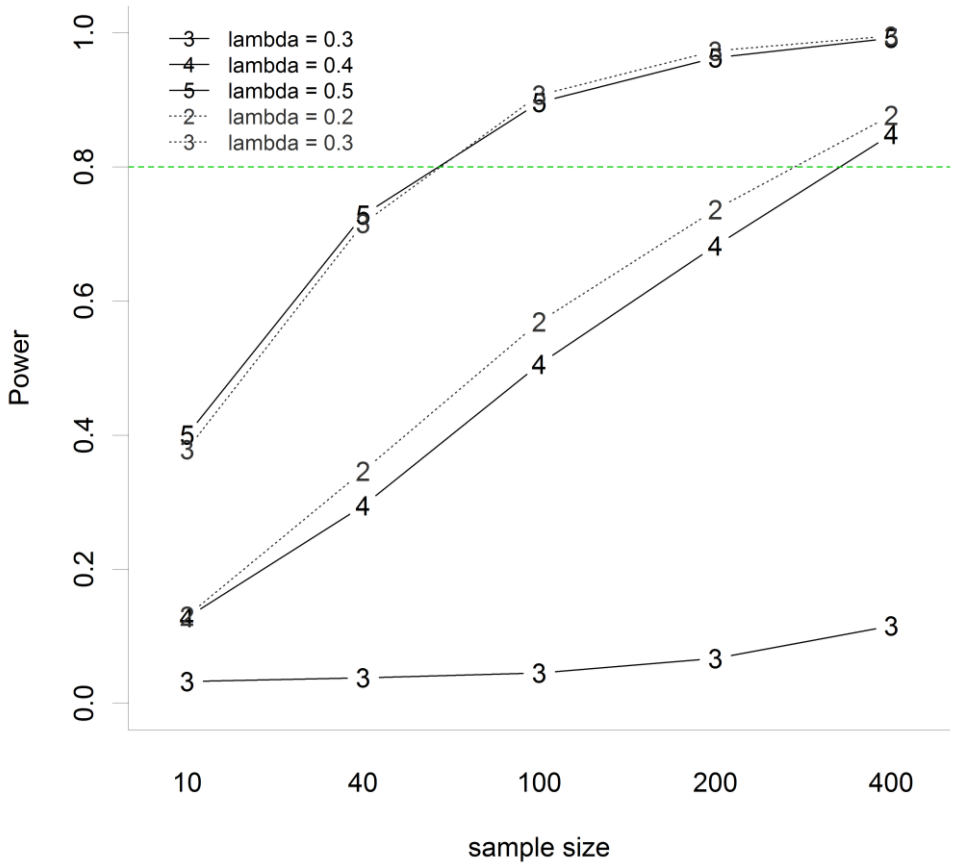

Figure 4.12 Power rates for increasingly coactivated models across sample sizes tested at percentile range 5 - 50%. Black full lines represent equal distributions conditions, and gray dotted represent different distributions conditions, both with plausible strength of coactivation *λ* values. (Significance level at 0.05).

As for the overall type I errors, sample size have a beneficial effect on the test performance: an increase in sample size generally results in a raise of power. This is even more accented for larger rates of coactivation. For coactivation models with RSE's

comparable to the race models, power is in a range of 10 to 40% for sample size 10 but well above 80% for sample size 400, when testing in the range of 5 – 50% (see Figure 4.12). An exception of this is the coactivation model with the lowest (plausible) coactivation strength (λ = 0.3 in the equal distributions condition). There, power never reaches beyond 10%.

Interestingly, an increase in the significance level is not very effective: even for the maximally coactivated model, a relaxing of the inner α from 0.01 to 0.10 increased the power at the most by 10% (see Figure 4.13). Keeping in mind, that an increase of the significance level affects overall type I errors disproportionally, this strongly advises against such a procedure to increase the power.

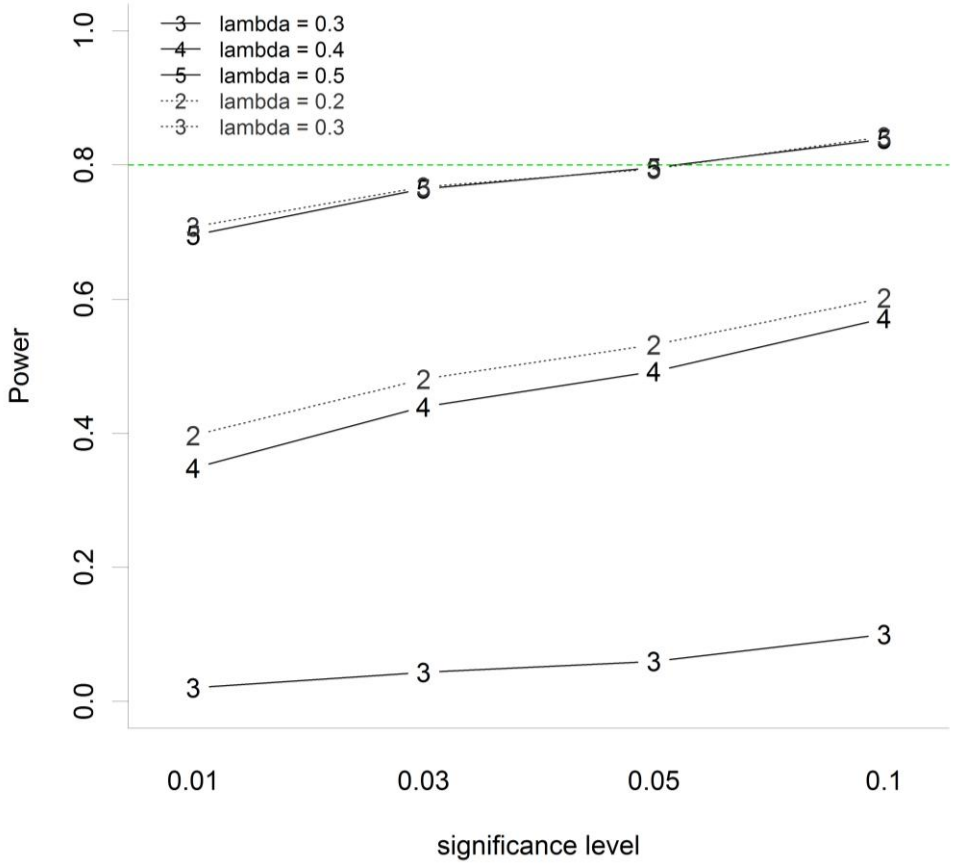

Figure 4.13 Power rates for increasingly coactivated models across significance levels tested at percentile range 5 - 50%. Black full lines represent equal distributions conditions, and grey dotted lines represent different distributions conditions, both with plausible λ values. (Significance level at 0.05).

The effect of subject size is also not as prominent as for the overall type I errors. In the equal distributions condition an increase of subjects from 8 to 40 results in a 3 to 8% rise in power (see Figure 4.14). As this procedure concurrently reduces overall type I errors, it should however be taken into account more seriously, than an increase of the significance level.

