The influence of red stimuli on cognitive performance in achievement context settings

Adam Pedley

Thesis submitted for the award of Doctor of Philosophy

June 2016

School of Psychology

UNIVERSITY OF SURREY
Abstract

The influence of colour on cognition has a long and varied history in experimental psychology. Research areas include the influence of colour on; mate selection; food evaluation; mood; sporting prowess; and cognitive performance. However, few reliable colour based effects have emerged. This thesis examines a recently proposed effect in the domain of colour and cognitive performance; the “Red Achievement-Context Cognitive Effect” (RACE). This effect, proposed by Elliot et al., (2007), proposes that in an achievement context (i.e. one with a pass or fail outcome) red stimuli evoke avoidance motivation that, mediated through local processing, impedes cognitive ability. Although the RACE is recognised in the research literature and is underpinned by established theory, few direct replications of the RACE have been published and, applying criteria proposed by Pashler and Harris (2012), the literature that describes the RACE is susceptible to being skewed by publication bias. Additionally, there is a debate regarding whether the RACE emerges in both genders. This thesis contributes to the literature by addressing four aims; assessing the RACE in an applied context (aim 1); investigating the gender discrepancy within the literature (aim 2); providing independent, direct replications of the RACE in addition to conceptual replications (aim 3); and conducting a meta-analysis of published and grey literature (aim 4). A wide range of research methods were implemented including primary research using experimental methods in applied, online and laboratory settings and secondary research methods including meta-analysis and observational analysis of existing data sets. This thesis concludes that the RACE is a fragile effect that only emerges in very specific experimental settings and, even when the RACE does emerge, direct replications using identical methods do not always reproduce the effect. Whilst the RACE may manifest in controlled laboratory settings, it is unlikely to have any influence in applied settings.
Declaration of Originality

This thesis and the work to which it refers are the results of my own efforts. Any ideas, data, images or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to their originator in the text, bibliography or in footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the University has the right to submit my work to the plagiarism detection service Turnitin, UK for originality checks. Whether or not drafts have been so-assessed, the University reserves the right to require an electronic version of the final document (as submitted) for assessment as above.

Signed: [Signature]

Date: 01/06/16
Acknowledgements

I would like to thank:

My doctoral supervisors Dr. Alexandra Grandison and Dr. Paul Sowden for their advice, support and hard work throughout this process.

Prof. Jane Ogden for overseeing the doctoral programme.

Prof. Andrew Elliot for his support and advice.

Mr. Nigel Woodger for his expertise and patience in producing materials.

Mr. Andrew Barnes for his help regarding psychometric measures and other equipment.

Dr. Adrian Banks and Dr. Rachel Avery for their insights and advice.

Dr. Henriette Hogh for her help in research methods, statistics and professional development.

Mrs. Natalie Astle and Miss. Leonie Smikle for their help with participant recruitment.

Miss. Fatima Khan, Miss. Amarah Anthony and Miss. Faiza Imran for their help in collecting data.

Prof. Alex Wade, Miss. Serena Simmons, Dr. Maria Kontogianni and Dr. Ben Oldfield for their guidance towards a PhD in experimental psychology.

Prof. Alan Palmer and Dr. Padraig Kitterick for their support in writing up the thesis.
# Contents

Abstract 2

Declaration of originality 3

Acknowledgements 4

Contents 5

List of figures 21

List of tables 22

Chapter 1: General introduction to colour and cognition 26

1.1. Overview 26

1.2. The science of colour vision 26

   1.2.1. What is colour? 26

   1.2.2. How do we see colour? 28

   1.2.3. Why do we see colour? 28

   1.2.4. Could colour influence cognition? 30

1.2.5. Pre-21st century empirical colour research

1.2.6. Contemporary analysis of pre-21st century empirical colour research 12
1.2.7. Summary of pre-21st century empirical colour research 12

1.2.8. Contemporary colour research 12

1.2.8.1. Colour and evaluation 12

1.2.8.2. Colour and attraction 12

1.2.8.2.1. Males evaluating females 12

1.2.8.2.2. Females evaluating males 12

1.2.8.2.3. Non-heterosexual mate evaluations 12

1.2.8.2.4. Colour and attraction summary 12

1.2.8.3. Colour and sporting performance 12

1.2.8.3.1. Colour and sporting performance summary 12

Chapter 2: The influence of red stimuli on cognitive performance 32

2.1. Introduction to the 'Red Achievement-context Cognitive Effect' (RACE) 32

2.1.1. Understanding the field of achievement motivation 33

2.1.2. The development of Elliot et al., (2007) 34

2.2. The theory that underpins the RACE 36
2.2.1. Does the perception of red stimuli in an achievement context induce avoidance motivation? 36

2.2.2. Does avoidance motivation impair cognitive performance? 39

2.2.3. How does colour induced avoidance motivation impair cognitive performance? 41

2.2.4. The Colour-in-Context Theory 45

2.3. Empirical evidence of the RACE 49

2.3.1. Non-independent replications and extensions of the RACE 49

2.3.2. Independent replications and extensions of the RACE 51

2.4. The need for further research 82

2.4.1. The need to assess the RACE in an applied context 82

2.4.2. The need to investigate if the RACE varies between genders 84

2.4.3. The need for replication 85

2.4.3.1. Why is replication important? 85

2.4.3.2. The need for direct replications 87

2.4.3.3. The need for independent replications 89
2.4.3.4. Identifying the RACE as a field that is susceptible to publication bias

2.4.3.5. Existing replications of the RACE

2.4.4. The need for meta-analysis

2.5. The aims of this thesis

2.6. Outline of experimental chapters

2.6.1. Chapter 3

2.6.2. Chapter 4

2.6.3. Chapter 5

2.6.4. Chapter 6

2.6.5. Chapter 7

Chapter 3: Using secondary research methods to assess the influence of red stimuli on cognitive performance in an applied setting

3.1. Experiment 1: The influence of uniform colour on academic achievement in primary schools

3.1.1. Introduction

3.1.2. Method
3.1.2.1 Sampling method

3.1.2.1.1. Primary schools

3.1.2.1.2. Key Performance Indicators (KPI's)

3.1.2.1.3. School uniform colour

3.1.3. Results

3.1.3.1. Analytical approach to results within this thesis

3.1.3.2. Descriptive statistics and null hypothesis significance testing (NHST) analysis

3.1.3.3. Estimation analysis

3.1.3.3.1. Estimation analysis- all comparisons of red to blue and green

3.1.3.3.2. Estimation analysis - assessing gender specific comparisons

3.1.3.4. Is the prevalence of red uniforms higher in lower performing LEAs?

3.1.4. Discussion

Chapter 4: The influence of red stimuli on cognitive performance in an online setting
4.1. Experiment 2: Replication of the RACE in an online environment (part I) 123

4.1.1. Introduction 123

4.1.2. Method 126

4.1.2.1. Participants 126

4.1.2.2. Materials and measures 128

4.1.2.3. Colour stimuli 129

4.1.2.4. Design and procedure 130

4.1.3. Results 132

4.1.3.1. Analyses of verbal analogy performance 132

4.1.3.1.1. Dichotomous analysis 132

4.1.3.1.2. Estimation analysis 134

4.1.3.1.3. Estimation analysis of male participants 135

4.1.3.2. Analysis of post-experiment questionnaire 137

4.1.4. Discussion 138

4.1.4.1. Summary of results 138
4.1.4.2. Experiment 2: Findings, limitations and considerations for further research

4.1.4.3. Experiment 2 summary

4.2. Experiment 3: The influence of red stimuli on cognitive performance in an online setting (part II)

4.2.1. Introduction

4.2.2. Method

4.2.2.1. Participants

4.2.2.2. Materials and measures

4.2.2.3. Colour stimuli

4.2.2.4. Design and procedure

4.2.3. Results

4.2.3.1. Analysis of verbal analogy performance

4.2.3.1.1. Dichotomous analysis

4.2.3.1.2. Estimation analysis

4.2.3.1.3. Estimation analysis in male participants
4.2.3.2. Analysis of post-experiment questionnaire 150

4.2.4. Discussion 151

4.2.4.1. Summary of results 151

4.2.4.2. Experiments 2 and 3: Findings, limitations and considerations for further research 153

4.2.4.3. Experiments 2 and 3 summary 154

Chapter 5: The influence of red on cognitive performance in the classroom 155

5.1. Experiment 4: The influence of red on verbal ability in an applied classroom setting 155

5.1.1. Introduction 155

5.1.2. Method 160

5.1.2.1. Participants 161

5.1.2.2. Materials and Measures 163

5.1.2.3. Colour Stimuli 164

5.1.2.4. Design and Procedure 165

5.1.3. Results 167

5.1.3.1. Analysis of Verbal Ability Performance 167
6.2.2.1. Participants  

6.2.2.2. Materials and Measures  

6.2.2.3. Colour Stimuli  

6.2.2.4. Design and Procedure  

6.2.3. Results  

6.2.3.1. Analyses of verbal analogy performance  

6.2.3.1.1 Dichotomous analysis  

6.2.3.1.2. Estimation analysis  

6.2.3.2. Post experiment questions  

6.2.4. Discussion  

6.2.4.1. Summary of results  

6.2.4.2. Findings, limitations and further research  

6.2.4.3. Experiment 6 summary  

Chapter 7: A meta-analysis of the RACE  

Experiment 7: A meta-analysis of the RACE  

7.1. Introduction
7.2. Method

7.2.1. Inclusion and exclusion criteria

7.2.2. Moderator and mediator analyses

7.2.3. Search strategy

7.2.3.1. Published literature

7.2.3.2. Grey literature

7.2.4. Summary of included studies

7.2.5. Statistical methods

7.3. Results

7.3.1. Model 1: Full analysis

7.3.2. Model 2: Analysis split on whether the experiment shares authorship with Elliot et al., (2007)

7.3.3. Model 3: Meta-analysis split on comparison colour

7.3.4. Model 4: Analysis split on whether the colour attributes were matched at presentation

7.3.5. Model 5: Analysis split on whether the literature was published or grey literature
7.4. Discussion

7.4.1. Summary of results

7.4.2. Findings, limitations and considerations for further research

7.4.3. Experiment 7 summary

Chapter 8: General Discussion

8.1. Overview

8.2. Summary of experimental chapters

8.2.1. Chapter 3: Using secondary research methods to assess the RACE in an applied setting

8.2.2. Chapter 4: The influence of red stimuli on cognitive performance in an online setting

8.2.3. Chapter 5: The influence of red stimuli on cognitive performance in a classroom setting

8.2.4. Chapter 6: The influence of red stimuli on cognitive performance in a laboratory setting

8.2.5. Chapter 7: Using meta-analytic methods to investigate the RACE

8.3. Chapter 8: Main findings
8.3.1. The gender discrepancy in the RACE 237

8.3.2. Fragility of the RACE 239

8.3.2.1. The evidence for the fragility of the RACE 239

8.3.2.2. The implications of the fragility of the RACE 241

8.3.2.3. Why do the findings in this thesis contradict those in the published literature on the RACE? 242

8.3.2.3.1. Publication bias 242

8.3.2.3.2. Moves towards minimising publication bias 246

8.4. Future directions for research 247

8.4.1. Direct replications 247

8.4.2. Extending and clarifying the findings of experiment 247

8.4.3. Does the fragility of the RACE discredit it as an effect worth researching? 248

8.5. Conclusion 249

References 250

Appendix A: Colour categorisation, measurement and production 274
A.1 How do we categorise colour?

A.1.1. Perceptual categorisation of colour

A.1.2. Colorimetry

A.1.2.1. Early attempts at systematically defining colour (=1758 to =1929)

A.1.2.2. Defining colour for the purpose of displaying on televisions and computer monitors (=1930 to =1980)

A.1.2.3. Defining colour to approximate human vision (=1931 - =2002)

A.2. Applying colour measurement and categorisation to the literature on RACE

A.2.1. Equating chroma and lightness when assessing hue

A.2.2. Ensuring colours are typical representations

A.3. Defining colour stimuli within the present experiments

A.3.1. Identifying the appropriate red

A.3.2. Producing colours in-house

A.3.3. Purchasing existing colour stimuli

Appendix B: Example of online experiment advertisement for experiment 2

Appendix C: Screenshots of each webpage of experiment 2
| Appendix D: Examples of online experiment advertisement for experiment 3 | 299 |
| Appendix E: Screenshots of each webpage of experiment 3 | 300 |
| Appendix F: Examples of posters and emails used for recruitment in experiment 4 | 307 |
| Appendix G: Experiment booklet from experiment 4 | 309 |
| Appendix H: Information sheet from experiment 4 | 318 |
| Appendix I: Script used in experiment 4 | 318 |
| Appendix J: Experiment pack used in experiments 5 and 6 | 322 |
| Appendix K: Information sheet used in experiments 5 and 6 | 335 |
| Appendix L: Experiment script for experiments 5 and 6 | 335 |
| Appendix M: Example of email sent to researchers to request grey literature | 342 |
| Appendix N: List of authors the email requesting grey literature was sent to | 343 |
| Appendix O: Data and forest plot for model 1 of meta-analysis | 344 |
| Appendix P: Data and forest plot for model 2 of meta-analysis | 345 |
| Appendix Q: Data and forest plot for model 3 of meta-analysis | 346 |
| Appendix R: Data and forest plot for model 4 of meta-analysis | 347 |
| Appendix S: Data and forest plot for model 5 of meta-analysis | 348 |
List of figures

**Figure 4.1.** Experiment 2 mean VA task scores for each colour condition. Confidence intervals (95%) are indicated by error bars. 133

**Figure 4.2.** Experiment 3 mean VA task scores for each colour condition. Confidence intervals (95%) are indicated by error bars. 146

**Figure 5.1.** Experiment 4 Mean VA task scores for each colour condition. Confidence intervals (95%) are indicated by error bars. 169

**Figure 6.1.** Experiment 5 Mean VA task scores for each colour condition. Confidence intervals (95%) are indicated by error bars. 189

**Figure 6.2.** Experiment 6 Mean VA task scores for each colour condition. Confidence intervals (95%) are indicated by error bars. 200
List of Tables

Table 2.1. Number of male and female participants in studies that uncovered the RACE. 70

Table 3.1. Location, level of performance, overall rank and number of primary schools within each of the selected LEAs. 102

Table 3.2. Description of each KPI, its KPI code and its assigned assessment category. 103

Table 3.3. The number of primary schools within each LEA and the number of uniform colours acquired for each of those selected LEAs. 105

Table 3.4. Colour terms drawn from school websites, their assigned colour group and the frequency of instances where the colour is required in schools. 108

Table 3.5. Sample size, mean, standard deviation and standard error of each uniform colour for each KPI. 111

Table 3.6. Sample size, degrees of freedom, Kruskall-Wallis value and associated p-value for each KPI. 113

Table 3.7. Mann-Whitney values, z-score, p values and r effect size comparing the blue and green colour conditions to the red condition. 115

Table 3.8. Mean difference, SE and 95% CI for each colour comparison across each KPI. 115

Table 3.9. A list of the 3 comparisons and their mean difference (red vs. comparison colour) that were not in the hypothesised direction. 117
| Table 3.10. KPI's that assess male and female students separately. | 118 |
| Table 3.11. Mean percentage of schools requiring red uniform in each performance category of LEA. | 119 |
| Table 4.1. Lightness, chroma and hue values for the four colour conditions. | 130 |
| Table 4.2. Mean VA task score, SD and SE of colour conditions. | 134 |
| Table 4.3. Mean difference, Cohen's d effect size, 95% CI's of the colour comparisons. | 134 |
| Table 4.4. Sample size, M, SD and SE for male participants by colour condition. | 136 |
| Table 4.5. Mean difference, Cohen's d effect size, 95% CI's of the colour comparisons. | 136 |
| Table 4.6. Lightness, chroma and hue values for the four colour conditions. | 145 |
| Table 4.7. Mean VA task score, SD and SE of colour conditions. | 148 |
| Table 4.8. Mean difference, upper limit 95% confidence interval and lower limit 95% confidence interval. | 148 |
| Table 4.9. Sample size, M, SD and SE for male participants by colour condition. | 149 |
| Table 4.10. The mean difference, Cohen’s d effect size and the associated 95% confidence intervals. | 150 |
| Table 5.1. Comparison of Elliot et al., (2007)^3 and experiment 4 from this thesis. | 160 |
| Table 5.2. Lightness, chroma and hue values for the four colour conditions. | 165 |
Table 5.3. Mean VA task score, SD and SE of colour conditions.  

Table 5.4. Mean difference, Cohen's d effect size, 95% CI's of the colour comparisons.  

Table 5.5. Mean score of each colour condition in the present experiment (UK College) and Elliot et al., (2007); German College).  

Table 6.1. Comparison of Elliot et al., (2007) and experiment 4 from this thesis.  

Table 6.2. Lightness, chroma and hue values for the two colour conditions.  

Table 6.3. Mean VA task score, SD and SE of colour conditions.  

Table 6.4. Mean difference, Cohen's d effect size and 95% CI's of the red and white colour comparisons.  

Table 6.5. Mean VA task score, SD and SE of colour conditions.  

Table 6.6. Mean difference, Cohen's d effect size and 95% CI's of the red and white colour comparisons.  

Table 6.7. Lightness, chroma and hue values for the three colour conditions.  

Table 6.8. Mean VA task score, SD and SE of colour conditions.  

Table 6.9. Mean difference, Cohen's d effect size and 95% CI's of the red and white colour comparisons.
Table 7.1. Studies selected for further consideration from the initial literature search. The reason for exclusion of those studies not retained is also given. 214

Table 7.2. Studies uncovered through searching literature reviews of the RACE. 217

Table 7.3. Studies uncovered through a call for grey literature of research that examines the RACE. 217

Table 7.4. Each comparison included in the meta-analysis, its effect size (Cohen’s d), the dependent variable used, whether the colour stimuli were matched at perception, and if the paper shares authorship with Elliot et al., (2007). 218

Table A.1. A description of comparison colours used in a selection of studies and whether or not these colours were matched on chroma and lightness to the red stimuli. 283

Table A.2. A list of colour values for stimuli used previously in the RACE literature. 285

Table A.3. Lightness, Chroma and Hue values for the colour stimuli produced ‘in-house’. 287

Table A.4. Lightness, chroma and hue colour values for purchased colour stimuli. 289
Chapter 1: General introduction to colour and cognition

1.1. Overview

This thesis explores the influence of colour perception on cognition. Colour, along with properties such as form and motion, constitutes a visual system that enables organisms to process information contained in visible light.

Colour research includes the physics of colour, the physiology of colour vision and the adaptive value of colour vision to organisms. This thesis explores a relatively understudied area of colour research; its effect on cognitive functioning with a particular focus on the influence of red stimuli on cognitive performance in an achievement setting.

Before the central theme is explored, this thesis will describe what is known about colour, much of which is important in assessing the influence of colour on cognition. The following key questions will be briefly addressed:

What is colour?

How do we see colour?

Why do we see colour?

1.2. The science of colour vision

1.2.1. What is colour?

Colour is a component of vision that enables organisms to distinguish and categorise wavelengths of electromagnetic radiation (EMR) that range between 400nm and
700nm, termed the 'visible light spectrum' or simply 'light' (Solomon & Lennie, 2007). Colour does not reside in an object's surface nor is it contained in EMR; it is instead an internal, cognitive representation of EMR that has been reflected from or emitted by an object and into an eye. An object appears the colour of EMR that it reflects/emits, e.g. a banana will absorb all EMR aside from EMR with a wavelength of ≈570-590nm (perceived as yellow). EMR that is absorbed by an object is converted to heat (Gegenfurtner & Sharp, 2000).

Light is no different in substance to any other form of EMR (e.g. FM radio waves or X-rays); all are built of photons formed as a result of changing electron positions within an atom through energy absorption (Grace, Lovett & Grace 2008). The only difference between an FM radio wave and visible light is the frequency of the waveform.

However, the human eye cannot detect EMR outside of the visible light spectrum. A thought experiment by Feynman (1981) illustrates this. A typical room is 'full' of EMR of varying wavelength; light, radio and television waves, mobile phone and Wi-Fi signals and RADAR from passing planes. All these forms of EMR are ever present around us but a human can only perceive the light in the room and nothing else. We know EMR is ever present because when we turn on a radio we hear a broadcast and when we connect to a Wi-Fi router we access the internet. Just as light does not contain colour, radio waves do not contain sound, so a receiver (an eye or a radio antennae) and a processor (a visual cortex or a radio tuner) are needed to process the raw EMR information to produce a representation of colour or sound in an otherwise 'empty' room. Feynman poetically summarises this as 'the inconceivable nature of nature' (Feynman, 1981, 45.50).
1.2.2. How do we see colour?

