# BEHAVIOURAL MEASURES OF INTERFERENCE AND FACILITATION IN AN AUDIOVISUAL COLOUR-WORD STROOP MATCHING TASK

by

IDO SHLOMO BORNSTEIN

B.A., The University of Haifa, Israel, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Audiology and Speech Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2015

## Abstract

Previous audiovisual Stroop studies used spoken colour-words mainly as ignored distractors when performing the visual Stroop task. Previous matching Stroop studies in the visual domain provided opposing views regarding whether interference effects resulted from conflicting semantic representations or conflicting responses. This study's main objective was to explore how a written word distractor affects audiovisual matching of a spoken colour-word and font colour. I presented colour-words written in congruent or incongruent font colours simultaneously with spoken colour-words. Participants pressed buttons to indicate whether the spoken word and font colour were "same" or "different", while ignoring written word meaning. I recorded response times and accuracy to measure interference and facilitation effects between experimental and control conditions. I hypothesised that incongruent written words (e.g., *red*) would interfere with "same" responses (e.g., font colour = blue, spoken word = /blue/) but facilitate "different" responses (e.g., font colour = green, spoken word = /blue/); and that congruent written words (e.g., *red*) would facilitate "same" responses (e.g., font colour = red, spoken word = /red/) but interfere with "different" responses (e.g., font colour = red, spoken word = /blue/, or font colour = blue, spoken word = /red/). The results showed large interference effects but no facilitation effects on audiovisual judgements. While incongruent written words interfered with "same" responses, congruent written words interfered with "different" responses. The largest interference effect occurred when the written word was incongruent with both task-relevant dimensions, while smaller effects occurred when the written word was congruent with either task-relevant dimension. Consistent with previous Stroop findings, my audiovisual matching task showed that in the case of cross-modal colour judgements, written word meaning predominantly interferes with but does not facilitate performance. The pattern of results showed

that a conflict between the outcome of the relevant matching task and the outcomes of two mistakenly performed matching tasks involving the written word produced interference effects, rather than a conflict among the semantic representations activated by the three stimulus dimensions.

## Preface

The idea for the research presented in this thesis was conceived by the author. The experimental design and paradigm, data collection and analysis, the interpretation of the results and writing of this thesis were all carried out by the author.

Dr. Anthony Herdman contributed to the design and programming of the experiment and to data analysis, and provided comments on previous versions of this thesis.

Parts of the study described in this thesis will be presented in a poster in the annual meeting of the Cognitive Neuroscience Society in San Francisco, March 28-31, 2015 under the title: “Audiovisual Colour-Word Stroop Matching Task: Interference but not Facilitation from Written Word Meaning”.

This study was reviewed and approved by the Behavioural Research Ethics Board of the University of British Columbia under the title: “Behavioural Measures and Electrophysiological Correlates of Interference and Facilitation in an Audiovisual Colour-Word Stroop Matching Task” (Certificate # H13-02396).

## Table of Contents

Abstract ..... ii
Preface ..... iv
Table of Contents ..... v
List of Tables ..... vii
List of Figures ..... viii
List of Abbreviations ..... ix
Acknowledgements ..... x
Dedication ..... xii
1 Introduction ..... 1
1.1 The Stroop Effect ..... 3
1.2 Source of Interference ..... 4
1.3 The Matching Stroop Task ..... 6
1.4 Audiovisual Versions of the Stroop Task ..... 9
1.5 My Thesis: Audiovisual Matching Stroop Task ..... 11
2 Methods ..... 16
2.1 Participants ..... 16
2.1.1 Handedness and Language Questionnaires ..... 16
2.2 Materials ..... 17
2.2.1 Visual Stimuli ..... 17
2.2.2 Auditory Stimuli ..... 18
2.2.3 Audiovisual Stimulus Combinations ..... 19
2.3 Procedure ..... 22
2.4 Data Analysis ..... 25
2.4.1 Accuracy ..... 28
2.4.2 Response Time ..... 32
3 Results ..... 35
3.1 Accuracy ..... 35
3.2 Response Time ..... 41
4 Discussion ..... 49

4.1 Interference from Conflicting Outcomes ....................................................................49
4.2 No Facilitation from Concurring Outcomes...............................................................54
4.3 Conflict Additivity.......................................................................................................58
4.4 Stimulus Set Membership ...........................................................................................60
4.5 Which Congruent Information Interferes More with a “Different” Response? ............61
4.6 “Same” versus “Different” Judgements.......................................................................65
4.7 Practice Effects.............................................................................................................67
5 Conclusions ......................................................................................................................68
References ............................................................................................................................70
Appendices ...........................................................................................................................79
Appendix A: List of Exclusion Criteria for Participation .......................................................79
Appendix B: Written Instructions Presented to Participants ..................................................80
Appendix C: ..........................................................................................................................81
C.1 Histograms of RTs by *Condition* ..............................................................................81
C.2 Histograms of inverse transformed RTs by *Condition*...............................................81
Appendix D: ..........................................................................................................................82
D.1 Comparison of regression models fitted to accuracy data with *Stimulus Combination* as fixed factor ..........................................................................82
D.2 Comparison of regression models fitted to accuracy data with *Condition* as fixed factor.........................................................................................................82
Appendix E: ..........................................................................................................................83
E.1 Comparison of regression models fitted to RT data with *Stimulus Combination* as fixed factor ..........................................................................83
E.2 Comparison of regression models fitted to RT data with *Condition* as fixed factor.......................................................................................................83

## List of Tables

Table 2.1 Study conditions and audiovisual stimulus combinations ................................21

Table 3.1 Number and proportion of correct responses by condition...............................35

Table 3.2 Means, medians and skewness of RT of correct responses by condition..........41

## List of Figures


Figure 2.1 Time course of experimental and practice trials ....................24

Figure 3.1 Model-predicted percent correct scores by condition....................37

Figure 3.2 Differences in model-predicted mean percent correct scores between experimental and control conditions....................39

Figure 3.3 Model-predicted mean RT by condition....................42

Figure 3.4 Differences in model-predicted mean RT between experimental and control conditions....................44

Figure 3.5 Model-predicted mean RT by trial-block ....................47

Figure 3.6 Trial-block contrasts....................48

## List of Abbreviations

| | |
|---|---|
| AC | All Congruent |
| AI | All Incongruent |
| CC | Colour Congruent |
| EEG | Electroencephalography |
| NCC | Neutral Colour Congruent |
| NCI | Neutral Colour Incongruent |
| NWC | Nonword Congruent |
| NWI | Nonword Incongruent |
| RT | Response Time |
| VC | Visual Congruent |
| WC | Word Congruent |

## Acknowledgements

The last two years have probably been the most laborious and least gratifying of my life. Thousands of hours of thinking, rethinking, speculating, calculating, formulating, formatting, programming, entering data, sketching, plotting, adjusting, modifying, writing, editing, and finalising have culminated to three short words, succinctly expressed in one phrase: “submit as is”. Spoken from the mouth of the person who guided me throughout the whole process, those words were all I could ever hope to hear after just finishing my thesis Defence. Those words made everything worthwhile, the countless hours of labour and the lack of gratification forgotten.

This thesis started with an idea conceived a long time ago. I had carried it with me across an ocean and presented it one morning to Dr. Tony Herdman, who seemed compelled enough and agreed to supervise me throughout the process. For that I am thankful to him, for that and for all the hours he put into helping me develop the idea and turn it into a research project. I also thank him for reading countless earlier versions of this thesis and for providing invaluable comments that helped me focus my writing and make it more concise.

I would like to thank my supervisory committee members, Dr. Mario Liotti and Dr. Lawrence Ward, for reading and commenting on my thesis and for their suggestions for future research. I am thankful to them for asking stimulating questions that turned my Defence into a fun and memorable experience.

I would also like to thank past and present members of the BRANE Lab, and in particular to Osamu Takai, for being there week-in week-out in lab meetings, listening to my research challenges and helping me overcome obstacles.

I would like to thank Rick White from the Statistical Consulting and Research Laboratory (SCARL) at UBC for his help with designing and formulating the statistical analyses.

Last, but by no means the least, I would like to thank my family who supported me throughout this long journey. In particular, none of this would have materialised without my best friend for life who agreed to set everything aside and join me on this journey, and without our beloved son who, since coming to our world, became the primary reason for our existence. Although I can no longer say it in person, I would have liked to thank my parents for always making me feel loved.

*To those without whom this would have not been accomplished*

# 1 Introduction

When literate individuals are asked to read a colour-word (e.g., say “red” for *red* in blue font) they can easily ignore its font colour when it is incongruent with the word’s meaning. However, when the task is changed to naming the word’s font colour (e.g., say “red” for *blue* in red font), readers have considerable difficulty in ignoring the word’s meaning when it is incongruent with the colour’s name (Stroop, 1935; for a review, see MacLeod, 1991).

Traditionally, this asymmetry in the ability to inhibit the irrelevant information (i.e., word meaning vs. font colour) was attributed to the relative speeds of (e.g., Morton, 1969; Morton & Chambers, 1973; Posner & Snyder, 1975), and degree of practice in (e.g., MacLeod & Dunbar, 1988), word reading versus colour naming. That is, because word reading is a faster and more practiced process than colour naming, a vocal reading response becomes available before a colour-name response. Therefore, words disrupt colour naming but colour names do not disrupt word reading. These views proposed that interference occurs at an output stage of processing, after stimulus evaluation has been completed. Other views proposed that interference occurs at an input or intermediate stage of processing. In these views, the word is involuntarily processed to a stage that its meaning interferes with the perceptual (Hock & Egeth, 1970), semantic (Luo, 1999), or conceptual (Seymour, 1977; Simon & Berbaum, 1988) representation of font colour, prior to response selection.

When readers attended to and matched one dimension of the colour-word while ignoring the other (e.g., the font colour or word meaning of *red* in blue font), to a reference, word-only (e.g., *blue* in black font), or colour-only (e.g., *xxxx* in red font) stimulus, task-irrelevant font colour and word meanings were shown to equally interfere with matching (e.g., Dyer, 1973;

Treisman & Fearnley, 1969). This finding challenged the speed-of-processing view because the supposedly slower dimension (i.e., font colour) interfered with the faster dimension (i.e., word meaning). Evidently, interference occurred in cross-dimensional comparisons of word meaning and font colour (i.e., colour-to-word and word-to-colour) regardless of the nature of the distractor. Consequently, later models of Stroop and Stroop-like tasks attributed processing delays to an additional translation process, preceding cross-dimensional comparisons. According to these models, interference occurs when the translated dimension encounters conflicting information in the dimension to which it is translated (Glaser & Glaser, 1989; Sugg & MacDonald, 1994; Virzi & Egeth, 1985). Two conflicting views proposed that interference in the matching Stroop task results from a semantic-conflict (Luo, 1999), versus a response-conflict (Goldfarb & Henik, 2006).

Spoken colour-words were mostly incorporated into the visual colour naming task as distractors (e.g., hear /green/ and see *blue* in red font, say “red”). Although the first such report showed larger interference from the combined audiovisual distraction than from visual or auditory distraction alone (Cowan & Barron, 1987), subsequent studies have not replicated this additive audiovisual effect (Appelbaum, Donohue, Park, & Woldorff, 2013; Elliott et al., 2014; Miles, Madden, & Jones, 1989). These findings may indicate that auditory distractors can be ignored in tasks that direct attention to a visual dimension (i.e., font colour), and that interference is solely driven by conflicting information present within the visual stimulus. However, none of the audiovisual Stroop studies asked participants to perform a truly audiovisual judgement of whether the meanings of auditory (e.g., /red/) and visual (e.g., *blue* in red font) stimuli were congruent or incongruent, while ignoring the written word meaning. Therefore, the possibility of evaluating the effects of written distraction on simultaneously attending to both modalities could not be assessed. The main objective of this thesis was to explore these effects.

## 1.1 The Stroop Effect

For almost eighty years, the Stroop task has been one of the most popular experiments in cognitive psychology, mostly employed as a measure of attention (MacLeod, 1992). In the task's original version (Stroop, 1935), participants viewed a list of colour-words written in incongruent font colours and either read the word or named its font colour, while ignoring the other dimension. Stroop showed that reading colour-words was equally fast whether the words were printed in incongruent font colours (e.g., say "red" for *red* in blue font) or in a black font (e.g., say "red" for *red* in black font). Conversely, naming font colours was significantly slowed down by an incongruent word meaning (e.g., say "red" for *blue* in red font) as compared to a meaningless control stimulus (e.g., say "red" for *xxxx* in red font). The interference from word meaning to colour naming is commonly referred to as the **Stroop effect**. The finding that reading colour-words is not slowed down by incongruent font colours is commonly referred to as a lack of a **reverse Stroop effect**.

For its robustness and replicability, over the years since Stroop's (1935) original paper research has focussed on the Stroop effect; i.e., on the interference from word meaning to colour naming. Later studies, introducing single-item versions of the task (as opposed to item-list versions), showed that congruent word meaning could also facilitate font colour naming (e.g., say "red" for *red* in red font) (e.g., Hintzman et al., 1972). However, the facilitation effect has been reported less frequently, and its magnitude has always been considerably smaller, than the interference effect (for a review see MacLeod, 1991).

## 1.2 Source of Interference

Two main hypotheses were proposed as the source of interference to cognitive processing in the Stroop task, namely **stimulus-conflict** and **response-conflict**. Stimulus-conflict accounts pointed to an input or intermediate stage of processing as the source of interference (e.g., Hock & Egeth, 1970; Luo, 1999; Seymour, 1977; Simon & Berbaum, 1988). In these views, font colours and written words are internally represented by separate systems. When font colour and word meaning are semantically incongruent, two different representations are activated and interference effects reflect the additional time required to resolve the semantic conflict, before executing the appropriate response. However, why do words interfere with font colour naming, but font colours do not interfere with word reading? According to Roelofs (2005), words access semantics via a lexical code whereas colours can directly access semantics. In a task that requires a vocal response (and thus the generation of a lexical code prior to motor output), word reading can proceed without semantic access, and thus font colours do not interfere. However, naming font colours encounters interference from the lexical code activated by the written word. Stimulus-conflict accounts also gain support from experiments that show that the degree of interference in the Stroop task depends on the degree of semantic relatedness between the colour to be named and the irrelevant written word (e.g., Klein, 1964).

Response-conflict accounts proposed that the two processes of word reading and font colour naming compete for a single response channel (e.g., Morton, 1969; Morton & Chambers, 1973; Posner & Snyder, 1975). In this view, interference is unidirectional (i.e., from irrelevant words to font colour naming but not vice versa) because the faster process of word reading reaches a response stage before the slower process of colour naming. A closely related explanation was provided by the concept of automaticity (e.g., MacLeod & Dunbar, 1988;

Posner & Snyder, 1975). In this view, interference occurs because, over years of experience, the process of reading has become automatic due to a greater degree of practice than the process of colour naming. Ultimately, both the relative speed-of-processing account and the automaticity account propose that Stroop interference and facilitation occur at an output stage of processing, after stimulus evaluation has been completed.

Stimulus- and response-conflict hypotheses viewed the processing involved in the Stroop task as sequential, implying that each processing stage had to be completed before the next could begin. More recent models highlighted the parallel nature of processing leading to different outcomes in the Stroop task (e.g., Cohen, Dunbar, & McClelland, 1990; Logan, 1980). These models maintain that interference occurs due to the intersection of processes that run in parallel and make use of similar cognitive resources. This intersection is not limited to a single or specific processing stage and thus can occur anywhere along a processing pathway after sensation. In these views, the concept of automaticity operates on a continuum such that learning and practice make a processing pathway stronger. Attention in these models modulates processing in a specific pathway in a way that is favourable to the performed task.

The issue of stimulus-response compatibility between a written word stimulus and reading in a vocal response paradigm has been suggested as an inherent limitation of the Stroop task (e.g., Treisman & Fearnley, 1969). This limitation stems from the fact that a response in the form of a spoken word is required in both colour naming and word reading tasks. Reading is not interfered by an irrelevant font colour because a word stimulus is compatible with a reading response, whereas a colour stimulus is incompatible with a reading response. This limitation was part of the motivation in designing matching versions of the Stroop task.

## 1.3 The Matching Stroop Task

In the matching Stroop task, a reference stimulus containing either physical (e.g., ███ or *xxxx* in red font) or lexical (e.g., *red* in black font) colour information, was paired with another stimulus, containing both physical and lexical colour information (e.g., *blue* in red font), of which only one was task-relevant (e.g., Caldas et al., 2012; Dyer, 1973; Goldfarb & Henik, 2006; Luo, 1999; Machado-Pinheiro et al., 2010; Mascolo & Hirtle, 1990; Simon & Berbaum, 1988; Tecce & Happ, 1964; Treisman & Fearnley, 1969; Zysset, Müller, Lohmann, & von Cramon, 2001). By varying the task-relevant dimension in separate experiments, both the effects of irrelevant word meaning and the effects of irrelevant font colour on matching could be evaluated. Word-to-word and colour-to-colour matches were regarded as within-dimensional matching, and word-to-colour and colour-to-word as cross-dimensional matching.

These studies showed that interference occurred in cross-dimensional matching regardless of the nature of the task-irrelevant dimension (i.e., font colour or word meaning). That is, colour-to-colour and word-to-word matching was faster than colour-to-word and word-to-colour matching. Both cross-dimensional matching directions were equally slow, indicating that a faster or more automatic reading response cannot account for processing delays in both match directions. This finding challenged response-competition accounts of the Stroop task because the supposedly slower and less automatic font colour dimension and the faster and more automatic word meaning dimension interfered with matching to the same degree.

Consequently, Virzi and Egeth (1985) proposed a translational model of the Stroop task. In this model, one stimulus dimension must be translated to another dimension when the form of responding is different from the form of the relevant dimension; i.e., when colour-words (irrelevant font colour) are matched with colour patches (e.g., match the meaning of *blue* in red font with the colour of ■ in blue), and when font colours (irrelevant word meaning) are matched with colour-words (e.g., match the colour of *blue* in red font with the meaning of *red* in black font) (see also Glaser & Glaser, 1989; Sugg & MacDonald, 1994). Interference occurs when the translated dimension encounters conflicting information in the dimension to which it was translated. Thus, matching the meaning of a colour-word (e.g., *blue* in red font) to a reference colour-only stimulus (e.g., ■ in blue) is interfered by the word's font colour (e.g., red), because translation of word meaning into colour is interfered by the word's own font colour. Equally, matching the font colour of a colour-word (e.g., *blue* in red font) to a reference word-only stimulus (e.g., *red* in black font) is interfered by the word's meaning (e.g., *blue*), because translation of font colour into meaning is interfered by the word's own meaning. Dyer (1973) also suggested that a translation stage is required prior to cross-dimensional matching. He, however, used a verbal "same" and "different" responding scheme in two cross-dimensional matching tasks (i.e., colour-to-word and word-to-colour). Thus, regardless of the task used (i.e., congruence judgement vs. matching to a colour patch or colour-word), the findings indicated that a translational stage preceded cross-dimensional matching.