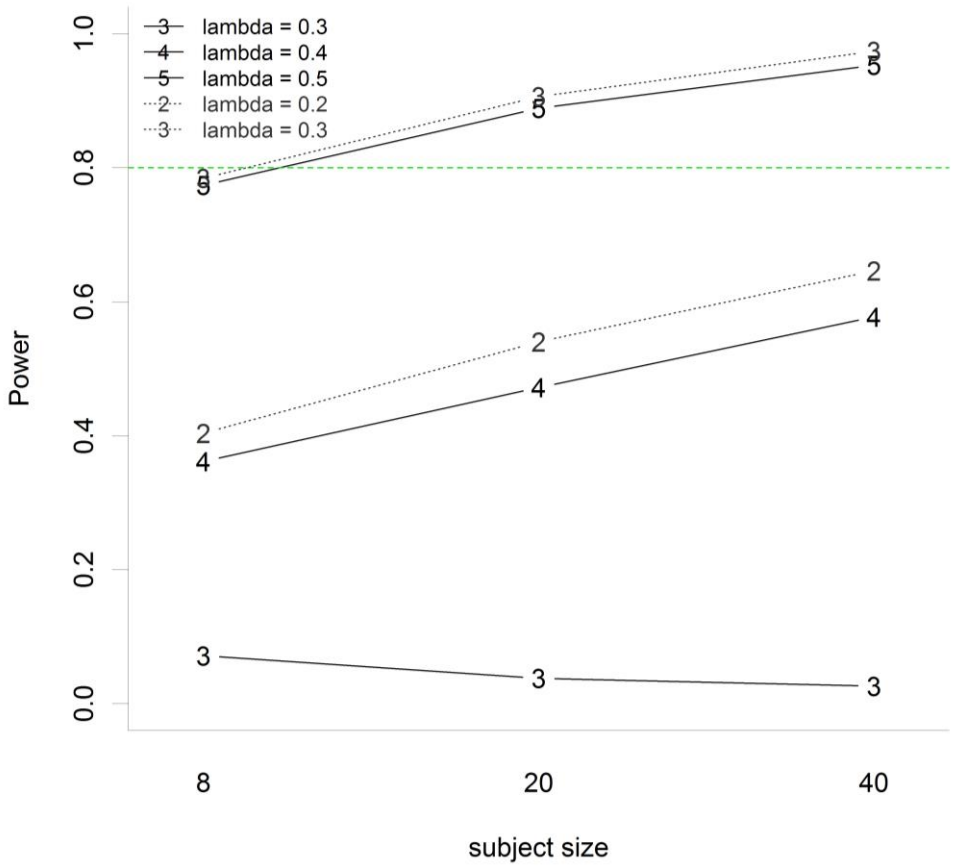

Figure 4.14 Power rates for increasingly coactivated models across significance levels tested at percentile range 5 – 50%. Black full lines represent equal distributions conditions, and grey dotted lines represent different distributions conditions, both with plausible λ values. (Sample size 100).

Unlike for the overall type I errors, a strong distribution relation effect on power is found. On average the equal distributions condition displays worse power rates, than the different distributions condition (see Figure 4.15, aggregated over coactivation strength). In the different distributions condition, power rates of around 80% can be reached, even for testing only 8 subjects (sample size 200 and 400). For the equally distributed conditions, power rates remains below the desirable 80% mark, even for the highest sample and subject size conditions. Theoretical explanations of this distribution relation effect are elaborated in the discussion.

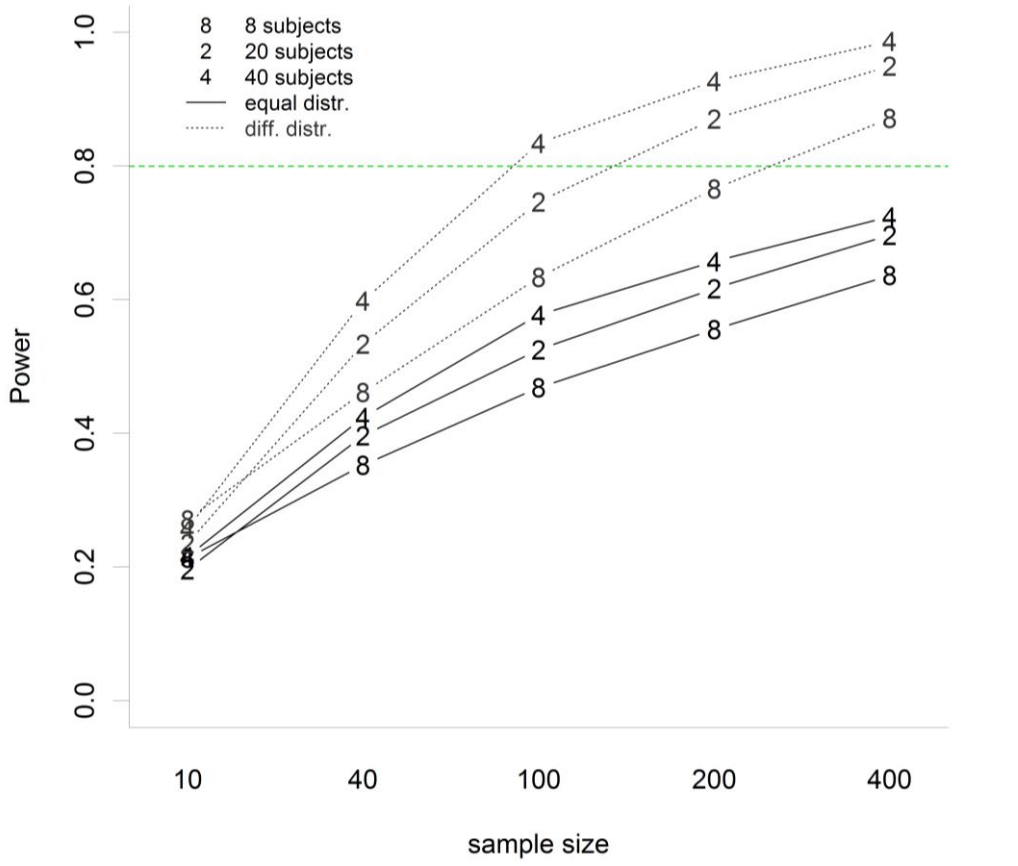

Figure 4.15 Power rates for increasingly coactivated models across sample sizes at percentile range 5 - 50%. Black full lines represent equal distributions conditions, and grey dotted lines represent different distributions conditions, both with plausible λ values. Subject sizes of 8, 20, and 40 are denoted by 8, 2, and 4. (Significance level at 0.05.)

Concerning the effect of base time variance on the power of the test, the simulations can confirm the theoretical considerations by Townsend & Honey (2007). An increase of the

base time variance can push the RMI test from a healthy detection rate to near undetectability (see Figure 4.16). Looking at plausible λ ranges reveals interesting results. In the equal distributions conditions (see Fig 4.16, full lines) an increase of base time variance from 50 to 200 means a drop in power rate of 10%, 45% and 35% for λ = 0.3, 0.4 and 0.5. Similarly, in the different distributions conditions (Fig 4.16., dotted lines) this base time variance increase results in a drop of 45% and 30% for λ = 0.2 and 0.3. Table 4.8 summarizes the results of the power analysis with respect to all manipulated factors.