The perceptual pathway of colour vision is complex and for the purpose of this thesis only a brief description will be provided based on a comprehensive review from Solomon and Lennie (2007). Light enters the eye through the pupil and is concentrated onto the retina that contains two types of photoreceptor cells; rods, used for night vision, and cones, used for colour vision. Cones contain one of three types of photopigments that are sensitive to different wavelengths of light; ‘short wavelengths’ (‘S-cones’, ~430 nm or roughly ‘blue’), ‘medium wavelengths’ (‘M-cones’, ~530 nm or roughly ‘green’) and ‘long wavelengths’ (L-cones’, ~560 nm or roughly ‘red’). These photopigments absorb photons from their specified wavelength range (although there is some overlap) and send a signal through the visual system’s complex arrangement of retinal cells, the optic nerve and the lateral geniculate nucleus to the visual cortex, which is responsible for processing visual information. Although complex, the description of the physiological structure and processes that enable colour vision represent; “the most complete and quantitative description of a perceptual pathway currently available to science” (Wade, Augath, Logothetis & Wandell, 2008, p. 6).

1.2.3. Why do we see colour?

Colour vision is thought to be adaptive, it allows properties of objects (e.g. whether fruit is ripe or poisonous) to be recognised and understood giving more information about an object aside from form and motion (Neitz, Carroll & Neitz, 2001; Osorio & Vorobyev, 2008). There are a number of competing theories regarding the catalyst of colour vision evolution including distinguishing between nutritious and toxic berries (Sumner and Mollon, 2000), identifying ripe fruit (Regan et al., 1998) and edible plant leaves (Dominoy & Lucas, 2001),
recognising sunburn or blood in urine, or assessing properly cooked meat (Neitz, Carroll & Neitz, 2001). The debate continues within the literature but the general consensus that colour vision holds an evolutionary advantage remains.

Colour vision is thought to have evolved through a series of adaptive genetic mutations. Research by Neitz, Carroll & Neitz, 2001 suggests that humans and other trichromatic primates share a common ancestor with only one type of cone cell that could distinguish between dark (analogous to 'black') and light (analogous to 'white'). A genetic mutation duplicated the single cone gene and its duplicate developed a different spectral sensitivity that could distinguish between blue and yellow light before a second gene duplication led to a third type of cone cell that could differentiate between red and green (Neitz, Carroll & Neitz, 2001). These extra photoreceptors increased the number of discernible colours exponentially from 200 (a single black-white cone cell) to 10,000 (the addition of the blue-yellow cone cell) to over 1,000,000 (the addition of the red-green cone cell). This increase in discernible colours demonstrates the immense adaptive value of additional cone cell variants to organisms.

Another reason why colour is so important to sensory perception is that humans often have a tendency to use vision (and hence colour) as the primary form of sensory information over the other senses, termed 'visual dominance' (Koppen, Alsius & Spence, 2007; Hecht & Reiner, 2008). This can lead humans to make perceptual mistakes. For example, participants wearing prism glasses (that distort vision) report straight objects as curved when asked to hold an object and report its shape (Gibson, 1933). This was developed further by Colavita (1974) who demonstrated that participants were more likely to report visual (relative to
audible) stimuli when presented simultaneously. Applied examples of visual dominance include white wine dyed red being described as tasting like red wine (Morrot, Brochet & Dubourdieu, 2001) and the 'ventriloquist effect' that explains how individuals watching films attribute the dialogue to be coming from the actors mouth and not the audio speakers (Alais & Burr 2004). This occasional bias towards relying on visual information make colour an important aspect of perception.

1.2.4. Could colour influence cognitive performance?

If the purpose of colour vision is to give more information about an environment could this visual information, consciously or unconsciously, influence cognitive performance? Previous research stretching back to the 19th century has assessed the influence of colour in related domains, including its influence on mood, sporting performance, mate selection, food and beverage evaluation. Although this thesis examines the effect of colour on cognitive performance, there are enough similarities to colour research in the domains mentioned above to warrant a brief examination of such research. Therefore chapter 1 concludes with a literature review of research that examines the influence of colour on mood, sporting performance, mate selection and food and beverage evaluation. For ease of interpretation this literature review will first attend to pre-21st century colour research before describing more recent colour research.

1.2.5. Pre-21st century empirical colour research

Much early work on colour and cognition regarded the subjective recording of colour preference in adults (Wells, 1910; Jastrow, 1897; both cited in Pressey, 1921) as well as its
development in children (Winch 1909 cited in Pressey 1921). These early findings are summarised as blue and red being the most preferred colours with yellow being the least preferred and that children are more likely to prefer bright colours with this preference diminishing through age.

The first noted work to examine the effect of colour upon cognitive performance was conducted by Pressey (1921) who details an interesting and well thought out set of experiments. Quite understandably the colour presentation and general experimental procedure are lacking compared to modern techniques, but these experiments were influential in establishing research that examines the influence of colour on cognitive performance. Pressey desired to move away from the introspective and anecdotal accounts of colour effects that dominated thinking at the time towards objective measurements of cognitive ability and physiological responses. It is probable that Pressey would be somewhat aggrieved to discover that introspective, anecdotal and erroneous accounts of the influence of colour on cognition still dominate modern culture and media (for examples Wright, 2015; Haller, 2014). Pressey examined the influence of colour on motor skills (rate of rhythmic finger-tapping), mental arithmetic, rate of free association, short term memory and reaction speed. He found brightness was positively correlated with performance in motor skill and mental arithmetic and also uncovered a Yerkes-Dodson curve in the performance of reaction speed as a function of brightness. No effects were found regarding the rate of free association or short term memory. With no inferential statistics and numerous research design issues, the legacy of this paper is not the results, but rather the paradigm that it instigated.
One of the more influential studies to examine the link between colour perception and performance was conducted by Goldstein (1942). Based on anecdotal insights from his work as a psychiatrist he theorised that arousal was a function of colour wavelength; longer wavelength colours cause more arousal with the opposite shown for shorter wavelength colours. Subsequent work has shown this simplistic view not to be the case (O'Connor, 2011).

An interesting area of research in the middle of the 20th century examined if pink paint could subdue prisoners and drunks in holding cells. Luscher (1969 cited in Schauss, 1979) noted that colour preference varied as a function of emotional state and hypothesised that these changes were mediated through the endocrine system. Schauss (1979) asked if this pathway could be reversed; could colour perception instigate changes in the endocrine system? This idea developed into the production of 'Baker-Miller Pink', a colour designed to lower arousal and calm an agitated person. Initial experiments at a military correctional institute found that 15 minutes exposure to Baker-Miller Pink reduced aggressive and violent behaviour (Schauss, 1979). This effect has since been discredited as subsequent research has not been able to replicate the results (Gilliam & Unruh, 1988). The work reported by Schauss is thought to be a product of weak experimental design coupled with experimenter effects (Genschow, Noll, Wanke & Gersbach, 2015). However, the idea that pink stimuli can calm a person prevails in modern culture. In 2004 the University of Iowa renovated their sports facilities and introduced pink lockers, showers and toilets to the already pink painted walls of the opponents’ dressing room. This proved controversial for many academics and students who opposed the reinforcement of gender and sexuality stereotypes that pink implies femininity, weakness and submission (New, 2014).
The work of Sinclair, Soldat and Mark (1997) examined if emotional responses to colour mediated changes in cognitive performance. They sought to combine two strands of research. The first is that 'happy moods' promote speculative cognitive processing that focuses on the ‘bigger picture’ and that 'sad moods' promote less speculative cognitive processing that focus on details in isolation (Sinclair & Mark, 1995). Tasks that require broad cognition tend to be improved by speculative processing but are impeded by detail based processing. The second strand regards research that examines emotional responses to colour. Jacobs and Blandino (1992) had previously reported that red stimuli promote positive ('happy') emotional responses and that blue stimuli promote negative ('sad') emotional responses. Soldat, Sinclair and Mark (1997) gave participants exam style questions, which required detail based processing, printed on red or blue paper. Results showed that those in the blue condition were more accurate than those in the red condition. Sinclair, Soldat and Mark (1998) replicated the effect by manipulating paper colour in an authentic mid-term university exam.

Although the findings of Soldat, Sinclair and Mark are similar to those reported by publications central to this thesis (see section 2.4.) the theory that underpins the effect is overly simplistic and has since been discredited. The idea that red conveys 'happiness' and that blue conveys 'sadness' ignores the complex, contradictory and context dependent nature of colour associations. For example red signifies both love (i.e. attraction) and danger (i.e. aversion) depending on the context. Similarly blue can signify both depression (an undesirable emotional state) and, from associations with blue sky, tranquillity (a desirable emotional state). Additionally, a growing body of work indicates that red stimuli promote
detail based processing and blue do not (Elliot et al., 2007; Mehta & Zhu, 2009; Hulshof, 2013; Elliot & Maier, 2014).

1.2.6. Contemporary analysis of pre-21st century empirical colour research

In their review of colour psychology, Elliot and Maier (2014) propose two limitations of pre-21st century work. The first limitation is that work on the subject emerged from applied concerns alone and lacks theoretical grounding. Examples include assessing what colour makes office workers more productive (Kwallek & Lewis, 1990) or what colour makes food more appealing (Imram, 1999). These issues were examined in isolation with little interest or insight of theoretical mechanisms (Elliot and Maier, 2014). The second limitation regards research design issues that span both generic problems of older psychology work (e.g. lack of statistical power [Clark-Carter, 1997] and publication bias [Nagakawa, 2004]) and issues specific to the measurement, production and presentation of colour (see chapter 3 for a detailed description of such issues). These two sets of limitations at best make the interpretation of previous literature difficult and at worst render the existing literature on colour and cognition essentially uninterpretable (Valdez & Mehrabian, 1994).

1.2.7. Summary of pre-21st century empirical colour research

The studies described above give a brief account of colour research conducted pre-21st century. Such work demonstrates the breadth of colour research taking in diverse topics such as colour preference, the impact of colour on the endocrine system and the influence of colour on mood. The extent to which previous work on colour and cognition did not attend to theory meant that work was conducted individually with little or no reference or
continuity between studies. This resulted in virtually no viable theoretical assumptions made concerning the effect of colour on cognition. One of the main issues is that researchers attempted to develop universal rules concerning the effect of colour on cognition with no regards to context, e.g. Soldat, Sinclair and Mark's (1998) assertion that red symbolises happiness even though in some contexts red can symbolise danger. This means any theory that does not consider context will not be able to accurately predict colour effects. Although much of the research conducted pre-21st century was flawed both in terms of theory and methods, these pioneers developed colour research into a scientific endeavour that is only beginning to be understood today.

1.2.8. Contemporary colour research

The central theme of this thesis is to examine how colour influences cognitive performance by studying the effect of colour on tasks that require cognitive skills such as knowledge, memory, problem solving and comprehension. However, there are other cognitive domains and outcome measures that have been studied in colour psychology that are also worth consideration. This section will briefly describe a selection of empirical research in cognitive domains other than raw cognitive performance.

One such cognitive domain is ‘evaluation’; the process of judging something’s quality, importance or value (Cambridge Dictionary of Psychology, [Matsumoto, 2009]). This differs from measures of cognitive performance in that there is no pass or fail element to making an evaluation. The literature that describes the effect of colour on evaluation is well formed and provides interesting methods and evidence that can be applied to work that examines cognitive performance.
One other outcome measure to consider is sporting performance. Whilst sporting performance is most often measured by physical outcomes (e.g. a ball crossing a line or a ‘knock-out’ punch) playing sport is underpinned by cognitive processes such as knowledge, attention, judgement, reasoning and decision making. Sporting performance therefore provides an alternative lens on how colour can influence cognition and demonstrates how colour effects can manifest in a range of settings.

Therefore, before focusing on the influence of red stimuli on cognitive performance in depth, this section will provide a brief snapshot of contemporary work on colour and cognition generally. This demonstrates the scope of colour psychology research and theory and will help set the experimental work of this thesis clearly in relation to the rest of the field. This section will comment upon:

Colour and mate evaluation. Describing the influence colour has on mate selection.

Colour and sport. Using mostly secondary methods this literature examines if sporting outcomes vary as a function of colour attire.

1.2.8.1. Colour and evaluation

The influence of colour on evaluation is not limited to mate evaluation but spans diverse topics such as food, beverage and advertising evaluation. However, these areas of research are less researched and the studies that have been conducted suffer from a lack of consistency, methodological rigour and thought regarding the underlying theory. For examples of work examining the role of colour on food and beverage evaluation see Shankar, Levitan and Spence (2010), Schuldt, (2013) and van Doorn, Wuillemin and Spence
For examples of work examining the role of colour on advertisement evaluation see Puccinelli, Chandrashekaran, Grewel and Suri (2013) and Kyrousi and Panigyrakis (2015).

Conversely, the influence of colour on sexual attraction is one of the most researched areas of colour psychology, its methods are well constructed and work has been done to examine mediating variables that can explain the effects. The work on colour and attraction also shares considerable overlap with the work on colour and sporting performance (most notably through links regarding red coloration and social status/dominance). Given that research on colour and evaluation is not central to this thesis (and is only illustrated here to serve as context for work on colour and cognitive performance) only work that examines the influence of colour on mate evaluation will be considered here.

One line of research in colour and mate evaluation regards the idea that red apparel and accessories can increase sexual attraction in heterosexual mate evaluations. Interestingly, separate mechanisms are thought to drive the effect between genders; red increases the perceived fertility of evaluated females but increases the perceived social dominance of evaluated males.

1.2.8.2. Colour and attraction

1.2.8.2.1. Males evaluating females

Elliot and Niesta (2008) drew on work that suggested red increases sexual attraction through associations that emerge from both biologically ingrained (e.g. reddening of skin in female primates during ovulation [Setchell, Wickings and Knapp, 2006]) and socially learnt (e.g. love hearts, roses, red light district) sources. They found that photos of women were
rated as more attractive when the photograph was placed in a red border or if the woman wore a red shirt. Subsequent work using observational methods, such as conversation proximity or frequency of asking intimate questions, replicated the effect (Niesta Kayser, Elliot & Feltman, 2010) as did work that found increased attraction when carrying a red laptop (Liu, 2014) or reading personality descriptions printed on red coloured paper (Hammett, Issler & Bashore 2014). Gueguen and Jacob (2012) placed fake dating profiles of women wearing different coloured t-shirts and recorded the number of contact requests they received. Those in the red t-shirts attracted significantly more requests from prospective sexual partners.

Pazda, Elliot and Greitemeyer (2012) report that the underlying mechanism driving the increase in female attractiveness ratings is perceived sexual receptivity. The same study ruled out other potential mate qualities such as perceived kindness and intelligence as mediating variables. Elliot and Pazda (2012) next asked if females (either consciously or subconsciously) signalled sexual receptivity through wearing red coloured apparel. Female participants were asked to imagine joining either a standard online dating website or an online dating website that facilitated casual sex. Participants then stated their preferred attire for a profile picture, including the colour of a t-shirt. Those who imagined they were interested in joining a casual sex dating website chose the red t-shirt on significantly more occasions. In two further experiments Elliot and Pazda analysed the profile photos from authentic online dating websites and found that those interested in casual sex were significantly more likely to have chosen a photograph of themselves clearly wearing red apparel.
Schwarz and Singer (2013) reasoned that the ultimate goal of increasing perceived sexual receptivity is impregnation and the continuation of life. If this was the case, Schwarz and Singer hypothesised that the red-attractiveness effect would cease to manifest when rating women who could no longer become pregnant (i.e. women post-menopause). Male participants were asked to evaluate the attractiveness of both a young female (≈24 years) and an older female around the onset of possible menopause (≈48 years). The photos were placed onto either a white or red surround, replicating the method of Elliot and Niesta (2008). Results showed that whilst the red-attraction effect manifested in evaluations of the younger woman, no effect was found regarding the older woman. This effect was consistent in two groups of male participants aged ≈25 and ≈55. Such findings support the notion that it is potential fertility and the desire to impregnate that drives the underlying effect of red stimuli increasing perceived sexual receptivity.

1.2.8.2.2. Females evaluating males

This work was extended to examine female perceptions of male attractiveness by Elliot et al., (2010). A comprehensive assessment of attraction across seven experiments found that women rated men as more sexually attractive if their photo was placed in a red border or if the subject was wearing a red t shirt. Their work also examined other desirable personality attributes such as honesty or sociability and found that these were not influenced by red stimuli. Social status was deemed to be the mediating variable that drove the increase in sexual attractiveness. This causal chain is also reported in non-human red-attractiveness studies. Just as red coloration in female animals can signify fertility, red coloration in non-human male animals can indicate social status. Red coloration has been shown to correlate
with dominance and good health (Changzi, Zhang & Shimodo, 2006) both of which are implicated as characteristics of a good potential mate of high social status (Nelson & Morrison, 2005). This effect of red coloration and increased social status has been found in non-human primates, fish and crustaceans (Setchell & Wickings, 2005). For example, for primates to have red skin coloration they must be in good health as skin vascularisation (the promotion of blood to the skin) is costly to produce as blood present here cannot be used to facilitate the functioning of other more vital organs such as the liver or kidneys.

1.2.8.2.3. Non-heterosexual mate evaluations

To date no research has been conducted to examine if these effects manifest in non-heterosexual mate evaluations. Perhaps the effect would be found in male-to-male attractiveness ratings given that the mediating variable is thought to be perceived status, a quality likely to be important to a relationship regardless of the orientation. It is possible however that the effect does not manifest in female-to-female attractiveness ratings given that the effect seems to be rooted in the biological desire to reproduce (an act not possible in an all-female relationship). However, of course non-heterosexual women still desire children and so it would be interesting to study the degree to which fertility compared to other parental qualities (i.e. nurturing, stability, emotional intelligence) influences the proposed red-attractiveness link within a same sex context. It must be noted that although female-female attractiveness ratings (Elliot et al., 2010) and male-male attractiveness ratings (Elliot & Niesta, 2008) were both recorded within studies these were administered to heterosexual participants and are not relevant to the above discussion. Null results were found in all instances of heterosexual same sex ratings.
1.2.8.2.4. Colour and attraction summary

The research regarding the influence of colour on mate evaluation represents the zenith of colour research, it is consistent and compelling and is underpinned by a coherent hypothetical argument. This introduction will now go on to assess a more mixed body of work, that of colour and sporting performance.

1.2.8.3. Colour and sporting performance

Research on the influence of colour on sporting performance is typically conducted by analysing if elite sporting outcomes vary as a function of uniform colour. When assessing this body of work it is worth noting that, like all secondary research, although these studies have high ecological validity there is little control regarding independent, dependent, confounding and extraneous variables and effects are typically evaluated in underpowered analyses. This variability means that contradicting findings are rife within the literature.

Hill and Barton (2005a) examined Olympic combat sport outcomes from the 2004 Olympic Games where participants were randomly assigned either red or blue apparel. They found across all combat styles (western boxing, taekwon do, Greco-Roman wrestling and freestyle wrestling) that those fighting in red apparel won more fights and more rounds across all weight classes. To explain these findings, Hill and Barton drew on examples in non-human animals that red coloration in males is indicative of increased social status that is mediated through perceived dominance and enacted aggression (see Little & Hill [2007] for a short review), from herein termed the 'red-dominance theory'. They propose that red attire in Olympic bouts influences performance outcomes by cueing the ‘receiver’ (i.e. the fighter in
blue) to imply that their opponent is more aggressive and dominant whilst simultaneously
cueing to the ‘sender’ (i.e. the fighter in red) to act more aggressively. Empirical evidence
for this comes from Feltman and Elliot (2011) who examined if viewing red caused a non-red
wearing combatant to fight worse (through a lack of aggression) or red caused a red wearing
combatant to fight better (through increased aggression). Participants were asked to
imagine they were to compete in a Taekwondo contest dressed in either red or blue (thus the
opponent would be dressed in the opposite colour) before completing measures of
perceived dominance and threat of both themselves and the opponent. Results showed that
both wearers of red perceived themselves to be more dominant and that opponents
wearing red were perceived to be more dominant, demonstrating that the proposed effect
could be bi-directional. This bi-directionality is analogous to that described in the mate
evaluation literature that demonstrates males find red on women sexually alluring (Niesta
Kayser, Elliot & Feltman, 2010) and also that women use red to make themselves appear
more sexually receptive (Elliot & Pazda, 2012).

Rowe, Harris and Roberts (2005) challenged the findings of Hill and Barton (2005a) by asking
if colour effects manifest in Olympic combat sports were red attire is not worn. They
assessed the outcome of Judo matches where opponents wear either a white or blue
‘Judogi’. They too found a colour effect on contest outcomes, those wearing blue attire won
significantly more bouts. Given that there is no obvious evolutionary explanation for blue
apparel improving sporting outcomes these findings are explained as a result of increased
visibility, i.e. the white Judogi was easier to visually track as it contrasted more with the
background and so conferred an advantage to the fighter in the blue Judogi (termed the
‘visibility theory’). From this assumption Rowe, Harris and Roberts dismiss any notion of the
‘red-dominance theory’. However, the arguments for visibility being central to the effect have several weaknesses. First, the assumption that, in a Judo match, white is more visible than blue is conjecture; it could easily be argued that a chromatic colour such as blue is more visible than an achromatic colour such as white. Second, it is plausible that red apparel is more visible than blue apparel within a combat sport context and so if increased visibility confers a disadvantage in combat sports then the visibility theory cannot explain the red-apparel advantage reported by Hill and Barton (2005a). Further analysis of Rowe, Harris and Roberts (2005) was conducted by Dijkstra and Preenen (2008) who found that the red effect found in the 2004 Judo Olympic competition could be explained better by the seeding system. Within the seeding system the first combatant in a pairing is automatically assigned the blue Judogi and this effectively makes the assignment of colour non-random. However, when the seeding system is not used colour based effects can be uncovered. Julio, Miarka, Rosa, Lima, Takito and Franchini (2015) showed across 1,233 non-seeded judo contests that wearing blue attire does indeed confer a sporting advantage across genders, level of competition and age range. Outside of Olympic combat sports, Pollet and Peperkoorn (2013) examined the fight outcomes of Mixed-Martial-Arts (MMA) bouts in the Ultimate Fighting Championship (UFC) where fighters can choose their own trunk colour and no seeding system is used. Results showed no benefit for fighters dressed in red trunks across 210 bouts. Ilie, Ioan, Zagrean and Moldovan (2008) extended the work of Hill and Barton (2005a) from elite hand-to-hand combat to elite virtual combat by analysing data from the world's top ten players of 'Unreal Tournament', a multi-player first-person-shooter online game. Data from 1,347 matches showed that those players randomly assigned to the 'red team' won significantly more (54%) of these contests. The effect of colour within combat
sports is inconsistent across studies and no firm conclusions regarding the effect of colour apparel on fight outcomes can be made.