Luo (1999) and Goldfarb and Henik (2006) used the matching Stroop task with a "same" and "different" button-press response to test if interference in this task was due to a semantic- or response-conflict. In both studies, colour-words written in varying font colours (e.g., *blue* in red font) were paired with colour patches (e.g., ■ in red). Participants compared word meaning to patch colour, while ignoring the word's own font colour (meaning decision). Luo also had participants compare the word's font colour with patch colour, while ignoring word meaning (visual decision). He found that irrelevant font colours disrupted meaning decisions; however, irrelevant word meanings did not affect visual decisions. He also found that "different" responses in the meaning decision task were slowed down more when the font colour and patch colour were incongruent (e.g., *blue* in red font paired with [redacted] in green) than when they were the congruent (e.g., *blue* in red font paired with [redacted] in red). He proposed that these findings supported the semantic-conflict hypothesis.

Goldfarb and Henik (2006) proposed a new approach to analyse the results of the meaning decision task. They pointed out that Luo's (1999) analysis only distinguished between "different" conditions wherein word meaning matched the word's own font colour (e.g., *red* in red font paired with [redacted] in blue) and those wherein word meaning did not match the word's own font colour. However, the latter condition actually consisted of two distinct conditions. In one, the word's font colour and patch colour matched (e.g., *blue* in red font paired with [redacted] in red) and in the other they mismatched (e.g., *blue* in red font paired with [redacted] in green). Incorporating this additional factor into the analysis revealed that "different" responses were slowed down the most by congruence between word meaning and its own font colour (e.g., *red* in red font paired with [redacted] in blue). However, delays were similar when font colour and patch colour matched (e.g., *blue* in red font paired with [redacted] in red) and when they mismatched (e.g., *blue* in red font paired with [redacted] in green). Based on these findings, Goldfarb and Henik proposed that participants erroneously matched the word meaning and its own font colour. That is, responding "same" (e.g., as in *blue* in red font paired with [redacted] in red) was interfered by mistakenly matching the word meaning to its own font colour (which elicited a "different" response). Responding "different" (e.g., as in *red* in red font paired with [redacted] in blue) was interfered by mistakenly matching the word meaning to its own font colour (which elicited a "same" response). They proposed that their findings supported a response-conflict hypothesis.

Important to the context of my thesis, matching versions of the Stroop task were only studied in the visual domain. Thus, the effects of distraction from an irrelevant dimension on matching performance could only be evaluated within the visual modality.

## 1.4 Audiovisual Versions of the Stroop Task

Most audiovisual versions of the Stroop task paired a spoken colour-word with a visual stimulus containing either physical (e.g., [redacted] in blue font) or lexical (e.g., *blue* in black font) colour information (e.g., Donohue, Appelbaum, Park, Roberts, & Woldorff, 2013; Elliott, Cowan, & Valle-Inclan, 1998; Hanauer & Brooks, 2003; Roelofs, 2005; Shimada, 1990). Thus, distraction was limited in these studies to information processed by separate modalities. These studies demonstrated crossmodal interference effects when the distracting dimension featured a colour-word (e.g., spoken or written). Thus, irrelevant spoken colour-words could not be ignored in a colour naming task or in a word reading task.

A few audiovisual studies explored whether spoken colour-words could act as auditory distractors during a visual Stroop task (e.g., Appelbaum et al. 2013; Cowan & Barron, 1987; Elliott et al., 2014; Miles, Madden, & Jones, 1989). In these studies, both written and spoken colour-words act as distractors to colour naming (e.g., hear /blue/ and see *blue* in red font, say "red"). Cowan and Barron (1987) showed that interference effects from the combined audiovisual distraction were larger than the interference from spoken or written distractors alone. Thus, they proposed that spoken word meaning could not be ignored when performing the visual Stroop task. However, Miles et al. (1989) could not replicate Cowan and Barron's findings and therefore questioned the actuality of a crossmodal Stroop effect. They theorised that in the Stroop task, colour names need not be stored in a short-term memory buffer and therefore should not be susceptible to interference from competing spoken words.

Recently, Elliot et al. (2014) tried to replicate Cowan and Barron's findings in a more carefully controlled study. They found that presentation of congruent spoken and written colour-word distractors that were both incongruent with font colour (e.g., hear /blue/ and see *blue* in red font, say "red") did not produce additional interference as compared to the interference from the written word alone. This finding may indicate that auditory distraction could be entirely ignored and that only visual distraction contributed to the interference effect. However, this interpretation might be limited by the fact that in all of their stimulus combinations the spoken and written colour words were semantically congruent. Thus, the potential interference from semantic incongruence between the two distractors (e.g., hear /blue/ and see *green* in red font, say "red") could not be evaluated. The findings of Appelbaum et al. (2013) also indicated that spoken colour-word distractors did not interfere with attending to font colour, when the written word meaning was congruent with font colour (e.g., hear /blue/ and see *red* in red font, say "red"). Thus, when the source of distraction was solely from audition, attention to font colour was not disrupted, although the written colour-word might have counteracted any auditory interference by enhancing the perception of the font colour.

All of the audiovisual Stroop studies described here have used some type of an identification task, wherein participants reported the identity of one stimulus dimension while other dimension/s were to be ignored. None of these studies asked participants to make a truly audiovisual congruence judgement based on the meaning of auditory and visual targets. Thus, the possibility of evaluating the effects on performance of simultaneously attending to both modalities could not be assessed. In light of this limitation, the main objective of my thesis was to explore the effects of written semantic distraction on audiovisual attention. To fulfil this objective, I designed an experiment that investigated the interference from written word meaning (e.g., *blue* in red font) on attending to and matching spoken word meaning (e.g., /red/) and font colour.

## 1.5 My Thesis: Audiovisual Matching Stroop Task

Colours can be presented auditorily only in their word form (except for individuals who experience synesthesia, e.g., Paulesu et al., 1995). Because visual colours and spoken colour-words are presented in different codes (i.e., physical vs. lexical), they cannot be compared based on physical (i.e., colour) or lexical (i.e., word) criteria. Therefore, I expect matching a font colour with spoken word meaning to impose a comparison of the internal representations of these stimuli. This comparison should be influenced by the meaning of the written word through the activation of an additional colour representation.

In the present task, font colours and spoken words were the task-relevant dimensions and written words were the task-irrelevant dimension. By varying the congruence among these three dimensions, five possible combinations can be formulated, two of which require a "same" response and three a "different" response:

1. Font colour and spoken word are congruent (i.e., "same") and the written word is congruent with both (e.g., hear /red/ and see *red* in red font).
2. Font colour and spoken word are congruent (i.e., "same") and the written word is incongruent with both (e.g., hear /red/ and see *blue* in red font).
3. Font colour and spoken word are incongruent (i.e., "different") and the written word is incongruent with both (e.g., hear /red/ and see *blue* in green font).
4. Font colour and spoken word are incongruent (i.e., "different") and the written word is congruent with the font colour but incongruent with the spoken word (e.g., hear /red/ and see *blue* in blue font).

5. Font colour and spoken word are incongruent (i.e., "different") and the written word is congruent with the spoken word but incongruent with the font colour (e.g., hear /red/ and see *red* in blue font).

Based on the findings of Luo (1999) and Goldfarb and Henik (2006) from visual matching Stroop tasks, semantic- and response-conflict accounts predict similar outcomes for "same" responses. Namely, a written colour-word interferes with matching more by being incongruent with both task-relevant dimensions (e.g., hear /red/ and see *blue* in red font) than by being congruent with both task-relevant dimensions (e.g., hear /red/ and see *red* in red font).

However, these accounts differ in their predictions for "different" responses. According to the semantic-conflict account, a written word should interfere with matching more by being incongruent with both task-relevant dimensions (e.g., hear /red/ and see *blue* in green font) than by being incongruent with either the font colour (e.g., hear /red/ and see *blue* in blue font) or the spoken word (e.g., hear /red/ and see *red* in blue font). This is because of the greater semantic-conflict exhibited when all dimensions are in conflict, as opposed to when two of them concur. According to the response-conflict account, the matching task entails three comparisons: a task-relevant comparison (i.e., spoken word and font colour) and two, mistakenly performed, task-irrelevant comparisons involving the written word (i.e., written word and spoken word, and written word and font colour). Interference is expected to occur when the outputs of these comparisons are in conflict (task-relevant comparison "different" ≠ task irrelevant comparison "same"; e.g., hear /red/ and see *blue* in blue font, or hear /red/ and see *red* in blue font) and no interference is expected when their outputs are in agreement (task-relevant comparison "different" = task-irrelevant comparison "different"; e.g., hear /red/ and see *blue* in green font).

Luo (1999) and Goldfarb and Henik (2006) did not include control stimuli in their visual matching tasks and therefore could not separate interference and facilitation effects. To quantify these effects in the present experiment, I included a condition featuring non-interfering nonword written stimuli (i.e., a series of #'s) as a point of reference. To separate congruency effects from lexicality effects (Brown, 2011), and to assess the effects of semantic relatedness (e.g., Klein, 1964), I included neutral written colour-word distractors (i.e., *white*) for which there was no corresponding font colour or spoken word in the task.

My hypotheses for the present study are more in line with the response-conflict view. I hypothesised that an outcome-conflict (Navon, 1985) between three separate comparisons (relevant: spoken word – font colour; irrelevant: written word – spoken word, written word – font colour) would modulate performance. That is, when the outcomes of all comparisons are in agreement (i.e., all are "same" or all are "different") performance should improve, and when these outcomes are in conflict (i.e., relevant "same" and irrelevant "different"; or relevant "different" and irrelevant "same") performance should degrade.

Specifically, I hypothesised that:

1. Written colour-words would facilitate "same" responses by being congruent with both task-relevant dimensions (e.g., hear /red/ and see *red* in red font).
2. Written colour words would interfere with "same" responses by being incongruent with both task-relevant dimensions (e.g., hear /red/ and see *blue* in red font).
3. Written colour-words would facilitate "different" responses by being incongruent with both task-relevant dimensions (e.g., hear /red/ and see *blue* in green font).

4. Written colour words would interfere with “different” responses by being congruent with either task-relevant dimension (e.g., hear /red/ and see *blue* in blue font, or hear /red/ and see *red* in blue font).

Additionally, I expected interference effects to be proportional to the number of outcomes of the irrelevant comparisons that conflicted with the outcome of the relevant comparison. Relevant comparisons indicating “same” would encounter interference from two irrelevant comparisons indicating “different” (e.g., hear /red/ and see *blue* in red font). Relevant comparisons indicating “different” would encounter interference from one irrelevant comparison indicating “same” (e.g., hear /red/ and see *blue* in blue font, or hear /red/ and see *red* in blue font). I expected the greater number of irrelevant outcomes conflicting with the relevant outcome to produce larger interference effects.

Furthermore, Klein (1964) showed that the more semantically related a written word was to the font colour, interference effects were larger. Therefore, I hypothesised that those written words that were also represented in the task-relevant stimulus set (e.g., red, green, and blue) would interfere with matching more than a neutral written colour-word (e.g., *white*).

Lastly, Roelofs (2005) showed that spoken and written colour-words interfered with attending to one another while patch colours did not interfere with attending to spoken colour-words. Also, the classic findings from the Stroop task show that font colours do not interfere with attending to written colour-words (e.g., Stroop, 1935). Therefore, I hypothesised that the congruence between written and spoken words (e.g., hear /red/ and see *red* in blue font) would interfere more with a “different” response than the congruence between written words and font colours (e.g., hear /red/ and see *blue* in blue font).

In this study, I asked participants to respond "same" or "different" based on the congruence between font colour and spoken word meaning, while ignoring written word meaning. My aim was to evaluate the effects of the written word meaning on audiovisual attention. By including nonword and neutral colour-word written stimuli as reference points, I was able to measure interference and facilitation effects separately, and the effect of the type of response ("same" vs. "different") with the same irrelevant written stimulus (e.g., *white*).

# 2 Methods

## 2.1 Participants

Thirty young adults volunteered to participate in this study. Two participants did not complete all experimental procedures and their data were excluded from further analyses. The mean age of the 28 participants (six male) included in data analysis was 24 years (*SD* = 3, Range 20 – 31). The study was reviewed by the Behavioural Research Ethics Board at the University of British Columbia and all participants signed their consent after reading a detailed description of the purpose of the study and experimental procedures. Participants received an honorarium to compensate for their time.

### 2.1.1 Handedness and Language Questionnaires

To assess hand dominance, an on-line handedness questionnaire (Cohen, 2008) adapted from the Edinburgh Handedness Inventory (Oldfield, 1971) was administered. A positive Laterality Index was taken to indicate a participant was right-handed and a negative Laterality Index was taken to indicate a participant was left-handed. Twenty five of the 28 participants (89%) were right-handed.

Languages status was assessed using the Canadian-English version of The Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld, & Kaushanskaya, 2007). Nineteen of the 28 participants (68%) reported being proficient in more than one language. All participants acquired English as their first language or one of their first languages.

### 2.1.2 Vision and Hearing Screening

Participants' visual acuity was screened using a Snellen chart. Participants named the letters in the line corresponding to 20/20 vision while standing at a distance of twenty feet from the chart. Colour vision was screened by asking participants to name the font colour of one pseudoword written in each one of the three font colours used in the experiment. Pseudowords were presented at the centre of a computer screen positioned 60-cm from participants' eyes and had the same font size as the visual stimuli used for the experiment.

Hearing screening was conducted in a sound attenuated audiometric booth. Hearing was screened in each ear at 20 dB Hearing Level (HL) at 250, 500, 1000, 2000, and 4000-Hz using an Interacoustics ® AC 40 clinical audiometer. Pure-tones were delivered through E-A-RTONE ® 3A (Auditory Systems, Indianapolis, IN) insert earphones. Participants raised a hand to respond. All participants included in data analyses passed all screening procedures successfully.

## 2.2 Materials

### 2.2.1 Visual Stimuli

Visual stimuli consisted of the written colour-words *red*, *green*, *blue* and *white*, and a string of four number signs (####) typed in red, green or blue lowercase Arial Bold font. Font colours were defined according to the RGB colour model with a digital 8 bit per channel notation, with red being 255, 0, 0, green being 0, 128, 0, and blue being 0, 0, 255. The green font colour was adjusted to have equal brightness as the other two font colours and equally legible against a black background, as judged by four observers. These specific font colours (red, green and blue) were chosen for the ease of digitally constructing them and for being easily identifiable by observers. The neutral colour-word (*white*) represented a control stimulus that had no corresponding font colour or spoken colour-word in this task. As such, it was chosen to create interference at the semantic category level, assuming it will require access to colour knowledge, but not to the specific colours relevant in this task (red, green and blue).

The word *white* was preferred over other colour-words for being monosyllabic and for not sharing phonological features such as initial or final phoneme with any of the other colour-words. The nonword control stimulus (*####*) was preferred over an English script control stimulus (e.g., *XXXX*) because it depicts a non-alphanumeric stimulus that is not likely to induce orthographic or phonological processing. It was also preferred to a patch of colour to allow for a certain degree of visual scanning in the perception of a string of characters.

Visual stimuli were presented for 500-ms at the centre of a 19" LCD monitor (DELL ®, 1908FPc) against a black background. Visual stimuli (words and nonwords) subtended 0.57 – 0.76º of visual angle vertically (based on 60-cm distance from the display and 6 – 8-mm letter height) and 1.34 – 2.58º of visual angle horizontally (1.4 – 2.7-cm word width).

### 2.2.2 Auditory Stimuli

Auditory stimuli consisted of the colour-words (/red/, /green/, and /blue/) spoken by a native English-speaking male audiologist with an average vocal fundamental frequency for the three words of 85-Hz. Spoken colour-words were recorded in an acoustically insulated audiometric booth using an electret condenser microphone (Sony ® ECM-MS907) and saved onto a single channel Pulse-Code Modulation (PCM) wave file using a 44.1-kHz sampling rate and 16-bits-per-sample resolution. Recorded words were temporally constrained using a time stretch function (Adobe ® Audition ® V. 3.0.1) to 500-ms total duration, and their peak intensities were normalised to a common reference level.

Spoken colour-words were delivered binaurally at a level of 70 dB HL through E-A-RTONE ® 3A (Auditory Systems, Indianapolis, IN) insert earphones using an Interacoustics ® AC-40 clinical audiometer. The peak sound pressure level (SPL) for each one of the spoken colour-words for each participant was measured in a 2-cc coupler using a QUEST ®

TECHNOLOGIES SoundPro ® SE Type I sound level metre immediately after the experiment was completed, using a fast time response and a flat (Z) frequency weighting. Mean (*SD*) peak SPLs were 72.3 (0.3), 70.8 (0.5), and 69.5 (0.5) dB Z, for /red/, /green/, and /blue/, respectively.

### 2.2.3 Audiovisual Stimulus Combinations

A total of 45 audiovisual stimulus combinations were available for presentation from the five visual stimuli (**R**ed, **G**reen, **B**lue, **W**hite, and **N**onword), three font colours (**R**ed, **G**reen, and **B**lue) and three auditory stimuli (**R**ed, **G**reen, and **B**lue). For convenience, audiovisual stimulus combinations will be notated by a three letter string indicating the spoken word first, the written stimulus second and the font colour third. For example, the combination of the spoken word /red/ with the written stimulus *white* in green font will be notated RWG.

To present each stimulus combination in equal probability in each block of trials, certain combinations had to be presented twice as often as others. This was because the number of combinations that could be generated in each condition was limited by the congruence among the three stimulus dimensions. For example, when all the dimensions matched, only three combinations could be generated (e.g., RRR, GGG, and BBB); however, when all the dimensions mismatched, six stimulus combinations could be generated (e.g., RGB, RBG, GRB, GBR, BRG, and BGR). Consequently, a total of 54 stimulus combinations were available for presentation. These combinations were grouped into nine conditions, varying by the congruence or incongruence among spoken word meaning, written word meaning, and font colour. In four conditions, font colour and spoken colour-word were congruent and thus required a “same” button response. These conditions were: (1) *Nonword Congruent*, (2) *Neutral Colour Congruent*, (3) *All Congruent*, and (4) *Colour Congruent*. In the other five conditions, font colour and spoken colour-word were incongruent and thus required a “different” button response. These conditions were: (5) *Nonword Incongruent*, (6) *Neutral Colour Incongruent*, (7) *All Incongruent*, (8) *Visual Congruent*, and (9) *Word Congruent*. The study conditions, inter-stimulus congruency, the required response, and all possible stimulus combinations are presented in Table 2.1.