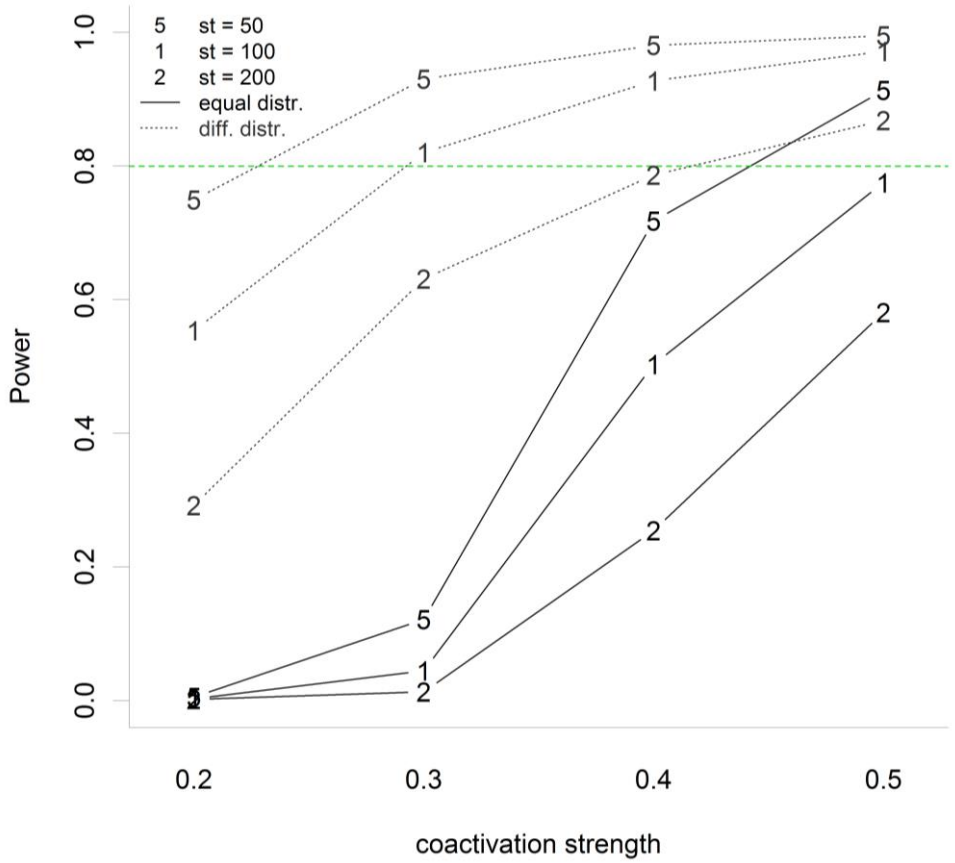

Figure 4.16 Power rates for coactivated models with increasing base time variance across coactivation strength at percentile range 5 - 50%. Black full lines represent equal distributions conditions, and grey dotted lines represent different distributions conditions, both with plausible λ values. Base time variances s_t of 50, 100, and 200 denoted by 5, 1, and 2. (Significance level at 0.05).

Table 4.8

Summary of Simulation Results for Power Rates

Factor	Power analysis
Sample Size	Power can exceed 80% with highest Sample Size (400) (except for λ = 0.3 equal distributions, this remains below 10%)
Alpha Level	Increase from 0.01 to 0.1 leads to a Power Raise of 15% maximally
Subject Size	Increase from 8 to 40 leads to a Power Raise of maximally 10%
Distribution Relation	Different distributions generally higher than equal Distributions
Base Time Variance	s_t increase from 50 to 200 leads to a Power Drop of 10 to 45%

Notes. See the respective elaborations in the text for more details.

4.5 Discussion and Conclusion

Although the RMI test is ubiquitous in multimodal and intersensory research, it has not yet been exhaustively examined. With this simulation study, RMI testing was investigated by addressing the following specific questions: (a) what is the power of the RMI test, (b) what is the estimation bias for coactivated data, (c) what are the effects of (high) base time variance on overall type I errors, power and bias, and (d) are the previous findings (at least qualitatively) replicable with a mechanistic reaction time model, the RDM. The conceptual scope of the studies by Kiesel et al. (2007) and Townsend and Honey (2007) were integrated and extended, as (differently correlated) race models and coactivated models (based on a weighted drift rate summation) were incorporated. The performance of the RMI test was analyzed under different statistical, experimental and reaction time parameters to ensure generalizability.

The results confirm the efficacy of the implemented manipulations (seven fully crossed factors), reflected in the means and standard deviations of the synthesized data. Not only do the race and coactivation models behave in an empirically plausible way, the entire simulation framework represents a potent and reliable tool for researchers of the RMI. It enables them to probe the performance of the RMI test for specific experimental and statistical setups and/ or parameter sets a priori or a posteriori. This can help optimize the RMI test and thus solidify inferences drawn from its outcome - be it pro or contra race models. Apart from addressing the research questions, it can be viewed as a proof of concept, independent of the specific parameter choices of the simulations. Naturally, the findings are conditioned on the parameter choices (for both the RDM parameters and statistical/ experimental parameters) and on the chosen coactivation architecture. Due to the modular design of the simulation framework, specific components and steps are exchangeable and lend themselves to further investigations and validations.

After a compact summary of the results, the distribution relation effect, potential reasons for differential results will be addressed and concrete implications for the empirical research will be sketched out.

4.5.1 Power Analysis of Latent Coactivation

What was missing in Kiesel et al.'s study (2007) was a look at the power of the RMI test. The simulation framework of the present study features a coactivation model based on a weighted drift rate summation between the channels *X* and *Y*. It can thus produce controllably coactivated channels. The boundary condition to pick "realistic" coactivation

strengths was to choose the ones, that produce RSE's comparable to the race models. The power analysis of the RMI in the simulations revealed both known and novel findings. For "realistic" coactivation strengths (λ = 0.3, 0.4, and 0.5 for equal and λ = 0.2 and 0.3 for different distributions condition), the power is rather low. For those λ's a healthy 80% power is achievable only for high sample sizes and low base time variance, while testing at multiple points. Higher λ's were able to reach that threshold with a lower sample size and for higher base time variances, but produce too large RSE's, to be considered empirically plausible and were not reported (for the full data set visit the authors homepage). This marks the first simulation study to report quantitative estimates of the power rates of RMI testing.

4.5.2 Alpha-Accumulation only for negatively correlated Race Models

The analysis of overall type I-errors attested only critical α-accumulation for the highly negatively correlated race models (*corr* = -1). For uncorrelated and moderately correlated racers (*corr* = -0.5, and 0), the inner significance level could be successfully retained, if passably high sample and subject sizes are chosen (greater than 40 samples per condition and 40 subjects). This is in contrast to the results by Kiesel et al. (2007), but may well be due to the differences in simulation parameters and/ or models used to generate the reaction times.

An important practical implication of this correlation effect is to find ways to estimate the channel correlation. If one were to exclude high negative correlations between the single target channels, the RMI test could more clearly separate race from coactivation architectures, as it would be turned less conservative. Also, problems with estimation bias and α-accumulation are drastically reduced. Estimating correlation between channels though, is in itself not unproblematic. Ulrich and Giray (1986) have shown analytically and via simulations that the estimation of cross-channel correlation is biased, in the presence of (high) base time variance. Further modeling and simulation studies are necessary to address this issue.

4.5.3 Estimation Bias is neglectable

Concerning the estimation bias, the simulation in general could not confirm the patterns in Kiesel's study. The race bias curves in Kiesel et al. (2007) have a stretched inverted U-shape curve, while the ones of the present study show this pattern only for the lowest sample size condition 10. Apart from that condition, all the others follow no systematic shape and are nearly flat. Furthermore, the biases (both for race and coactivation models) are bounded by 0.5 ms for both sides, once more than 10 samples per condition are taken.

For plausible λ choices in both the equal and different distributions conditions, there are biases favoring violations, albeit in a probability range where (due to the construction of the RMI) no violations can be found.

There is an interaction of correlation and base time variance for the estimation base in race models, which produces biases against violations for lower probability points (p = 0.2, …, 0.4) and pro violations for higher probability points (p = 0.5, …, 0.9) in the smallest sample size conditions (10). The latter part is also true for the coactivation models (again, only sample size 10).