Atrill, Gresty, Hill and Barton (2007) assessed if the advantage of wearing red attire in sport was present in elite association football in England over 55 seasons. They found significantly more league champions playing in red than any other colour and a significant win advantage in home games (where home teams choose the strip colour) for those playing in red. Additional analyses paired teams by proximity (teams within the same city to account for fan base sizes) and found that the team playing in red had significantly more success in seven of eight cities. Allen and Jones (2014) replicated this work to include only the first 20 seasons of the Premier League (1992 to 2011) and found that whilst teams who played in red were generally more successful there was no home advantage for teams playing in red as Atrill et al., (2007) had described. Subsequent research did not replicate the effect in German (Kocher & Sutter, 2008) or Spanish (Garcia-Rubio, Picazo-Tadeo & Gonzalez) association football. Research into rugby football also describes contradictory findings. Piatti, Savage and Torgler (2010) examined the role of colour kit on the outcome of Rugby League Football matches across 30 seasons in the Australasian National Rugby League. Their analysis shows limited support for the idea that red apparel increases sporting prowess. Descriptive statistics and initial analysis show an advantage of wearing red, however this effect is reduced to non-significance once additional variables (such as game round, crowd size and opposition team) are added to the model. Further analyses showed strong support for the red effect when the difference between two examined hues is larger (e.g. red is more distinct from green than it is from yellow) even after the additional variables are added in.
However, the notion that colour can influence sporting outcomes has permeated association football to some degree. Cardiff City F.C. changed (and then subsequently re-changed) their strip from blue to red and this *did* coincide with their first promotion to top flight football for 51 years (BBC, 2012). Southampton F.C. dropped their traditional red and white stripes for a red only kit (Augustus, 2014) during their recent resurgence to top level Association Football. There have also been calls from teams to fans to create a 'wall of colour' within stadia to further accentuate home advantage (The Football League, 2014) in much the way that German club Borussia Dortmund’s fans create the famous 'Die Gelbe Wand' (The Yellow Wall). Although the empirical and anecdotal evidence presented here suggests that kit colour can significantly influence team performance further work examining association football reveals a more contradictory body of research.

1.2.8.3.1. Colour and sporting performance summary

The influence of colour on both individual and team sports described here is unclear, inconsistent and even contradictory in some cases. It is possible that a third variable is influencing the outcome in sport contests but given the low control over extraneous variables within the analysis of secondary research it is difficult to highlight possible candidates. This mixed body of work contrasts with the consistent results found in the work regarding colour and mate evaluation. One explanation for this contrast is that the work on colour and mate evaluation relies more on primary research methods and thus has more control over research design. This control over extraneous and confounding variables results in a much more consistent identification and analysis of colour effects. The fact that colour
effects are less consistent when using secondary research methods could signify that colour
effects are fragile and are susceptible to being influenced or negated by other variables.

Chapter 1 highlighted the importance of colour to human psychological functioning and
briefly assessed some basic tenets of colour science. In doing so chapter 1 suggests that, due
to its value as a source of information, colour could influence cognitive functioning. Chapter
1 concluded with a description of both early and contemporary colour research in domains
such as mate evaluation, food evaluation and sporting performance. Chapter 2 focuses in
more depth on the influence of colour, specifically red, on cognitive performance in an
achievement setting. This effect will form the basis for empirical work in this thesis.
Chapter 2: The influence of red stimuli on cognitive performance

Chapter 1 provided a brief overview of colour research that explores basic tenets of what colour is, how colour is perceived and why colour vision developed before exploring colour research in areas other than its effect on cognitive performance. This background information is important to understanding how and why colour could influence cognition.

The research in this thesis focuses on the effect of red stimuli on cognitive performance. The decision to examine one specific colour reflects the body of literature as the influence of red stimuli on psychological functioning is by far the most examined topic in this area (see section 2.2 and appendices A and B for a review of such work). Additionally, literature suggests that red seems to be a 'special' colour; red is the colour with the longest EMR wavelength that humans can perceive (Koendernik, 2010), red coloration is used extensively by non-human animals as a signal (Milinski & Bakker, 1990; Setchell & Wickings, 2005), humans are able to distinguish small changes in the red coloration of faces (Tan & Stephen, 2013), and the categorisation of red as a distinct hue appears in every recorded culture that distinguishes chromatic colours (Kay, Berlin, Maffi, Merrifield & Cook, 2009). Loreto, Muhkerjee and Tria (2012) theorise that it is the ability to easily distinguish small changes in red coloration that is the key reason for why red is universally the first colour category to be agreed by consensus within a culture.

2.1. Introduction to the 'Red Achievement Context Cognitive Effect' (RACE)

The linchpin of current research into the effect of red stimuli on cognitive performance is work by Elliot, Maier, Moller, Friedman and Meinhardt (2007). This paper details four
experiments demonstrating that, **in an achievement context, the presence of red stimuli prompts avoidance motivation, which results in impaired cognitive performance.** This definition represents the effect that will be the central issue of this thesis and from herein is termed the 'Red Achievement Context Cognitive Effect' (RACE). Such a specific definition of the effect helps differentiate it from other colour based effects, e.g. the effect of red on the evaluation of potential mates (Elliot & Niesta, 2008) or the effect of red on sporting performance (Hill & Barton, 2008). These colour effects differ from that described by the RACE as a function of the underlying mechanism that facilitates the effect and also the fact that red coloration seems to facilitate favourable evaluation compared to impairing cognitive performance (see 2.2). Whilst the effect of red stimuli on mate evaluation and sporting performance is dependent on aspects of perceived dominance in human conspecifics, the RACE depends more on aspects of danger and infection in a wider ecological context.

Before considering the findings of Elliot et al., (2007) in depth, it is important first to understand the theoretical background from where the work emerged. Understanding this background information aids in assessing the mechanisms that underpin the RACE.

2.1.1. Understanding the field of achievement motivation

The work in the Elliot et al., (2007) paper was conducted in the field of **achievement motivation**; a field that examines how individuals are ‘energised’ towards an evaluation of competence (Elliot, 1997). Typical research questions within achievement motivation include what factors predict task competence and what consequences arise as a function of different motivational strategies. Two key concepts of achievement motivation are
particularly important when assessing the RACE. The first is the concept of an 'achievement context', a setting that is characterised by a clear pass or fail element to task completion. It is this setting in which achievement motivation is examined. The second is the concept of motivational direction, which comprises a continuum from approach motivation to avoidance motivation. This concept describes the degree to which an individual adopts an approach based goal (e.g. attempting to score more than 10 to pass an exam) or an avoidance based goal (e.g. avoiding a score of 9 or less and fail the exam) during achievement pursuit. This directional distinction has been shown to predict performance in cognitive tasks with avoidance based goals generally resulting in poorer performance (Elliot, Shell, Henry & Maier, 2005). In addition to these explicit avoidance/approach goals, research has shown that more subtle, internal cognitive processes and external cues can prompt more general implicit approach/avoidance motivation and this implicit motivation can impair performance in a similar fashion to that of explicit goals (Custers & Aarts, 2011).

2.1.2. The development of the RACE

The team of researchers behind the RACE (Elliot et al., 2007) were interested specifically in how external cues could instigate implicit avoidance/approach based motivation (and thus changes in performance) in participants; for example the use of 'STOP' or 'CAUTION' signs in an experimental setting, (Prof. Elliot, personal communication, November 19th, 2011).

Often, these types of sign are coloured red. This line of enquiry led the researchers to ask if the colour red alone could signal danger and thus instigate implicit avoidance motivation and impede performance. After a pilot investigation showed promise, Elliot and colleagues went on to develop the experiments detailed in the seminal work Elliot et al., (2007).
In the first experiment of Elliot et al., (2007), participants were asked to ostensibly check that page numbers (written in red, green or black ink) were correct on test papers containing anagram puzzles. Those who had checked page numbers written in red ink scored significantly lower solving the anagrams than those in the other two conditions.

Experiment 2 gave participants a verbal analogy task set in a psychology laboratory. Immediately before taking the task they turned a page to reveal a 12cmx18cm square coloured either red, green or white, which they viewed for 5s. Those who had viewed the red square scored significantly worse than those who had viewed the green or white square.

In the third experiment the setting changed to a German high school. Participants were tested in a group, the white condition was changed to grey and the colour exposure duration was shortened to 2s. Participants in the red condition scored worse than those in the green or grey conditions. The fourth experiment replicated the third with the exception that the dependent measure was changed to a numerical based task, again those in the red condition scored significantly worse than those in either of the other two groups. Two other experiments examining the role of red stimuli in promoting avoidance motivation are also presented in Elliot et al., (2007) and these will be discussed further in assessing the theory of the RACE (see section 2.3.). The work by Elliot et al., (2007) clearly demonstrates a reliable and provocative effect of colour on cognitive performance that, unlike previous research (e.g. Goldstein, 1942; Kwallek & Lewis, 1990; Soldat, Mark & Sinclair, 1998) is based on a strong, established theoretical foundation.

The publication of Elliot et al., (2007) led to the emergence of two research strands. The first examines the influence of the RACE on cognitive performance (i.e. the RACE as an absolute effect on cognition) whilst the second examines the theoretical processes that underpin the
RACE. This thesis will first consider the theoretical processes that underpin the RACE as doing so will provide a good basis for understanding and critiquing the influence of the RACE on cognitive performance later.

2.2. The theory that underpins the RACE

In short; the perception of red stimuli, in an achievement context, prompts avoidance motivation that impairs cognitive performance. This causal chain requires the consideration of three key questions:

1. Does the perception of red stimuli in an achievement context prompt avoidance motivation?

2. Does avoidance motivation impair cognitive performance?

3. How does colour induced avoidance motivation impair cognitive performance?

The last question is important as although the direction of motivation has been shown to influence performance in a variety of settings, the precise mediating variables that achieve this vary between settings.

2.2.1. Does the perception of red stimuli in an achievement context induce avoidance motivation?

To assess this first question researchers examined if colour stimuli influenced established measures of avoidance motivation. This was assessed as early as the Elliot et al., (2007) paper in an experiment that used a classic risk taking task to measure avoidance motivation.
Participants were asked to select the number of easy and moderately difficult questions they would like to answer in an IQ test, a measure previously used to assess avoidance motivation. Immediately before giving their answer, participants viewed a red, green or grey square for 2s. Those participants in the red condition requested more ‘easy’ questions than those in the other two conditions. Next, the researchers moved away from self-report methods of avoidance motivation towards measuring neural correlates of avoidance behaviour. Davidson, Schwartz, Saron, Bennett and Goleman (1979) had previously found that avoidance motivation increases activity in the right frontal cortex, a finding that has been replicated over 100 times (Elliot et al., 2007). Elliot and colleagues fitted an EEG cap to participants and sat them at a computer to ostensibly take an IQ test. As in the previous experiment participants were shown a coloured square for 2s prior to taking the test. Those participants who viewed the red square showed significantly greater right frontal activation compared to those in the grey or green conditions. Thus, Elliot et al., (2007) were able to tie the onset of avoidance motivation (as measured through both self-report and brain imaging techniques) specifically to the perception of red stimuli.

To further examine this link Elliot, Maier, Binser, Friedman and Pekrun (2009) published a paper that explicitly considered if red stimuli in an achievement context induced avoidance behaviour. In the first experiment participants were placed in either an ‘achievement context’ or an ‘evaluative context’ by being asked to either take a test or rate whether other students would like a test. In each of these groups the students were given a folder that contained the instructions for the test they would be taking/rating. This folder served as the independent variable and contained the colour manipulation; a square coloured either red or green placed immediately before the instructions. Following this colour presentation
participants were asked to walk to a different laboratory down the hall to take/rate the appropriate test. Upon arrival they found a closed door with a sign reading ‘Please Knock’. The total number of door knocks was used to determine whether or not avoidance motivation was present within participants (fewer knocks indicating greater avoidance motivation). In the evaluative context group there was no difference in the number of door knocks between colour conditions. However, in the achievement context those in the red condition knocked on the door significantly fewer times than those in the green group. In a second experiment participants were fitted with an inclinometer that measured a person’s movement towards or away from a stimulus. Previous research has described how avoidance motivation manifests in individuals physically moving away from a stimulus (Barsalou, Niedenthal, Barbey, & Ruppert, 2003) and so provides an excellent non-self-report measure of avoidance motivation. In this instance the inclinometer measured the direction and distance that a person moved as they sat at a computer to take an IQ test. Immediately before the IQ test the participants were shown a title page for 2s that was coloured either red, green or grey before being shown a fixation cross for 5s before the test started. Measurement from the inclinometer started 5s before the test page was shown and continued for 12s (until the start of the IQ test). Analysis showed that in the seconds after the colour presentation, those in the red condition moved significantly further back in their seated position than those in either of the green or grey conditions, thus demonstrating enacted avoidance behaviour. This work not only assesses avoidance motivation using several complementary measures, it also highlights the importance of context setting; in the evaluative context no avoidance motivation was recorded but in the achievement context
those who were exposed to the same red stimuli did display avoidance motivation. Clearly
achievement context setting is important in further research.

Tanaka and Tokuno (2011) used the same risk taking measure as Elliot et al., (2007) that
asked participants to indicate the number of easy and moderately difficult questions they
would like to answer. In this experiment the colour manipulation took the form of a t-shirt
(red, green or white) worn by the experimenter under a partially buttoned up white
laboratory coat (i.e. only a small portion of the t-shirt was visible). Participants were tested
individually and in total the exposure to the colour manipulation was approximately 90s,
although no measures were taken to indicate if the participant had indeed noticed the
colour of the shirt. Participants who took the task with a red shirted experimenter chose
significantly easier question items. No significant differences were found between the green
and white conditions. What perhaps is most surprising about this finding is the covert and
seemingly inconsequential nature of the red coloured t-shirt.

In summary, there seems to be convincing evidence that the presence of red stimuli in an
achievement context elicits avoidance motivation. The research presented here
demonstrates a breadth of self-report and observed avoidance measures that are
influenced by quite subtle presentations of red stimuli.

2.2.2. Does avoidance motivation impair cognitive performance?

The work that examines the effect of avoidance motivation on cognitive performance is
voluminous and complex. Only a basic understanding is required for the purpose of this
thesis and is presented here for proof of concept only. This section will go on to briefly describe work that shows avoidance motivation impairs cognitive performance.

Elliot, Macgregor and Gable (1999) assessed the direction of motivation through established questionnaire measures in 164 undergraduate students and asked if exam performance on their mid-term papers could be predicted by the motivation direction. Results showed that those who had adopted approach goals performed better in the exam than those who had adopted avoidance goals. Later, Elliot & McGregor (2001) found that adopting approach based goals led to increased performance in short answer exams, multiple choice exams and overall exam performance compared to adopting avoidance based goals. The researchers of both these papers implicated study strategies as an important mediating variable of the effect. In general they state that approach motivation lends itself to study strategies that focus on understanding the information through deep processing, whilst avoidance motivation is more likely to result in study strategies that simply aim to pass the examination, characterised by the shallow processing of surface information. A striking feature of the two studies described above is that the examinations were authentic university assessments giving a high level of ecological validity.

Payne, Youngcourt and Beuabien, (2007) conducted a meta-analysis of 178 experiments examining the role of motivational direction on performance related outcomes. The meta-analysis revealed that goals pursued by avoidance motivation were negatively correlated with learning, academic performance, task performance and job performance. Dweck (1986) assessed avoidance motivation in children and found that those engaging in avoidance motivation tended to display negative affect (e.g. anxiety, worry) and negative
self cognitions that can have profound effects on cognitive performance. Dweck makes the important assumption that engaging in either approach or avoidance based motivation is not a reflection of absolute intellectual ability but instead impacts on the processes and strategies used to engage in cognitive tasks. This subtle influence of motivation on a wide range of cognitive processes means that stimuli can influence motivation outside of an individual’s awareness.

There are other instances where avoidance motivation (prompted by external cues similar to colour) have been shown to impede cognitive functioning that are relevant here. Ciani and Sheldon (2010) used similar methods to those of Elliot et al., (2007) to show that participants whose exam booklets had the letter 'F' printed on the front perform worse than those with exam booklets that displayed the letter 'A'. In an ingenious set of experiments Schuler, Brandstatter and Bauman (2013) asked if the numbers '1' and '6' could produce similar results. They ran their experiment in both Germany (where school assigned grades run from '1' [best] to '6' [worst]) and Switzerland (where the school grades run from '6' [best] to '1' [worst]). Their results showed that in both countries booklets that displayed the failure related number performed worse in a verbal analogy measure and scored higher in measures of avoidance motivation and fear of failure. This demonstrates that it is not only biologically ingrained associations of failure that can instigate avoidance motivation but socially learnt associations can have powerful implications too.

In summary, the wide range of research briefly described here does indicate that avoidance motivation can impede cognitive performance and this influence, despite being potent, is often subtle and can take place outside of an individual’s awareness.
2.2.3. How does colour induced avoidance motivation impair cognitive performance?

The previous two sections demonstrated that; the perception of red stimuli in an achievement context can prompt avoidance motivation and; that avoidance motivation can negatively influence task performance. However it is important to uncover how exactly avoidance motivation achieves such an effect because previous work in achievement motivation has found that mediating variables vary between settings. For example, Elliot, Macgregor and Gable (1999) found that avoidance motivation was mediated through the adoption of contrasting study strategies that in turn resulted in impaired exam performance. Brodish and Devine (2009) identified worry as the mediating variable that impaired performance as a consequence of avoidance motivation. Elliot and Mcgregor (1999) noted that general test anxiety and fear-of-failure are strong mediating variables regarding the impact of avoidance motivation on exam performance. During their initial investigation of the RACE, Elliot et al., (2007) considered numerous mediating variables (e.g. threat, challenge appraisals of the task, general arousal and perceived competence) but found that none of these could account for the increase in avoidance motivation. Elliot et al., (2007) also administered an achievement goals questionnaire to measure the explicit adoption of approach or avoidance goals and also found that this could not explain the increase in avoidance motivation. Therefore Maier, Elliot and Lichtenfeld (2009) proposed that participants’ impeded performance was not mediated through the explicit adoption of avoidance based goals but instead through implicit general avoidance motivation that is not mediated by any of the explicit measures reported previously in the literature. Therefore the mediating variable must be an implicit variable that does not influence goal adoption.
per se, but rather increases the availability and potency of avoidance motivation that underlies cognition.

Maier, Elliot and Lichtenfeld (2009) posited ‘attentional scope’ as a possible mediating variable. Attentional scope had previously been identified as an implicit aspect of perception that is influenced by the direction of motivation (Derryberry & Tucker, 1994). Attentional scope can be described as a continuum that ranges from narrow attention that focuses on small, concrete details (termed ‘local processing’) to broad attention that encompasses more general and contextual details (termed ‘global processing’). Maier, Elliot and Lichtenfeld propose that avoidance motivation facilitates local processing with the opposite true for approach motivation. Critically, local processing facilitates rigid thinking that utilises a narrow range of mental constructs, whilst global processing facilitates flexible thinking utilising a broad range of mental constructs. Tasks that require broad, flexible thinking (i.e. numerical or verbal problem solving) benefit from global processing whilst tasks that require narrower, rigid thinking (e.g. proof reading, search and rescue) benefit from local processing.

Thus, the proposition from Maier, Elliot and Lichtenfeld is that the perception of red instigates avoidance motivation that, mediated through local processing, leads to impaired cognitive performance in cognitive tasks that benefit from global processing (e.g. verbal analogy tasks or numerical sequencing tasks).