**Table 2.1 Study conditions and audiovisual stimulus combinations**

| Response | Comparison | | | Condition Name (abbr.) | Stimulus Combination | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Relevant** Spoken word + Font colour | **Irrelevant** Spoken word + Written word | **Irrelevant** Font colour + Written word | | A V | A V | A V | A V | A V | A V |
| Same | Congruent | NA | NA | Nonword Congruent (NWC) | R #### | G #### | B #### | | X2 | |
| | Congruent | Incongruent* | Incongruent* | Neutral Colour Congruent (NCC) | R white | G white | B white | | X2 | |
| | Congruent | Congruent | Congruent | All Congruent (AC) | R red | G green | B Blue | | X2 | |
| | Congruent | Incongruent | Incongruent | Colour Congruent (CC) | R green | R blue | G Red | G blue | B red | B green |
| Different | Incongruent | NA | NA | Nonword Incongruent (NWI) | R #### | R #### | G #### | G #### | B #### | B #### |
| | Incongruent | Incongruent* | Incongruent* | Neutral Colour Incongruent (NCI) | R white | R white | G white | G white | B white | B white |
| | Incongruent | Incongruent | Incongruent | All Incongruent (AI) | R green | R blue | G Red | G blue | B red | B green |
| | Incongruent | Incongruent | Congruent | Visual Congruent (VC) | R green | R blue | G Red | G blue | B red | B green |
| | Incongruent | Congruent | Incongruent | Word Congruent (WC) | R red | R red | G green | G green | B blue | B blue |

Note. * represents the neutral colour-word *white* which had no corresponding spoken word or font colour. NA – Not applicable
A – Auditory stimulus, V – Visual stimulus, R – /red/, G – /green/, B – /blue/

## 2.3 Procedure

All testing was conducted in the BRANE lab facility at the University of British Columbia. Participants were seated in a comfortable chair in a dimly lit sound-attenuated booth, approximately 60-cm across from a 19" LCD monitor (DELL ®, 1908FPc). A properly fitted EEG head cap (Electro-Cap International, Eaton, OH) with leads was placed on a participant's head, and insert earphones were placed in their ear canals. A standard computer keyboard was placed on a table in front of the participant.

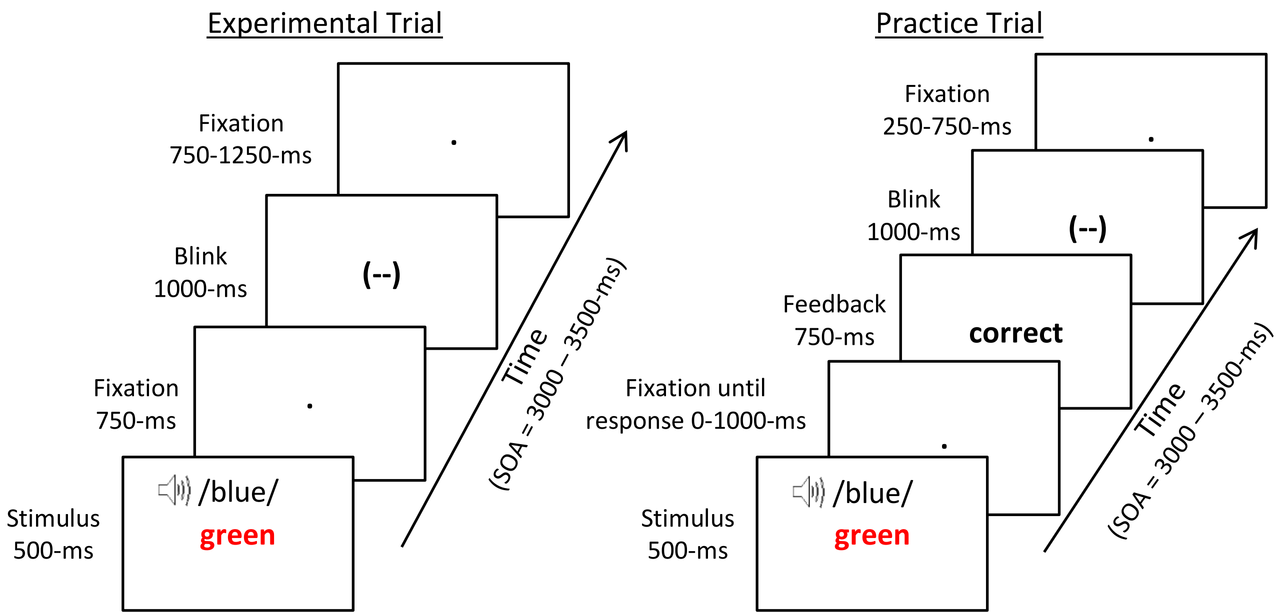
Participants responded by pressing designated keys on the keyboard. The specific response-to-key correspondence was randomly interchanged between participants. The right-arrow key was designated as "YES" ("same") and the left-arrow key as "NO" ("different") in 15 participants, and the assignment was reversed in 13 participants. This procedure aimed to prevent systematic bias in accuracy or RT between "same" and "different" responses related to key assignment. Participants were instructed to press a key as quickly and as accurately as possible, using the index and middle finger of their dominant hand. The "YES" key was to be pressed for trials in which the spoken colour-word and the font colour were congruent, and the "NO" key for trials in which the spoken colour-word and the font colour were incongruent. Participants were also instructed to ignore the meaning of the written words. The use of a matching task removed the need to include a response acquisition phase because participants did not need respond by a specific colour output, but to judge whether stimuli were the same or different. The complete written instructions are described in Appendix B.

Visual and auditory stimulus presentation was controlled using Neurobehavioral Systems Presentation ® software (Version 14.5, www.neurobs.com). Visual and auditory stimuli were presented with simultaneous onsets and offsets. Stimulus onset asynchronies (SOA) between trials were randomised between 3000 – 3500 milliseconds. A grey fixation point was presented at the centre of the screen throughout the inter-trial period to orient participants to the location of visual stimuli. A designated blinking period, marked by two adjacent hyphens placed within brackets (i.e., (--)) appearing for one second at centre screen, was inserted in each inter-trial period. Participants were encouraged to blink only within that designated time window. This was done to reduce the number of trials that might be rejected because they contained blink artefacts in EEG recordings. To minimise systematic influences of sequential effects (i.e., priming) on the results, trial order was randomised between trial blocks and between participants.

Each of the 54 available audiovisual stimulus combinations was first presented in a practice block. This block served to acquaint participants with the task and to practice response-key assignments. Stimulus presentation was the same as that described above with the exception of the insertion of a visual feedback after a participant's response. The written word *correct* or *incorrect* was shown for 750-ms immediately after a key was pressed. This feedback provided participants with verification of their performance and allowed them to modify subsequent responses accordingly. Participants' accuracy in the practice block was analysed immediately after its completion to allow for reinstruction. Performance of 80% or higher in the practice block (44 correct of 54 trials) was chosen as the minimum required to proceed to experimental blocks. Lower performance indicated reinstruction and re-administration of the practice block. Of the 28 participants, two did not meet the 80% criterion in the first run and the practice block was repeated. Mean accuracy across participants in the practice block was 93.1% ($SD$ = 4.7).

Experimental blocks consisted of 108 trials, yielding twelve presentations per condition. Feedback for correct or incorrect performance was not given in these blocks. Seven blocks were presented to each participant, for a total of 756 trials, 84 per condition. A one minute break was given between consecutive blocks, after which participants continued once they were ready. The time course of experimental- and practice-trials is presented in Figure 2.1.

**Figure 2.1 Time course of experimental and practice trials**




Note. SOA = Stimulus Onset Asynchrony.

## 2.4 Data Analysis

To test my hypotheses, I fitted separate mixed-effects regression models to accuracy and RT data. Four fixed factors and a single random factor were defined in these models. Each fixed factor was a categorical independent variable with a certain number of levels, corresponding to the number of categories present in that variable. For each fixed factor, I defined specific contrasts between this factor's levels in accordance with my hypotheses and with the nature of that variable. The factors I included in regression models were:

1. *Response Mapping* is a between-participant factor specifying whether the participant responded "YES" and "NO" with the right and left arrow keys, respectively, or vice versa. This factor thus had two levels and one contrast: YES-right arrow key versus YES-left arrow key. This factor was included to test if a systematic bias in performance was related to response-key assignment.
2. *Condition* is a within-participant factor defining the nine conditions that were used in the experiment. This was the main variable of interest because it was used to test differences in performance between conditions. This factor had nine levels, corresponding to the nine conditions: *Nonword Congruent*, *Neutral Colour Congruent*, *All Congruent*, *Colour Congruent*, *Nonword Incongruent*, *Neutral Colour Incongruent*, *All Incongruent*, *Visual Congruent*, and *Word Congruent*. I defined seven contrasts for this factor, such that each experimental condition was contrasted with the nonword control condition that required the same response (i.e., "same" or "different"). Contrasting conditions in this way incorporated the actual effects produced by the written word into the model, while accounting for the inherent RT and accuracy differences between "same" and "different" responses. Findings of matching tasks often demonstrate a *fast-"same"* effect and a *false-"different"* effect. That is,

participants typically respond “same” faster than they respond “different”, and also, but not as uniformly as the speed difference, they make more errors in “same” trials (incorrectly responding “different”) than in “different” trials (incorrectly responding “same”) (e.g., Farrel, 1985; Luce, 1986: 445; Proctor, 1981; Ratcliff, 1985, 1987). These effects and their origins have been the focus of considerable research, but are outside the scope of my study. However, for the present analysis, evaluating the effects of the irrelevant written word on correctly responding “same” or “different” must take into account the inherent differences between these two judgements. Therefore, I used the two nonword control conditions (*Nonword Congruent* and *Nonword Incongruent*) as separate reference points for measuring interference and facilitation effects from the written word on making each type of judgement. After accounting for the “same” - “different” disparity and isolating the effect of the irrelevant written word, I could compare the effect of the irrelevant written word between conditions that required a different response.

3. *Stimulus Combination* is a within-participant factor describing the stimulus combination presented in each trial in terms of spoken word (R/G/B), written stimulus (R/G/B/W/N), and font colour (R/G/B). This factor had 45 levels, corresponding to all possible audiovisual stimulus combinations across the nine conditions (i.e., RRR, GGG, BBB; RGR, RBR, GRG, GBG, BRB, BGB; RWR, GWG, BWB; RNR, GNG, BNB; RGB, RBG, GRB, GBR, BRG, BGR; RRG, RRB, GGR, GGB, BBR, BBG; BRR, BGG, GRR, GBB, BRR, BGG; RWG, RWB, GWR, GWB, BWR, BWG; RNG, RNB, GNR, GNB, BNR, BNG). I defined 36 contrasts for this factor, such that each stimulus combination that belonged to an experimental condition was contrasted with the stimulus combination that belonged to the nonword control condition that featured the same spoken word and font colour, and that

required the same response. Thus, contrasted pairs differed only by their written word component. Contrasting stimulus combinations in this way incorporated into the regression model the actual interference and facilitation effects produced by the written word, while accounting both for the required response, and for specific spoken colour-words and font colours. For example, the stimulus combination RNG was subtracted from RBG. Both required the same response ("different") and differed only by the written component being a colour-word in the experimental condition (*blue*) and a nonword (*####*) in the control condition. I included this factor to test if within each condition, certain written words showed significantly greater effects on performance than others, as compared to the respective nonword stimulus. I expected these tests to show that the effect of a specific written word (e.g., *blue*) within each condition was not significantly different than the effect of others (e.g., *red*), a finding that in turn would indicate that merging stimulus combinations into conditions is statistically justifiable.

4. *Block* is a within-participant factor describing the block number to which each trial belonged, and thus had seven levels. I defined six contrasts for it, comparing each block with the previous block. This contrasting intended to capture block-by-block changes in performance.
5. *Participant* is a random factor I included in all fitted models to capture the dependence inherent in repeated observations made on the same individual. The quantity estimated in fitted models for random factors is their variance. Controlling for the variance associated with the performance of a participant assures that when testing fixed factors, an effect that shows significance is indeed significant after the within-participant variability in performance was taken into account (Jaeger, 2008: 444).

To test if response-key assignments and trial-blocks affected performance differently in each condition, I included in all fitted regression models two two-way interactions, *Response Mapping* X *Condition* and a *Block* X *Condition*.

For each regression model that had the same composition of fixed factors, I fitted three alternative models, varying by the definition the random factor term: one included only random intercept, to capture the overall variability in performance for each participant; the second included random intercept and random slope for *Condition*, to capture the variability in overall performance and in condition-specific performance for each participant; and the third included only random slope for *Condition*, to capture the variability in condition-specific performance for each participant, without accounting for variability in overall performance. Omission of the intercept term is plausible if the variability associated with the factor of interest (*Condition*) is primarily accountable for the overall variability.

All Statistical analyses were performed using R Software for Statistical Computing (R Core Team, 2014) Version 3.1.1, "Sock it to Me".

### 2.4.1 Accuracy

Behavioural responses were monitored and recorded on-line while participants performed the task. Responses were considered as correct if participants pressed the button corresponding to the correct response no sooner than 200-ms after stimulus onset and no later than the presentation of the subsequent trial. Trials in which no key was pressed within the inter-trial period ($n$ = 12) were counted as incorrect responses.

In two-alternative forced choice tasks, as was used in the present experiment, accuracy is typically quantified by calculating a by-participant percent correct score for each condition, followed by averaging across participants to get the by-condition percent correct score. Using linear statistical models, such as the Analysis of Variance (ANOVA), over proportions has been shown to be untenable on several grounds. First, proportions take on values that are bounded by 0 and 1, whereas linear models can take on values from an infinite range. Second, ANOVAs assume normal distribution of the outcome variable, whereas binary outcome variables are binomially distributed. Third, proportions are inherently heteroscedastic (and thus violate the homogeneity of variance assumption of ANOVA) because the variance of a binomially distributed variable is mathematically related to its mean, such that the variance is maximal when the mean equals 0.5 and approaches 0 as the mean approaches 0 or 1 (e.g., Agresti, 2002: 120; Dixon, 2008; Jaeger, 2008).

Alternatively, binomially distributed outcome variables can be analysed using logistic regression. In logistic regression models, proportions are converted into a linear scale by a logarithmic transformation of the odds. The odds are the ratio of success (correct response) versus failure (incorrect response) ($odds = \frac{p}{1-p}$) and thus transform proportions into an infinite positive scale. Logits, the natural logarithm of the odds ($logit = \ln \frac{p}{1-p}$) transform odds into an infinite scale, which can then be described by a linear combination of independent variables.

To analyse accuracy data, I used a mixed-effects logistic regression model. This model is a type of a Generalized Linear Mixed Model (GLMM, Breslow & Clayton, 1993), a statistical model that describes an outcome variable as a linear combination of fixed and random factors. Logistic regression models are specifically designed for analysing binomially distributed outcome variables, while using a mixed-effects model allows capturing the within-participant performance dependence inherent in repeated measures experiments.

I fitted regression models to accuracy data using the function *glmer* in the R package *lme4* (Bates, Maechler, Bolker, & Walker, 2014). This function fits a GLMM using a logit link and estimates model parameters using Restricted Maximum Likelihood (REML), which is a goodness-of-fit measure describing the degree of deviance of the fitted model from the observed data. I used the function *anova* from the same R package to compare the deviance of regression models that had different definitions of the random factor. This function contrasts fitted models using an Analysis-of-Deviance chi-square test, by comparing the difference in deviance between models as a function of the difference between their degrees of freedom. The degrees of freedom are equivalent to the total number of parameters estimated by the model. A significant result indicates that the addition of terms into the model significantly reduces the deviance and hence provides a better fit. A nonsignificant result indicates that additional terms do not change the deviance and thus do not contribute to the fit.

After identifying the best fitting model, I tested fixed factors and factor interactions from that model for significance using the function *Anova* from the R package *car* (Fox & Weisberg, 2011). This function uses Type-II Wald chi-square Analysis-of-Variance tests to determine if factors and factor interactions have a significant effect on the outcome variable. Then, I omitted nonsignificant factors and interactions and refitted the model, incorporating significant factors only, to come up with the most parsimonious model to account for the observed data.

I performed pairwise comparisons between levels of significant factors using the function *glht* from the R package *multcomp* (Bretz, Hothorn, & Westfall, 2010; Hothorn, Bretz & Westfall, 2008). These comparisons use the parameters estimated by the most parsimonious regression model to test general linear hypotheses. Control for the family-wise error rate is implemented by this function using either Tukey's all-pairwise comparisons or the single-step method, which computes *p*-values based on the joint normal or *t* distribution of the linear function.

Before testing hypotheses regarding differences between conditions in accuracy, I tested if individual stimulus combinations within each experimental condition were comparable with respect to their effects on accuracy, as compared to their respective nonword control stimulus combinations. Therefore, I first fitted regression models with the *Stimulus Combination* factor rather than the *Condition* factor (a chi-square Analysis-of-Deviance table of the comparison between these models is presented in Appendix D1). I also included in these models the factors *Block* and *Response Mapping*, but no interaction terms to avoid model over-parameterisation due to the large number of stimulus combinations (45). Then, I performed pairwise comparisons of within-condition stimulus combinations. For example, within the condition *Colour Congruent*, I wanted to test if the effect on accuracy of the written word *red* in the stimulus combination GRG was equivalent to the effect of the same written word in the stimulus combination BRB. To compare their effects, I subtracted from the parameter estimated for each one of these stimulus combinations the parameter estimated for its respective control stimulus combination (i.e., GRG − GNG, and BRB − BNB). I then compared these differences, representing the actual effects produced by the irrelevant written word, with each other [e.g., (GRG − GNG) − (BRB − BNB)]. I performed a total of 81 pairwise comparisons, fifteen for each of the five experimental conditions that consisted of six stimulus combinations and three for each of the two experimental conditions that consisted of three stimulus combinations. All of these comparisons were nonsignificant (*p*'s > .05), indicating that there was no evidence for within-condition differences in accuracy as a function of stimulus combination. Therefore, *Stimulus Combination* was substituted with *Condition* in all subsequent analyses.

### 2.4.2 Response Time

Response times were recorded in milliseconds (to the one-tenth of millisecond) as the time elapsed from stimulus onset to key press. I included in the analysis only the RTs for correct responses ($n$ = 20374) and discarded the RTs for incorrect responses.

Previous research has shown that RT distributions have a typical positively skewed shape, characterised by a normally shaped left tail and an extended right tail. This shape can be simply explained by the fact that theoretically, responses are limited only by how fast, but not by how slow, they can be. From a statistical standpoint, this characteristic shape implies that a typical sample of RTs is neither symmetric nor normally distributed, and thus might be misrepresented by common measures of central tendency and dispersion. Consequently, using statistical models that assume normality to make inferences from mean RT performance may lead to spurious findings and misinterpretation (e.g., Heathcote, Popiel, & Mewhort, 1991; Ratcliff, 2012; Van Zandt, 2002; Van Zandt & Townsend, 2012).

Several methods have been suggested to address the issue of nonnormality when making inferences about RT distributions. Trimming outlying data points by discarding those that exceed a predefined criterion or a certain number of standard deviations from the mean, may serve the goal of normalising the distribution and of increasing statistical power, but at the same time may exclude valid extreme scores that could have resulted from the process of interest (e.g., Ratcliff, 1993; Ulrich & Miller, 1994). Using medians instead of means is justifiable because medians are less sensitive to extreme data points. However, medians tend to have higher between-participant variability than means, and therefore may reduce statistical power, leading to concealment of genuine effects (Ratcliff, 2012; Van Zandt & Townsend, 2012). Another method to reduce the skew of RT distributions is to rescale the data using a logarithmic ($Log_e$RT) or inverse (1/RT) transformation. These nonlinear transformations normalise the distribution and make it more symmetric by increasingly lowering data points as their values increase. Using this method is beneficial, as potentially valid data points need not be discarded while statistical power is not severely affected (Ratcliff, 1993).