At least for these RDM parameter choices, the estimation biases both for race and coactivation models become negligible, once reasonably high sample sizes are chosen (again more than 10 samples per condition). As low and unequal sample sizes for the single and redundant channels or other RT distribution functions and coactivation models were not investigated, systemic biases cannot be excluded for those combinations. It might also be the case that for other parameters sets and reaction time models, systematic patterns may still emerge. This is now empirically addressable, as follow-up simulations can investigate these variants with help of the simulation framework.

4.5.4 High Base Time Variance reduces both Type I Errors and Power

Referring to Townsend and Honey's (2007) results, the simulation can basically replicate the base time variance effect on type I errors and power - namely a reduction of both. The only exception being, that an increase of sample size always helps the power, while the authors found the opposite phenomenon for the high base time variance condition. There an increase in sample sizes worsens the power of the test.

Another new finding of the simulation study is the interaction of high negative correlations and base time variance for overall type I errors. Low base time variances (s_t = 50 and 100) and the maximal negative correlation (*corr* = -1) produce the highest overall type I errors, well above the inner significance level of the *t*-test. If the correlation is less negative (*corr* = 0 or -0.5) or the base time variance is highest (s_t = 200), then the inner significance level can almost be retained.

4.5.5 Distribution Relation Effect

An interesting, although theoretically deducible result is the distribution relation effect: power is considerably higher (up to 20%) in the different distributions conditions, than in the equal distributions conditions. As the RMI bound is calculated as a sum of the single target

channel CDFs (or estimates thereof, see equation 8 or chapter 2.3.5), the slower channel CDF drags the curve to the left and thus facilitates violations. Similar to Kiesel et al.'s results (2007), the effect on the overall type I error is less pronounced, so that in this case, the power is not bought at the expense of higher type I errors. This also provides a concrete empirical recommendation to increase the power: use stimuli with different feature contrasts to induce "rather different" single target channel CDFs. However, there should be follow-up studies/ simulations to fit empirical RSE data to this and other coactivation model candidates (for example, a nondecisional or combined model) to validate and further strengthen the results.

4.5.6 Alternative Architectures producing Violations

Whenever researchers fail to find violations of the Race model inequality in a redundant signals paradigm, doubts remain, whether this is due to data actually stemming from a race model architecture or the test being vastly underpowered. Mordkoff & Miller (1993) showed that robust violations can be observed when interstimulus contingencies, favoring redundant-target trials are present although the data was generated by a race architecture. Also, when measures are taken to exclude fast guesses, the RMI test is biased to favor race models (Miller & Lopes, 1991). Apart from these experimental and methodical issues, there are also alternative architecture that are capable to produce violations of the RMI: serial exhaustive (Townsend & Nozawa, 1997) or interactive race models (Mordkoff & Yantis, 1991). Zehetleitner, Krummenacher and Müller (2009a) tested these accounts experimentally and could eliminate them as potential architectures for the visual RSE (see chapter 2.2 for a review of the literature).

The findings of this study cannot without a doubt exclude an unfavorable choice of experimental parameters and (as a consequence) low power of the test, when researchers fail to find violations of the RMI. There are paradigms and specific experimental settings that for theoretic reasons should not induce a coactivated processing of the stimuli (see Feintuch & Cohen, 2002, Schröter, Ulrich, & Miller, 2007, Experiment 1). The simulation study contributes to the debate by uncovering nontrivial properties of the RMI test and providing a flexible framework that can improve RMI testing in a well-grounded manner.

4.5.7 How should RMI Testing look from now on?

A general recommendation to be inferred from the simulation is to perform an a priori simulation of the RMI test before "jumping to conclusions". By that, parameters like the subject and sample size, probability test points and violation criterion can be set, so that

overall type I errors and (potential) estimation biases are minimized while power is maximized. An increase of the significance level is strongly disadvised, as it has a disproportional effect on power and overall type I errors in the simulation. The most satisfying results are achieved for high sample and subject sizes, when high base time variance and negative correlation can be excluded.

Furthermore a power analysis provides the community (and the respective journal) with quantitative estimates of the overall type I errors and power rates and thus helps to solidify results (or null results). This is also in line with the new statistical guidelines (concerning the problem area of publication bias and reporting power analyses) publishers and journals are currently implementing (for example, Open Science Collaboration, 2012). Another more diagnostic use of the simulation framework can be to re-analyze “failed attempts” to detect violations and check, whether this was due to an underpowered design or potential estimation biases.

On a practical level, the question arises how to find the right coactivation strength for specific data sets. One pragmatic approach is to use programs that allow for estimation of the RDM parameters. One, the *EZ diffusion model* (Wagenmakers et al., 2007) uses a simplified version of diffusion models and provides an interactive homepage, where researchers can do a fast parameter fitting (http://www.ejwagenmakers.com/EZ.html). An alternative is the *fastdiffusion model analysis* program (Voss & Voss, 2007). With both, one can estimate the RDM parameters of the single channels and then simulate different λ’s to see which can produce the observed empirical RSE. More computationally intensive methods involve numerical fitting of the CDFs X, Y and XY_C to a drift rate constrained coactivation model and find the parameters rendering the best goodness-of-fit (for concrete fitting steps see chapter 3.2 or Ratcliff, 2002 or Gomez et al., 2007).

4.5.8 Conclusion

Summing up, the simulation study gives the following answers to the explicated research questions. The RMI test indeed is a very conservative test, as it misses out on (latent) coactivation in most empirically plausible cases. When coactivation models are used that produce a comparable RSE, violations can only be achieved for high sample and subject sizes and under the assumption that the base time variance is low. When the data is afflicted with a high base time variance, the chances of detecting violations of the RMI are yet worse.

The amount of estimation bias is negligible for both race and coactivation models as it is confined to a narrow -0.5 to 0.5 ms band for plausible probability ranges and experimental conditions (i.e. high enough sample sizes per experimental condition).

Concerning the effect of a highly varying base time component on the power, the simulations reaffirmed the results by Townsend and Honey (2007). A high base time variance decreases both the power and the type I errors of the RMI test. An extension of their results is the phenomenon that low base time variances produce large type I errors if the channel correlation is maximally negative. Thus, empirically it is important to find a means to estimate the channel correlation and at best, exclude the maximum negative case.

A novel finding is that the distribution relation affects the detectability of RMI violations: the more pronounced the difference in distributions, the higher the chances to detect coactivation, if it is the source of the RSE.

Overall, this study integrates and extends the studies of Towsend and Honey (2007) and Kiesel et al. (2007) in form of the simulation framework. Using state-of-the art models of decision making, it enables researchers now to further investigate and assess the RMI test for various experimental tasks, conditions and test related parameters. Also it provides a tool to optimally set up the RMI test, depending on the intention of the researcher (be it minimize α accumulation or maximize power of the test). Albeit, the RMI test alone is not able to settle the question of cognitive architecture in the redundant signals paradigm, it is the critical first step to rule out a large class of otherwise self-evident models.

5 GENERAL DISCUSSION

The application of formal mathematical models to empirical cognitive and Experimental psychology has paved the way for numerous revelations on the inner workings of the mind (e.g. the serial-parallel dilemma Townsend & Wenger, 2004). One prevalent enigma is the question about cognitive architecture in contexts where an integration of signals is necessary to carry out perceptual-motor tasks (e.g. the redundant signals paradigm). A prime example of how mental chronometry can help answer this question is the distributional bound on reaction times for the redundant signals effect, presented by Miller (1982). It established a pivotal non-parametrical test to exclude one class of explanatory model – race models. Furthermore it minimizes the risk of model mimicry as it is formulated on a distributional level and thus accounts for a larger proportion of the data.