The onset of local/global processing occurs implicitly outside of an individual's awareness, thus explaining the absence of self-reported avoidance motivation in Elliot et al., (2007). Elliot et al., (2007) had previously shown that implicit measures of avoidance behaviour
were instigated through the perception of red stimuli (e.g. number of door knocks or the increased activation of the right frontal cortex) and Maier, Elliot and Lichtenfeld (2009) aimed to extend these findings through an attentional focus paradigm. They used a visual matching task drawn from Kimchi and Palmer (1982) that showed participants a figure of a target shape constructed from other shapes (e.g. a large square made up of smaller triangles). Participants were then asked to decide which of two comparison shapes (in this example a square or a triangle) the target shape was more similar to. Participants focussing their attention on the square were displaying global processing whilst participants focussing on the triangles were displaying local processing, as previously demonstrated by Gasper (2004). Maier, Elliot and Lichtenfeld also measured avoidance motivation through an established explicit indicator of avoidance motivation; the State-Trait Anxiety Inventory. Participants were given a folder containing an IQ test preceded by the colour stimuli. However, in this experiment the implicit and explicit measures of avoidance motivation were placed between the colour stimuli and IQ test. As expected, participants in the red (relative to grey) colour condition demonstrated higher levels of local processing in the matching task (the implicit measure) but no differences were found regarding the State-Trait Anxiety Inventory of avoidance motivation (the explicit measure). Next, the researchers aimed to demonstrate that local processing can impede cognitive performance without colour stimuli being present. Participant’s attentional scope was manipulated using the same visual matching stimuli in the previous experiment but instead of giving a free choice to select the local or global features (and thus measure attentional scope), participants were instead instructed to attend to either the local or global features of each picture (thus instigating changes in attentional scope). Following this manipulation a
numeric IQ test (as used in Elliot et al., 2007) was administered to the participants. Those participants who attended to the local figures performed significantly worse than those who had their attentional focus upon the global figures. The experiments described in Maier, Elliot and Lichtenfeld (2009) show that; (1) red stimuli can impede cognitive performance, (2) red stimuli facilitates local processing and (3) that local processing impedes cognitive performance. Finally, Maier, Elliot and Lichtenfeld enacted a 'measurement-of-mediation' approach that binds all factors described above into a single experiment. Participants viewed colour stimuli, completed a measure of attentional scope and took a numeric IQ test. Those who had viewed the red (relative to grey) stimuli demonstrated significantly more local processing and scored significantly worse on the numeric IQ test.

The systematic examination of the underlying mechanisms that drive the RACE suggest they are consistent and well understood. These theoretical experiments set a firm foundation on which to build a reliable body of work that examines the influence of red stimuli on cognitive performance. The inclusion of existing theory (i.e. that of ‘attentional scope’) helps set this work firmly in the context of other work regarding external cues and achievement motivation.

2.2.4. The Colour-in-Context Theory

In addition to the relevant achievement motivation theory described above, Elliot and Maier (2012) published the Colour-in-Context Theory (CIC), which describes six premises that describe colour effects. The CIC theory classifies context as a set of physical or psychological circumstances that encapsulate the environment the colour is presented in. However, these
premises can be seen less as a predictive theory and more a loose framework that describes
basic tenets that must be true if colour based effects exist.

The six premises set out by Elliot and Maier (2012) are:

1. Colour carries meaning.

The notion that colour carries semantic meaning serves to illustrate that colour has value
beyond aesthetics, i.e. colour acts as a communication signal as opposed to simply eliciting a
level of preference in an individual. Work demonstrating that, in an achievement context, 
red is intrinsically associated with failure and green with success support this premise
(Moller, Elliot & Maier, 2009; Rutchick, Slepian & Ferris, 2010).

2. Viewing colour influences psychological functioning.

This premise represents the next stage of colour effects by stating that the semantic
meaning that colour communicates (premise 1) interacts with and influences an individual's
psychological functioning. Several lines of work have shown that viewing red in an
achievement context (thus representing 'failure') causes an individual to engage in
avoidance based behaviours (in an attempt to avoid failure), such as choosing easier tasks
(Elliot et al., 2007), physically reclining from red stimuli or knocking fewer times on a door
(Elliot et al., 2009). However, this premise is fairly self-evident for a theory whose purpose is
to describes colour’s effect on psychological functioning. If this premise were not true then
there would be no need for the CIC theory to describe it.
3. Colour effects are automatic.

The effect of colour on psychological functioning occurs outside of an individual’s awareness. Post experiment questions work that examines the RACE has shown that participants were unaware of colour influencing them and in most cases participants were not overtly aware of colour being presented (Elliot et al., 2007; Lichtenfeld et al., 2009). No experiment that has found an influence of colour on cognition has uncovered data that would suggest participants were aware that colour was instigating changes in their cognition.

4. Colour meanings have two sources; learning and biology.

This assumption states that colour meanings drawn from learning and biology interact and change over time. For example, red may be associated with potential failure owing to its evolutionary associations with attack readiness in a superior opponent (Pryke, Andersson, Lawes & Piper, 2001) and associations with blood, injury and infection (Gerend & Sias, 2009). Subsequent pairing of red with 'stop signs' and educational mistakes (i.e. ‘red pen’) may have further compounded the semantic meaning of the colour. Research has demonstrated that these learnt associations are malleable; Zhang and Han (2014) examined Chinese stockbrokers, who work in an industry where red indicates a stock rise (a success outcome) and green represents a fall in stock value (a negative outcome). Zhang and Han found that the perception of red prompted approach based motivation in Chinese stockbrokers whilst in Chinese university students (who did not have experience of the Chinese stock market) the opposite effect was found.
5. Relations between colour perception and affect, cognition, and behaviour are reciprocal.

The notion that the effect of colour on cognition is reciprocal is the least developed of the six premises of the CIC. A body of research not related to colour perception has shown that visual perception is influenced by internal psychological states. Studies have shown that visual perception (specifically feature detection and shape/form evaluations) is influenced by the psychological state (i.e. the affect, cognition and behaviour) of an individual (Barsalou, 2008). Although theory suggests that this is true for colour perception too, little empirical evidence has been offered to support the idea (Elliot and Maier, 2012). Even though this premise is less developed, its importance to the model is clear if, as expected, colour perception is influenced by an individual’s cognitive state. It may be that measures of affect, behaviour and cognition are required to minimise unsystematic variation within data and identify colour effects.

6. Colour meanings and effects are context specific.

The final premise is key to understanding work on colour and cognition as it allows colours (e.g. red) to signify different and even contradictory meanings in different contexts. For example, in an achievement context red signifies danger and thus avoidance motivation. Yet in a mate evaluation context red signifies increased attraction (e.g. red love hearts, lipstick, roses) and thus approach motivation. The idea that colour holds different, often contradictory, meanings in different contexts has been well documented (Jung, Kim, & Han, 2011; Meier, D’Agostino, Elliot, Maier & Wilkowski, 2012) and so a context dependent module is key to any model that attempts to describe or predict colour effects on cognition.
Although the CIC theory accurately and succinctly describes the work regarding colour (specifically red) in both achievement and mate evaluation contexts its predictive power as a theory is lacking as the 6 premises are all very broad. Therefore the CIC theory can be seen more as a post-hoc model that describes a limited number of established colour effects rather than a model that can predict the effect of a range of colours in a range of contexts. Further work on this nascent idea may well develop the CIC theory into one that has predictive power but given the complexities of colour perception and achievement motivation across different contexts this may prove difficult.

The next section will describe replications and extensions of Elliot et al., (2007) that examine the RACE and a small selection of other relevant empirical studies that assess related topics. This section will first assess non-independent replications/extensions (i.e. those that have share authorship of Elliot et al., 2007) of the RACE before going to assess independent replications/extensions of the RACE.

2.3. Empirical evidence of the RACE

2.3.1. Non-independent replications and extensions of the RACE

Following the publication of Elliot et al., (2007), researchers from the same group (Maier, Elliot & Lichtenfeld, 2008) replicated experiment 4 from Elliot et al., (2007) twice. In the first replication they found that those in the red condition scored less than the grey condition and in the second replication they found that those in the red condition scored less than the green condition, replicating the findings in Elliot et al., (2007).
This group of researchers (Lichtenfeld, Maier, Elliot & Pekrun, 2009) went on to demonstrate that the RACE could be elicited not with chromatic stimuli but instead by presenting participants with the word 'red'. In the first of four experiments they placed participants into either the 'rot' (German for 'red') or 'ort' (German for 'place') group and completed the same verbal analogy measure used in Elliot et al., (2007). Those assigned to the red group scored significantly lower than those in the neutral group. In the next experiment participants were given a verbal analogy test. In this test the correct answer to one of the 20 questions was either (in reference to a tree) 'Red-Alder' or 'Grey-Alder'. Those who had viewed the red answer performed significantly worse on the whole test than those who had viewed the grey answer. In the third experiment the colour word was placed in a copyright label ("© by Hogrefe Series of Tests Red/Grey, 1978") located in the bottom right hand corner of the first three pages of the experiment materials before a numerical sequencing test (as used in Elliot et al., 2007). Again, those in the red condition scored significantly worse than those in the grey condition. In the final experiment this finding was replicated with the alteration of changing the word 'Grey' to 'Green' within the copyright label. What is most surprising about these findings is the subtlety and seemingly inconsequential placing of the colour stimulus and the large effects that were recorded. At no point were the participants instructed to view the colour stimuli and it was presented covertly. An interesting methodological consequence of this lexical presentation is that the use of a colour word negates many of the methodological issues associated with accurately reproducing colours (see appendix A for a description of such issues) and, if the effect persists, could be a cheap, consistent and effective method to uncover colour driven cognitive effects.
Elliot, Payen, Brisswalter, Cury and Thayer (2011) examined differences in heart rate variability (HRV) during an IQ test that followed the presentation of red stimuli. HRV is a natural process most commonly demonstrated by respiratory sinus arrhythmia (RSA) - the extent to which heart rate speeds up and slows down naturally during respiration. Previous research had shown that taxing cognitive tasks and experimentally manipulated negative emotions (induced through aversive film clips for example) both decrease HRV. The researchers here wanted to examine if a red stimulus (acting as a subtle threat cue) seen before an IQ test can also decrease HRV. Using the folder method from Elliot et al., (2007) results showed that red stimuli, relative to blue and grey, was associated with decreased HRV and, central to this discussion, decreased performance on the IQ test. An important note for this study is that the experimental method is somewhat protracted in relation to that of previous work in the area (e.g. Elliot et al., 2007; Maier, Elliot & Lichtenfeld, 2008) as the IQ test did not immediately follow the colour presentation. Instead the colour presentation was followed by a three minute HRV assessment and a mood and general activation questionnaire before the central measure of cognitive performance was finally undertaken. This gives some clues regarding the robustness and duration of the RACE decay following colour exposure as the RACE was still present following a three minute physiological assessment and a self-report cognitive assessment.

2.3.2. Independent replications and extensions of the RACE

A related experiment to the work of Lichtenfeld, Maier, Elliot and Pekrun, (2009) that also used colour words within the procedure was conducted by Ioan et al., (2007) in a study that was published around the same time of Elliot et al., (2007). This group of researchers viewed
colour effects not directly through the lens of approach/avoidance motivation theory but instead from an evolutionary biological viewpoint. Specifically from the viewpoint that the redness of skin in males is thought to signify social dominance through increased testosterone and blood flow that acts as a cue to other males (Drummond, 1997). Empirical evidence examining the influence of red uniform on sporting outcomes supports this notion (Hill and Barton, 2008; Atrill, Gresty, Hill and Barton, 2008) and so Ioan et al. aimed to extend this work within the laboratory. Thus, in a 'competitive context' Ioan and colleagues hypothesised that red cues (both natural and artificial) act as a negative distracter, particularly for men given their biological vulnerability to this effect. They used a classic Stroop test (Stroop, 1935), where participants are shown one of three colour words (here red, blue and green) that is printed in one of the three colours. The participant’s task is to report the colour a word is printed in. The colours were presented on a computer screen and were equated on brightness and saturation, although the screen they were presented on was not calibrated to display colour in a standardised manner. Results showed that, in males, Stroop interference (SI) was ≈50ms greater during red colour naming compared to the pooled green and blue colour naming conditions. Surprisingly females were found to be significantly faster (≈55ms) in red colour naming compared to the other two conditions. Whilst these findings could be explained by the dominance theory explanation it is also justifiable to apply the RACE theory to this work. The red stimuli could instigate avoidance behaviour thus impeding Stroop test performance for that condition. However, there are three discrepancies that are in contrast to the findings of Elliot and colleagues described previously. First, the negative effect of red stimuli is only found in males, a finding not present in the work conducted by Elliot and colleagues, despite the majority of participants
being female within their studies (see 2.5.3. for a description of gender within the work). Researchers examining the impact of red on dominance in humans (Wiedemann, Burt, Hill & Barton, 2015) and in animals (Milinski & Bakker, 1990; Setchell & Wickings, 2005) have also uncovered similar gender effects. The second point, linked to the first, is that females showed an advantage for viewing red stimuli, an effect not found in any other work regarding the RACE. Finally, Lichtenfeld et al., (2009) found that the word 'red' can elicit the RACE even when printed in another colour. This finding poses a problem when using a Stroop test as the dependent variable given the colour-word interaction. Therefore the interplay between the lexical and chromatic representations of the colour red could be more complex than they first appear. There is possibly some form of hierarchy that determines whether motivation orientation is determined by chromatic or lexical representations of colour when both are presented.

A similar gender effect of colour and cognitive performance was uncovered by Gnambs, Appel and Batinic (2010) who provide a conceptual replication of Elliot et al., (2007). Gnambs, Appel and Batinic asked participants to complete a 40 item online general knowledge test drawn from a standardised general knowledge test. Topics included biology, literature and history. In addition to this a multiple choice vocabulary test served as a general ability covariate. A coloured progress bar (either red or green) in the upper right hand side of the page served as the independent variable. Participants (drawn from an Austrian University) took the test from their home computers. Colours were defined using the HSL colour model and were equated on saturation and lightness using properly calibrated hardware when designed. However, as the test was taken remotely the participants would have viewed quite different colours owing to different monitor settings.
Results showed that those in the red condition performed worse in the general knowledge measure than the green group. However, this difference in means was driven by an underlying gender effect as males exhibited a stronger and significant effect of colour compared to females. This gender effect was somewhat unexpected given that the previous research by Elliot et al., (2007), Maier, Elliot and Lichtenfeld, (2008) and Lichtenfeld, Maier, Elliot and Pekrun, (2009) did not.

To further explore these gender effect Gnambs, Appel and Batinic (2010) completed a second experiment with the following alterations to the design: the comparison colour was changed from green to blue; the colour stimuli was changed from a progress bar to a 'forward' button; the general knowledge test was increased to 89 items and finally a third, mixed colour condition was added. In this third condition the first forward button was coloured red whilst the 15 subsequent forward buttons were coloured blue. This third condition was used to assess how 'strong' the RACE is, i.e. once instigated does it override other chromatic stimuli, or can its influence be easily varied through additional chromatic stimuli? The findings for this condition are particularly important for assessing the degree to which the RACE is prevalent in real world settings given the variety of colour stimuli present during ecologically valid cognitive tasks. As with the first experiment in Gnambs, Appel and Batinic only the male participants exhibited the RACE whereas females did not. Further analysis of the male data revealed that whilst those in the red condition scored lower than the blue condition there was no difference between the means of the mixed colour condition and either the red or blue conditions.
This conceptual replication of Elliot et al., (2007) demonstrates a robustness of the effect not before demonstrated. The effect was found with very little in the way of colour control (i.e. varying saturation and brightness due to the remote nature of the experiment) using a measure (general knowledge) that relies more on exposure to learning and encoding material, as opposed to innate ability (e.g. from verbal fluency tasks). The finding that the RACE only manifests in male participants replicates the gender effect from Ioans et al., (2007) although Gnambs, Appel and Batinic found no increase in performance from female participants as Ioans and colleagues did.

One of the first independent studies to conceptually replicate and extend the work on the RACE was conducted by Mehta and Zhu (2009) who assessed the impact of colour on performance in a variety of tasks. In particular they wanted to examine if a blue stimulus could facilitate approach motivation (mediated through global processing) and increase task performance, specifically tasks requiring creative cognition. In the first experiment they asked participants to solve three types of anagrams; approach based anagrams (e.g. tudeavnre = 'adventure'), avoidance based anagrams (e.g. neprevt= 'prevent') and neutral anagrams (e.g. procumet = 'computer), presented on either a red, blue or neutral colour background on a computer monitor. Mehta and Zhu do not report the total number of correct answers (data that would be beneficial comparing this to previous work on the RACE) but instead report the speed that correct answers were given. Results showed that avoidance based anagrams were solved quicker on a red background than the blue or neutral background with the opposite shown for approach based anagrams. Mehta and Zhu conclude that this implied stronger approach/avoidance motivation activation when completing the task. There are some issues with this experiment. The first regards the fact
that no specific criteria were given for deciding which words were to be used for the approach, avoidance and neutral anagram words and these seem to be subjective. For example, the word 'Olympic' is used as an approach based word, yet a survey (Survation, 2012) found that prior to the London 2012 Olympic Games, whilst 51.4% of respondents were positive towards the Olympics, 17.1% were negative and 30.4% were indifferent. Another issue regards the fact that no details were given concerning the number of anagrams that participants got correct, a key measure of performance. It is likely that there were no significant differences across conditions otherwise they would have been reported. A final issue, noted by Steele (2013) is that the anagrams (only three in each category) varied by word length and this acts as a confound variable given that shorter anagrams are solved quicker than longer anagrams. So although the research questions that Mehta and Zhu aim to set out are original and valid, the execution of the research and subsequent analysis means the results and implications are of questionable validity and reliability.

Steele (2014) attempted to directly and independently replicate the experiment reported in Mehta and Zhu (2009). Results showed no colour and word orientation interaction but did find that, after collapsing across the word orientation variable, anagrams were solved more quickly in the red condition than the blue but not the white condition. It is hard to integrate this finding into the RACE literature centred on Elliot et al, (2007) as that work focuses on absolute performance, not the speed of giving correct answers. It is plausible that avoidance motivation was causing participants to rush through the anagrams due to red's hypothesised threat cue properties but this is yet to be explored. An additional explanation is that the red and white coloured backgrounds were easier to read than the blue condition. Further work examining this effect has not been undertaken as the majority of research that
examines the RACE (including that within this thesis) measures performance as a measure of absolute performance.

In another experiment reported by Mehta and Zhu (2009) they identified a short term memory exercise as a 'detail-oriented' task that they thought would benefit from a local processing. Elliot et al., (2007) had previously shown that red stimuli can prompt local processing and so Mehta and Zhu asked if this could improve performance in a task that benefited from a narrow attentional scope. Participants were given two minutes to memorise a list of 36 words, displayed on a computer screen on top of either a red, blue or grey background. After a 20 minute delay participants were asked to recall the list. Those who studied the list on the red \( (M = 15.89, SD = 5.90) \) background made more correct recalls than the blue condition \( (M = 12.31, SD = 5.48) \). Surprisingly no mention is made within the manuscript of the grey condition; both its descriptive values and associated inferential statistics (regarding its difference to the red and blue conditions) are not reported. However, a figure at the foot of the document shows that the grey condition fell somewhere between the red and blue conditions (very approximately between 13.5 to 14.5 correct recalls) and is not likely to be significantly different from either chromatic condition. This poses an issue with Mehta and Zhu's position that red improves performance and that blue impedes performance in this memory task - if neither are different from the neutral condition then can the effect in question actually be present? Although the blue condition led to significantly more false recalls than both the red and grey conditions there was no difference in the total number of recalls between colour conditions. Also, Rollinson & Sowden (1997) found that participants required to read two passages from the Wechsler Logical Memory Test presented in one of four colours (red, yellow, green or blue) on a grey
background showed better immediate and delayed (24hrs) recall when the passages were presented in green or blue compared to red or yellow. This is the opposite of Mehta and Zhu’s finding with respect to red and blue and suggests a lack of consistency in the literature. To further explore the idea that performance in detail oriented tasks can be increased through viewing red stimuli, Mehta and Zhu identified another task that requires local processing; a proof reading task. Participants were given five sets of names and addresses that were either identical or very similar. Participants were tasked with identifying if the pairs were identical or not. Those in the red condition scored better than those in the blue condition and the neutral condition.

Moving away from tasks that require local processing towards those that benefit from global processing, Mehta and Zhu next looked at measures of creativity. Participants were given a Remote Associates Test (RAT) used to measure creativity presented on a coloured screen. A RAT gives four related words (e.g. 'shelf', 'read' and 'end') that are related to an overall concept (e.g. 'book'). Results showed that more correct answers were given when presented on a blue screen compared to the red or grey screen. However, more recent work has shown that a RAT, despite being used for decades as a measure of creativity, is actually a measure of analytical and convergent thinking, as opposed to divergent or 'creative' cognitive processing (Lee, Huggins & Therriault, 2014). Such a realisation is beneficial for this thesis as it aligns the results from the RAT measure more closely to the verbal and numerical fluency measures used in other papers such as Elliot et al., (2007). This interpretation of the RAT indicates that the red condition is not impeding performance in the way that the RACE has been previously shown to do so. Instead what this experiment shows is that red is not significantly decreasing cognitive performance (relative to grey) but
that blue is increasing cognitive performance relative to both red and grey stimuli. Such a finding is inconsistent with the majority of literature on the RACE.