To asses normality in RT data, I tested the RT distribution of every participant in each condition (Total of 252 distributions = 28 participants x 9 conditions) using the Shapiro-Wilk test. Of the 252 distributions, 219 (87%) significantly deviated from normality ($p < .05$). Therefore, I inverse transformed RT data by dividing one by the RT (in seconds) of correct responses. Using this transformation is theoretically sensible as it converts a time-based measure (RT) to a speed-based measure and can be simply understood as the number of correct responses per second a certain participant made under different conditions. Further, Ratcliff (1993) showed that using this type of transformation was superior to the logarithmic transformation in maintaining higher statistical power. The Shapiro-Wilk test of inverse transformed RTs showed that of the 252 distributions, 78 (31%) significantly deviated from normality ($p < .05$). Histograms of RTs and inverse transformed RTs as a function of condition and collapsed across participants are presented in Appendix C and D, respectively. These figures show the typical positive skewness of the RT distribution and the more symmetrical bell-shaped distribution of inverse transformed RTs.

To analyse RT data, I fitted mixed-effects linear regression models to inverse transformed RTs using the function *lmer* in the R package *lme4* (Bates et al., 2014). I used the function *anova* from the same R package to compare the deviance of regression models that had different definitions of the random factor, and to select the model that provided the best fit. I then tested factors and interactions from that model for significance using the function *Anova* from the R package *car* (Fox & Weisberg, 2011). I refitted the model after omitting nonsignificant factors and interactions to come up with the most parsimonious model to account for the data. I performed pairwise comparisons between levels of significant factors using the function *glht* from the R package *multcomp* (Bretz et al., 2010; Hothorn et al., 2008).

As was previously described for accuracy data, initially I wanted to test if individual stimulus combinations within each experimental condition were comparable with respect to their effects on RT, as compared to their respective nonword control stimulus combinations. Therefore, I first fitted regression models with the *Stimulus Combination* factor rather than the *Condition* factor (a chi-square Analysis-of-Deviance table of the comparison between these models is presented in Appendix E1). I also included in these models the factors *Block* and *Response Mapping*. Then, I performed pairwise comparisons of within-condition stimulus combinations using the same procedure that was previously described for accuracy data. All but one of these comparisons were statistically nonsignificant ($p$'s $> .05$), indicating that with one exception, there was no evidence for within-condition differences in RT as a function of stimulus combination. The one exception was in the *Neutral Colour Congruent* condition, where the effect on RT of the neutral written word *white* in the stimulus combination GWG, as compared to its respective nonword control (GNG), was significantly greater than the effect of the same written word in the stimulus combination RWR, as compared to its respective nonword control (RNR), $z = 3.56$, $p < .05$, d = 0.063, 95% CI [0.003, 0.123]. Assuming that this difference could not have altered the general pattern of results, *Stimulus Combination* was substituted with *Condition* in all subsequent analyses.

# 3 Results

## 3.1 Accuracy

The number and proportion of correct responses by condition, collapsed across participants, are presented in Table 3.1.

A chi-square Analysis-of-Deviance table of the comparison between the three models fitted with *Condition* as fixed factor is presented in Appendix D2. This comparison revealed that Model 3, which incorporated only random slope for *Condition*, provided the best fit and thus was selected for further analysis. A Type-II Wald chi-square test of fixed factor terms included in that model showed significant main effects of *Condition*, $\chi^2$ (7, $N$ = 21168) = 57.19, $p < .001$, and of *Response Mapping*, $\chi^2$ (1, $N$ = 21168) = 17.84, $p < .001$. The *Block* factor, $\chi^2$ (6, $N$ = 21168) = 4.61, $p = .595$, the *Block* X *Condition* interaction, $\chi^2$ (42, $N$ = 21168) = 46.32, $p = .299$, and the *Response Mapping* X *Condition* interaction, $\chi^2$ (7, $N$ = 21168) = 10.3, $p = .172$, were nonsignificant. Thus, Model 3 was refitted after omission of nonsignificant factors and interactions.

**Table 3.1 Number and proportion of correct responses by condition**

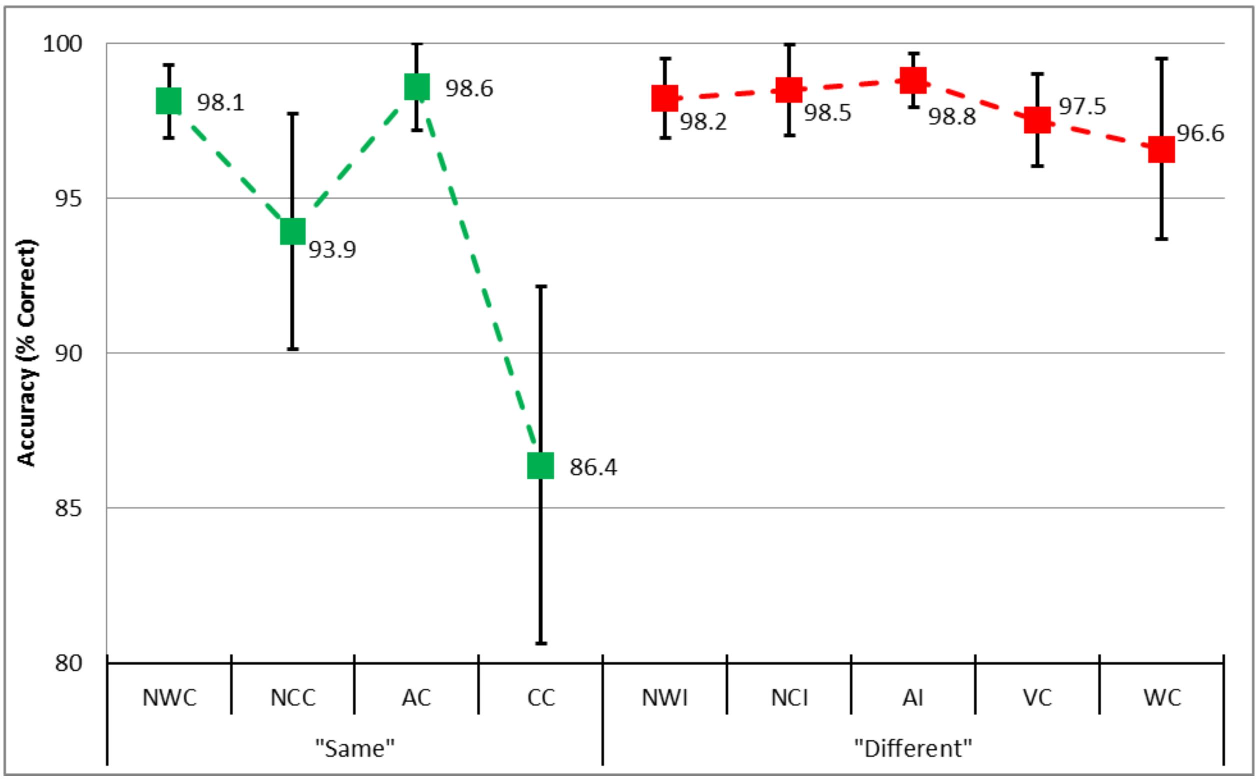
| Response | Condition | *N* Trials | *N* Correct | *N* Incorrect | % Correct | *SD* |
|---|---|---|---|---|---|---|
| Same | Nonword Congruent | 2352 | 2302 | 50 | 97.87 | 1.81 |
| | Neutral Colour Congruent | 2352 | 2203 | 149 | 93.66 | 4.81 |
| | All Congruent | 2352 | 2313 | 39 | 98.34 | 2.01 |
| | Colour Congruent | 2352 | 2026 | 326 | 86.14 | 7.14 |
| Different | Nonword Incongruent | 2352 | 2314 | 38 | 98.38 | 1.75 |
| | Neutral Colour Incongruent | 2352 | 2319 | 33 | 98.60 | 1.89 |
| | All Incongruent | 2352 | 2327 | 25 | 98.94 | 1.27 |
| | Visual Congruent | 2352 | 2297 | 55 | 97.66 | 2.26 |
| | Word Congruent | 2352 | 2273 | 79 | 96.64 | 3.82 |

Collectively, these results indicated that accuracy significantly differed between conditions and between response-key assignments. However, there was no evidence that accuracy significantly differed between trial-blocks in general or as a function of condition, or that accuracy significantly differed between response-key assignments as a function of condition. The nonsignificant *Response Mapping* X *Condition* interaction indicated that key assignment could not have changed the pattern of differences between conditions observed in the data collapsed across assignments. Apparently, one key-assignment was in general associated with more errors than the other.

Model-predicted percent correct scores as a function of condition are presented in Figure 3.1. To test the effect of the written word on accuracy, I contrasted each experimental condition with the nonword condition that required the same response (7 comparisons). To test the effect of the type of response ("same" vs. "different") without any written lexical distraction, I contrasted the two nonword conditions. These comparisons showed that *Colour Congruent* decreased accuracy significantly as compared to *Nonword Congruent*, $z = -5.24$, $p < .001$, $d = -2.37$, 95% CI [−3.57, −1.17], and that *Word Congruent* decreased accuracy significantly as compared to *Nonword Incongruent*, $z = -2.83$, $p < .05$, $d = -1.05$, 95% CI [−2.03, −0.07]. All other comparisons were nonsignificant ($p$'s > .05).

These results indicated that as compared to the nonword condition, interference effects to accuracy were statistically significant in two conditions while there was no evidence for facilitation effects. Accuracy was significantly decreased when the written word was incongruent with both the spoken word and font colour and the response was "same" (e.g., hear /red/ and see *blue* in red font); and, when the written word was congruent with the spoken word and the response was "different" (e.g., hear /red/ and see *red* in blue font). In general, these results supported the outcome-conflict hypothesis because accuracy was significantly decreased only when the "different" outcomes of both irrelevant comparisons conflicted with a "same" response, and when a "same" outcome of the irrelevant spoken word–written word comparison conflicted with a "different" response. There was no evidence that without written word distraction, one type of response (i.e., "same" or "different") was associated with more errors.

**Figure 3.1 Model-predicted percent correct scores by condition**



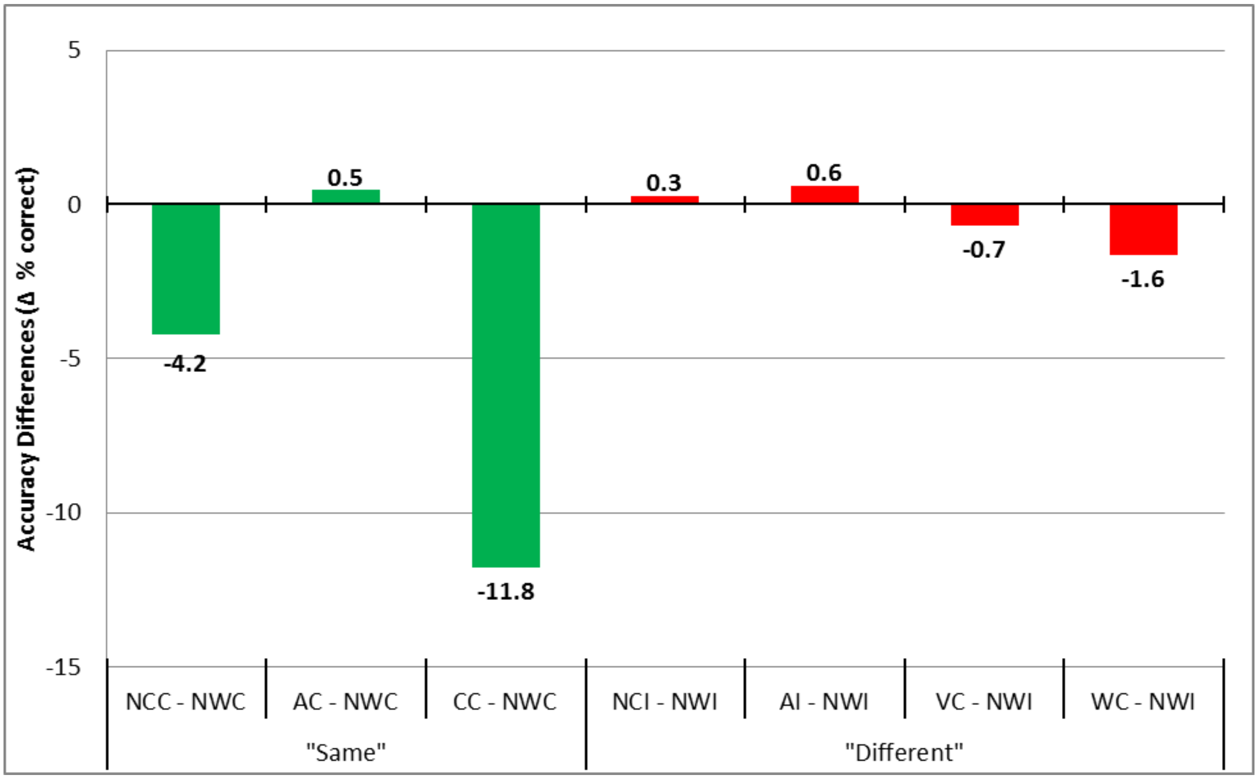

Note. Error bars represent ±1 standard deviation from the mean model-predicted percent correct score. See text for details.
AC = All Congruent, AI = All Incongruent, CC = Colour Congruent, NCC = Neutral Colour Congruent, NCI = Neutral Colour Incongruent, NWC = Nonword Congruent, NWI = Nonword Incongruent, VC = Visual Congruent, WC = Word Congruent.

The differences in model-predicted mean percent correct scores between experimental and control conditions are presented in Figure 3.2. To compare the effects of the written word on accuracy between experimental conditions, I performed 21 additional pairwise comparisons.

These included nine within-judgement contrasts (three "same" and six "different") and twelve between-judgement contrasts ("same" vs. "different"). In within-judgement contrasts, I compared the parameters estimated for experimental conditions that required the same response. In between-judgement contrasts, I subtracted from the parameter estimated for each experimental condition the parameter estimated for its respective control condition, to account for differences related to the type of response (i.e., "same" vs. "different"), and then compared the residual effects.

These comparisons showed that within "same" conditions, *Colour Congruent* significantly decreased accuracy as compared to *Neutral Colour Congruent*, $z = -4.09$, $p < .001$, $d = -1.18$, 95% CI [−2.00, −0.35], and *All Congruent*, $z = -4.40$, $p < .001$, $d = -3.15$, 95% CI [−5.19, −1.10]. The decrease in accuracy from *Neutral Colour Congruent* as compared to *All Congruent* was marginally significant, $z = -2.75$, $p = .07$, $d = -1.97$, 95% CI [−4.02, 0.08]. Within "different" conditions, *Word Congruent* significantly decreased accuracy as compared to *All Incongruent*, $z = -2.94$, $p < .05$, $d = -0.97$, 95% CI [−1.91, −0.03]. Between "same" and "different" conditions, *Colour Congruent* significantly decreased accuracy as compared to *Visual Congruent*, $z = -3.98$, $p < .001$, $d = -1.91$, 95% CI [−3.29, −0.54], *All Incongruent*, $z = -3.91$, $p < .05$, $d = -2.29$, 95% CI [−3.96, −0.62], and *Neutral Colour Incongruent*, $z = -3.39$, $p < .05$, $d = -2.26$, 95% CI [−4.16, −0.35]. All the remaining comparisons were nonsignificant ($p$'s > .1).

**Figure 3.2** **Differences in model-predicted mean percent correct scores between experimental and control conditions**




Note. Downward bars represent decrements in accuracy as compared to the respective nonword condition. Error bars are not provided because comparisons of the differences were performed on logit transformed accuracy scores. See text for details.
AC = All Congruent, AI = All Incongruent, CC = Colour Congruent, NCC = Neutral Colour Congruent, NCI = Neutral Colour Incongruent, NWC = Nonword Congruent, NWI = Nonword Incongruent, VC = Visual Congruent, WC = Word Congruent.

These results indicated that when the written word was incongruent with both task-relevant dimensions and the response was “same” (e.g., hear /red/ and see *blue* in red font), accuracy was lower than in all the other conditions, except the condition in which the written word was congruent with the spoken word and the response was “different” (e.g., hear /red/ and see *red* in blue font). When the written word was congruent with the spoken word and the response was “different” (e.g., hear /red/ and see *red* in blue font), accuracy was lower than when the written word was incongruent with both task-relevant dimensions and the response was “different” (e.g., hear /red/ and see *blue* in green font). Lastly, when the written word was the neutral colour-word *white* and the response was "same" (e.g., hear /red/ and see *white* in red font), accuracy was marginally significantly lower than when the written word was congruent with both task-relevant dimensions (e.g., hear /red/ and see *red* in red font).

In general, these findings supported the outcome-conflict hypothesis because accuracy was lower in "different" responses when the irrelevant spoken word – written word comparison indicated "same", as compared to when both irrelevant comparisons indicated "different". Also, these findings showed that the effects of outcome-conflicts were additive, because accuracy was lower when the outcomes of both irrelevant comparisons conflicted with the response, as compared to when only the outcome of the irrelevant font colour – written word comparison conflicted with the response. Moreover, these findings indicated that the degree of semantic relatedness of the written word to the task-relevant dimensions affected accuracy. This was because written words that were represented in the same stimulus set as the task-relevant dimensions (i.e., *red*, *green*, and *blue*) decreased accuracy more than the neutral colour-word (i.e., white).

The comparison between the two levels of *Response Mapping* showed that participants that responded "same" with the left-arrow key and "different" with the right-arrow key had significantly lower accuracy ($M$ = 95%, $SD$ = 5.4%) than those that responded using the reverse key assignment ($M$ = 97.4%, $SD$ = 3.6%), $z = -3.7$, $p < .001$, $d = -0.67$, 95% CI [−1.02, −0.31]. As noted before, the nonsignificant *Response Mapping* X *Condition* interaction indicated that the effect of response-key assignment on accuracy was similar across conditions, and therefore comparisons between conditions did not need to account for the differences associated with key assignment.

## 3.2 Response Time

Means, medians and skewness of RT of correct responses by condition are presented in Table 3.2.