The present study makes a vital contribution to this endeavor as it combines state-of-the-art models of decision making to the rigor of distributional testing. This combination enables researchers to analyze and evaluate the performance of the RMI test by means of computer intensive methods, such as Monte Carlo simulations and numerical fittings. The present study pursued objectives on a theoretical and methodological level. On a theoretical level it investigated the locus of coactivation for two bimodal detection tasks (with focused and divided attention). There, the quantile proportion fitting revealed that a combined decisional and nondecisional coactivation model was in best agreement to the data of Experiment 1 and 2. Interestingly, for the divided attention detection task in Experiment 2, the motor coactivation contributed more strongly to the redundant signals effect, than the decisional component.

On a methodological level, the study improved the understanding and value of the RMI test as a statistical tool, as the interplay of various psychological and experimental parameters on type I errors and power rates was explicated. Novel findings are the strong vulnerability of the test in terms of type I error accumulation when the data is maximally negatively correlated, and the pronounced distribution relation effect, which can impact the power of the test substantially. The present study is also the first to implement and test a mechanistic model of coactivation that allows for controlling the strength of integration. The study could show that such coactivation models frequently fail to produce RMI violations and healthy power rates are only attainable, if high base time variance can be excluded and both sample and subject sizes are maximized. The here developed simulation framework furthermore provides a well-grounded diagnostic tool for the RMI test, as it enables researchers to optimally set up the parameters of the test to their specific requirements. Due

to its modular design, it can be adapted to different RMI tests, decision models and experimental and statistical parameters.

In this last chapter, limitations and theoretical and methodological implications of the present study are discussed.

5.1 Limitations of the present Study

5.1.1 Theoretical

From a theoretical stance, it is important to clarify, that the RSE can be conceived as an umbrella term, enveloping different mechanisms, depending on the specific experimental situation at hand (that is, task, modalities and stimuli involved). This notion was already brought forward by Miller and Reynolds (2003). Although applying different experimental tasks and setups to the question of signal integration is necessary to come to a conclusive and robust answer concerning the structure and integration mechanism, it entails the risk of invalidly and unconsciously pooling across inherently different phenomena and mechanisms. In order to thoroughly understand the mechanisms and architecture necessary for the cognitive system to produce the data patterns for each experimental paradigm, a systematic and exhaustive inquiry is inevitable. A classification system for both the paradigms and explanatory models of the RSE is an important conceptual step (see Töllner et al., 2011). The diffusion model analysis (as adopted in the present study) will further help disentangle different accounts of coactivation models and thus shed light on the source(s) of the RSE. By applying this modeling approach to further experimental tasks and modality combinations, it can be determined whether the RSE has to be divided into subphenomena and according to what categories (e.g. locus of coactivation and contribution of loci). The present study provides a first step, examining simple detection tasks in the audiovisual domain.

On a critical note, it is theoretically possible that this program might fail because of an untestable presupposition. The formulation of the RMI bound and its application to empirical data rests heavily on the assumption of context independence. Context independence assumes that the processing rate of each channel is invariant, across the single and redundant signals conditions. It is crucial for the RMI test, as it justifies the comparison between data gathered in the single and redundant conditions. It is however untestable, at least with respect to mental chronometry. Similar to this barrier, is the equivalent in the motor regime. It assumes that the base times are equal across conditions (and thus independent of the decision stage). If any of these assumptions do not hold, the whole RMI test is theoretically unhinged and invalid. (Solutions to this caveat will be discussed in 5.3).

5.1.2 Methodological

The present study used an implementation of the Ratcliff diffusion model to generate and fit the data for race and coactivation models. Ratcliff diffusion models, as an instance of sequential sampling models, allow for a direct mapping of psychological variables to the model parameters and have shown to exhibit excellent fits to empirical data on behavioral and neuronal level. In contrast to descriptive (and atheoretic) models of reaction times, sequential sampling models give a full account of the observers performance, as it provides accuracy and reaction time distribution for both correct and incorrect responses.

It is conceivable, that these models will be refined and modified over time as further research will necessitate this. Certainly, the results of the simulations in the present study depend on the reaction time generating model. The use of alternative models of decision making (and thus, reaction times) can come to differential results. Due to the modular design and flexibility, the generality and transferability of results can be explicitly investigated. If for example, researchers want to test the RMI bound using other models for decision making or different distribution functions, only one module in the simulation framework needs to be adapted. Also the application of different violation criteria or alternative RMI tests is facilitated by the modular design of the framework. Alternative tests of the RMI have been proposed (see Colonius & Diederich, 2006; Maris & Maris, 2003), however the current study has focused on investigating and implementing only the most widespread version of RMI testing. It is imaginable that one of these alternative test will feature more favorable statistical properties (in terms of type I errors and power rates) for specific boundary conditions compared to the standard RMI test. This now is testable as the simulation framework allows for implementing different tests and test criteria.

With regards to the performance of the fitting procedure in chapter 3, further theoretical and analytical investigations are necessary. One question is whether, one can draw conclusions from the general fitting performance (in terms of failed fits) to the aptitude of a specific model for the data. The fitting results of Experiment 1 and 2 exhibit this remarkable pattern, as the model with the best overall fit was at the same time the model with the highest rate of failed fits.

Another aspect of the fitting procedure that must be looked into to further strengthen the results of the locus analysis is a series of validation simulations. Data generated by known parameters have to be fit by the used procedure to see, whether it is able to recover the instantiated parameters. This would serve as a proof for the validity of the fitting results. One way to assess this parameter recovery – a parametric bootstrap method - has been presented by Wagenmakers, Ratcliff, Gomez and Iverson (2004).

Concerning the goodness-of-fit measure, a comparative analysis of in the fitting procedure could help optimize the computational effort and fitting precision for different cost functions. Also, procedures to deal with contaminant reaction times could provide a clearer distinction between models (see Ratcliff & Tuerlinckx, 2002).

A combination of methods would also prove beneficial for the fitting procedure (improving the rate of successful fits), as different measures (e.g. physiological or response related measures) might exclude or further constrain parameters of the models. Especially in Experiment 1, an estimate of the base time variance would be obtainable, using the dual response paradigm by Ulrich and Stapf (1984; see chapter 2.2.6). Adapting this experiment by introducing another contrast level for the auditory and visual stimulus would allow the application of equation 5 to estimate s_t for the fitting.

5.2 Implications for Theory and Methodology

On a meta-level, the present study strongly promotes and advocates the extended use of sequential sampling models for accounting empirical data of decisions across a wide band. These current models combine an explicit formulation with a theoretical foundation. The formalization invites researchers to effortlessly implement these models and make use of them in large scale simulations. The theoretical foundation guarantees that parameters shifts can be interpreted in a psychologically sound way and reveal changes in the way the cognitive system processes the information. The use of sequential sampling models also concurrently acknowledges the accuracy and the reaction time distribution determined by the same set of model parameters.

Applied to the redundant signals paradigm, the following natural follow-up questions come to mind. For the redundant signals paradigm the modeling approach of the present study, provides a means to systematize the RSE with respect to its generating mechanism and locus. Using the coactivation model variants in the present study to analyze the fit data of previous RSP studies (in form of a re- or meta-analysis) might indicate at task or modality combinations that rather induce a decisional or nondecisional integration of signals or even retain race models. This can effectively help disentangle the umbrella term of the RSE.