The notion that colour can influence creativity has also been examined by Lichtenfeld, Elliot, Maier and Pekrun (2012). Whilst this paper does not directly assess the RACE it is worth mentioning given the closeness of its methods and theory. This paper comes from a similar perspective as demonstrated in Mehta and Zhu (2009) in that Lichtenfeld and colleagues contend that if red stimuli induce avoidance motivation then other colours (green in this case) could induce approach motivation and improve creative task performance. In the first experiment participants were recruited through a crowd sourced internet workforce (Amazon M-Turk) and were required to take an ‘alternate uses test’ (where participants must report alternate uses for everyday objects) from a remote location. Immediately before taking the task the participants were ostensibly shown an identification number coloured either red, green or white. Similar to the Mehta and Zhu (2009) findings, whilst the total number of creative responses was similar across all conditions the quality of responses in the green condition was independently graded as being more creative. In three subsequent studies taking place in German college classrooms, participants were shown to be more creative when primed with green (relative to red, blue, white, grey) colour stimuli. Two interesting points are found in the results of these experiments. First, the researchers, unlike Mehta and Zhu (2009), found no improvement of creativity for the blue condition over the grey condition. Second, the researchers found no influence of red on creativity compared to the grey condition. This is despite a hypothesised difference due to the narrow attentional scope prompted by red stimuli limiting creative thought (Friedman and Forster, 2001). This finding represents the first published failure to replicate by the authors of Elliot
et al., (2007) with regard to the RACE. However, this failure to replicate is not stated, rather they conclude that although red stimuli impedes analytic thought (e.g. measures of verbal fluency) it has no impact on creative thought. Additionally they posit that it is green that influences creativity and not blue as Mehta and Zhu (2009) had found. Whether this explanation is a product of post-hoc rationalisations or has some a priori basis is unclear.

Moving away from measures of creativity, Hulshof (2013) examined the role of colour and scent on cognitive performance. In particular Hulshof wanted to determine how high arousal colours and aromas (red and peppermint in this study) and low arousal colours and scents (here blue and sandalwood) compared in their influence on cognitive performance. The colour manipulation took the form of coloured fabric sheets hung from the wall giving an environment that was not ecologically valid, both in terms of lighting and decor. Participants, drawn from the Dutch working population, took two measures of cognitive performance; a 10 item IQ-style test and a separate memory and recall test. Regardless of room scent, those in the red condition scored worse than those in the blue condition on the IQ test. However, the opposite effect was found for the memory and recall test with those in the blue condition performing worse than those in the red condition. Interestingly this effect is similar to that reported in Mehta and Zhu (2009) who also found that red stimuli seemed to facilitate improved recall from memory. How this fits with existing theory requires further examination, especially given the Mehta and Zhu (2009) finding that (in reference to the neutral condition) it is the blue stimuli that are impeding task performance, not the red stimuli improving performance. Regardless, it does seem that the RACE could be task dependent as well as context dependent and that some specific tasks could be unaffected by RACE. One factor that may have biased results is that it is likely the
experimenters were not blind to the participant’s condition and so experimenter bias could be a factor. This is not addressed explicitly in the manuscript but it is stated that participants were tested in blocks of scent condition (i.e. all 30 blue sandalwood participants then all 30 red sandalwood participants) meaning that randomisation would be very difficult to achieve. Additionally it would be difficult for the experimenter to avoid noticing the overt colour/scent manipulation present unless a second researcher, blind to the research aims, carried out the research.

Another experiment in the Netherlands by Buiks (2013) gave participants a line graph and a bar graph with either a blue or red coloured background and asked them multiple choice questions regarding interpretation of the information. The experiment took place on computers in large groups and colours were equated on brightness and chroma values. No details were given regarding if the monitors were calibrated to display colour consistently. Results showed that those exposed to the red background were significantly worse at interpreting the data on the graph compared to those presented with a blue background. It is plausible that a constriction of attentional focus as a result of the RACE would result in worse performance on this task although it is also possible that a narrow attentional scope would improve performance given the detail oriented nature of the task. Further work examining tasks that require broad/narrow attentional processing is required to further explore this.

Jung, Kim, and Han, (2011) assessed the RACE and the role of colour on mate evaluation in a single experiment. They presented a single colour that preceded both an achievement context task and a mate evaluation task. Participants (both male and female) watched a
video of a woman (dressed in either a red, blue or grey t-shirt) read out a list of 10 words. Participants then took a word recognition task before rating the female reader in terms of attractiveness. Results showed that those in the red condition rated the female as more attractive than those in the blue condition. No details are given regarding the attractiveness ratings from the grey condition, however the figure provided shows that those in the grey condition rated the female more attractive than the blue condition but less attractive than the red condition. Regarding the measure of cognitive performance, those in the blue condition performed significantly better than those in both other conditions and, unexpectedly, those in the red condition scored significantly better than those in the grey condition. Such findings do not align with the general view of the RACE that states that red stimuli leads to a decrease in performance compared to all colours, especially as grey is often used as a comparison colour within existing work (e.g. Elliot et al., 2007; Maier, Elliot & Lichtenfeld, 2008).

Several researchers have examined the influence of paper colour on performance within exams, with the majority of studies showing no effect. Whilst not specifically using the paradigm set out in Elliot et al., (2007), these papers are deemed similar enough to warrant inclusion within this section of the literature review. Schmidt, Ruskell and Kohl (2013) gave participants examination questions printed on shop bought paper coloured either grey, blue, green or cherry. Participants completed three mandatory exams for an electricity and magnetism module on a physics course at the Colorado School of Mines. Crucially, the Colorado School of Mines already used these paper colours to differentiate between versions of exam papers so this method demonstrated excellent ecological validity as authentic materials were used and all participants would be highly motivated to pass the
examination. Results showed no differences between colour conditions on any measure of absolute performance or the rate of improvement over examinations. A number of factors may be important when critiquing this experiment; first it may be that the student's high levels of natural motivation overrode the environmental motivations influenced by colour. Another factor is that the colours (in particular cherry) used were not adequate to invoke the avoidance motivation (i.e. cherry did not elicit responses of 'threat' or 'danger'). The fact that heightened motivation may negate any colour effects means that there may be very limited applications of the RACE to real world situations.

Smajic, Merritt, Banister and Blinebury (2014) aimed to replicate the RACE but in a true applied examination setting in a similar way to Schmidt, Ruskell and Kohl (2013). They manipulated a real examination in a U.S. university that contributed towards the degree mark for the 137 business students. Half the students received a test booklet with a red cover and half with a green cover (details of the specific colour values or how they were produced are not given). The examination itself was from an organisational behaviour course and consisted of 50 multiple choice questions. Results showed no difference of exam performance as a function of cover colour. In a second experiment Smajic et al., (2014) took participants from three separate courses (two from psychology and one from business) and gave them realistic (but not real) examinations. Instead of colour manipulation being present only on the front cover they also placed colour stimuli on each page of the examinations thus giving a longer and repeated exposure to the colour cue. Again no effect of colour on performance was found. Measures of anxiety taken both post and during the experiments also revealed no effect of colour.
Pedley and Wade (2013) aimed to examine both the effect of the RACE on cognitive performance and also the underlying theory that the RACE is mediated through local processing. They used a psychophysical paradigm using an established glass pattern task that had previously been used to investigate local and global shape detection in early visual perception (Mandelli & Kiper, 2005). Glass patterns are constructed by layering an image of random dots over an identical, but slightly rotated, image of random dots that thus give rise to a 'global' shape form. Glass patterns can be used to investigate local and global aspects of vision by asking participants to attend to either the 'local' dots or the 'global' shape. Work on neural correlates of visual perception support this statement (Wilson, Wilkinson & Assad, 1997). Pedley and Wade examined participants’ ability to discriminate between changes in global form or local dot contrast. These tasks were constructed to include a colour cue (either red or blue instructions and fixation cross) and this meant it was possible to determine if colour cues could influence local/global aspects of form detection. Prior research suggested that, in the red condition, performance should be higher in the local processing task and lower in the global processing task. Results showed no significant difference between performances in either task as a function of colour. Nor were any differences found in the rate of improvement of the tasks as a function of colour. Pedley and Wade proposed that whilst previous research indicates that the RACE can influence performance in high-level cognitive tasks (e.g. measures of verbal fluency), no such effect is present in tasks that rely on neuronal signal processing in the early visual system.

A paper using similar methods to Pedley and Wade was published by Sorokowski and Szmajke (2011). They gave participants one of three simple 'arcade style' video games to play. In one, players had to shoot balls that descended from the screen, in another players
had to avoid falling balls hitting a panel located at the bottom of the screen and in another had to escape from balls attacking them from all sides. For all three games each had a red, blue and black condition that corresponded to the colour of the balls. This paper aimed to assess the 'visibility theory' put forward by Hill and Barton (2005), not the avoidance based theory that grounds the RACE literature. However, its methods and implications are better discussed here given the application to cognitive performance, especially because the first game described stems from an approach-based goal (shooting balls) whilst the last two games stem from avoidance-based goals (avoiding balls). The achievement motivation theory would imply that in games requiring approach motivation red stimuli would impede performance whilst in games requiring avoidance based motivation red stimuli would improve performance. Results here suggest that neither is true. In the approach based games (shooting balls) those in the red condition outperformed those in the other colour conditions, whereas there was no colour based difference in either of the avoidance based games.

The recent paper from Larsson and von Stumm (2015) represents the first time the RACE has been explored in a UK population to assess high level cognitive tasks. They recognise the possible importance of the RACE in assessing intelligence but highlight four limitations of previous research. First, the relatively small sample size in previous work is highlighted, however the authors make no mention of power, a much more relevant concept to address given that the appropriateness of a sample size is dependent on the expected effect size and the number of independent variables levels. Second, the authors note that the majority of previous work has tested college or student populations. Larsson and von Stumm suggest that if the RACE, either wholly or in part, is a result of education based achievement context
associations (i.e. red pen feedback), then other populations may not be as sensitive to the RACE. Third, the authors highlight that no other work in the field has addressed how crystallised (opposed to fluid) measures of intelligence are influenced by the RACE and that this form of knowledge based intelligence is often important in applied contexts. The authors fail to note Gnambs, Appel and Batinic’s (2010) paper that explicitly replicates the influence of the RACE on crystallised intelligence in an online setting. Finally, the authors wanted to uncover if participants were aware that the colour red was influencing performance.

To examine this, based on their previous work on intelligence, Larsson and von Stumm proposed that self-estimated ability of task performance would also vary as a function of colour cue. Larsson and von Stumm report a well-executed experiment based partially on the first experiment from Elliot et al., (2007). They tested 20 participants at a time in a lecture hall at a UK university. Each participant was given testing booklets that, in the top right hand corner of each page, contained the participant's table number made up of a nonsense letter and number combination. This number was written in either red or blue ink on pages containing the dependent measures. Participants first completed five measures of verbal ability (a fluid intelligence measure) before taking a five minute break and then completed a 13 subject area knowledge quiz covering topics such as art, music, science, fashion and sport (a crystallised intelligence measure). After each measure of intelligence participants indicated how well they thought they had done. Finally participants completed a personality measure and gave background information in a separate booklet. This testing took around two hours and participants were rewarded through non-performance dependent monetary compensation. Results showed no differences between either colour
condition for both the fluid and crystallised measures of intelligence. Additionally no differences were found for participant's self-estimated performance in either measure nor for intelligence as a whole. Larsson and von Stumm conclude that their findings reflect the fact that in an adult population the RACE is not present as adults have been away from education (and thus educational feedback) for too long and have since become desensitised to the RACE. One issue with Larsson and von Stumm’s study is the number and duration of the intelligence tests, which took participants two hours to complete. Such demands can leave participants feeling fatigued, which could mask any colour based effects.

The work described above demonstrates that the publication of Elliot et al., (2007) has been influential in the emergence of a research movement that focuses on the RACE. One important distinction to make is that whilst the non-independent replications and extensions of Elliot et al., (2007) have all found significant effects the picture is less clear when assessing the independent replications and extensions of Elliot et al., (2007). It is unclear whether this is due to differences in research design, participants, colour stimuli or another unknown factor. Whilst the body of literature as a whole is impressive it is still in a nascent stage and several issues, discussed in the next section, require further research that will form the basis of this thesis.

2.4. The need for further research

2.4.1. The need to assess the RACE in an applied context

The research documented in the literature review demonstrates that colour can influence cognitive performance in some settings. Clearly such a reliable and significant influence on
cognition is provocative and, if the effect manifests in an applied setting, has great importa
tance to many aspects of everyday life. Consider all the achievement contexts (i.e. set
tings that contain a pass or fail outcome) where colour induced impaired cognitive perfor
mance could be detrimental to an individual; a student's education at school, health
care professionals or business strategists making important decisions; elite sports men and women preparing mentally for an event; a coach driver making decisions regarding speed and lane selection; or when pursuing a hobby such as playing video games or chess. The number of instances that pair achievement pursuit with some kind of colour stimuli is immeasurable in the human experience and such an effect must be understood. If viewing a red stimulus can reliably and significantly impair cognitive performance, such applications could be brought to the general public quickly and relatively cheaply; for example the exclusion (or reduction) of red items within schools, workplaces and in vehicles. However promising the RACE seems, the literature describes a still emerging field, the potential of which (in terms of application) has not yet been outlined. The majority of studies, especially those with higher impact, have to date been conducted in controlled laboratory settings (e.g. Elliot et al., 2007; Mehta & Zhu, 2009). Therefore, further work assessing if the RACE manifests in more ecologically valid settings is essential to the progression of the body of literature. If the research in this thesis concludes that the RACE is a robust effect that does manifest in applied settings then suggestions can be made regarding the most efficient method to minimise the RACE.
2.4.2. The need to investigate if the RACE varies between genders

Another issue that requires further research regards the gender effect discrepancy noted between studies. In general, the research suggests the RACE is present in both genders (e.g. Elliot et al., 2007; Maier, Elliot and Lichtenfeld, 2008; Lichtenfeld, et al., 2009 and Elliot et al., 2011). However, a handful of studies, most notably the work of Gnambs, Appel and Batinic (2010) and to an extent that of Ioans et al., (2007), state that such a gender difference does exist and that the RACE is only present in males. Research in other domains such as sporting performance and animal studies also suggests that the negative effect on performance emerges only in males (Hill & Barton, 2005).

To add to the debate, some research has uncovered the RACE using female only samples or samples that contain a majority of female participants. Such work suggests that the RACE present in mixed gender samples is not driven by the male participants. See table 2.1. for a brief description of the gender split within the body of research. Given that a convincing argument to rectify this incongruence has yet to be offered, such basic discrepancies within the literature only serve to weaken its standing. Therefore the resolution of such disagreements is paramount. This thesis will examine gender differences in each experiment to aid in resolving this issue.
Table 2.1. Number of male and female participants in studies that have uncovered the RACE.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Male Participants</th>
<th>Female Participants</th>
<th>Ratio of F to M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elliot et al., (2007), Exp. 1.</td>
<td>18</td>
<td>53</td>
<td>2.9</td>
</tr>
<tr>
<td>Elliot et al., (2007), Exp. 2.</td>
<td>4</td>
<td>42</td>
<td>10.5</td>
</tr>
<tr>
<td>Elliot et al., (2007) Exp. 3.</td>
<td>5</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>Elliot et al., (2007) Exp. 4.</td>
<td>25</td>
<td>32</td>
<td>1.28</td>
</tr>
<tr>
<td>Maier, Elliot &amp; Lichtenfeld (2008), Exp. 1.</td>
<td>11</td>
<td>9</td>
<td>1.2</td>
</tr>
<tr>
<td>Lichtenfeld et al., (2009), Exp. 1.</td>
<td>33</td>
<td>16</td>
<td>0.48</td>
</tr>
<tr>
<td>Lichtenfeld et al., (2009), Exp. 2.</td>
<td>0</td>
<td>44</td>
<td>N/A</td>
</tr>
<tr>
<td>Lichtenfeld et al., (2009), Exp. 3.</td>
<td>26</td>
<td>14</td>
<td>0.53</td>
</tr>
<tr>
<td>Lichtenfeld et al., (2009), Exp. 4.</td>
<td>14</td>
<td>6</td>
<td>0.42</td>
</tr>
<tr>
<td>Elliot et al., (2011) Exp. 1</td>
<td>30</td>
<td>3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Note: ‘Exp’ = ‘Experiment’. 'F' = female, 'M' = male.

2.4.3. The need for replication

"Confirmation comes from repetition. Any attempt to avoid this statement leads to failure and more probably to destruction" (Tukey, 1969; pp. 84).

Whilst some authors contend that the RACE has already been replicated a satisfactory number of times (Meier, Hill, Elliot & Barton, 2015) this thesis argues that a thorough examination of the current literature highlights that further direct, independent replications of the RACE are required.

2.4.3.1. Why is replication important?
Replication is key to Psychology as it can aid in identifying, diagnosing, and minimizing many issues concerning the reliability and reproducibility of published effects (Makel, Plucker & Hegarty, 2012). Jasny, Chin, Chong and Vignieri (2011) deem replication the 'gold standard' of scientific practice but also highlight that replication is complicated through large data sets, the complexity of questions being asked, longitudinal and expensive studies that are difficult to run and increased pressure to advance research (as opposed to merely verifying existing work) in the typical 'publish-or-perish' working paradigm (Mcgrail, Rickard & Jones, 2006). Academia is more competitive than ever and academic employers value high impact publications more than they do unpublished, valiant, failed attempts to replicate an existing effect. This is evident in UK academic institutions requiring the explicit use of bibliometric measures for both new employment and internal promotions (van Dalen and Henkens, 2012). This practice is so ingrained within academia that early career researchers can take the view that: ‘You can fail to do everything else as long as you have lots and lots of papers’ (Anderson, Ronning, De Vrie & Martinson, 2007, pp. 443), and that what you write no longer matters. Instead how often, where and with whom you write is much more important to an academic institution given that university rankings rely heavily on objective research impact factors and not on the subjective intellectual worth of research (van Dalen & Henkens, 2012).

Such competition and pressures to publish manifest in 'publication bias'; the tendency for journals to publish positive results (they, like any other publication, are subject to the financial pressures of gaining subscriptions and so often only want to publish 'exciting' or 'novel' findings) and subsequently for researchers to submit only those papers that find effects, whilst quietly filing away those papers that found null effects (termed the 'file-
drawer problem’). This resistance from journals to publish null effects and the subsequent reluctance in researchers to submit such findings serve to feed each other in exacerbating the problem. Not only are journals often hesitant to publish non-significant results they also often base their decision on arbitrary criteria (most notably the fabled $p<0.05$ value), even when such null hypothesis significance testing has long been discredited as a means to decide if a finding is significantly meaningful (Rozeboom, 1960; Gill, 1999; Nickerson, 2000).

Additionally, journals are more likely to publish novel results rather than successful replications, even when they explore the same effect (Pashler & Harris, 2012). There is however, a current trend towards increasing the number of replications in Psychology. Between the year 2000 and 2012 the number of replications in the top 100 psychology journals has increased from 1.07% (1900 to 1999) to 2.39% (Makel, Plucker & Hegarty, 2012) and replications have been the focus of a number of special issues of journals focussing solely on replication (e.g. Perspectives on Psychological Science, 7, [6], 2012; Science, 334, [6060], 2011). However, even with this increase, a rate of only 2.39% of replications within the whole of psychology is a remarkably low figure for a practice deemed to be integral to the scientific method.

An important distinction is required between direct replications (those that mimic the methods and statistical analysis of existing work exactly) and conceptual replications (those that examine the same effect but with a tweaked method, e.g. a different presentation of colour stimuli or a different measure of performance).

2.4.3.2. The need for direct replications
Some researchers, even senior psychologists, argue that conceptual replications are more effective than direct replications as they not only assess the reliability of the finding but also they gauge the generalisability of the effect (Carpenter, 2012). However, Pashler and Harris (2012) advocate that this is not the case. They argue that a body of literature containing few direct replications and many conceptual replications can skew the reality of an effect as conceptual and direct replications are not equal in their likelihood of being formally published or disseminated through more informal means. Consider the consequences of finding a null result in both direct and conceptual replications. Neither are likely to be published (Makel, Plucker & Hegarty, 2012), however null results in a direct replication are much more likely to be disseminated through informal channels within the field (e.g. personal emails or conference discussions) than a null result in a conceptual replication. Pashler and Harris (2010) state one reason for this is that null results from a conceptual replication will typically be explained and assimilated into existing literature through the argument that the replicator has moved too far from the original method and has introduced some confounding variable. Consequently, although the researcher may conclude that the original effect is not robust or generalisable to other methods/contexts/populations, they are not as likely to question the original study had the researcher employed a direct replication. Next consider the consequences of finding a significant result. A conceptual replication is more likely to be published given that it has extended the original experiment and has advanced the field. A successful direct replication however is less likely to be published given that it has merely confirmed what was already known and thus, its publication is likely to elicit a lot less 'gossip' within academic circles.
This means that a conceptual replication is more likely to be published than a direct replication regardless of whether an effect is found. Given the aforementioned publish-or-perish mind frame, researchers are more likely to skip a direct replication and, for their own and their institution's reputation, engage in conceptual replications. Pashler and Harris (2012) conclude that this tendency to perform conceptual replications over direct replications interacts with and 'amplifies the publication bias problem in a rather insidious fashion' (pp. 533).