**Table 3.2 Means, medians and skewness of RT of correct responses by condition**

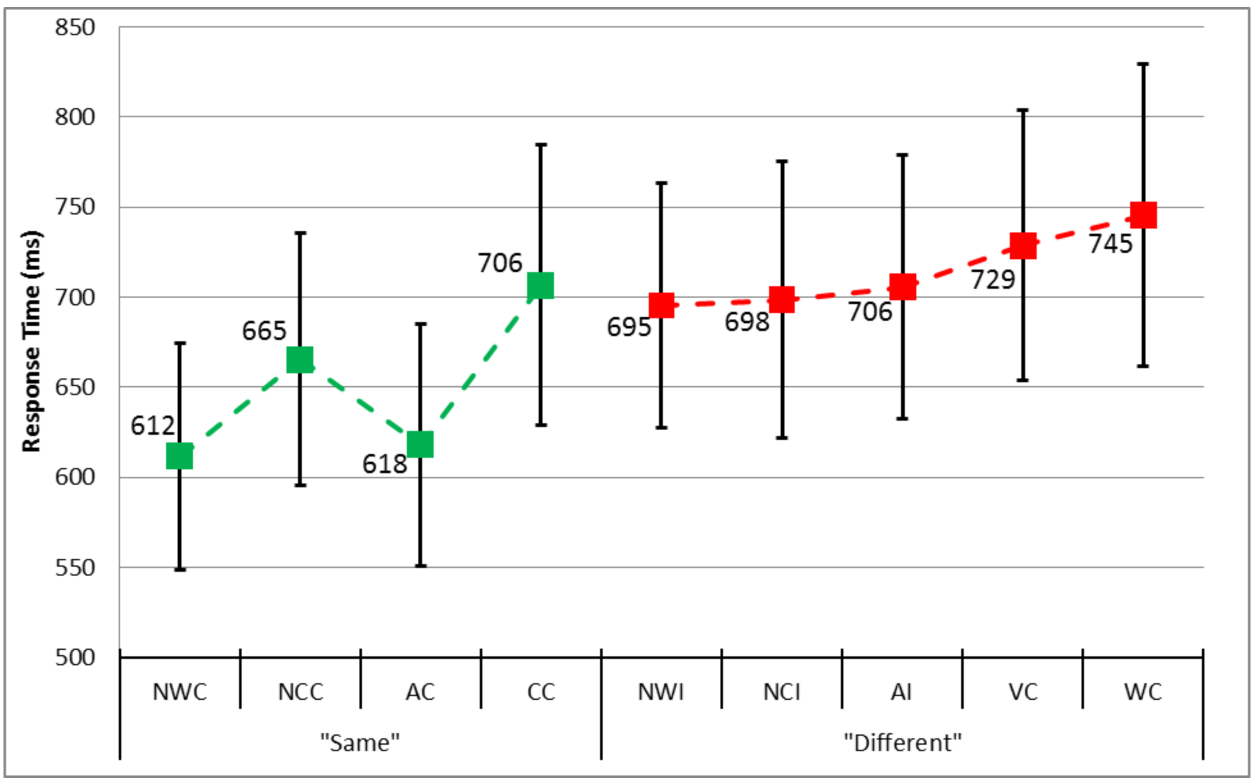
| Judgement | Condition | *M (SD)* | *Md* (*SD*) | *Skewness* (*Range*) |
|---|---|---|---|---|
| Same | Nonword Congruent | 625.78 (64.68) | 608.58 (60.39) | 1.50 (0.16 – 3.55) |
| | Neutral Colour Congruent | 688.11 (71.25) | 666.83 (68.50) | 1.14 (0.02 – 2.71) |
| | All Congruent | 634.95 (71.09) | 613.19 (67.85) | 1.58 (−0.01 – 5.82) |
| | Colour Congruent | 733.30 (80.44) | 709.37 (81.47) | 0.77 (0.05 – 1.74) |
| Different | Nonword Incongruent | 713.94 (70.00) | 690.95 (65.59) | 1.28 (−0.39 – 2.86) |
| | Neutral Colour Incongruent | 718.12 (79.99) | 694.87 (75.52) | 1.06 (−0.32 – 2.02) |
| | All Incongruent | 728.44 (77.15) | 703.12 (71.11) | 1.19 (−0.14 – 2.88) |
| | Visual Congruent | 754.08 (79.03) | 722.18 (74.45) | 1.14 (0.33 – 2.05) |
| | Word Congruent | 770.33 (89.56) | 742.86 (87.68) | 1.19 (−0.17 – 2.21) |

Note. All times are in milliseconds.

A chi-square Analysis-of-Deviance table of the comparison between the three models fitted with *Condition* as fixed factor is presented in Appendix E2. This comparison revealed that Model 3, which incorporated only random slope for *Condition*, provided the best fit and thus was selected for further analysis. A Type-II Wald chi-square test of fixed factors included in Model 3 showed significant main effects of *Condition*, $\chi^2$ (7, *N* = 20374) = 175.73, *p* < .001, and of *Block*, $\chi^2$ (6, *N* = 20374) = 627.39, *p* < .001. The *Response Mapping* factor, $\chi^2$ (1, *N* = 20374) = 0.49, *p* = .485, the *Block* X *Condition* interaction, $\chi^2$ (42, *N* = 20374) = 41.28, *p* = .502, and the *Response Mapping* X *Condition* interaction, $\chi^2$ (7, *N* = 20374) = 8.41, *p* = .298, were nonsignificant. Thus, Model 3 was refitted after omission of nonsignificant factors and interactions. Collectively, these results indicated that RTs significantly differed between conditions and between trial-blocks. However, there was no evidence that RTs significantly differed between response-key assignments in general or as a function of condition, or that RTs significantly differed between trial-blocks as a function of condition.

Model-predicted mean RTs as a function of condition are presented in Figure 3.3. To test the effect of the written word on RT, I contrasted each experimental condition with the nonword condition that required the same response (7 comparisons). To test the effect of the type of response ("same" vs. "different") without any written word distraction, I contrasted the two nonword conditions.

**Figure 3.3 Model-predicted mean RT by condition**



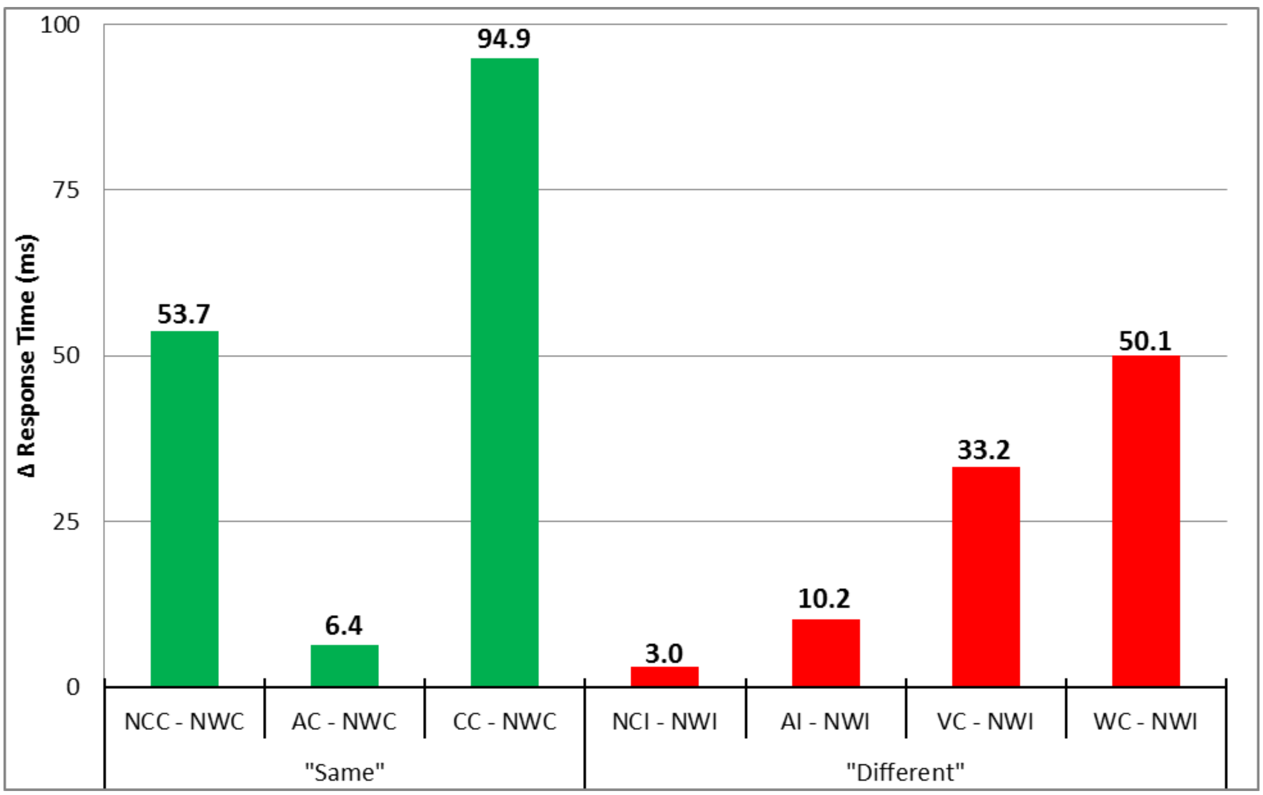

Note. Error bars represent ±1 standard deviation from the model-predicted mean RT. See text for details. AC = All Congruent, AI = All Incongruent, CC = Colour Congruent, NCC = Neutral Colour Congruent, NCI = Neutral Colour Incongruent, NWC = Nonword Congruent, NWI = Nonword Incongruent, VC = Visual Congruent, WC = Word Congruent.

These comparisons showed that, as compared to *Nonword Congruent*, responses were significantly delayed in *Colour Congruent*, *z = −9.17*, *p* < .001, *d* = −0.12, 95% CI [−0.16, −0.09], in *Neutral Colour Congruent*, *z* = −9.93, *p* < .001, *d* = −0.09, 95% CI [−0.12, −0.07], and in *All Congruent*, *z* = −3.4, *p* < .05, *d* = −0.03, 95% CI [−0.05, −0.01]. As compared to *Nonword Incongruent*, responses were significantly delayed in *Visual Congruent*, *z* = −4.61, *p* < .001, *d* = −0.04, 95% CI [−0.06, −0.02], and in *Word Congruent*, *z* = −4.85, *p* < .001, *d* = −0.06, 95% CI [−0.10, −0.03]. The comparison between the two nonword conditions showed that responses were significantly faster in *Nonword Congruent* as compared to *Nonword Incongruent*, z = 5.34, p < .001, *d* = 0.05, 95% CI [0.02, 0.07], reflecting a *fast-"same"* effect, and thus justifying the use of a different baseline for measuring interference and facilitation effects for each type of response. The remaining comparisons were nonsignificant (*p*'s > .05).

Collectively, these results indicated that as compared to the nonword conditions, none of the experimental conditions showed evidence for facilitation effects. Interference effects however, were evident in "same" and "different" conditions. "Same" responses were significantly delayed whenever a written word was presented, regardless of its congruence with the task-relevant dimensions. This result was unexpected for the condition in which the written word was congruent with both task-relevant dimensions (e.g., hear /red/ and see *red* in red font), because that condition was expected to facilitate, rather than interfere with, performance. "Different" responses were significantly delayed when the written word was congruent with either task-relevant dimension (e.g., hear /red/ and see *blue* in blue font, and hear /red/ and see *red* in blue font). In general, these results supported the outcome-conflict hypothesis, because responses were significantly delayed whenever comparing the written word and either one or both of the task-relevant dimensions produced outcomes that conflicted with the response. That is, when the comparison of the written word with both task-relevant dimensions resulted in a "different" outcome, a "same" response was significantly delayed. And, when the comparison of the written word with either task-relevant dimension resulted in a "same" outcome, a "different" response was significantly delayed.

The differences in model-predicted mean RT between experimental and control conditions are presented in Figure 3.4. To compare the effects of the irrelevant written word on RT between experimental conditions, I performed 21 additional pairwise comparisons as previously described for accuracy data.

**Figure 3.4 Differences in model-predicted mean RT between experimental and control conditions**




Note. Upward bars represent increased RT as compared to the respective control condition. Error bars are not provided because comparisons of the differences were performed on inverse transformed RTs. See text for details. AC = All Congruent, AI = All Incongruent, CC = Colour Congruent, NCC = Neutral Colour Congruent, NCI = Neutral Colour Incongruent, NWC = Nonword Congruent, NWI = Nonword Incongruent, VC = Visual Congruent, WC = Word Congruent.

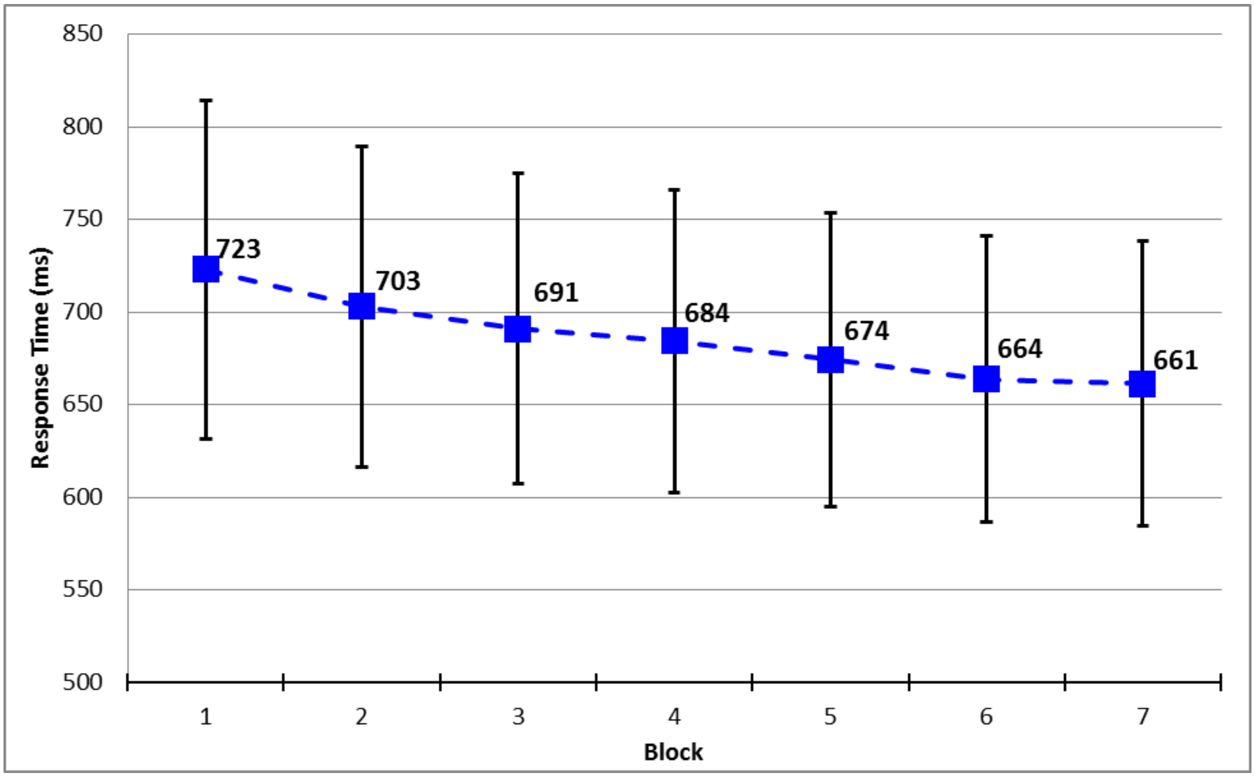
These comparisons showed that within "same" conditions, *Colour Congruent* and *Neutral Colour Congruent* significantly delayed responses as compared to *All Congruent*, $z = -7.19$, $p < .001$, $d = -0.10$, 95% CI [−0.14, −0.06], and $z = -6.34$, $p < .001$, $d = -0.07$, 95% CI [−0.10, −0.04], respectively. Within "different" conditions, *Visual Congruent* significantly delayed responses as compared to *All Incongruent*, $z = -6.22$, $p < .001$, $d = -0.05$, 95% CI [−0.07, −0.03] and *Neutral Colour Incongruent*, $z = -5.11$, $p < .001$, $d = -0.04$, 95% CI [−0.07, −0.02]. *Word Congruent* significantly delayed responses as compared to *All Incongruent*, $z = -6.8$, $p < .001$, $d = -0.08$, 95% CI [−0.11, −0.04] and *Neutral Colour Incongruent*, $z = -5.81$, $p < .001$, $d = -0.07$, 95% CI [−0.11, −0.04]. Between "same" and "different" conditions, *All Congruent* significantly delayed responses as compared to *All Incongruent*, $z = -3.59$, $p < .05$, $d = -0.04$, 95% CI [−0.07, −0.01] and marginally significantly as compared to *Neutral Colour Incongruent*, $z = -2.89$, $p = .052$, $d = -0.03$, 95% CI [−0.07, 0.0001]. *Colour Congruent* significantly delayed responses as compared to *Neutral Colour Incongruent*, $z = -7.83$, $p < .001$, $d = -0.13$, 95% CI [−0.18, −0.08], *All Incongruent*, $z = -8.59$, $p < .001$, $d = -0.14$, 95% CI [−0.18, −0.09], *Visual Congruent*, $z = -5.44$, $p < .001$, $d = -0.09$, 95% CI [−0.13, −0.04], and *Word Congruent*, $z = -3.24$, $p < .05$, $d = -0.06$, 95% CI [−0.12, −0.01]. *Neutral Colour Congruent* significantly delayed responses as compared to *Neutral Colour Incongruent*, $z = -7.69$, $p < .001$, $d = -0.10$, 95% CI [−0.14, −0.06], *All Incongruent*, $z = -8.48$, $p < .001$, $d = -0.11$, 95% CI [−0.14, −0.07], and *Visual Congruent*, $z = -4.97$, $p < .001$, d = −0.06, 95% CI [−0.09, −0.02]. The remaining comparisons were nonsignificant ($p$'s > .1).

These results indicated that when the written word was incongruent with both task-relevant dimensions and the response was "same" (e.g., hear /red/ and see *blue* in red font), responses were significantly delayed as compared to all the other conditions, except the condition in which the written word was the neutral word *white* and the response was “same” (e.g., hear /red/ and see *white* in red font). “Different” responses were significantly delayed when either task-irrelevant comparison indicated “same” (e.g., hear /red/ and see *blue* in blue font, and hear /red/ and see *red* in blue font), as compared to when both task-irrelevant comparisons indicated “different” (e.g., hear /red/ and see *blue* in green font, and hear /red/ and see *white* in blue font). When the written word was the neutral word *white* and the response was “same” (e.g., hear /red/ and see *white* in red font), responses were significantly delayed as compared to all the “different” conditions, except the condition in which the written word was congruent with the spoken word (e.g., hear /red/ and see *red* in blue font).

Collectively, these findings supported the outcome-conflict hypothesis because “different” responses were delayed more when either task-irrelevant comparison indicated “same” as compared to when both irrelevant comparisons indicated “different”. Also, these results showed that additional interference occurred when the outcomes of both irrelevant comparisons conflicted with the response, as compared to when the outcome of either irrelevant comparison conflicted with the response. However, these findings did not indicate that a “different” response was delayed more from congruence between the written word and spoken word than from congruence between the written word and font colour. Also, these results did not indicate that response delays were related to the degree of semantic relatedness between the written word and task-relevant dimensions. This is because the neutral written word *white* delayed responses to the same degree as the written colour words that were represented in the task-relevant stimulus set (i.e., *red*, *green*, and *blue*).

Model-predicted mean RTs as a function of trial-block are presented in Figure 3.5.