Another application would be the modeling of the RSE for more than two racers. As there is no direct analogon to the RMI test for the multivariate case (see Colonius & Diederich, 2006), a fitting procedure could help specify the mechanisms for the tri-racers in more detail.

The simulation framework enables researchers of the RSE to perform elaborate a priori or a posteriori analyses of the RMI test. These may include power or type I error accumulation analyses. Based on these diagnostic tests, the outcomes of the RMI test in specific contexts can be solidified and statistically ensured.

The simulation study itself revealed the detrimental effect of a high negative correlation between the signal channels on type I errors. For practitioners it would be vital to exclude this extreme race model in their empirical data. The fitting procedure of the present study might offer a test for that. Researchers could first fit the individual single signals trials to a diffusion model and then perform a fitting of the redundant data, where the correlation coefficient is fitted. This fitting can be constrained by the best fitting parameters of the single signals conditions. Previous simulations with atheoretic reaction time models suggest that the interchannel estimation might be affected by the base time variance (see Ulrich & Giray, 1986). Within the simulation framework of the present study however, this research question is directly addressable.

On a macroscopic time scale, a diffusion model analysis might further help illuminate the development of the RSE. Fitting empirical data of observers of different age groups (while accounting for standard findings from developmental psychology) might further localize and test the generating mechanisms of the effect. On more fine grained time scales, the learning or practice effect of the RSE could by analyzed, employing a diffusion model analysis. Experiments featuring many blocks over an extended period of time could form the basis to analyze changes of the mechanisms (or the locus) generating the effect. Both the aging and practice effect have already been successfully investigated with help of the Ratcliff diffusion model (see Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009; Ratcliff, Thapar, et al., 2004).

Beyond the paradigm of mental chronometry, diffusion model analyses can be naturally combined with other physiological or imaging measures. To illustrate the generality and potency of this modeling approach, the study by Jepma, Wagenmakers and Nieuwenhuis (2009) combined electrophysiological measures with a diffusion model analysis in the accessory stimuli paradigm to pinpoint the source of the accessory stimuli effect. The diffusion model analysis suggested that the parameters of the decision process were not affected by the presence of an accessory stimulus. The EEG component analysis revealed a stimulus-dependent modulation and point to a speed-up at the level of stimulus encoding. This signifies a conclusion that would not have been possible, when the analysis were limited to either method alone and illustrates the synergies possible for combining diffusion model analyses to prevalent methods and measures in experimental psychology like EEG, fMRI, PET, etc.

When turning to more complex perceptual decisions, it will be inevitable to model multi-stage decision models or decision chains with help of sequential sampling models. Again, burrowing from the history of mental chronometry will prove fruitful for this endeavor. Using the logic of Donders and Sternberg together with diffusion model analyses, will help model complex behavior. Concretely this entails the partition of complex decisions to simple(r) decisions. First a fitting analysis of these initial simple decisions (like the detection stage) is performed. With help of the “first link model” a restricted fitting for the subsequent decision stages is rendered possible. This approach is invariant to the specific paradigm and context of the decision.

Overall, the theoretical and formal framework provided by current models of decision making allow for new ways of probing the architecture and organization of cognitive processes. Due to the psychological plausibility and interpretability of sequential sampling models, an interface to other investigation methods and research paradigms is provides (for a neural application, see the review by Gold & Shadlen, 2007).

5.3 Overall Conclusion

The present study combined a classical paradigm of mental chronometry to state-of-the-art models of decision making. The study improved the understanding of coactivation as it could identify a combined decisional and nondecisional locus for the redundant signals effect by a diffusion model analysis. A large scale simulation was performed to investigate the statistical performance of the most important distributional test to infer on the cognitive architecture involved in decisions based on evidence integration. The simulation revealed that coactivation is frequently missed for standard experimental and statistical parameters. Highly negative race models are prone to produce large amounts of type I error accumulation and base times hinder the detection of violations for both race and coactivation models.

On a methodological level, the presented simulation framework and the fitting procedures will directly help researchers conducting their experiments and adapting their analyses in the redundant signals paradigm. A critical assessment of the results as well as theoretical and empirical implications of the present study were provided. Building on these elaborations, an outline of future research ideas was presented and concludes the present study.

REFERENCES

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

Atkinson, R.C., Holmgren, J.E., & Juola, J.F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics, 6*(6), 321-326.

Balota, D.A., & Yap, M.J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20*(3), 160-166.

Billingsley, P. (2008). *Probability and measure*: John Wiley & Sons.

Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Brown, S., Steyvers, M., & Wagenmakers, E.J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology, 53*(6), 453-462.

Bucur, B., Allen, P.A., Sanders, R.E., Ruthruff, E., & Murphy, M.D. (2005). Redundancy gain and coactivation in bimodal detection: Evidence for the preservation of coactive processing in older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60*(5), P279-P282.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: Lawrence Erlbaum.

Collaboration, Open Science. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science, 7*, 657-660.

Colonius, H. (1986). Measuring channel dependence in separate activation models. *Attention, Perception, & Psychophysics, 40*(4), 251-255.

Colonius, H. (1988). Modeling the redundant signals effect by specifying the hazard function. *Perception & psychophysics, 43*(6), 604-606.

Colonius, H. (1990a). A note on the stop-signal paradigm, or how to observe the unobservable. *Psychological Review, 97*(2), 309.

Colonius, H. (1990b). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology, 34*(3), 253-275.

Colonius, H. (1999). A theorem on parallel processing models with a generalized stopping rule. *Mathematical Social Sciences, 38*(3), 247-258.

Colonius, H., & Diederich, A. (2006). The race model inequality: Interpreting a geometric measure of the amount of violation. *Psychological Review, 113*(1), 148.

Colonius, H., & Ellermeier, W. (1997). Distribution inequalities for parallel models of reaction time with an application to auditory profile analysis. *Journal of Mathematical Psychology, 41*(1), 19-27.

Colonius, H., & Vorberg, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology, 38*(1), 35-58.

Corballis, M.C. (1998). Interhemispheric neural summation in the absence of the corpus callosum. *Brain, 121*(9), 1795-1807.

Diederich, A. (1992). *Intersensory facilitation: Race, superposition, and diffusion models for reaction time to multiple stimuli*: Peter Lang (Frankfurt am Main and New York).

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology, 39*(2), 197-215.

Diederich, A., & Busemeyer, J.R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics, 68*(2), 194-207.

Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? *Psychological Research, 49*(1), 23-29.

Donders, F.C. (1969). On the speed of mental processes. *Acta psychologica, 30*, 412.

Donkin, C., Brown, S.D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review, 16*(6), 1129-1135.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review, 16*(6), 1026-1036.

Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology, 2*(2), 312-329.

Egeth, H. E., & Mordkoff, J. T. (1991). Redundancy gain revisited - evidence for parallel processing of separable dimensions. *Perception of Structure*, 131-143.

Egeth, H.E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics, 1*(4), 245-252.

Einstein, A. (1905). Investigations on the Theory of the Brownian Movement. *Annalen der Physik*.

Feintuch, U., & Cohen, A. (2002). Visual attention and coactivation of response decisions for features from different dimensions. *Psychological Science, 13*(4), 361-369. doi: 10.1111/1467-9280.00465

Forstmann, B.U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D.Y., Ridderinkhof, K.R., & Wagenmakers, E.J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences, 105*(45), 17538-17542.

Gardiner, C.W. (2004). *Handbook of stochastic methods*. Berlin, Germany: Springer.