2.4.3.3. The need for independent replications

A related problem to that of the need for direct replications is the need for independent replications that share no overlap of authors from the original paper. In an analysis of 500 replication articles Makel, Plucker and Hegarty (2012) calculated that whilst independent replications were successful in 64.6% of cases, those with at least one author from the original study were successful in 91.7% of articles and that this is likely to stem from both intentional and unintentional experimenter bias. It is therefore essential that direct replications of the RACE are conducted outside of the authorship of Elliot et al., (2007) to rule out experimenter effects.

2.4.3.4. Identifying the RACE as a field that is susceptible to publication bias

Pashler and Harris (2012) go on to suggest properties of a field of research that is susceptible to the issues regarding the interaction of publication bias and conceptual replications. They suggest published effects that are both easy to uncover (i.e. a short study with no expensive equipment) and exciting (i.e. a novel or implausible effect) tempt
hundreds of researchers and students to attempt to conceptually replicate the effect. Regardless of whether the effect exists or not some of these studies will produce significant results and these are likely to be published. Additionally, such conditions would promote the use of small, low-powered conceptual replication attempts and these practices have also been found to increase the probability of attaining ‘\(p < 0.05\)’ (and thus have a higher probability of being published), regardless of whether the effect actually exists (Bakker, van Dijk & Wicherts, 2012).

The properties described by Pashler and Harris are clearly present in the RACE literature; the effect is novel and exciting and requires very few materials to replicate. Therefore the body of literature that describes the RACE is susceptible to conceptual replications and publication bias skewing the validity, reliability and applicability of the RACE. With this knowledge, care must be taken in ensuring that sufficient direct replications of the RACE are conducted before extensions to the effect are sought.

2.4.3.5. Existing replications of the RACE

Within the literature there are a lack of independent, direct replications of the original experimental paradigm set out in Elliot et al., (2007). For clarity it is the methods from experiments 2, 3 and 4 that are being replicated. The colour presentation technique from experiment 1 (hand written pen on paper) is not consistent enough in its application to be replicated. The three experiments in question constitute a pre-task presentation of a printed colour square (for between 2s and 5s) and a dependent measure of fluid intelligence (verbal or numerical fluency tasks) with the results showing that the red condition impedes task performance compared to the other conditions across both genders.
A cursory glance at the literature regarding the RACE would imply that the RACE described in Elliot et al., (2007) has been successfully replicated on numerous occasions. This is formalised in Meier, Hill, Elliot and Barton (2015) who state that: *Other researchers successfully replicated the red effect by using additional chromatic controls and cognitive tasks (e.g., verbal reasoning, working memory, attentional interference, detection, creativity, and language proficiency; Buiks, 2013; Elliot et al., 2011; Gnambs, Appel, and Batinic, 2010; Houtman and Notebaert, 2013; Hulshof, 2013; Ioan et al., 2007; Jung, Kim, and Han, 2011; Maier et al., 2008, Thorstenson*, submitted; Yamazaki and Eto, 2011) *. pp. 12.

However further investigation of this literature reveals that these studies are either conceptual replications of the RACE or direct replications that share authorship with Elliot et al., (2007). It is not that researchers are attempting to pass off experiments as being direct and independent replications but rather that they do not place importance on such research practices. This is often implicitly communicated. In the quote given above the researchers clearly make no attempt to suggest the listed replications are direct or independent replications. But the studies listed are provided as strong evidence for successful replications of the RACE despite the problems associated with shared authorship and conceptual replications in a field that is susceptible to publication bias. A thorough investigation of the extant literature on the RACE suggests that an independent, direct replication of the RACE is yet to be conducted on the effect described in Elliot et al., (2007).
2.4.4. The need for meta-analysis

Meta-analysis combines effect sizes from several studies and, by 'weighting' studies as a function of statistical power, allows the size of the effect to be determined with greater precision (Cummings, 2013). In essence, a meta-analysis seeks to reduce unsystematic variation in individual experimental analyses to provide a smaller estimation interval of the 'true' effect (Begg, Cooper & Hedges, 1994). To reduce the impact of publication bias and the file-drawer problem (see 2.5.2.) a meta-analysis includes 'grey literature'. Grey literature is defined as unpublished literature that has not been disseminated for a variety of reasons. These include the work having been submitted but not yet published, work that does not meet publication criteria (i.e. statistical analyses that result in \( p > 0.5 \)) or experiments that were conducted for student projects (Conn, Valentine, Cooper & Rantz, 2003).

Although a meta-analysis is a powerful tool for aggregating research (Cooper & Hedges, 1994) and is becoming ever more popular in Psychology (Cummings, 2013) it is not a panacea able to draw conclusions from a complex and disorganised body of research. Lispey and Wilson (2001) compare both strengths and weaknesses of the method. They found that whilst meta-analysis has the potential to represent findings in a sophisticated, objective and differential manner, meta-analyses are still prone to misinterpreting effect sizes for a variety of reasons. These include using 'blemished studies', comparing studies that are not equivalent (the degree to which two studies are equal is often subjective) and the problems associated with identifying the extent of the unknown grey literature. However, these issues can be avoided to some degree by conducting a range of meta-analyses using different
inclusion criteria and also by recognising that a meta-analysis does not give a definitive answer and can only be used as a tool to interpret and evaluate a body of research.

A meta-analysis of the work on the RACE is yet to be conducted. A well conducted meta-analysis of the RACE is required to both examine the experiments already published in a quantitative manner and to further establish the field as a credible area of interest for future researchers. In doing so it may be possible to uncover how robust the RACE is and in turn the degree to which it is a meaningful effect in applied contexts.

2.5. The aims of this thesis

The four issues that describe the need for further research will be addressed throughout this thesis. In short the experiments will collectively seek to provide data to achieve the following aims:

1. To assess the RACE in an applied context.

2. To investigate the gender discrepancy within existing work.

3. To provide an independent, direct replication of the RACE as well as providing other conceptual replications where relevant.

4. To construct and run a meta-analysis of published and grey literature.

In addressing these four aims the thesis will contribute to the RACE literature using a range of research methods that span both primary research experiments and secondary research natural experiments (i.e. those using observational data). In particular the thesis will draw
on contemporary statistical methods (e.g. interval estimation, meta-analysis) and will attend specifically to the accurate production and measurement of colour stimuli. In doing so the thesis will contribute to the debate by investigating if the RACE can be reliably replicated in both laboratory and applied contexts and if gender influences the onset or strength of the RACE.

2.6. Outline of experimental chapters

2.6.1. Chapter 3

Chapter 3 describes a natural experiment that assesses if the RACE influences primary school students who wear red uniform. Using secondary research methods experiment 1 examines if schools that require red colour uniform are more likely to record lower measures of academic achievement (drawn from the Department for Education). The theory that underpins the RACE (see 2.2.) suggests that in an achievement context, such as that found in school examinations, the presence of red coloured stimuli (e.g. uniform) could impede performance in examinations through the onset of avoidance motivation. Experiment 1 examines the RACE in an applied context (aim 1) and by assessing gender specific outcome measures of academic achievement also investigates the gender discrepancy within existing work (aim 2). The analysis demonstrates a weak, non-significant, but consistent effect of red uniform colour negatively influencing measures of academic success (27 of 29 outcome measures) in UK primary schools. However, further analysis of the data show that lower performing schools were more likely to require red coloured uniform and so the direction of the (non-significant, very small) effects is unclear.
Chapter 4 comprises two experiments that move towards primary research experimental methods. Experiments 2 and 3 assess if red coloured stimuli (e.g. the background on a computer display) impede performance in a measure of verbal analogies taken in an online setting (accessed remotely). These experiments examine the RACE in an online (i.e. applied) setting (aim 1) and specific analyses assess whether or not gender influenced the RACE (aim 2). The experimental methods used to determine these aims represent a conceptual replication of the RACE (aim 3).

Chapter 5 examines the RACE in an experiment designed to more precisely control stimulus colour than the previous online experiments, whilst also recruiting from a more homogenous participant pool. Experiment 4 gave participants (recruited from a homogenous college student population) a verbal analogy task in a classroom setting. In this setting much more control over extraneous colour stimuli was possible. Experiment 4 examined the RACE in an applied context (i.e. an ecologically valid college classroom; aim 1) and provided a conceptual replication of the RACE (aim 3). Too few male participants were recruited to investigate the gender discrepancy in previous work. Following experiment 4, a consultation with Prof. Elliot (lead author of Elliot et al., 2007) highlighted several experimental method changes that were implemented in experiment 5.
2.6.4. Chapter 6

Chapter 6 describes two experiments that examine the RACE in a controlled laboratory setting that implement the recommendations made by Prof. Elliot. Experiments 5 and 6 assessed if a red (relative to white) stimulus impeded performance in a verbal analogy task. These changes include removing external performance contingencies, making immediate feedback salient and minimising stereotype threat. These experiments move to a controlled laboratory setting to ensure that no extraneous colour stimuli could interfere with the independent variable colour stimuli. Therefore experiments 5 and 6 can be considered a direct, independent replication of the second experiment by Elliot et al., (2007), thus investigating aim 2 of this thesis.

2.5.4. Chapter 7

Chapter 7 describes a natural experiment that assesses and interprets a broad range of research that examine the RACE. Experiment 7 employs a meta-analysis of both published and unpublished research on the RACE to determine the most likely effect size estimate. This investigates the fourth aim of this thesis; to construct and run a meta-analysis of published and grey literature.
Chapter 3: Using secondary research methods to assess the RACE in an applied setting

3.1. Experiment 1: The influence of uniform colour on academic achievement in primary schools

3.1.1. Introduction

The previous three chapters described general colour research, research that specifically examines the RACE and method considerations of work that attends to the RACE. These three chapters served to determine avenues for further research (see section 2.5. for a summary) as well as highlighting possible pitfalls in research design that could potentially undermine findings. This chapter, and the four subsequent chapters describe the empirical work undertaken for this thesis.

This thesis aims to use a wide range of research methods to examine the RACE including both primary and secondary research methods. At the beginning of a research project it is secondary research methods that have the fewest barriers to entry as ethical clearance is normally not required and the complexities of recruiting and running participants are avoided. Therefore, secondary research represents the optimal method of initiating research as it can be undertaken whilst other, more time consuming, aspects of subsequent experiments (e.g. gaining ethical clearance, creating stimuli) can be undertaken. Thus, this strategy represents the most efficient use of time to undertake doctoral research.

Within colour psychology there are numerous examples of secondary research methods being used. One example is using data from elite sport contests to determine if uniform colour can predict performance outcomes. Sports that have been shown to confer a colour
based performance advantage include boxing, tae kwon do, greco-roman and freestyle wrestling (2005a), judo (Rowe, Harris & Roberts, 2005), English Premier League association football (Atrill et al., 2007; Allen & Jones, 2014) and virtual 'first-person-shooter' competitions (Ilie et al., 2008). However, these colour effects are not consistent across all sports or even across studies assessing the same sport. Using similar methods to the research described above, no effect of colour was uncovered in Spanish association football (Garcia-Rubio, Picazo-Tadeo & Gonzalez-Gomez, 2011) or German association football (Kocher and Sutter, 2008) or in the fight outcomes of Mixed-Martial-Arts (MMA) bouts in the Ultimate Fighting Championship (UFC; Pollet & Peperkoorn, 2013). See appendix B for a detailed description of colour psychology that examines sporting outcomes.

Secondary research methods have also been employed in the colour-attraction domain. Elliot and Pazda (2012) examined dating websites to determine if 'relationship intentions' (i.e. is an online profile interested in casual sex or interested in serious relationships) varied as a function of female clothing colour in profile pictures. They found women wearing red clothing in profile pictures were significantly more likely to indicate they were interested in casual sex.

Although colour effects have been examined using secondary research methods in the sporting and attraction domains, no study has used secondary research methods to assess the RACE.

One instance where cognitive performance is paired with overt colour 'branding' is in UK schools where students are required to wear uniform typically consisting of grey trousers, a white shirt and a coloured sweatshirt or blazer. This uniform colour varies between schools.
It is likely that this uniform branding is more overt when wearing a sweatshirt compared to a blazer as a sweatshirt provides a ‘block’ of colour, whereas blazers are often worn open at the front, making the white shirt and tie more prominent. Therefore, when assessing the impact of uniform colour on academic achievement, it is optimal to examine schools that require sweatshirts be worn. In the UK 94% of primary schools require sweatshirts are worn, compared to only 25% of secondary schools (Davies, 2015) and so primary schools represent the optimal sample to be examined. The colour of the sweatshirt varies between schools and so represents a randomly occurring independent variable that can be used within statistical analysis of a natural experiment.

All primary schools in the UK follow the Key Stage 2 (KS2) curriculum that is taught to students between the ages of 7 and 11 (Education Act, 2002). At the end of KS2 students sit examinations in English, mathematics and science and are also given separate 'teacher assessed' marks. These results are published as Key Performance Indicators (KPI's) that describe a school's academic performance. These KPI’s include both student examination results and teacher assessment scores. These KS2 KPI’s can serve as dependent variables in an analysis that asks if KPI’s vary as a function of school uniform colour.

As these KPI’s emerge from a setting that can be characterised as an ‘achievement context’ (see 2.1. for a description of the importance of an achievement context), previous work on the RACE would predict that schools requiring a red uniform perform worse than those wearing other colour uniforms. Whether the RACE emerges in such an applied, ecologically valid environment replete with other colour stimuli and other motivational sources (e.g. parental input, quality of teaching) is of primary importance to understanding the
robustness of the RACE. Additionally, it is of particular interest to see if any colour based effects vary as a function of gender, given the gender effect discrepancy within the current literature (see 2.4.4.). For example, whether the RACE emerges only in male students, or if the effect size is larger in males when comparing the performance of genders.

Based on previous research in the RACE literature it is expected that schools requiring red uniforms would perform worse than schools requiring other uniform colours in measures of academic achievement (the KPI’s) and that no differences are observed between any other colours. This study informs two aims of the thesis set out in section 2.5. By using existing data drawn from ecologically valid settings it is possible to assess the RACE in an applied context (aim 1) for all genders as well as assessing the effect on each gender separately (aim 2).

If the RACE is uncovered from this analysis it is easy and cheap to apply to UK schools by altering school uniform requirements. The effect of school uniform colour is especially interesting given that the UK is something of an anomaly regarding enforced school uniform; its requirement is rare in both mainland Europe and Northern America (Northen, 2011). If the enforcement of coloured uniform does reliably impact academic achievement then its application within UK schools will require reconsideration.

This chapter will go on to describe an analysis that determines if UK primary school KPI’s vary as a function of primary school uniform colour.
3.1.2. Method

3.1.2.1. Sampling method

To obtain the relevant information a data set ("National curriculum assessments at key stage 2 in England: 2012") containing statistics for 'Key Stage 2' (KS2) curriculum assessments and review outcomes in the 2011/2012 academic year was obtained from the Department for Education (2013). This data set contained 135 assessment based key performance indicators (KPI's) for 15,917 primary schools split across 149 local education authorities (LEAs).

3.1.2.1.1. Primary schools

To ensure the sample reflected the wide range of primary schools in the UK six LEAs were randomly selected; two from the 1st quartile (high performing LEAs), two from the inter-quartile range (typical performing LEAs) range and two from the 4th quartile (low performing LEAs). The rankings were drawn from the Department of Education (2013) national statistics and are calculated using pooled KPI’s.

Randomisation was achieved through a random number generator using numbers that related to each LEAs rank. Table 3.1 provides the location, level of performance, overall rank and number of primary schools within each of the selected LEAs.
Table 3.1. Location, level of performance, overall rank and number of primary schools within each of the selected LEAs.

<table>
<thead>
<tr>
<th>LEA</th>
<th>Performance</th>
<th>Rank</th>
<th>Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bromley</td>
<td>High performing LEA</td>
<td>16</td>
<td>120</td>
</tr>
<tr>
<td>Greenwich</td>
<td>High performing LEA</td>
<td>31</td>
<td>97</td>
</tr>
<tr>
<td>Swindon</td>
<td>Typical performing LEA</td>
<td>65</td>
<td>95</td>
</tr>
<tr>
<td>Sunderland</td>
<td>Typical performing LEA</td>
<td>90</td>
<td>81</td>
</tr>
<tr>
<td>Coventry</td>
<td>Low performing LEA</td>
<td>125</td>
<td>89</td>
</tr>
<tr>
<td>Kingston upon Hull</td>
<td>Low performing LEA</td>
<td>147</td>
<td>81</td>
</tr>
</tbody>
</table>

Note: LEA = ‘Local Education Authority’.

3.1.2.1.2. Key Performance Indicators (KPI’s)

Within the data set 135 KPI’s of student assessment were available. Of these, 35 KPI’s reported relevant achievement outcomes for students. The other 100 KPI’s were not relevant to the analysis as they described student assessments for sub-populations within the data set (e.g. ethnicity, religion, or whether English was an additional language) or were not measures of absolute performance (i.e. measures of progress). Of these 35 KPI’s, 6 were comprised of other KPI’s within the data set and were removed from the data set to ensure data was independent. For example ‘Pupils achieving level 3 or below in both English and mathematics’ is the composite score of ‘Pupils achieving level 3 or below in English’ and ‘Pupils achieving level 3 or below in mathematics’. However, composite KPI’s that related to previous academic years were retained as the individual KPI’s that determine these composite KPI’s were unavailable within the data set and so no conflict of independence
occurred. The interpretation of these composite KPI’s must be carefully considered as combining independent data into a composite score can mask or reverse effects that are present in the separate measures, i.e. 'Simpsons Paradox' (Blyth, 1972).

Overall, these criteria resulted in 29KPI’s retained for the analysis. These measures that span assessments in English, mathematics and science can be split into three assessment categories; measures of attainment (exam assessed), measures of attainment (teacher assessed) and historical measures of attainment (exam assessed). The three sub-categories and their respective KPI’s are listed in table 3.2. The majority of KPI’s have positive valence (i.e. a higher value means better performance) although several have negative valence (i.e. a lower value means better performance).

Table 3.2. Description of each KPI, its KPI code and its assigned assessment category.

<table>
<thead>
<tr>
<th>KPI</th>
<th>KPI code</th>
<th>Assessment Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pupils achieving level 4 or above in English</td>
<td>KPI-1</td>
<td>Measures of attainment (EA)</td>
</tr>
<tr>
<td>2. Pupils achieving level 5 or above in English</td>
<td>KPI-2</td>
<td>Measures of attainment (EA)</td>
</tr>
<tr>
<td>3. Pupils achieving level 3 or below in English*</td>
<td>KPI-3*</td>
<td>Measures of attainment (EA)</td>
</tr>
<tr>
<td>4. Pupils achieving level 4 or above in mathematics</td>
<td>KPI-4</td>
<td>Measures of attainment (EA)</td>
</tr>
<tr>
<td>5. Pupils achieving level 5 or above in mathematics</td>
<td>KPI-5</td>
<td>Measures of attainment (EA)</td>
</tr>
<tr>
<td>6. Pupils achieving level 3 or below in mathematics*</td>
<td>KPI-6*</td>
<td>Measures of attainment (EA)</td>
</tr>
<tr>
<td>7. Boys achieving level 4 or above in both English and mathematics</td>
<td>KPI-7</td>
<td>Measures of attainment (EA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>8.</td>
<td>Girls achieving level 4 or above in both English and mathematics</td>
<td>KPI-8</td>
</tr>
<tr>
<td>9.</td>
<td>Boys achieving level 5 or above in both English and mathematics</td>
<td>KPI-9</td>
</tr>
<tr>
<td>10.</td>
<td>Girls achieving level 5 or above in both English and mathematics</td>
<td>KPI-10</td>
</tr>
<tr>
<td>11.</td>
<td>Pupils achieving level 4 or above in English TA</td>
<td>KPI-11</td>
</tr>
<tr>
<td>12.</td>
<td>Pupils achieving level 5 or above in English TA</td>
<td>KPI-12</td>
</tr>
<tr>
<td>13.</td>
<td>Pupils achieving level 3 or below in English TA</td>
<td>KPI-13*</td>
</tr>
<tr>
<td>14.</td>
<td>Pupils achieving level 4 or above in mathematics TA</td>
<td>KPI-14</td>
</tr>
<tr>
<td>15.</td>
<td>Pupils achieving level 5 or above in mathematics TA</td>
<td>KPI-15</td>
</tr>
<tr>
<td>16.</td>
<td>Pupils achieving level 3 or below in mathematics TA*</td>
<td>KPI-16</td>
</tr>
<tr>
<td>17.</td>
<td>Pupils achieving level 4 or above in science TA</td>
<td>KPI-17</td>
</tr>
<tr>
<td>18.</td>
<td>Pupils achieving level 5 or above in science TA</td>
<td>KPI-18</td>
</tr>
<tr>
<td>19.</td>
<td>Pupils achieving level 3 or below in science TA*</td>
<td>KPI-19*</td>
</tr>
<tr>
<td>20.</td>
<td>Pupils achieving level 4 or above in reading TA</td>
<td>KPI-20</td>
</tr>
<tr>
<td>21.</td>
<td>Pupils achieving level 5 or above in reading TA</td>
<td>KPI-21</td>
</tr>
<tr>
<td>22.</td>
<td>Pupils achieving level 3 or below in reading TA</td>
<td>KPI-22*</td>
</tr>
<tr>
<td>23.</td>
<td>Pupils achieving level 4 or above in writing TA</td>
<td>KPI-23</td>
</tr>
<tr>
<td>24.</td>
<td>Pupils achieving level 5 or above in writing TA</td>
<td>KPI-24</td>
</tr>
</tbody>
</table>
25. Pupils achieving level 3 or below in writing TA* KPI-25* Measures of attainment (TA)

26. Pupils achieving level 4 or above in both English and mathematics in 2009 KPI-26 Historical measures of attainment (EA)

27. Pupils achieving level 4 or above in both English and mathematics in 2010 KPI-27 Historical measures of attainment (EA)

28. Pupils achieving level 4 or above in both English and mathematics in 2011 KPI-28 Historical measures of attainment (EA)

29. Pupils achieving level 4 or above in both English and mathematics in 2012 KPI-29 Historical measures of attainment (EA)

Note: EA = Exam Assessed, TA = Teacher Assessed,* = negative valence KPI.