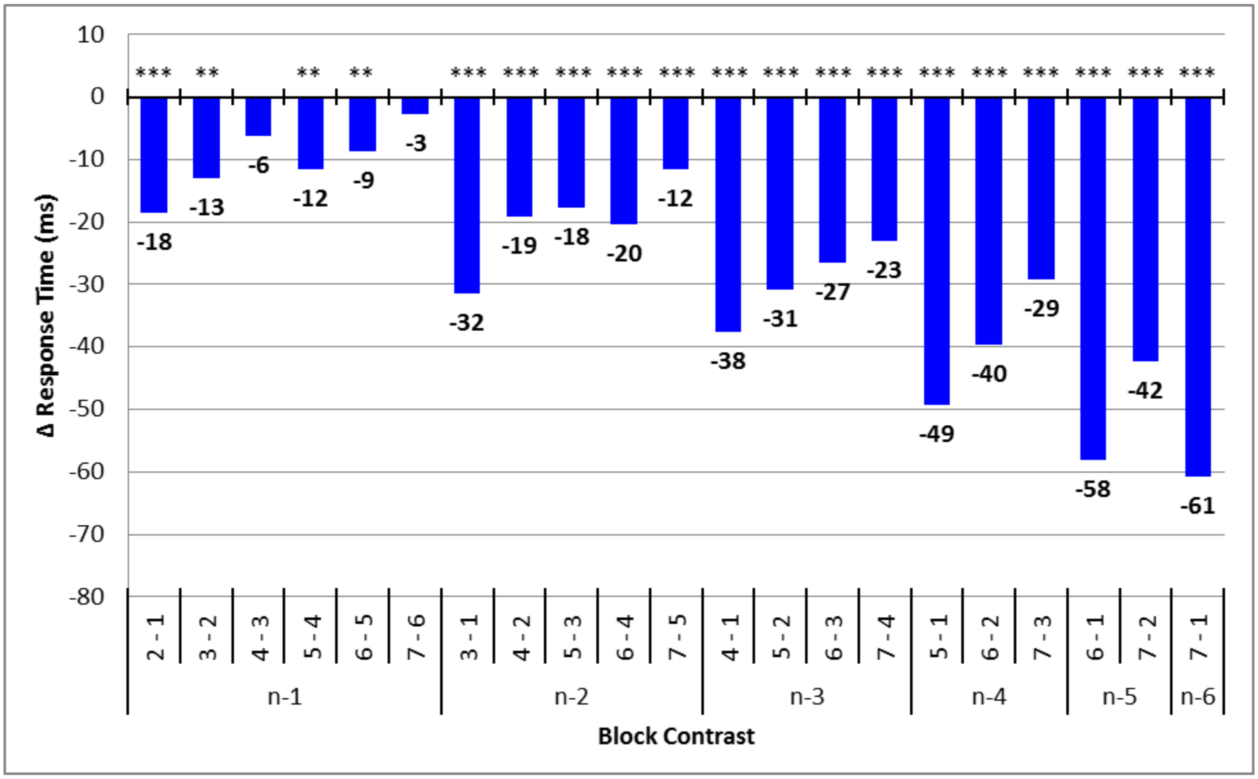
**Figure 3.5 Model-predicted mean RT by trial-block**




Note. Error bars represent ±1 standard deviation from the mean model-predicted RT.

To explore how performance changed over the course of the experiment, I performed 21 pairwise comparisons between consecutive (e.g., 2 − 1, 3 − 2, etc.) and equally separated (e.g., 3 − 1, 4 − 2, etc.) blocks. Differences in model-predicted mean RT between the contrasted pairs of blocks are presented in Figure 3.6. The results of these comparisons showed that overall, RTs significantly decreased over the course of the experiment, with the exception of the difference between blocks 3 and 4, and between blocks 6 and 7, where the comparison was nonsignificant ($p$’s > .05). There was no evidence for an increase in RT between any pair of consecutive or equally separated blocks. These results indicated that in general, performance improved over the course of the experiment. The pattern of decrease in RTs showed that the performance improvements became smaller as the experiment progressed.

**Figure 3.6 Trial-block contrasts**




Note. Downward bars represent RT decreases. ** $p < .05$, *** $p < .001$. *p*-values reflect significant differences between pairs of trial-blocks based on pairwise comparisons performed on parameters estimated from the regression model. Error bars are not provided because comparisons were performed on inverse transformed RTs. See text for details.

# 4 Discussion

Previous audiovisual Stroop studies used spoken colour words mainly as to-be-ignored distractors when performing the classic visual Stroop task. My study aimed to explore the effects of written word distractors on attending to and matching stimuli presented to the auditory and visual modalities. To address this aim, I specifically designed an auditory-visual matching version of the classic colour-word Stroop task.

## 4.1 Interference from Conflicting Outcomes

The findings of my study indicated that the meaning of the written word significantly decreased the speed and accuracy within which spoken words and font colours were judged to be congruent or incongruent. The pattern of findings demonstrated that a conflict between the outcomes of the relevant matching task (spoken word – font colour) and two irrelevant matching tasks (spoken word – written word, font colour – written word) produced the observed interference effects. These findings are consistent with those of Goldfarb and Henik (2006) from a visual matching Stroop task and weigh against those of Luo (1999). Thus, my findings support the view that a response-conflict, rather than a semantic-conflict produces interference effects.

Navon (1985) coined the term *outcome-conflict* to reflect a state where the output of one task modifies (and potentially interferes with) a variable that is relevant to the performance of a concurrent task. The processing costs associated with outcome-conflicts might be expressed as decreased accuracy and/or delayed responses. I propose that in my study, outcome-conflicts occurred whenever the relevant matching task and the two mistakenly performed matching tasks produced conflicting outputs (i.e., “same” vs. “different”). Interference effects were large and significant only in conditions that featured such a conflict. This interpretation can be substantiated by the results of my study.

For example, the *All Incongruent* condition presented three different colour values (e.g., hear /red/ and see *blue* in green). The semantic-conflict hypothesis predicts that interference in this condition would be the greatest because the three colour values activate three distinct semantic representations of colour. The outcome-conflict hypothesis does not predict any interference in this condition because all three comparisons between colour representations indicate "different" outcomes. My results did not show any evidence for interference effects in this condition as compared to the nonword control condition. Thus, my results are more consistent with the outcome-conflict hypothesis than the semantic-conflict hypothesis.

Additional evidence against the semantic-conflict hypothesis comes from the comparison between the *All Incongruent* and *All Congruent* conditions. The *All Incongruent* condition (e.g., hear /red/ and see *blue* in green) created a three-way semantic-conflict (all representations incongruent) with no outcome-conflict (all comparisons "different"), whereas the *All Congruent* condition (e.g., hear /red/ and see *red* in red) created neither a semantic-conflict (all representations congruent) nor an outcome-conflict (all judgements "same"). Thus, these two conditions differed by the type of response ("different" vs. "same") and by a semantic-conflict present in *All Incongruent* and absent in *All Congruent*. Comparison of the RTs of these conditions showed that responses in *All Incongruent* were on average 88-ms slower than in *All Congruent*. This delay can be attributed to the difference in the type of response, to semantic-conflict, or to both. However, comparison of the two nonword conditions (*Nonword Incongruent* and *Nonword Congruent*) showed that "different" responses were on average 83-ms slower than "same" responses. Thus, the relatively equal delays observed in these two comparisons shows that the entire delay between *All Incongruent* and *All Congruent* was likely attributed to the difference in the type of response (i.e., "different" vs. "same"), and that there was no additional RT interference due to semantic-conflict.

Additional support for the outcome-conflict hypothesis comes from the comparison between the *Neutral Colour Congruent* (e.g., hear /red/ and see *white* in red) and the *Neutral Colour Incongruent* (e.g., hear /red/ and see *white* in blue) conditions. In both conditions, the written word *white* semantically conflicted with the spoken word and font colour. These conditions differed by the type of response (i.e., "same" vs. "different"), and by an outcome-conflict present in *Neutral Colour Congruent* (relevant match "same" and irrelevant matches "different") and not in *Neutral Colour Incongruent* (all matches "different"). As previously noted, the comparison between the two nonword conditions showed that without written word distraction, "different" responses were on average 83-ms slower than "same" responses. However, with the word *white* as distractor, "different" responses were on average only 33-ms slower than "same" responses. Thus, the RT difference between "same" and "different" responses evidently became smaller by the same distracting written word. This decrease can be attributed only to the delay in "same" responses as compared to the nonword condition (54-ms), because "different" responses were hardly affected as compared to the nonword condition (3-ms). Because the written word semantically conflicted with the spoken word and font colour in both neutral conditions, the only available explanation to the slowing down being restricted to "same" responses is outcome-conflict. That is, for "same" responses, matching the written word and the spoken word and font colour resulted in an outcome-conflict, whereas for "different" responses, these matches resulted in concurring outcomes.

Lastly, in the two "different" conditions where the written word was congruent either with the spoken word (*Word Congruent*, e.g., hear /red/ and see *red* in blue) or with the font colour (*Visual Congruent*, e.g., hear /red/ and see *blue* in blue), responses were delayed (by 50- and 33-ms, respectively) as compared to the nonword condition (*Nonword Incongruent*; e.g., hear /red/ and see #### in blue). Thus, although two of the colour representations were semantically congruent (i.e., no semantic-conflict), responding “different” in these conditions was considerably delayed. One can reason that the incongruence between the written word and the other relevant dimension produced the interference, but that explanation does not align with the finding that no interference occurred when all representations were incongruent (i.e., in *All Incongruent*). Therefore, it must have been the outcome-conflict between the “same” output of the irrelevant matching and the “different” response that produced interference effects.

The findings of my audiovisual matching task are also consistent with the *confluence model* proposed by Eviatar, Zaidel, and Wickens (1994), based on their findings from a visual matching task. According to their model, in matching tasks, all stimulus dimensions are processed automatically and simultaneously regardless of task-relevance, and the outputs of all processors reach a point of confluence before a response is selected. This model proposes that interference occurs at a stage preceding the response stage, where the outputs of all processes are compared. According to this notion, in my study, auditorily and visually presented colour information were processed to a point at which their representations could be compared. The congruence or incongruence among the outputs of these comparisons determined the likelihood of and delay in selecting the correct response. The interpretation provided by the confluence model is similar to that proposed by Navon’s (1985) outcome-conflict; however, the confluence model is specifically pertaining to matching tasks, and it is also more explicit with regard to the processing stage to which interference is attributed.

In light of the confluence model, the similarities between the findings of my audiovisual matching Stroop task and those of visual matching Stroop tasks (e.g., Dyer, 1973; Goldfarb & Henik, 2006) might indicate that the mechanisms that produce interference effects in this task are similar regardless of the sensory channel through which the information is perceived. This, in turn, might indicate that the comparison processes involved in matching are amodal, and achieved by high-level processing systems capable of executive functions such as conflict detection and resolution and in decision making.

The behavioural results of Caldas et al. (2012) replicated those of Goldfarb and Henik (2006) and thus were also consistent with a response-conflict hypothesis of interference in a matching Stroop. Their electrophysiological data, however, supported a semantic-conflict hypothesis. They showed that conflict related brain activity, as indicated by a greater frontal negativity (N450), was not observed in “different” conditions that featured conflicting irrelevant “same” outputs. Rather, N450 amplitude was greater in “different” conditions that featured incongruent font and bar colours (e.g., *red* in red font paired with [redacted] in blue, and red in green paired with [redacted] in blue) than in the “different” condition that featured congruent font and bar colours (e.g., *red* in blue font paired with [redacted] in blue). The discrepancy between the behavioural and electrophysiological findings led Caldas et al. to propose that conflicts at both the response and nonresponse levels produce interference in the matching Stroop task. Future analysis of electrophysiological data collected for my study might provide further evidence with regard to the manifestation of different sources of conflict in brain activity, and extend previous findings by examining the effects of cross-modality on conflict processing in the brain.

## 4.2 No Facilitation from Concurring Outcomes

The findings of my study showed no evidence for facilitation effects in any of the experimental conditions, as compared to the nonword conditions. This finding was contrary to my expectation that facilitation would occur when the outcomes of the irrelevant matching tasks agreed with the response to the relevant matching task (i.e., when all were "same" or "different"). The results showed that in the three conditions in which all outcomes agreed (i.e., *All Congruent*; e.g. hear /red/ and see *red* in red, *All Incongruent*; e.g. hear /red/ and see *green* in blue, and *Neutral Colour Incongruent*; e.g. hear /red/ and see *white* in blue), the effects were in general minimal and with one exception, nonsignificant. The exception was in the *All Congruent* condition, where responses were significantly slower as compared to the nonword condition, although this effect was nominally small, only amounting to 6-ms. These findings indicated that when the spoken word and the font colour were congruent (i.e., "same"), additional non-conflicting information in the form of a congruent written word did not facilitate the perception of a non-conflict. Similarly, when the spoken word the font colour were incongruent (i.e., "different"), additional conflicting information in the form of an incongruent written word did not facilitate the perception of a conflict.

The asymmetry between the large interference effects and the lack of facilitation effects is consistent with findings from the classic Stroop task that show that facilitation effects, where present, are always much smaller in magnitude than interference effects (for a review see MacLeod, 1991). This asymmetry challenges the notion that interference and facilitation result from complementary processing mechanisms. That is, that interference stems from diverging information present in incongruent trials and that facilitation stems from converging information present in congruent trials (e.g., Cohen et al., 1990; Melara & Algom, 2003). MacLeod and

MacDonald (2000) proposed a different hypothesis to explain the inconsistent findings regarding facilitation in the classic Stroop task. In their view, facilitation effects, where reported, result from inadvertent reading of the colour-word that is incorrectly counted as a correct colour naming response. Because congruent trials lead to the same response whether the word is read or the colour named, it is impossible to tell which response was selected on a given trial, and therefore naming and reading times are intermixed. Because there are within- and between-study differences in how individuals adhere to task instructions, inconsistent reporting of facilitation effects would be expected.

The findings of my study indicated that participants unlikely responded based on inadvertently reading the words. If they had done so, facilitation effects would have been evident, because matching stimuli that share a code (i.e., lexicality) is faster than matching stimuli that do not share a code (e.g., Treisman & Fearnley, 1969; Virzi & Egeth, 1985). The lack of facilitation showed that participants adhered to the instructions and first represented font colour in a form that could be compared with the spoken word and only then matched the two dimensions. At that point in time, reading the colour-word could no longer speed up the response. This idea might also explain the slightly but nonsignificantly improved accuracy in the conditions in which all the outcomes concurred, as compared to the nonword conditions. In these conditions, the written word provided additional evidence toward the correct response and thus augmented accuracy, but this additional information preceded the point in time at which the task-relevant information accumulated to allow the correct response, and thus could not speed it up.

A possible explanation for the lack of facilitation is that given the task instructions of attending to the spoken word and font colour, participants attempted to ignore the written meaning. Evidently, their attempts were not efficient enough to prevent interference when the written word provided evidence that conflicted with the response. However, in trials where the written meaning provided evidence supporting the correct response, the attempts to ignore it might have eliminated the potential benefits of reading the word. In a colour naming task, Lowe and Mitterer (1982) showed that increasing the proportion of congruent trials increased interference from incongruent trials independently of the proportion of incongruent trials. This finding indicates that because processing the written word was beneficial to performance in congruent trials, when an incongruent trial was presented, the inclination to process the written word increased interference. In my study, trials in which the written word meaning could have benefited performance occurred in one-third of the overall trials (i.e., in three of the nine conditions). Thus, both the task instructions and the majority of the trials indicated that written word meaning is merely a distraction. It is not surprising then that on the relatively smaller proportion of trials in which the written meaning could have benefited performance, in reality it did not.

My findings showed that in conditions that featured concurring outcomes, responses were slightly slower as compared to the nonword conditions. Although only one of these conditions actually slowed down responses significantly (*All Congruent*), the general pattern indicated that the extra processing associated with an additional stimulus dimension can slow down the response. This finding is consistent with electrophysiological studies that showed increased anterior cingulate cortex activation, indicative of conflict detection, in both incongruent and congruent Stroop words, as compared to a nonword condition, even though there was no evidence for interference in congruent trials in behavioural measures (Bench et al., 1993; Carter, Mintun, & Cohen, 1995). Thus, the mere presence of an additional meaningful dimension might produce some interference, regardless of its semantic relationship to task-relevant dimensions.

Future analysis of electrophysiological data collected in my study might provide further evidence with regard to the presence or absence of conflict related brain activity in conditions that did not show evidence for interference in behavioural performance.

In my study, I used a nonword written stimulus as a reference point for measuring interference and facilitation effects. Brown (2011) proposed that selecting that type of stimulus as reference confounds the comparison of interference and facilitation effects. This is because the total interference from incongruent words, as compared to nonwords, is composed of a *lexicality effect*, i.e., interference from the written stimulus being a word (regardless of its meaning), and a *congruency effect*, i.e., interference from the written word being semantically incongruent with font colour. While nonwords differ from congruent and incongruent colour-words both in lexicality and in congruency, colour-words differ from each other only in congruency. Brown suggested that using both a nonword as reference and a neutral meaningful word as reference might help delineate lexicality effects (the difference between nonwords and neutral words) and congruency effects (the difference between neutral words and colour-words).

Applying Brown's (2011) ideas to my results revealed that indeed, comparing "same" conditions using the neutral word *white* as reference, instead of a nonword, equalised congruency effects. The mean neutral-incongruent interference amounted to 41-ms and the mean neutral-congruent facilitation amounted to 47-ms. Furthermore, comparing neutral and nonword "same" conditions showed that the mean lexicality effect amounted to 54-miliseconds. This finding shows that a larger portion of the interference from incongruent colour-words, as compared to nonwords, could be attributed to lexicality. However, the neutral word selected for my experiment might have also confounded the comparison between interference and facilitation, because it was in itself an incongruent colour-word. Thus, it might reflect a biased reference point, because it belonged to the same semantic category as the other distracting colour-words (i.e., *red*, *green*, and *blue*), and therefore its neutrality was valid only within this task.

Comparison of “different” conditions poses some challenges to Brown’s ideas. Lexicality effects were essentially absent (e.g., 3-ms) when comparing the nonword “different” condition and the neutral “different” condition. Thus, apparently there were no processing costs associated with the lexicality of the irrelevant written word when a “different” response was required. Congruency effects cannot be directly measured from “different” conditions because none of them featured a congruency between task-relevant dimensions (and hence they all required a “different” response). However, the similarity in RT between the nonword and neutral word stimuli implies that with either reference point, the magnitude of interference and facilitation effects in “different” conditions would turn out to be essentially the same.

By incorporating other forms of neutral written stimuli, such as non-colour-words and orthographically legal nonwords (pseudowords), future audiovisual Stroop studies might further elucidate and characterise the factors that comprise facilitation and interference effects.