Gilchrist, W. (2000). *Statistical modelling with quantile functions*: Chapman & Hall/CRC.

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.

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

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of experimental psychology: General, 136*(3), 389-413. doi: 10.1037/0096-3445.136.3.389

Gondan, M., Götze, C., & Greenlee, M.W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Attention, Perception, & Psychophysics, 72*(6), 1692-1709. doi: 10.3758/APP

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics* (Vol. 1): Wiley New York.

Grice, G.R., Canham, L., & Boroughs, J.M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Attention, Perception, & Psychophysics, 35*(5), 451-463.

Hazen, A. (1914). Storage to be provided in impounding municipal water supply. *Transactions of the American Society of Civil Engineers, 77*(1), 1539-1640.

Heijden, A.H.C., Heij, W., & Boer, J.P.A. (1983). Parallel processing of redundant targets in simple visual search tasks. *Psychological Research, 45*(3), 235-254.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology, 63*(3), 289.

Ho, T.C., Brown, S., & Serences, J.T. (2009). Domain general mechanisms of perceptual decision making in human cortex. *The Journal of Neuroscience, 29*(27), 8675-8687.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.

Hubel, D.H., & Livingstone, M.S. (1985). Complex–unoriented cells in a subregion of primate area 18. *Nature, 315*(6017), 325-327.

Hubel, D.H., & Livingstone, M.S. (1987). Segregation of form, color, and stereopsis in primate area 18. *The Journal of Neuroscience, 7*(11), 3378-3415.

Hyndman, R.J., & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician, 50*(4), 361-365.
Iacoboni, M., & Zaidel, E. (2003). Interhemispheric visuo-motor integration in humans: the effect of redundant targets. *European Journal of Neuroscience, 17*(9), 1981-1986.
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.
Kecs, W., & Giurgiuţiu, V. (1982). *The convolution product and some applications*: Editura Academiei.
Kiesel, A., Miller, J.O., & Ulrich, R. (2007). Systematic biases and type I error accumulation in tests of the race model inequality. *Behavior Research Methods, 39*(3), 539-551.
Kinchla, R.A. (1974). Detecting Target Elements in Multielement Arrays - Confusability Model. *Perception & Psychophysics, 15*(1), 149-158.
Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*(4), 219-227.
Koene, A.R., & Zhaoping, L. (2007). Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom–up saliency map in V1. *Journal of Vision, 7*(7).
Krummenacher, J., Müller, H.J., & Heller, D. (2001). Visual search for dimensionally redundant pop-out targets: Evidence for parallel-coactive processing of dimensions. *Attention, Perception, & Psychophysics, 63*(5), 901-917.
Krummenacher, J., Müller, H.J., & Heller, D. (2002). Visual search for dimensionally redundant pop-out targets: parallel-coactive processing of dimensions is location specific. *Journal of Experimental Psychology: Human Perception and Performance, 28*(6), 1303.
LaBerge, D. (1994). Quantitative models of attention and response processes in shape identification tasks. *Journal of Mathematical Psychology, 38*(2), 198-243.
Laming, D.R.J. (1968). *Information theory of choice-reaction times*. Oxford, England: Academic Press.
Laughlin, S.B., & Sejnowski, T.J. (2003). Communication in neuronal networks. *Science, 301*(5641), 1870-1874.
Link, S.W., & Heath, R.A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40*(1), 77-105.
Livingstone, M.S., & Hubel, D.H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *The Journal of Neuroscience, 4*(1), 309-356.
Livingstone, M.S., & Hubel, D.H. (1987). Connections between layer 4B of area 17 and the thick cytochrome oxidase stripes of area 18 in the squirrel monkey. *The Journal of Neuroscience, 7*(11), 3371-3377.
Livingstone, M.S., & Hubel, D.H. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science, 240*(4853), 740-749.
Luce, R.D. (1991). *Response times: Their role in inferring elementary mental organization*. Oxford, England: Oxford University Press, USA.
Maris, G., & Maris, E. (2003). Testing the race model inequality: A nonparametric approach. *Journal of Mathematical Psychology, 47*(5), 507-514.
Marzi, C.A., Smania, N., Martini, M.C., Gambina, G., Tomelleri, G., Palamara, A., . . . Prior, M. (1996). Implicit redundant-targets effect in visual extinction. *Neuropsychologia, 34*(1), 9-22.
Matzke, D., & Wagenmakers, E.J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16*(5), 798-817.
McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*(4), 287-330.
Miller, J.O. (1978). Multidimensional same - different judgments: evidence against independent comparisons of dimensions. *Journal of experimental psychology: Human perception and performance, 4*(3), 411.

Miller, J.O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*(2), 247-279.
Miller, J.O. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics, 40*(5), 331-343.
Miller, J.O. (2004). Exaggerated redundancy gain in the split brain: A hemispheric coactivation account. *Cognitive Psychology, 49*(2), 118-154.
Miller, J.O. (2007). Contralateral and ipsilateral motor activation in visual simple reaction time: a test of the hemispheric coactivation model. *Experimental Brain Research, 176*(4), 539-558.
Miller, J.O., Beutinger, D., & Ulrich, R. (2009). Visuospatial attention and redundancy gain. *Psychological Research, 73*(2), 254-262.
Miller, J.O., & Lopes, A. (1991). Bias produced by fast guessing in distribution-based tests of race models. *Attention, Perception, & Psychophysics, 50*(6), 584-590.
Miller, J.O., & Reynolds, A. (2003). The locus of redundant-targets and nontargets effects: evidence from the psychological refractory period paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 29*(6), 1126.
Miller, J.O., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology, 46*(2), 101-151. doi: 10.1016/S0010-0285(02)00517-0
Mordkoff, J.T., & Miller, J.O. (1993). Redundancy gains and coactivation with two different targets: the problem of target preferences and the effects of display frequency. *Perception & psychophysics, 53*(5), 527-535.
Mordkoff, J.T., Miller, J.O., & Roch, A.C. (1996). Absence of coactivation in the motor component: evidence from psychophysiological measures of target detection. *Journal of Experimental Psychology: Human Perception and Performance, 22*(1), 25.
Mordkoff, J.T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17*(2), 520.
Murdock, B.B. (1971). Four-channel effects in short-term memory. *Psychonomic Science*.
Nelder, J.A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7*(4), 308-313.
Pachella, R.G. (1973). The Interpretation of Reaction Time in Information Processing Research: DTIC Document.
Palmer, E.M., Horowitz, T.S., Torralba, A., & Wolfe, J.M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 58.
Patching, G.R., & Quinlan, P.T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance, 28*(4), 755.
Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.
Pike, A.R. (1966). Stochastic models of choice behaviour: response probabilities and latencies of finite markov chain systems. *British Journal of Mathematical and Statistical Psychology, 19*(1), 15-32.
Pike, A.R. (1973). Response latency models for signal detection. *Psychological Review, 80*(1), 53-68.
Pollmann, S., & Zaidel, E. (1999). Redundancy gains for visual search after complete commissurotomy. *Neuropsychology, 13*(2), 246.
Posner, M.I. (2005). Timing the brain: Mental chronometry as a tool in neuroscience. *PLoS Biology, 3*(2), e51.
Purcell, B.A., Heitz, R.P., Cohen, J.Y., Schall, J.D., Logan, G.D., & Palmeri, T.J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review, 117*(4), 1113.
Raab, D.H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*.