3.1.2.1.3. School uniform colour

The uniform colour required for each primary school was identified by accessing the schools’ websites and searching for a uniform list. This information was not available for all schools. Reasons for not being able to acquire this information include; the schools not having a website, the website being unavailable (i.e. 'under construction') at the time of research, the information not being available from the website, the school not having a required uniform or the school using different colours for different year groups within a school. The number of school uniform colours acquired for each LEA are given in table 3.3.

Table 3.3. The number of primary schools within each LEA and the number of uniform colours acquired for each of those selected LEAs.

<table>
<thead>
<tr>
<th>LEA</th>
<th>Schools</th>
<th>Uniform colours identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bromley</td>
<td>110</td>
<td>43</td>
</tr>
<tr>
<td>Greenwich</td>
<td>97</td>
<td>34</td>
</tr>
<tr>
<td>Swindon</td>
<td>95</td>
<td>39</td>
</tr>
<tr>
<td>Sunderland</td>
<td>81</td>
<td>72</td>
</tr>
</tbody>
</table>
The colour terms used to describe uniform colour varied greatly between schools. For example; blue, mid-blue, navy blue and royal blue to describe various blue hues. To increase the sample size of each colour condition colours were grouped into general colour terms. In general, uniform colours were grouped by their hue, e.g. 'green', 'bottle green' and 'emerald green' were all considered 'green'. However, an exception was made regarding maroon and burgundy uniform colours. These were not placed in the red hue category, despite being a dark red hue. The reason for this is that a key assumption of the RACE is that red stimuli impede performance relative to all other colours, therefore it was of specific importance that the red uniform colours used within the analysis were consistent and appropriate whilst other colours could be grouped together. The specific red hue that has previously been used to elicit the RACE is a relatively bright red (see section appendix A for a description of colours used within the RACE literature) that is dissimilar to dark red maroon and burgundy hues. Additionally, the theoretical assumptions that ground the RACE posit that the red based association is drawn from associations with vivid or light reds, e.g. blood, infection (Gerend & Sias, 2009) and changes in facial coloration (Pryke, Andersson, Lawes & Piper, 2001).

Grouping the colour terms in this way does mean uniform colours that differ on saturation and brightness are considered functionally equivalent. For example; dark green is less bright
than emerald green but both are considered 'green' for the purpose of this analysis. Whilst some literature has argued that saturation and brightness influences performance in tasks (e.g. Elliot et al., 2007; Maier, Elliot & Lichtenfeld, 2008) the RACE has been uncovered on numerous occasions where saturation and brightness have varied (e.g. Elliot et al., 2007; Gnambs, Appel & Batinic, 2010; Lichtenfeld, Elliot & Maier, 2012) including when using secondary research methods (e.g. Atrill, Gresty, Hill & Barton, 2007; Allen & Jones, 2014). See appendix A for a description of work that has uncovered the RACE whilst not controlling for saturation and brightness.

For a list of all colour terms uncovered, their assigned colour group and their frequency see table 3.4. Only schools that required red (72 schools), blue (139 schools) and green (38 schools) colour uniforms were retained in the subsequent analyses. This was because the other colour groups: burgundy (14 schools); purple (15 schools); grey (5 schools); black (4 schools); brown (3 schools); and teal (1 school) contained far fewer schools and so could have biased statistical analyses due to small sample sizes. This inclusion of only schools that require red, blue and green uniform increases the statistical power of the analysis and also ensures a greater degree of accuracy in the calculation of descriptive statistics. This also reflects the methods of previous work on the RACE that has typically used green or blue as a chromatic comparison to red (e.g. Elliot et al., 2007; Maier, Elliot & Lichtenfeld, 2008; Gnambs, Appel & Batinic, 2010).

Table 3.4. Colour terms drawn from school websites, their assigned colour group and the frequency of instances where the colour is required in schools.
<table>
<thead>
<tr>
<th>Colour term</th>
<th>Colour group</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>Red</td>
<td>68</td>
</tr>
<tr>
<td>Scarlet</td>
<td>Red</td>
<td>1</td>
</tr>
<tr>
<td>Pillar box red</td>
<td>Red</td>
<td>1</td>
</tr>
<tr>
<td>Bright red</td>
<td>Red</td>
<td>2</td>
</tr>
<tr>
<td>Blue</td>
<td>Blue</td>
<td>41</td>
</tr>
<tr>
<td>Royal Blue</td>
<td>Blue</td>
<td>36</td>
</tr>
<tr>
<td>Mid blue</td>
<td>Blue</td>
<td>1</td>
</tr>
<tr>
<td>Light Blue</td>
<td>Blue</td>
<td>1</td>
</tr>
<tr>
<td>Navy Blue</td>
<td>Blue</td>
<td>60</td>
</tr>
<tr>
<td>Green</td>
<td>Green</td>
<td>22</td>
</tr>
<tr>
<td>Jade Green</td>
<td>Green</td>
<td>3</td>
</tr>
<tr>
<td>Emerald Green</td>
<td>Green</td>
<td>1</td>
</tr>
<tr>
<td>Bottle Green</td>
<td>Green</td>
<td>9</td>
</tr>
<tr>
<td>Dark green</td>
<td>Green</td>
<td>2</td>
</tr>
<tr>
<td>Forest Green</td>
<td>Green</td>
<td>1</td>
</tr>
<tr>
<td>Burgundy</td>
<td>Burgundy</td>
<td>6</td>
</tr>
<tr>
<td>Maroon</td>
<td>Burgundy</td>
<td>9</td>
</tr>
<tr>
<td>Purple</td>
<td>Purple</td>
<td>15</td>
</tr>
<tr>
<td>Grey</td>
<td>Grey</td>
<td>5</td>
</tr>
<tr>
<td>Black</td>
<td>Black</td>
<td>4</td>
</tr>
<tr>
<td>Brown</td>
<td>Brown</td>
<td>3</td>
</tr>
<tr>
<td>Teal</td>
<td>Teal</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>292</strong></td>
</tr>
</tbody>
</table>
3.1.3. Results

3.1.3.1. Analytical approach to results within this thesis

The results section here and hereafter throughout the thesis interprets data through both traditional dichotomous analysis (e.g. NHST) and contemporary estimation analysis (e.g. confidence intervals on mean differences). The implementation of estimation analysis in addition to NHST aids in assessing a more accurate picture of data analysis given that NHST (and in particular the rigid application of using \( p < 0.05 \) as criteria) has long been questioned as a sole means to draw meaningful conclusions from data (Rozeboom, 1960; Gill, 1999; Nickerson, 2000; Cummings, 2013).

Estimation analysis will be expressed using both the raw mean difference between conditions and Cohen's \( d \) effect size. The mean difference analysis allows intuitive interpretation of the results within a paper, e.g. a mean difference of 2 can be interpreted directly as participants correctly answering 2 more verbal analogies. The Cohen's \( d \) analysis allows interpretation of results between experiments that use different measures. This allows an experiment using a 20 item verbal ability measure to be compared directly to an experiment using a 60 item numerical fluency measure. The Cohen's \( d \) effect sizes will be interpreted through the approximate criteria given in Cohen (1988), e.g. \( d > 0.8 \) = a 'strong effect'. However, both Cohen (1988) and Baguley (2009) warn that the interpretation of 'canned effect sizes' is fraught with many dangers and that context is important in assessing the effect sizes. As well as considering context (i.e. raw mean differences), the use of 95% confidence intervals (95% CI) of these values allows a more precise estimation of their value.
and these will be used in conjunction with the canned effect sizes that Cohen provides to provide a more measured estimation.

Where appropriate both standard deviations and standard errors will be given so that readers can understand both the shape of the data in the sample and the probable accuracy of the mean in relation to the population, without any mental calculation. Where appropriate the data will be presented both graphically and in table form. This allows a reader to first draw a quick interpretation about the data (by considering the graph) and then, if required, to delve deeper into the data by assessing the raw values. The addition of raw values ensures transparent statistical reporting and also allows quick access to values required for meta-analysis of this data.

3.1.3.2. Descriptive statistics and null hypothesis significance testing (NHST) analysis

Of the 29 KPI's only 4 conformed to parametric assumptions; KPI18, KPI21, KPI24 and KPI27. Additionally, the sample sizes of each colour group were uneven. Therefore, all 29 KPI's will be analysed using a non-parametric Kruskall-Wallis test of differences. Any significant differences will be addressed through running post-hoc Mann-Whitney tests. From herein the different coloured uniform groups will be referred to as; the red condition, the blue condition and the green condition.

Several KPI's were missing from the data set drawn from the Department of Education. The reasons for these omissions were due to industrial action taking place that year, the KPI's being inapplicable to the school (e.g. measures of boys’ performance in 'all-girl' schools) or were unstated. Where values are not included, replacement values were not calculated and
this is reflected in the varying sample sizes between KPI's. Table 3.5. lists the sample size, mean, standard deviation and standard error of each KPI by colour condition. Table 3.6. lists the results of the Kruskall-Wallis analysis for each KPI.

Table 3.5. Sample size, mean, standard deviation and standard error of each uniform colour for each KPI.

<table>
<thead>
<tr>
<th>KPI</th>
<th>Colour</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPI-1</td>
<td>Red</td>
<td>72</td>
<td>82.86%</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-1</td>
<td>Blue</td>
<td>139</td>
<td>85.17%</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-1</td>
<td>Green</td>
<td>38</td>
<td>84.92%</td>
<td>0.10</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-2</td>
<td>Red</td>
<td>72</td>
<td>33.79%</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-2</td>
<td>Blue</td>
<td>139</td>
<td>37.45%</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-2</td>
<td>Green</td>
<td>38</td>
<td>37.42%</td>
<td>0.17</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-3*</td>
<td>Red</td>
<td>72</td>
<td>15.43%</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-3*</td>
<td>Blue</td>
<td>139</td>
<td>12.94%</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-3*</td>
<td>Green</td>
<td>38</td>
<td>13.18%</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-4</td>
<td>Red</td>
<td>71</td>
<td>80.59%</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-4</td>
<td>Blue</td>
<td>138</td>
<td>82.64%</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-4</td>
<td>Green</td>
<td>38</td>
<td>84.50%</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-5</td>
<td>Red</td>
<td>72</td>
<td>39.00%</td>
<td>0.20</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-5</td>
<td>Blue</td>
<td>139</td>
<td>44.91%</td>
<td>0.23</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-5</td>
<td>Green</td>
<td>38</td>
<td>43.87%</td>
<td>0.23</td>
<td>0.04</td>
</tr>
<tr>
<td>KPI-6*</td>
<td>Red</td>
<td>71</td>
<td>17.85%</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-6*</td>
<td>Blue</td>
<td>138</td>
<td>16.67%</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-6*</td>
<td>Green</td>
<td>38</td>
<td>14.53%</td>
<td>0.10</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-7</td>
<td>Red</td>
<td>72</td>
<td>67.19%</td>
<td>0.24</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-7</td>
<td>Blue</td>
<td>137</td>
<td>70.55%</td>
<td>0.22</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-7</td>
<td>Green</td>
<td>38</td>
<td>72.89%</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-8</td>
<td>Red</td>
<td>72</td>
<td>77.81%</td>
<td>0.16</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-8</td>
<td>Blue</td>
<td>137</td>
<td>79.96%</td>
<td>0.19</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-8</td>
<td>Green</td>
<td>38</td>
<td>82.47%</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-9</td>
<td>Red</td>
<td>72</td>
<td>20.40%</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-9</td>
<td>Blue</td>
<td>137</td>
<td>23.60%</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-9</td>
<td>Green</td>
<td>38</td>
<td>26.29%</td>
<td>0.18</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI</td>
<td>Color</td>
<td>Value</td>
<td>Percentage</td>
<td>Baseline</td>
<td>Delta</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>-------</td>
<td>------------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>KPI-10</td>
<td>Red</td>
<td>72</td>
<td>22.78%</td>
<td>0.15</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-10</td>
<td>Blue</td>
<td>137</td>
<td>24.69%</td>
<td>0.18</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-10</td>
<td>Green</td>
<td>38</td>
<td>28.37%</td>
<td>0.21</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-11</td>
<td>Red</td>
<td>72</td>
<td>83.17%</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-11</td>
<td>Blue</td>
<td>139</td>
<td>84.77%</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-11</td>
<td>Green</td>
<td>38</td>
<td>84.47%</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-12</td>
<td>Red</td>
<td>72</td>
<td>31.54%</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-12</td>
<td>Blue</td>
<td>139</td>
<td>36.81%</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-12</td>
<td>Green</td>
<td>38</td>
<td>36.63%</td>
<td>0.18</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-13</td>
<td>Red</td>
<td>72</td>
<td>15.63%</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-13</td>
<td>Blue</td>
<td>139</td>
<td>13.63%</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-13</td>
<td>Green</td>
<td>38</td>
<td>13.61%</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-14</td>
<td>Red</td>
<td>72</td>
<td>83.06%</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-14</td>
<td>Blue</td>
<td>139</td>
<td>84.63%</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-14</td>
<td>Green</td>
<td>38</td>
<td>86.00%</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-15</td>
<td>Red</td>
<td>72</td>
<td>33.83%</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-15</td>
<td>Blue</td>
<td>139</td>
<td>38.40%</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-15</td>
<td>Green</td>
<td>38</td>
<td>41.08%</td>
<td>0.20</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-16</td>
<td>Red</td>
<td>72</td>
<td>15.85%</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-16</td>
<td>Blue</td>
<td>139</td>
<td>13.86%</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-16</td>
<td>Green</td>
<td>38</td>
<td>12.29%</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-17</td>
<td>Red</td>
<td>72</td>
<td>84.18%</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-17</td>
<td>Blue</td>
<td>139</td>
<td>86.74%</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-17</td>
<td>Green</td>
<td>38</td>
<td>85.66%</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-18</td>
<td>Red</td>
<td>72</td>
<td>32.33%</td>
<td>0.15</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-18</td>
<td>Blue</td>
<td>139</td>
<td>39.29%</td>
<td>0.20</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-18</td>
<td>Green</td>
<td>38</td>
<td>35.68%</td>
<td>0.20</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-19*</td>
<td>Red</td>
<td>72</td>
<td>14.63%</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-19*</td>
<td>Blue</td>
<td>139</td>
<td>11.59%</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-19*</td>
<td>Green</td>
<td>38</td>
<td>12.92%</td>
<td>0.10</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-20</td>
<td>Red</td>
<td>72</td>
<td>84.25%</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-20</td>
<td>Blue</td>
<td>139</td>
<td>85.23%</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-20</td>
<td>Green</td>
<td>38</td>
<td>85.13%</td>
<td>0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-21</td>
<td>Red</td>
<td>72</td>
<td>38.61%</td>
<td>0.17</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-21</td>
<td>Blue</td>
<td>139</td>
<td>42.10%</td>
<td>0.17</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-21</td>
<td>Green</td>
<td>38</td>
<td>45.42%</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-22*</td>
<td>Red</td>
<td>72</td>
<td>14.79%</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI</td>
<td>N</td>
<td>df</td>
<td>Kruskall-Wallis (H)</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>----</td>
<td>---------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>KPI-22*</td>
<td>Blue</td>
<td>139</td>
<td>13.35%</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-22*</td>
<td>Green</td>
<td>38</td>
<td>13.74%</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-23</td>
<td>Red</td>
<td>72</td>
<td>78.39%</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-23</td>
<td>Blue</td>
<td>139</td>
<td>79.72%</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-23</td>
<td>Green</td>
<td>38</td>
<td>79.55%</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-24</td>
<td>Red</td>
<td>71</td>
<td>29.82%</td>
<td>0.19</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-24</td>
<td>Blue</td>
<td>139</td>
<td>34.44%</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-24</td>
<td>Green</td>
<td>38</td>
<td>33.97%</td>
<td>0.20</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-25*</td>
<td>Red</td>
<td>72</td>
<td>26.50%</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-25*</td>
<td>Blue</td>
<td>139</td>
<td>26.14%</td>
<td>0.24</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-25*</td>
<td>Green</td>
<td>38</td>
<td>24.00%</td>
<td>0.22</td>
<td>0.04</td>
</tr>
<tr>
<td>KPI-26</td>
<td>Red</td>
<td>72</td>
<td>68.17%</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-26</td>
<td>Blue</td>
<td>135</td>
<td>71.24%</td>
<td>0.18</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-26</td>
<td>Green</td>
<td>38</td>
<td>75.29%</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-27</td>
<td>Red</td>
<td>53</td>
<td>61.23%</td>
<td>0.25</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-27</td>
<td>Blue</td>
<td>100</td>
<td>59.60%</td>
<td>0.31</td>
<td>0.03</td>
</tr>
<tr>
<td>KPI-27</td>
<td>Green</td>
<td>26</td>
<td>62.85%</td>
<td>0.30</td>
<td>0.06</td>
</tr>
<tr>
<td>KPI-28</td>
<td>Red</td>
<td>64</td>
<td>71.73%</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-28</td>
<td>Blue</td>
<td>119</td>
<td>72.92%</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-28</td>
<td>Green</td>
<td>34</td>
<td>78.00%</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-29</td>
<td>Red</td>
<td>63</td>
<td>76.41%</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>KPI-29</td>
<td>Blue</td>
<td>116</td>
<td>79.95%</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>KPI-29</td>
<td>Green</td>
<td>33</td>
<td>80.36%</td>
<td>0.10</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Note: * = negative valence. KPI. ‘N’ = sample size, ‘M’ = mean, ‘SD’ = standard deviation, ‘SE’ = standard error.

Table 3.6. Sample size, degrees of freedom, Kruskall-Wallis value and associated p-value for each KPI.
KPI-11 249 2 2.21 .331
KPI-12 249 2 4.90 .086
KPI-13* 249 2 2.26 .323
KPI-14 249 2 3.19 .203
KPI-15 249 2 6.19 .045
KPI-16 249 2 3.80 .149
KPI-17 249 2 4.68 .097
KPI-18 249 2 6.60 .037
KPI-19* 249 2 4.83 .089
KPI-20 249 2 1.38 .501
KPI-21 249 2 3.04 .218
KPI-22* 249 2 1.43 .488
KPI-23 249 2 1.07 .586
KPI-24 248 2 2.42 .299
KPI-25* 249 2 1.00 .605
KPI-26 245 2 6.82 .033
KPI-27 179 2 0.61 .738
KPI-28 217 2 6.07 .048
KPI-29 212 2 3.95 .139

Note: * = negative valence KPI. Boldface indicates p-values below 0.05. ‘N’ = sample size, ‘df’ = degrees of freedom.

In total, four KPI’s contained colour conditions that were significantly different; KPI-21, KPI-24, KPI-32 and KPI-36. All significant differences were in the hypothesised direction, i.e. the red condition score is lower than both the green and blue condition scores. Mann-Whitney tests were used to follow up these significant findings. These results are listed in table 3.7. Given that the theory of the RACE is one directional (i.e. red stimuli impede performance) and only predicts differences between red stimuli and any other colour (i.e. no expected differences between green and blue conditions), one tailed p-values are reported for the comparisons between the red condition and each of the other two conditions. All differences are in the hypothesised direction, although only 6 of the 8 comparisons are statistically significant. The effect sizes calculated represent a small effect size (Field, 2008). No obvious pattern emerges to describe those non-significant findings.
Table 3.7. Mann-Whitney values, z-score, p values and r effect size comparing the blue and green colour conditions to the red condition.