## 4.3 Conflict Additivity

The results of my study indicated that the magnitude of interference effects depended on the number of outcomes of the irrelevant matching tasks that conflicted with the response to the relevant matching task. That is, when both outcomes conflicted with the response, interference effects were larger than when only one outcome conflicted with the response. And, when one outcome conflicted with the response, interference effects were larger than when all outcomes agreed. An example from my results illustrates this graded interference. The delay in responding to the *Colour Congruent* condition, as compared to the nonword “same” condition (95-ms) was approximately the sum of the delays in responding to the *Word Congruent* condition (50-ms) and in responding to the *Visual Congruent* condition (33-ms) as compared to the nonword "different" condition. This example shows that interference effects were additive, such that the number of irrelevant outcomes that conflicted with the response determined the degree to which responses were delayed.

This pattern of findings is in line with the previously mentioned concept of outcome-conflict (Navon, 1985; Navon & Miller, 1987). According to Navon & Miller (1987), "Outcome conflicts may either degrade the performance of tasks that are processed in parallel or call for serial processing to avoid such degradation" (p. 435). In this conceptualisation, performance in the present task was determined by a conflict of outcomes between three separate processes, each one resulting in a binary "same" or "different" outcome. When the three processes concurred and produced similar outcomes (i.e., all were "same" or "different"; as was in the *All Congruent*, *All Incongruent*, and *Neutral Colour Incongruent* conditions), no delay in processing was observed as compared to the nonword condition. When one irrelevant matching outcome conflicted with the response (i.e., the response was "different", one irrelevant task was "same" and the other "different"; as was in the *Word Congruent* and *Visual Congruent* conditions), some delay in processing was observed as compared to the nonword condition. When both irrelevant matching outcomes conflicted with the response (i.e., the response was "same" and both irrelevant tasks were "different; as was in *Colour Congruent* and *Neutral Colour Congruent*), the largest delays in processing were observed, as compared to the nonword condition. Thus, in light of Navon and Miller's idea, as the number of outcome-conflicts became larger, performance was in general more prone to errors; and, to maintain higher accuracy, participants might have resorted to serial processing of separate comparisons, which in turn might have produced additional response delays.

To note, in the *Neutral Colour Congruent* condition, where both irrelevant matching outcomes conflicted with the response, even though the distraction was from the neutral written word *white*, response delays, as compared to the nonword condition, were greater than in the *Visual Congruent* and in the *Word Congruent* conditions. This finding indicated that interference effects were driven more by the number of irrelevant outcomes that conflicted with the response than by whether or not the irrelevant word was also represented in the task-relevant stimulus set.

## 4.4 Stimulus Set Membership

Comparing the *Colour Congruent* (e.g., hear /red/ and see *blue* in red) and the *Neutral Colour Congruent* (e.g., hear /red/ and see *white* in red) conditions allowed evaluating how interference was affected by whether or not the written word was also a member in the task-relevant stimulus set. Both conditions required a "same" response, but in *Colour Congruent* the written word was from the same set of colours presented in the font colour and the spoken word dimensions (i.e., red, green and blue), and in *Neutral Colour Congruent* the written word was the neutral word *white*. This comparison showed that a written word that belonged to the same set of colours significantly decreased accuracy, but did not slow down responses, as compared to a neutral colour word. This finding indicated that the processing costs associated with semantic relatedness did not manifest in additional delays, but in an increase of the likelihood of responding incorrectly. The distinction between accuracy and RT measures in two conditions that featured the same degree of outcome-conflict (i.e., in both conditions, two "different" outcomes of the irrelevant matching task conflicted with a "same" response), but a different degree of semantic-conflict, might reflect on a qualitative difference between the two types of conflict. That is, while outcome-conflicts mainly slowed down responses, an additional semantic-conflict was expressed also in decreased accuracy. This interpretation, however, does not align with the finding that "different" responses were equally unaffected by a distractor from the same stimulus set and a neutral word distractor, as exemplified by the comparison between the *All Incongruent* (e.g., hear /red/ and see *green* in blue) and *Neutral Colour Incongruent* (e.g., hear /red/ and see *white* in blue) conditions.

Semantic gradients have been consistently reported in the Stroop task literature (e.g., Klein, 1964). However, they were not documented in an audiovisual Stroop task where spoken words served as distractors (Cowan & Barron, 1987). They found that spoken colour-words interfered with naming the font colour of written colour-words, but spoken non-colour-words did not, and therefore proposed that interference to font colour naming was sensitive to the phonemic or semantic similarity of the distractor to the target stimulus.

The comparable effects on response delays of a colour-word and a neutral colour-word distractor, along with the greater interference to accuracy from colour-words might suggest that additional time is required to overcome outcome-conflicts regardless of the semantic relatedness of the targets and distractor. However, once that conflict is resolved, there might be an additional cost to semantic relatedness that is expressed in accuracy deficits. Also, the finding that a colour-word did not significantly delay responses more than a neutral colour-word might stem from the fact that both words belonged to the same semantic category (i.e., colours). In this context, future investigations should include written non-colour-words as distractors to further elucidate the effects of semantic similarity on interference.

## 4.5 Which Congruent Information Interferes More with a "Different" Response?

By comparing the effects of the *Visual Congruent* (e.g., hear /red/ and see *blue* written in blue font) and *Word Congruent* (e.g., hear /red/ and see *red* written in blue font) conditions, I intended to test if interference was differently affected by whether the irrelevant word matched the spoken word or the font colour. These conditions required a “different” response, but in the *Visual Congruent* condition the written word matched the font colour and in the *Word Congruent* condition the written word matched the spoken word. Thus, in one condition, congruent information from the same visual object interfered with a “different” response, while in the other condition, congruent information from auditorily and visually presented words interfered with that response. The results showed that even though the *Word Congruent* condition slowed down responses more than the *Visual Congruent* condition (by 17-ms), and produced more errors (by 1 %), these differences were not statistically significant. This finding indicated that responding “different” was not differently affected by a match between the written word and font colour or by a match between the written word and the spoken word. Thus, attention to the task-relevant information (i.e., font colour and spoken word) was not shown to be differentially diverted by distraction from the same visual object (written word meaning matching font colour) or from separate objects that shared a representational code (written word meaning matching spoken word meaning).

Even though there was no evidence that interference effects between these conditions differed when directly compared, as compared to the nonword condition, both conditions significantly delayed responses but only the *Word Congruent* condition also significantly decreased accuracy. Thus, activating a semantic representation of the font colour dimension delayed processing to a similar degree regardless of which task-relevant dimension was congruent with the written word; however, once this representation was available for comparison with the spoken word, this comparison was more likely to result in an error when the written word matched the spoken word.

I expected that congruence between the written word and spoken word would interfere with making a “different” judgement more than congruence between the written word and font colour. Because attention in the present task was directed to both modalities, I expected that a conflicting “same” judgement that was made based on information presented cross-modally would be more distracting than a conflicting “same” judgement that was made based on information present within the written word. Furthermore, Dyer (1973) proposed that in matching of words and colours, the task-relevant colour is transformed to a form that is closer to a word to allow comparison with the task-relevant word, and not vice versa. In the present experiment, the distracting stimulus was a word. Therefore, I expected that erroneously matching the written word and the spoken word would be harder to inhibit than erroneously matching the written word and the font colour. This idea is also related to Roelofs’s (2005) *word production architecture* account of the colour-word Stroop asymmetry. This account asserts that words interfere with colour naming but colours do not interfere with reading because words access pronunciation before meaning whereas the reverse is true for colours. In the present experiment, although a verbal response was not required, I expected the mutual code that spoken and written words share to present a harder-to-inhibit source of interference.

Dyer (1973) showed that in “different” conditions, as compared to a control condition, responses were slowed down the most by congruence between the irrelevant word meaning and relevant font colour (parallels the *Visual Congruent* condition in my study), less by a total mismatch among the dimensions (i.e., *All Incongruent*), and even less by congruence between the irrelevant word and the reference relevant word (i.e., *Word Congruent*). Goldfarb and Henik (2006) showed that “different” responses were slowed down more by congruence between the written word and its font colour, and less by congruence between the font colour and bar colour, which did not differ from a total mismatch among the dimensions. In both studies, interference effects to “different” responses were largest when the information within the same visual object, by indicating “same”, conflicted with the response. Goldfarb and Henik proposed that because attention is drawn to the dimensions of the same object, interference effects are larger when the conflicting information is featured in the same object. The pattern of interference effects found in those studies is discrepant with that obtained in my study, because their effects seemed to be governed by the source of the conflicting information rather than by the presence or absence of outcome-conflicts. Possibly then, when attention is focussed on visual targets, there is more interference from conflicts within the same object than interference from conflicts between physically separate stimuli, regardless of the presence of outcome-conflicts. However, when attention is divided between auditory and visual targets, the relationship between the dimensions of the same visual object becomes less distracting, and the relationships among all representations become equally distracting. Then, interference emerges when comparisons between representations produce conflicting outcomes, but no interference results when all outcomes concur.

The distinction between the results of my study and visual matching Stroop studies (e.g., Dyer, 1973; Goldfarb & Henik, 2006) indicates that the nature of the irrelevant information might play a part in shaping the pattern of interference effects. A question that arises in this context then, is if and how this pattern changes when task-relevant dimensions are altered. In my study, to maximise the effects and to more closely resemble the classic Stroop task, font colour was chosen as a task-relevant dimension and written word as a task-irrelevant dimension. By varying the task-relevant dimensions to include matching between written and spoken words (irrelevant font colours) and between written words and font colours (irrelevant spoken words), future audiovisual Stroop tasks might reveal whether attending to stimulus dimensions within the same modality or across modalities shows a different pattern of interference.

### 4.6 "Same" versus "Different" Judgements

The results of my study indicated that without written distraction (i.e., a nonword), "same" responses were significantly faster than "different" responses. Accuracy, however, was not differently affected by the two types of response. These results justified the use of a separate reference point to evaluate the effects of the irrelevant written word on making each type of judgement, at least for RT measures. Faster "same" than "different" responses can be interpreted as the effect of the semantic congruence between the task-relevant stimuli. However, the finding that responses in fact slowed down by the presence of a semantically congruent written word (i.e., in the *All Congruent* condition) might show that the effects of congruence operate differently depending on task-relevance.

Typically, "same" - "different" experiments show a *fast-"same"* effect regardless of the level of processing required to make the judgement. That is, this effect has been documented in matching of stimuli based on physical criteria such as shape and colour (e.g., Egeth, 1966; Williams, 1974), a letter order criterion (Bamber, 1969), and a nominal criterion such as letter name (e.g., Eviatar et al., 1994; Proctor, 1981). The apparent ubiquity of the fast-"same" effect has led researchers to propose that "same" and "different" judgements stem from different processes. While "same" judgements are based on holistic processing, "different" judgements are more analytic in nature. Furthermore, the two judgement types were shown to be differently affected by the number of irrelevant dimensions (e.g., Miller & Bauer, 1981). As the number of irrelevant dimensions increases, RTs to "same" judgements rises monotonically while "different" judgements are minimally affected (for a review see Farrell, 1985). This pattern has been associated with the characterisation of "same" judgements as *exhaustive* and of "different" judgements as *self-terminating*. In simple terms, judging two objects to be the "same" requires them to be identical on all relevant dimensions, while "different" judgements can be made as soon as one difference is detected. Given that irrelevant information cannot be completely ignored, less of it will be processed in a self-terminating "different" judgement than in an exhaustive "same" judgement. This explanation fits well with the relative insensitivity of "different" judgements and with the susceptibility of "same" judgements to the number of irrelevant judgements.

In my study, the written word was the only irrelevant dimension; however, it created conflicts with neither of, either of, or both of the relevant dimensions. In "same" conditions, the effect of the written word conflicting with both task-relevant dimensions (i.e., in *Colour Congruent*) greatly slowed down the response (by 95-ms), while the effect of the written word conflicting with neither of the task-relevant dimensions slowed it down only minimally (by 6-ms). In "different" conditions, the effect of the written word conflicting with either task-relevant dimension (i.e., in *Visual Congruent* and *Word Congruent*) slowed down the response to some degree (by 33-ms for font colour – written word congruence and by 50-ms for spoken word – written word congruence). The effect of the written word conflicting with both task-relevant dimensions (i.e., in *All Incongruent*) slowed it down only minimally (by 10-ms). These comparisons indicate that the written word differently affected each type of response, because the number of conflicts that the written word created with the relevant dimensions increasingly affected "same" responses but decreasingly affected "different" responses. This pattern of findings is consistent with the notion that "same" and "different" responses are differently affected by the presence of irrelevant information (e.g., Miller & Bauer, 1981). The demonstration of this pattern in the present task suggests that the distinction between “same” and “different” responses is preserved in crossmodal judgements that require semantic access.

## 4.7 Practice Effects

The results of my study indicated that RTs decreased over the course of the experiment, while no effects on accuracy were documented. Furthermore, the improvement in RTs over the course of the experiment was independent of condition, showing that participants were faster to respond regardless of the difficulty of a particular condition. This finding indicated that participants in general responded more rapidly over the course of the experiment, but not necessarily due to an improved ability to inhibit the effects of distraction. The comparison of adjacent and evenly separated trial-blocks indicated that participants’ improvement in performance became smaller as the experiment progressed.

# 5 Conclusions

The results of my study demonstrated that an audiovisual matching Stroop task exclusively produced interference effects due to a conflict of outcomes between a relevant matching task and two irrelevant matching tasks. By being obligatorily processed, the meaning of the written word was mistakenly compared with the task-relevant dimensions–font colour and spoken colour word. When the outcomes of these comparisons conflicted with the outcome of the relevant matching task, performance degraded. Conversely, when the outcomes of the irrelevant comparisons concurred with the outcome of the relevant matching task, there was no evidence that performance benefited. Thus, the perception of conflict was not facilitated by additional conflicting information, and the perception of non-conflict was not facilitated by additional non-conflicting information. Furthermore, my study provided evidence that semantic-conflict, by itself, did not interfere with the accuracy or speed of behavioural responses.

My results also showed that interference effects due to outcome-conflicts were additive. That is, as the number of outcome-conflicts between the relevant and irrelevant matching tasks became larger, so did the magnitude of interference effects. This finding might indicate that with a larger number of outcome-conflicts, more stimulus pairs needed to be processed serially rather than in parallel.

Interference effects related to the semantic relatedness of the distractor were documented in accuracy results but not in RT results. The distinction between the two measures might indicate that processing costs associated with semantic gradients are not expressed in increased processing time but in an increased likelihood of responding incorrectly.

Ultimately, I found no evidence that interference effects were greater due to a need to inhibit information that could be matched at a lexical level (words) than information that had to be matched at a semantic level (written word and colour). This finding indicated that the semantic representations derived from the three stimulus dimensions were activated regardless of the processing level at which the distracting dimension (i.e., written word) and task-relevant dimensions (i.e., font colour and spoken word) could be compared.

## References


Agresti, A. (2002). *Categorical data analysis* (2nd edition). New York, NY: John Wiley & Sons.

Appelbaum, L. G., Donohue, S. E., Park, C. J., & Woldorff, M. G. (2013). Is one enough? the case for non-additive influences of visual features on crossmodal stroop interference. *Frontiers in Psychology,* 4:799. doi: 10.3389/fpsyg.2013.00799

Bamber, D. (1969). Reaction times and error rates for "same"-"different" judgments of multidimensional stimuli. *Perception & Psychophysics, 6*(3), 169-174. doi: 10.3758/BF03210087

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.1-7). Retrieved from http://CRAN.R-project.org/package=lme4.

Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowiak, R. S. J., & Dolan, R. J. (1993). Investigations of the functional anatomy of attention using the stroop test. *Neuropsychologia, 31*(9), 907-922. doi: 10.1016/0028-3932(93)90147-R

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*(421), 9-25. doi: 10.1080/01621459.1993.10594284

Bretz, F., Hothorn, T., & Westfall, P. (2010). *Multiple comparisons using R*. CRC Press. Boca Raton.

Brown, T. L. (2011). The relationship between stroop interference and facilitation effects: Statistical artifacts, baselines, and a reassessment. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 85-99. doi:10.1037/a0019252

Caldas, A., Machado-Pinheiro, W., Souza, L., Motta-Ribeiro, G., & David, I. (2012). The stroop matching task presents conflict at both the response and nonresponse levels: An event-related potential and electromyography study. *Psychophysiology, 49*(9), 1215-1224. doi: 10.1111/j.1469-8986.2012.01407.x

Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An $H_2{}^{15}O$ PET study of stroop task performance. *Neuroimage, 2*(4), 264-272. doi: 10.1006/nimg.1995.1034

Cohen, M., S. (2008). Handedness Questionnaire. [On-line questionnaire]. Retrieved September 7, 2013 from: http://www.brainmapping.org/shared/Edinburgh.php

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the stroop effect. *Psychological Review, 97*(3), 332-361.

Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual stroop interference and possible implications for speech memory. *Attention, Perception, & Psychophysics, 41*(5), 393-401. doi: 10.3758/BF03203031

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*(4), 447-456. doi: 10.1016/j.jml.2007.11.004

Donohue, S. E., Appelbaum, L. G., Park, C. J., Roberts, K. C., & Woldorff, M. G. (2013). Cross-modal stimulus conflict: The behavioral effects of stimulus input timing in a visual-auditory stroop task. *PloS One, 8*(4), e62802. doi: 10.1371/journal.pone.0062802

Dyer, F. N. (1973). Same and different judgments for word-color pairs with "irrelevant" words or colors: Evidence for word-code comparisons. *Journal of Experimental Psychology, 98*(1), 102-108. doi: 10.1037/h0034278

Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, *1*(4), 245-252. doi: 10.3758/BF03207389

Elliott, E. M., Cowan, N., & Valle-Inclan, F. (1998). The nature of cross-modal color-word interference effects. *Attention, Perception, & Psychophysics*, *60*(5), 761-767. doi: 10.3758/BF03206061

Elliott, E. M., Morey, C. C., Morey, R. D., Eaves, S. D., Shelton, J. T., & Lutfi-Proctor, D. A. (2014). The role of modality: Auditory and visual distractors in stroop interference. *Journal of Cognitive Psychology*, *26*(1), 15-26. doi: 10.1080/20445911.2013.859133

Eviatar, Z., Zaidel, E., & Wickens, T. (1994). Nominal and physical decision criteria insame-different judgments. *Perception & Psychophysics*, *56*(1), 62-72. doi: 10.3758/BF03211691

Farell, B. (1985). "Same"–"different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, *98*(3), 419-456. doi: 10.1037/0033-2909.98.3.419

Fox, J., & Weisberg, S. (2011). An R Companion to Applied Regression, 2nd Edition. Thousand Oaks CA: Sage. Retrieved July 12, 2014 from: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*(1), 13-42. doi: 10.1037/0096-3445.118.1.13

Goldfarb, L., & Henik, A. (2006). New data analysis of the stroop matching task calls for a reevaluation of theory. *Psychological Science, 17*(2), 96-100. doi: 10.1111/j.1467-9280.2006.01670.x

Hanauer, J. B., & Brooks, P. J. (2003). Developmental change in the cross-modal stroop effect. *Attention, Perception, & Psychophysics*, *65*(3), 359-366. doi: 10.3758/BF03194567

Heathcote, A., Popiel, S. J., & Mewhort, D. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109*(2), 340-347. doi: 10.1037/0033-2909.109.2.340

Hintzman, D. L., Carre, F. A., Eskridge, V. L., Owens, A. M., Shaff, S. S., & Sparks, M. E. (1972). "Stroop" effect: Input or output phenomenon? *Journal of Experimental Psychology, 95*(2), 458-459. doi: 10.1037/h0033644

Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology, 83*(2, pt.1), 299-303. doi: 10.1037/h0028512

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346-363. doi: 10.1002/bimj.200810425

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446. doi: 10.1016/j.jml.2007.11.007

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental Organization*. Oxford University Press.