Raftery, A.E. (1986). Choosing models for cross-classifications. *American Sociological Review, 51*(1), 145-146.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review, 85*(2), 59-108.
Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*(3), 446.
Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88*(6), 552.
Ratcliff, R. (1988). Continuous versus discrete information processing modeling accumulation of partial information. *Psychological Review, 95*(2), 238-255.
Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review, 9*(2), 278-291.
Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of Neurophysiology, 90*(3), 1392-1407.
Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological review, 111*(1), 159.
Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347-356.
Ratcliff, R., & Rouder, J.N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology Human Perception and Performance, 26*(1), 127-140.
Ratcliff, R., & Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*(2), 333.
Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging, 19*(2), 278.
Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics, 65*(4), 523-535.
Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*(3), 438-481.
Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261-300.
Reuter-Lorenz, P.A., Nozawa, G., Gazzaniga, M.S., & Hughes, H.C. (1995). Fate of neglected targets: a chronometric analysis of redundant target effects in the bisected brain. *Journal of Experimental Psychology: Human Perception and Performance, 21*(2), 211.
Rudin, W. (2006). *Real and complex analysis*: Tata McGraw-Hill Education.
Schmiedek, F., Oberauer, K., Wilhelm, O., Suss, H.M., & Wittmann, W.W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136*(3), 414.
Schouten, J.F., & Bekker, J.A.M. (1967). Reaction time and accuracy. *Acta Psychologica, 27*, 143-153.
Schröter, H., Frei, L.S., Ulrich, R., & Miller, J.O. (2009). The auditory redundant signals effect: An influence of number of stimuli or number of percepts? *Attention, Perception, & Psychophysics, 71*(6), 1375-1384.
Schröter, H., Ulrich, R., & Miller, J.O. (2007). Effects of redundant auditory stimuli on reaction time. *Psychonomic Bulletin & Review, 14*(1), 39-44.
Schwarz, W. (1989). A new model to explain the redundant-signals effect. *Attention, Perception, & Psychophysics, 46*(5), 498-500.
Schwarz, W. (1994). Diffusion, superposition, and the redundant-targets effect. *Journal of Mathematical Psychology*.
Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, 33*(4), 457-469.

Schwarz, W., & Ischebeck, A. (1994). Coactivation and statistical facilitation in the detection of lines. *Perception, 23*, 157-157.
Sincich, L.C., & Horton, J.C. (2005). The circuitry of V1 and V2: integration of color, form, and motion. *Annual Review of Neuroscience, 28*, 303-326.
Smith, P.L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27*(3), 161-168.
Smith, P.L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology, 32*(2), 135-168.
Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta psychologica, 30*, 276-315.
Stone, M. (1960). Models for choice-reaction time. *Psychometrika, 25*(3), 251-260.
Thompson, J.R. (2009). *Simulation: a modeler's approach* (Vol. 522): Wiley-Interscience.
Thornton, T.L., & Gilden, D.L. (2007). Parallel and serial processes in visual search. *Psychological Review, 114*(1), 71.
Todd, J.W. (1912). *Reaction to multiple stimuli*. The Science Press: The Science Press.
Töllner, T., Zehetleitner, M., Krummenacher, J., & Müller, H.J. (2011). Perceptual basis of redundancy gains in visual pop-out search. *Journal of Cognitive Neuroscience, 23*(1), 137-150.
Townsend, J.T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics, 10*(3), 161-163.
Townsend, J.T., & Ashby, F.G. (1983). *The stochastic modeling of elementary psychological processes*: Cambridge University Press.
Townsend, J.T., & Colonius, H. (1997). Parallel processing response times and experimental determination of the stopping rule. *Journal of Mathematical Psychology, 41*(4), 392-397.
Townsend, J.T., & Honey, C.J. (2007). Consequences of base time for redundant signals experiments. *Journal of mathematical psychology, 51*(4), 242-265.
Townsend, J.T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*(4), 321-359.
Townsend, J.T., & Nozawa, G. (1997). Serial exhaustive models can violate the race model inequality: Implications for architecture and capacity. *Psychological Review, 104*(3), 595.
Townsend, J.T., & Wenger, M.J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review, 11*(3), 391-418.
Ulrich, R., & Giray, M. (1986). Separate-activation models with variable base times: Testability and checking of cross-channel dependency. *Attention, Perception, & Psychophysics, 39*(4), 248-254.
Ulrich, R., & Miller, J.O. (1997). Tests of race models for reaction time in experiments with asynchronous redundant signals. *Journal of Mathematical Psychology, 41*(4), 367-381.
Ulrich, R., Miller, J.O., & Schröter, H. (2007). Testing the race model inequality: An algorithm and computer programs. *Behavior Research Methods, 39*(2), 291-302.
Ulrich, R., & Stapf, K.H. (1984). A double-response paradigm to study stimulus intensity effects upon the motor system in simple reaction time experiments. *Perception & Psychophysics, 36*(6), 545-558.
Usher, M., & McClelland, J.L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*(3), 550-592.
Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review, 14*(6), 1011-1026.
Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods, 40*(1), 61-72.
Vandekerckhove, J., Tuerlinckx, F., & Lee, M.D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods, 16*(1), 44.

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.
Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics, 13*(1), 37-58.
Vickers, D. (1978). An adaptive module of simple judgements. *Attention and performance, VII*, 598-618.
Vickers, D. (1979). *Decision processes in visual perception*: Academic Press New York.
Vickers, D., Caudrey, D., & Willson, R. (1971). Discriminating between the frequency of occurrence of two alternative events. *Acta Psychologica, 35*(2), 151-172.
Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32*(7), 1206-1220.
Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods, 39*(4), 767-775.
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.
Wagenmakers, E.J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192-196.
Wagenmakers, E.J., Grasman, R.P.P.P., & Molenaar, P. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology, 49*(3), 195-204.
Wagenmakers, E.J., Ratcliff, R., Gomez, P., & Iverson, G.J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology, 48*(1), 28-50.
Wagenmakers, E.J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58*(1), 140-159.
Wagenmakers, E.J., Van Der Maas, H.L.J., & Grasman, R.P.P.P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3-22.
Ward, R., & McClelland, J.L. (1989). Conjunctive search for one and two identical targets. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 664-672.
Wickelgren, W.A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica, 41*(1), 67-85.
Wolfe, J.M. (1994). Guided search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202-238.
Zehetleitner, M., Koch, A.I., Goschy, H., & Müller, H.J. (2013). Salience-based selection: Attentional capture by distractors less salient than the target. *PLoS ONE, 8*(1), e52595.
Zehetleitner, M., Krummenacher, J., & Müller, H.J. (2009a). Co-activation – neither serial exhaustive nor interactive–models can explain violations of the race model in visual pop-out search. *Perception, Psychophysics, 71*(8), 1739-1759.
Zehetleitner, M., Krummenacher, J., & Müller, H.J. (2009b). The detection of feature singletons defined in two dimensions is based on salience summation, rather than on serial exhaustive or interactive race architectures. *Attention, Perception, & Psychophysics, 71*(8), 1739-1759.
Zehetleitner, M., & Müller, H.J. (2010). Salience from the decision perspective: You know where it is before you know it is there. *Journal of Vision, 10*(14).
Zehetleitner, M., Müller, H.J., & Krummenacher, J. (2008a). The redundant-signals paradigm and preattentive visual processing. *Frontiers in Bioscience, 13*, 5279.
Zehetleitner, M., Müller, H.J., & Krummenacher, J. (2008b). What the redundant-signal paradigm reveals about preattentive visual processing. *Frontiers in Biology, 14*, 5279-5293.