Logan, G. D. (1980). Attention and automaticity in stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*(4), 523-553. doi: 10.1016/0010-0285(80)90019-5

Luo, C. R. (1999). Semantic competition as the basis of stroop interference: Evidence from color-word matching tasks. *Psychological Science, 10*(1), 35-40. doi: 10.1111/1467-9280.00103

Machado-Pinheiro, W., Volchan, E., Vila, J., Dias, E. C., Alfradique, I., Oliveira, L. d., . . . David, I. A. (2010). Role of attention and translation in conflict resolution: Implications for stroop matching task interference. *Psychology & Neuroscience, 3*(2), 141-150. doi: 10.3922/j.psns.2010.2.003

MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203. doi: 10.1037/0033-2909.109.2.163

MacLeod, C. M. (1992). The stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General, 121*(1), 12-14. doi: 10.1037/0096-3445.121.1.12

MacLeod, C. M., & Dunbar, K. (1988). Training and stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 126-135. doi: 10.1037/0278-7393.14.1.126

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*(10), 383-391. doi: 10.1016/S1364-6613(00)01530-8

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language and Hearing Research, 50*(4), 940. doi: 10.1044/1092-4388(2007/067)

Mascolo, M. F., & Hirtle, S. C. (1990). Verbal coding and the elimination of stroop interference in a matching task. *The American Journal of Psychology*, *103*(2), 195-215. Retrieved May 30, 2013 from http://www.jstor.org/stable/1423142

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*(3), 422- 471. doi: 10.1037/0033-295X.110.3.422

Miles, C., & Jones, D. M. (1989). The fallacy of the cross-modal stroop effect: A rejoinder to cowan (1989). *Attention, Perception, & Psychophysics, 45*(1), 85-86. doi: 10.3758/BF03208038

Miles, C., Madden, C., & Jones, D. M. (1989). Cross-modal, auditory-visual stroop interference: A reply to cowan and barron (1987). *Perception & Psychophysics, 45*(1), 77-81. doi: 10.3758/BF03208036

Miller, J., & Bauer, D. W. (1981). Irrelevant differences in the “same”— “different” task. *Journal of Experimental Psychology: Human Perception & Performance, 7*(1), 196-207. doi: 10.1037/0096-1523.7.1.196

Morton, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology, 60*(3), 329-346. doi: 10.1111/j.2044-8295.1969.tb01204.x

Morton, J., & Chambers, S. M. (1973). Selective attention to words and colours. *The Quarterly Journal of Experimental Psychology, 25*(3), 387-397. doi: 10.1080/14640747308400360

Navon, D. (1985). Attention division or attention sharing? In M. I. Posner & O. S. M. Marin (Eds.), Attention and performance XI (pp. 133-146). Hillsdale, NJ: Erlbaum.

Navon, D., & Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 13*(3), 435-448. doi: 10.1037/0096-1523.13.3.435

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*(1), 97-113. doi: 10.1016/0028-3932(71)90067-4

Paulesu, E., Harrison, J., Baron-Cohen, S., Watson, J. D. G, Goldstein, L., Heather, J., Frackowiak, R. S. J., & Frith, C. D. (1995). The physiology of coloured hearing. A PET activation study of colour-word synaesthesia. *Brain: A Journal of Neurology,118 (3),* 661-676. doi: 10.1093/brain/118.3.661

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), Information processing and cognition: The Loyola symposium (pp. 55–85). Hillsdale, NJ: Erlbaum

Proctor, R. W. (1981). A unified theory for matching-task phenomena. *Psychological Review, 88*(4), 291-326. doi: 10.1037/0033-295X.88.4.291

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved May 20, 2014 from: http://www.R-project.org/

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review, 92*(2), 212-225. doi: 10.1037/0033-295X.92.2.212

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510-532. doi: 10.1037/0033-2909.114.3.510

Ratcliff, R. (1987). More on the speed and accuracy of positive and negative responses. *Psychological Review, 94*(2), 277-280. doi: 10.1037/0033-295X.94.2.277

Ratcliff, R. (2012). Response Time Distributions. In H. E. Cooper, (Editor-in-chief.), *APA Handbook of Research Methods in Psychology, vol 1: Foundations, planning, measures, and psychometrics.* (pp. 429-443). American Psychological Association. doi: 10.1037/13619-023

Roelofs, A. (2005). The visual-auditory color-word stroop asymmetry and its time course. *Memory & Cognition, 33*(8), 1325-1336. doi: 10.3758/BF03193365

Seymour, P. H. (1977). Conceptual encoding and locus of the stroop effect. *The Quarterly Journal of Experimental Psychology, 29*(2), 245-265. doi: 10.1080/14640747708400601

Shimada, H. (1990). Effect of auditory presentation of words on color naming: The intermodal Stroop effect. *Perceptual and Motor Skills, 70*(3c), 1155-1161. doi: 10.2466/pms.1990.70.3c.1155

Simon, J. R., & Berbaum, K. (1988). Effect of irrelevant information on retrieval time for relevant information. *Acta Psychologica, 67*(1), 33-57. doi: 10.1016/0001-6918(88)90023-6

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662. doi: 10.1037/h0054651

Sugg, M. J., & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 647-675. doi: 10.1037/0096-1523.20.3.647

Tecce, J. J., & Happ, S. J. (1964). Effects of shock-arousal on a card-sorting test of color-word interference. *Perceptual and Motor Skills, 19*(3), 905-906. doi: 10.2466/pms.1964.19.3.905

Treisman, A., & Fearnley, S. (1969). The stroop test: Selective attention to colours and words. *Nature, 222*, 437-439. doi: 10.1038/222437a0

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123*(1), 34-80. doi: 10.1037/0096-3445.123.1.34

Van Zandt, T. (2002). Analysis of response time distributions. In J. T. Wixted (Vol. Ed.) & H.Pashler (Series Ed.), *Stevens' handbook of experimental psychology: Vol.4.*

*Methodology in experimental psychology* (3rd ed., pp. 461-516). New York: Wiley. doi: 10.1002/0471214426.pas0412

Van Zandt, T., & Townsend, J. T. (2012). Designs for and analyses of response time experiments. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods: Vol.1. Foundations* (pp. 260-285). New York: Oxford University Press.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of stroop interference. *Memory & Cognition, 13*(4), 304-319. doi: 10.3758/BF03202499

Williams, C. (1974). The effect of an irrelevant dimension on “same-different” judgements of multi-dimensional stimuli. *The Quarterly Journal of Experimental Psychology, 26*(1), 26-31. doi: 10.1080/14640747408400384

Zhang, R., Hu, Z., Roberson, D., Zhang, L., Li, H., Liu, Q. (2013) Neural Processes Underlying the “Same”-“Different” Judgment of Two Simultaneously Presented Objects- An EEG Study. *PLoS ONE, 8*(12): e81737. doi:10.1371/journal.pone.0081737

Zysset, S., Müller, K., Lohmann, G., & von Cramon, D. Y. (2001). Color-word matching stroop task: Separating interference and response conflict. *Neuroimage, 13*(1), 29-36. doi: 10.1006/nimg.2000.0665

# Appendices

## Appendix A: List of Exclusion Criteria for Participation

**Sensory Impairments:**

Impaired vision (unless corrected using contact lenses or glasses)

Colour blindness

Impaired hearing

**Cognitive Challenges/Impairments:**

| | |
|---|---|
| ADD/ADHD | Autism |
| Depression | Fetal Alcohol Syndrome |
| Psychosis | Dyslexia |
| Down’s syndrome | Cerebral Palsy |
| Schizophrenia | Antisocial Personality Disorder or Conduct Disorder |
| Claustrophobia | Motor control disorder |
| Tourette syndrome | Concussion (recent or with long-lasting effects) |
| Agnosia | Aphasia |
| Brain tumour | Epilepsy |
| Speech/language impairments | Perception/production impairments |
| Traumatic brain injury | |

**Drugs/Medications:**

| | |
|---|---|
| Anti-depressants – Prozac, celexa, etc | Narcotics – codeine, morphine |
| Anti-anxiety – Paxil, etc | Barbiturates/sedatives – Valium, Ketamine |
| Anti-psychotics – Lithium, etc | Depressants – Alcohol |
| Hallucinogens – LSD, Magic Mushrooms, peyote | Marijuana |
| Stimulants – Cocaine, ephedrine, Ritalin, etc | Other mind altering substances |

Known skin reactions/sensitivity to standard surgical tape (Micropore)

## Appendix B: Written Instructions Presented to Participants

On this screen, in each trial you will see the word “red”, “green”, “blue” or “white” or the string of symbols “####”, written in red, green, or blue font colour. At the same time, you will hear a colour-word (red, green, or blue) through the earphones.

Using the index and middle finger of your dominant hand,

Press “YES” (left/right*-arrow key)

when the word you hear and the text colour match.

Press “NO” (right/left*-arrow key)

when the word you hear and the text colour do not match.

PLEASE IGNORE THE MEANING OF THE WRITTEN WORD

PLEASE BE AS QUICK AND AS ACCURATE AS YOU CAN

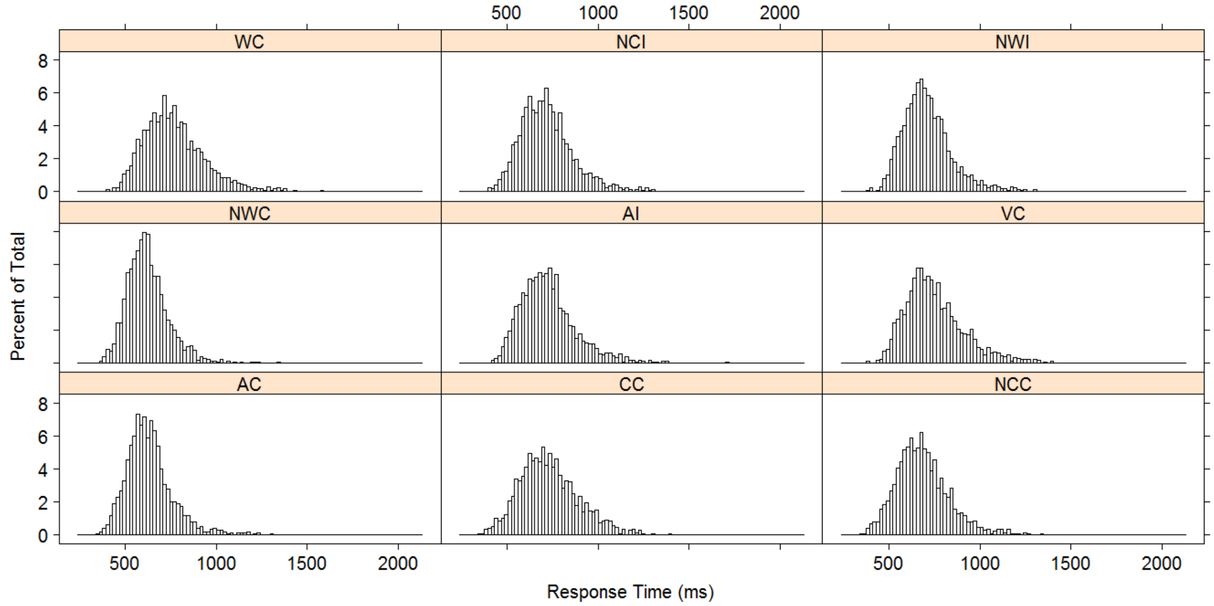
PLEASE BLINK ONLY WHEN YOU SEE (--)

Press “YES” or “NO” to continue

*response keys were interchanged for half the participants

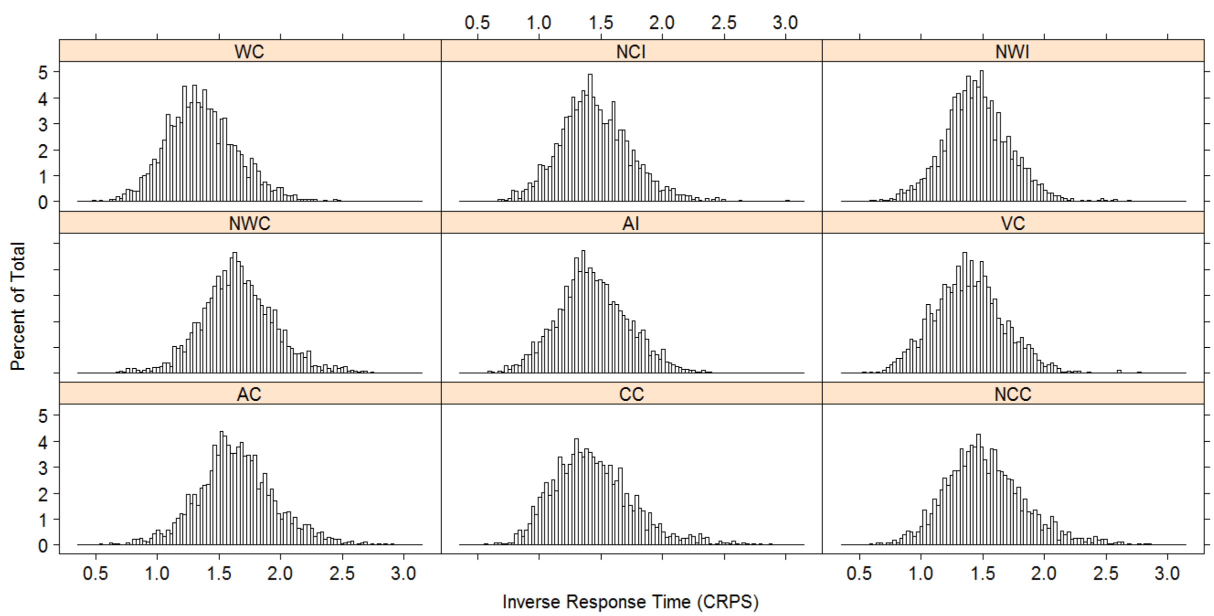
## Appendix C:

### C.1 Histograms of RTs by *Condition*




Note. AC = All Congruent, AI = All Incongruent, CC = Colour Congruent, NCC = Neutral Colour Congruent, NCI = Neutral Colour Incongruent, NWC = Nonword Congruent, NWI = Nonword Incongruent, VC = Visual Congruent, WC = Word Congruent.

### C.2 Histograms of inverse transformed RTs by *Condition*




AC = All Congruent, AI = All Incongruent, CC = Colour Congruent, NCC = Neutral Colour Congruent, NCI = Neutral Colour Incongruent, NWC = Nonword Congruent, NWI = Nonword Incongruent, VC = Visual Congruent, WC = Word Congruent. CRPS = Correct Responses per Second.

## Appendix D:

### D.1 Comparison of regression models fitted to accuracy data with *Stimulus Combination* as fixed factor

| | Model Formula | *df* | Deviance | $\chi^2$ | $\chi^2$ *df* | *p* ( > $\chi^2$) |
|---|---|---|---|---|---|---|
| Model 1: | Accuracy ~ Block + Response Mapping + Stimulus Combination + (1 \| Participant) | 45 | 6061.5 | | | |
| Model 2: | Accuracy ~ Block + Response Mapping + Stimulus Combination + (1 + Condition \| Participant) | 80 | 5957.9 | 103.583 | 35 | <.001 |
| Model 3: | Accuracy ~ Block + Response Mapping + Stimulus Combination + (0 + Condition \| Participant) | 89 | 5868.1 | 89.837 | 9 | <.001 |

Note. The outcome variable precedes the tilde operator and fixed factor terms follow. Random factor terms in parentheses. See text for additional description of the models. df = Degrees of Freedom.

### D.2 Comparison of regression models fitted to accuracy data with *Condition* as fixed factor

| | Model Formula | *df* | Deviance | $\chi^2$ | $\chi^2$ *df* | *p* ( > $\chi^2$) |
|---|---|---|---|---|---|---|
| Model 1: | Accuracy ~ Block + Response Mapping + Condition + Block * Condition + Response Mapping * Condition + (1 \| Participant) | 65 | 5989.0 | | | |
| Model 2: | Accuracy ~ Block + Response Mapping + Condition + Block * Condition + Response Mapping * Condition + (1 + Condition \| Participant) | 100 | 5925.4 | 63.577 | 35 | <.001 |
| Model 3: | Accuracy ~ Block + Response Mapping + Condition + Block * Condition + Response Mapping * Condition + (0 + Condition \| Participant) | 109 | 5869.3 | 56.112 | 9 | <.001 |

Note. The outcome variable precedes the tilde operator and fixed factor terms follow. Random factor terms in parentheses. See text for additional description of the models. df = Degrees of Freedom.

## Appendix E:

### E.1 Comparison of regression models fitted to RT data with *Stimulus Combination* as fixed factor

| | Model Formula | *df* | Deviance | $\chi^2$ | $\chi^2$ *df* | *p* ( > $\chi^2$) |
|---|---|---|---|---|---|---|
| Model 1: | Inverse RT ~ Block + Response Mapping + Stimulus Combination + (1 \| Participant) | 46 | 2521.8 | | | |
| Model 2: | Inverse RT ~ Block + Response Mapping + Stimulus Combination + (1 + Condition \| Participant) | 81 | 1425.4 | 1096.4 | 35 | <.001 |
| Model 3: | Inverse RT ~ Block + Response Mapping + Stimulus Combination + (0 + Condition \| Participant) | 90 | 1063.1 | 362.3 | 9 | <.001 |

Note. The outcome variable precedes the tilde operator and fixed factor terms follow. Random factor terms in parentheses. See text for additional description of the models. df = Degrees of Freedom.

### E.2 Comparison of regression models fitted to RT data with *Condition* as fixed factor

| | Model Formula | *df* | Deviance | $\chi^2$ | $\chi^2$ *df* | *p* ( > $\chi^2$) |
|---|---|---|---|---|---|---|
| Model 1: | Inverse RT ~ Response Mapping + Block + Condition + Response Mapping*Condition + Block*Condition + (1 \| Participant) | 66 | 3290.9 | | | |
| Model 2: | Inverse RT ~ Response Mapping + Block + Condition + Response Mapping*Condition + Block*Condition + (1 + Condition \| Participant) | 101 | 3160.0 | 130.89 | 35 | <.001 |
| Model 3: | Inverse RT ~ Response Mapping + Block + Condition + Response Mapping*Condition + Block*Condition + (0 + Condition \| Participant) | 110 | 1224.3 | 1935.72 | 9 | <.001 |

Note. The outcome variable precedes the tilde operator and fixed factor terms follow. Random factor terms in parentheses. See text for additional description of the models. df = Degrees of Freedom.