<table>
<thead>
<tr>
<th>KPI</th>
<th>Comparison Colour</th>
<th>U</th>
<th>z</th>
<th>p</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPI-15</td>
<td>Blue</td>
<td>4109.50</td>
<td>-2.128</td>
<td>.015*</td>
<td>-0.203</td>
</tr>
<tr>
<td>KPI-15</td>
<td>Green</td>
<td>1037.00</td>
<td>-2.081</td>
<td>.018*</td>
<td>-0.143</td>
</tr>
<tr>
<td>KPI-18</td>
<td>Blue</td>
<td>3917.50</td>
<td>-2.585</td>
<td>.005*</td>
<td>-0.246</td>
</tr>
<tr>
<td>KPI-18</td>
<td>Green</td>
<td>1236.00</td>
<td>-0.827</td>
<td>.204</td>
<td>-0.053</td>
</tr>
<tr>
<td>KPI-26</td>
<td>Blue</td>
<td>4145.00</td>
<td>-1.741</td>
<td>.041*</td>
<td>-0.166</td>
</tr>
<tr>
<td>KPI-26</td>
<td>Green</td>
<td>950.50</td>
<td>2.625</td>
<td>.005*</td>
<td>-0.181</td>
</tr>
<tr>
<td>KPI-28</td>
<td>Blue</td>
<td>3499.50</td>
<td>-.903</td>
<td>.118</td>
<td>-0.086</td>
</tr>
<tr>
<td>KPI-28</td>
<td>Green</td>
<td>774.00</td>
<td>2.345</td>
<td>.010*</td>
<td>-0.161</td>
</tr>
</tbody>
</table>

Note. *' indicates p values below 0.05. One tailed p-values are reported.

3.1.3.3. Estimation analysis

Having analysed the KPI's using NHST, the analysis will now examine underlying trends within the data using estimation analysis. Estimation analysis allows a more measured and contextually based approach to defining whether effects exist compared to the arbitrary cut off points used within NHST (Gill, 1999; Nickerson, 2000). To conduct estimation analysis the mean difference and its 95% CI was calculated for each of the red vs. blue and red vs. green comparisons across all the KPI's. These are reported in table 3.8.

Table 3.8. Mean difference, SE and 95% CI for each colour comparison across each KPI

<table>
<thead>
<tr>
<th>KPI</th>
<th>Colour</th>
<th>Mdiff</th>
<th>SE</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPI-1</td>
<td>Blue</td>
<td>-2.30%</td>
<td>1.87%</td>
<td>-5.99%</td>
<td>1.38%</td>
</tr>
<tr>
<td>KPI-1</td>
<td>Green</td>
<td>-2.06%</td>
<td>2.37%</td>
<td>-6.76%</td>
<td>2.64%</td>
</tr>
<tr>
<td>KPI-2</td>
<td>Blue</td>
<td>-3.66%</td>
<td>2.14%</td>
<td>-7.87%</td>
<td>0.55%</td>
</tr>
<tr>
<td>KPI-2</td>
<td>Green</td>
<td>-3.63%</td>
<td>3.04%</td>
<td>-9.65%</td>
<td>2.39%</td>
</tr>
<tr>
<td>KPI-3*</td>
<td>Blue</td>
<td>2.50%</td>
<td>1.34%</td>
<td>-0.15%</td>
<td>5.14%</td>
</tr>
<tr>
<td>KPI-3*</td>
<td>Green</td>
<td>2.25%</td>
<td>1.89%</td>
<td>-1.49%</td>
<td>5.99%</td>
</tr>
<tr>
<td>KPI-4</td>
<td>Blue</td>
<td>-2.05%</td>
<td>1.85%</td>
<td>-5.70%</td>
<td>1.59%</td>
</tr>
<tr>
<td>KPI-4</td>
<td>Green</td>
<td>-3.91%</td>
<td>2.39%</td>
<td>-8.64%</td>
<td>0.83%</td>
</tr>
<tr>
<td>KPI-5</td>
<td>Blue</td>
<td>-5.91%</td>
<td>3.15%</td>
<td>-12.11%</td>
<td>0.30%</td>
</tr>
<tr>
<td>KPI-5</td>
<td>Green</td>
<td>-4.87%</td>
<td>4.16%</td>
<td>-13.12%</td>
<td>3.38%</td>
</tr>
<tr>
<td>KPI-6*</td>
<td>Blue</td>
<td>1.17%</td>
<td>1.74%</td>
<td>-2.25%</td>
<td>4.59%</td>
</tr>
<tr>
<td>KPI</td>
<td>Color</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>KPI-6*</td>
<td>Green</td>
<td>3.32%</td>
<td>2.11%</td>
<td>-0.87%</td>
<td>7.51%</td>
</tr>
<tr>
<td>KPI-7</td>
<td>Blue</td>
<td>-3.36%</td>
<td>3.31%</td>
<td>-9.88%</td>
<td>3.16%</td>
</tr>
<tr>
<td>KPI-7</td>
<td>Green</td>
<td>-5.70%</td>
<td>4.50%</td>
<td>-14.62%</td>
<td>3.22%</td>
</tr>
<tr>
<td>KPI-8</td>
<td>Blue</td>
<td>-2.16%</td>
<td>2.59%</td>
<td>-7.26%</td>
<td>2.94%</td>
</tr>
<tr>
<td>KPI-8</td>
<td>Green</td>
<td>-4.67%</td>
<td>2.98%</td>
<td>-10.58%</td>
<td>1.24%</td>
</tr>
<tr>
<td>KPI-9</td>
<td>Blue</td>
<td>-3.20%</td>
<td>2.14%</td>
<td>-7.42%</td>
<td>1.03%</td>
</tr>
<tr>
<td>KPI-9</td>
<td>Green</td>
<td>-5.89%</td>
<td>3.13%</td>
<td>-12.10%</td>
<td>0.32%</td>
</tr>
<tr>
<td>KPI-10</td>
<td>Blue</td>
<td>-1.91%</td>
<td>2.49%</td>
<td>-6.81%</td>
<td>2.99%</td>
</tr>
<tr>
<td>KPI-10</td>
<td>Green</td>
<td>-5.59%</td>
<td>3.47%</td>
<td>-12.46%</td>
<td>1.28%</td>
</tr>
<tr>
<td>KPI-11</td>
<td>Blue</td>
<td>-1.60%</td>
<td>1.85%</td>
<td>-5.25%</td>
<td>2.05%</td>
</tr>
<tr>
<td>KPI-11</td>
<td>Green</td>
<td>-1.31%</td>
<td>2.34%</td>
<td>-5.94%</td>
<td>3.33%</td>
</tr>
<tr>
<td>KPI-12</td>
<td>Blue</td>
<td>-5.27%</td>
<td>2.20%</td>
<td>-9.62%</td>
<td>-0.93%</td>
</tr>
<tr>
<td>KPI-12</td>
<td>Green</td>
<td>-5.09%</td>
<td>3.04%</td>
<td>-11.11%</td>
<td>0.93%</td>
</tr>
<tr>
<td>KPI-13*</td>
<td>Blue</td>
<td>2.00%</td>
<td>1.45%</td>
<td>-0.86%</td>
<td>4.86%</td>
</tr>
<tr>
<td>KPI-13*</td>
<td>Green</td>
<td>2.02%</td>
<td>1.88%</td>
<td>-1.70%</td>
<td>5.74%</td>
</tr>
<tr>
<td>KPI-14</td>
<td>Blue</td>
<td>-1.58%</td>
<td>1.73%</td>
<td>-4.98%</td>
<td>1.83%</td>
</tr>
<tr>
<td>KPI-14</td>
<td>Green</td>
<td>-2.94%</td>
<td>2.26%</td>
<td>-7.42%</td>
<td>1.53%</td>
</tr>
<tr>
<td>KPI-15</td>
<td>Blue</td>
<td>-4.57%</td>
<td>2.15%</td>
<td>-8.80%</td>
<td>-0.34%</td>
</tr>
<tr>
<td>KPI-15</td>
<td>Green</td>
<td>-7.25%</td>
<td>3.19%</td>
<td>-13.56%</td>
<td>-0.93%</td>
</tr>
<tr>
<td>KPI-16</td>
<td>Blue</td>
<td>1.99%</td>
<td>1.37%</td>
<td>-0.71%</td>
<td>4.69%</td>
</tr>
<tr>
<td>KPI-16</td>
<td>Green</td>
<td>3.56%</td>
<td>1.91%</td>
<td>-0.23%</td>
<td>7.35%</td>
</tr>
<tr>
<td>KPI-17</td>
<td>Blue</td>
<td>-2.56%</td>
<td>1.96%</td>
<td>-6.43%</td>
<td>1.30%</td>
</tr>
<tr>
<td>KPI-17</td>
<td>Green</td>
<td>-1.48%</td>
<td>2.43%</td>
<td>-6.29%</td>
<td>3.33%</td>
</tr>
<tr>
<td>KPI-18</td>
<td>Blue</td>
<td>-6.96%</td>
<td>2.65%</td>
<td>-12.18%</td>
<td>-1.74%</td>
</tr>
<tr>
<td>KPI-18</td>
<td>Green</td>
<td>-3.35%</td>
<td>3.30%</td>
<td>-9.90%</td>
<td>3.20%</td>
</tr>
<tr>
<td>KPI-19*</td>
<td>Blue</td>
<td>3.04%</td>
<td>1.50%</td>
<td>0.07%</td>
<td>6.00%</td>
</tr>
<tr>
<td>KPI-19*</td>
<td>Green</td>
<td>1.70%</td>
<td>2.09%</td>
<td>-2.44%</td>
<td>5.85%</td>
</tr>
<tr>
<td>KPI-20</td>
<td>Blue</td>
<td>-0.98%</td>
<td>1.68%</td>
<td>-4.29%</td>
<td>2.33%</td>
</tr>
<tr>
<td>KPI-20</td>
<td>Green</td>
<td>-0.88%</td>
<td>1.95%</td>
<td>-4.75%</td>
<td>2.99%</td>
</tr>
<tr>
<td>KPI-21</td>
<td>Blue</td>
<td>-3.49%</td>
<td>2.52%</td>
<td>-8.45%</td>
<td>1.47%</td>
</tr>
<tr>
<td>KPI-21</td>
<td>Green</td>
<td>-6.81%</td>
<td>3.56%</td>
<td>-13.86%</td>
<td>0.24%</td>
</tr>
<tr>
<td>KPI-22*</td>
<td>Blue</td>
<td>1.44%</td>
<td>1.36%</td>
<td>-1.24%</td>
<td>4.12%</td>
</tr>
<tr>
<td>KPI-22*</td>
<td>Green</td>
<td>1.05%</td>
<td>1.78%</td>
<td>-2.47%</td>
<td>4.58%</td>
</tr>
<tr>
<td>KPI-23</td>
<td>Blue</td>
<td>-1.33%</td>
<td>2.05%</td>
<td>-5.36%</td>
<td>2.70%</td>
</tr>
<tr>
<td>KPI-23</td>
<td>Green</td>
<td>-1.16%</td>
<td>2.66%</td>
<td>-6.44%</td>
<td>4.11%</td>
</tr>
<tr>
<td>KPI-24</td>
<td>Blue</td>
<td>-4.62%</td>
<td>2.94%</td>
<td>-10.42%</td>
<td>1.17%</td>
</tr>
<tr>
<td>KPI-24</td>
<td>Green</td>
<td>-4.16%</td>
<td>3.86%</td>
<td>-11.82%</td>
<td>3.50%</td>
</tr>
<tr>
<td>KPI-25*</td>
<td>Blue</td>
<td>0.36%</td>
<td>3.33%</td>
<td>-6.21%</td>
<td>6.92%</td>
</tr>
<tr>
<td>KPI-25*</td>
<td>Green</td>
<td>2.50%</td>
<td>4.27%</td>
<td>-5.97%</td>
<td>10.97%</td>
</tr>
<tr>
<td>KPI-26</td>
<td>Blue</td>
<td>-3.07%</td>
<td>2.43%</td>
<td>-7.85%</td>
<td>1.71%</td>
</tr>
<tr>
<td>KPI-26</td>
<td>Green</td>
<td>-7.12%</td>
<td>2.85%</td>
<td>-12.77%</td>
<td>-1.47%</td>
</tr>
<tr>
<td>KPI-27</td>
<td>Blue</td>
<td>1.63%</td>
<td>4.98%</td>
<td>-8.20%</td>
<td>11.46%</td>
</tr>
<tr>
<td>KPI-27</td>
<td>Green</td>
<td>-1.62%</td>
<td>6.41%</td>
<td>-14.39%</td>
<td>11.15%</td>
</tr>
<tr>
<td>KPI-28</td>
<td>Blue</td>
<td>-1.18%</td>
<td>2.17%</td>
<td>-5.47%</td>
<td>3.11%</td>
</tr>
<tr>
<td>KPI-28</td>
<td>Green</td>
<td>-6.27%</td>
<td>2.80%</td>
<td>-11.83%</td>
<td>-0.70%</td>
</tr>
<tr>
<td>KPI-29</td>
<td>Blue</td>
<td>-3.54%</td>
<td>1.85%</td>
<td>-7.18%</td>
<td>0.11%</td>
</tr>
<tr>
<td>KPI-29</td>
<td>Green</td>
<td>-3.95%</td>
<td>2.54%</td>
<td>-9.00%</td>
<td>1.09%</td>
</tr>
</tbody>
</table>

*Note: * = negatively valenced KPI
3.1.3.3.1. Estimation analysis – All comparisons of red to blue and green

Of the 58 KPI comparisons of red to either green or blue, 55 comparisons (94.8%) were found to be in the hypothesised direction, i.e. those in the red condition performed worse than those in the blue or green conditions. The 3 comparisons that were not in the hypothesised direction are listed in table 3.9 (i.e. in these analyses the red condition scored higher than the blue or green condition).

Table 3.9. A list of the 3 comparisons and their mean difference (red vs. comparison colour) that were not in the hypothesised direction.

<table>
<thead>
<tr>
<th>KPI</th>
<th>KPI Description</th>
<th>Comparison Colour</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPI-22</td>
<td>Pupils achieving level 3 or below in mathematics TA*</td>
<td>Blue</td>
<td>1.99%</td>
</tr>
<tr>
<td>KPI-22</td>
<td>Pupils achieving level 3 or below in mathematics TA*</td>
<td>Green</td>
<td>3.56%</td>
</tr>
<tr>
<td>KPI-27</td>
<td>Pupils achieving level 4 or above in both English and mathematics in 2010</td>
<td>Blue</td>
<td>1.63%</td>
</tr>
</tbody>
</table>

*Note: EA = Exam Assessed, TA = Teacher Assessed,* = negatively valenced KPI.

These three comparisons stand out as anomalies. Other KPI’s that assess pupils achieving level 3 or below in mathematics (e.g. KPI-6), teacher assessed KPI’s below level 3 (e.g. KPI-19, KPI-22 and KPI-25) and historical measures of both English and Mathematics (e.g. KPI-28, KPI-29) do show the expected direction of the RACE.

3.1.3.3.2. Estimation analysis – assessing gender specific comparisons

As a key aim of this thesis is to examine the gender effect discrepancy in the current literature (see 2.5.) the 4 KPI’s that assess male and female students separately are listed in
table 3.10. Additionally, composite means of the male and female subpopulations are also listed in table 3.10.

**Table 3.10.** KPI’s that assess male and female students separately

<table>
<thead>
<tr>
<th>KPI</th>
<th>Gender</th>
<th>Comparison Colour</th>
<th>Mdiff</th>
<th>SE</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPI-7</td>
<td>Male</td>
<td>Blue</td>
<td>-3.36%</td>
<td>3.31%</td>
<td>-9.88%</td>
<td>3.16%</td>
</tr>
<tr>
<td>KPI-7</td>
<td>Male</td>
<td>Green</td>
<td>-5.70%</td>
<td>4.50%</td>
<td>-14.62%</td>
<td>3.22%</td>
</tr>
<tr>
<td>KPI-8</td>
<td>Female</td>
<td>Blue</td>
<td>-2.16%</td>
<td>2.59%</td>
<td>-7.26%</td>
<td>2.94%</td>
</tr>
<tr>
<td>KPI-8</td>
<td>Female</td>
<td>Green</td>
<td>-4.67%</td>
<td>2.98%</td>
<td>-10.58%</td>
<td>1.24%</td>
</tr>
<tr>
<td>KPI-9</td>
<td>Male</td>
<td>Blue</td>
<td>-3.20%</td>
<td>2.14%</td>
<td>-7.42%</td>
<td>1.03%</td>
</tr>
<tr>
<td>KPI-9</td>
<td>Male</td>
<td>Green</td>
<td>-5.89%</td>
<td>3.13%</td>
<td>-12.10%</td>
<td>0.32%</td>
</tr>
<tr>
<td>KPI-10</td>
<td>Female</td>
<td>Blue</td>
<td>-1.91%</td>
<td>2.49%</td>
<td>-6.81%</td>
<td>2.99%</td>
</tr>
<tr>
<td>KPI-10</td>
<td>Female</td>
<td>Green</td>
<td>-5.59%</td>
<td>3.47%</td>
<td>-12.46%</td>
<td>1.28%</td>
</tr>
<tr>
<td>KPI-7 &amp; 9</td>
<td>Male</td>
<td>N/A</td>
<td>-4.53%*</td>
<td>3.27%*</td>
<td>-11.01%*</td>
<td>1.93%*</td>
</tr>
<tr>
<td>KPI-8 &amp; 10</td>
<td>Female</td>
<td>N/A</td>
<td>-3.58%*</td>
<td>2.88%*</td>
<td>-9.27%*</td>
<td>2.11%*</td>
</tr>
</tbody>
</table>

*'=' = Mean values calculated from several KPI’s.

The decrease in performance is 0.95% larger in male students than female students in all 4 KPI’s across both colours. This effect is consistent across all gender based KPI’s, i.e. males in the red condition recorded larger drops in performance than females in all gender based KPI’s.

3.1.3.4. Is the prevalence of red uniforms higher in lower performing LEAs?

The estimation analysis suggests there is a consistent, but very small effect of uniform colour on academic performance that could be interpreted as red uniform impeding academic performance. This colour effect, if genuine, is too small to be detected with the given sample size using NHST. If an effect does exist it may not be that red coloured uniform decreases academic performance, but it may be that schools in lower performing LEAs are more likely to require red uniform.
To investigate this possibility a Pearson's chi-square test of association was run to determine if a significant association between LEA category and the percentage of schools requiring red uniform exists. See table 3.11 for a description of the percentage of schools requiring red uniform for each LEA. The analysis revealed no significant association between the percentage of schools requiring red uniform and the performance of the LEA, $\chi^2(2) = .942, p = .630$.

**Table 3.11.** Mean percentage of schools requiring red uniform in each performance category of LEA.

<table>
<thead>
<tr>
<th>LEA</th>
<th>LEA category</th>
<th>Percentage of schools requiring red uniform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bromley</td>
<td>High Performing LEA</td>
<td>21.62%</td>
</tr>
<tr>
<td>Greenwich</td>
<td>High Performing LEA</td>
<td>29.03%</td>
</tr>
<tr>
<td>Swindon</td>
<td>Typical Performing LEA</td>
<td>27.27%</td>
</tr>
<tr>
<td>Sunderland</td>
<td>Typical Performing LEA</td>
<td>31.66%</td>
</tr>
<tr>
<td>Coventry</td>
<td>Low Performing LEA</td>
<td>31.03%</td>
</tr>
<tr>
<td>Kingston Upon Hull</td>
<td>Low Performing LEA</td>
<td>33.33%</td>
</tr>
</tbody>
</table>

3.1.4. Discussion

In general there seems to be a consistent, small effect of colour on KPI's of academic achievement that supports the hypothesis that red uniform negatively impacts academic performance. However, analysis is underpowered and so caution must be exercised in the interpretation of these findings.

The differences in academic performance as a function of uniform colour were too small to be detected by null-hypothesis significance testing. However, the consistency with which red uniform schools perform worse than blue and green uniform schools is interesting; in 27
of the 29 KPI's the red uniform schools performed worse than the blue uniform schools and in 28 of 29 KPI's the red uniform schools performed worse than the green uniform schools. Although these differences are consistent, they are small and all 95% CI's cross zero, which implies that the effect could be smaller or even in the opposite direction. However, the majority of these 95% CI's are heavily weighted towards the expected direction of the RACE.

Additional analysis of the possible colour effects as a function of gender reveals that the decrease in performance is 0.95% larger in male students than female students. Whilst this finding suggests a gender difference in the RACE similar to that of Gnambs, Appel and Batinic (2010) further work must be conducted to clarify this given the small difference recorded in an underpowered analysis.

Although the recorded differences both within schools and within genders are small, such increments of improvement can have a meaningful impact on the reputation of a school. For example, a one percent increase in the number of students achieving level 5 or above in mathematics examinations would raise the median school's national ranking by 508 places. A school's ranking is of paramount importance to prospective parents of students as it serves as an objective and standardised measure to rate whether it is appropriate to send a child there. Research by Francis and Hutchings (2013) demonstrated that 57% of students rely on a schools ranking to choose a school and that 32% of 'professional parents' have moved areas so that they can be close to a good school. So, although an increase of 1% may seem trivial and 'statistically insignificant', when considered in context this finding could be of importance.
Additionally, these differences emerge from an observational analysis of naturally occurring data. When you consider the many variables that contribute towards primary school performance (e.g. teaching standards, student's socio-economic status) and the multitude of other non-uniform colour stimuli that student's encounter, it is remarkable that any consistent pattern (that was predicted a priori) emerges. It is likely that if the RACE is present, the systematic variation caused by uniform colour is in some way masked by the aforementioned unsystematic variation and so the differences as a function of uniform colour could be larger than reported here. Future work that factored in such covariates would be useful in exploring this finding further.

In addition to additional covariates in future work, it is also important to ensure statistical power is achieved, especially when the recorded effects are so small. To ensure statistical power is achieved a larger sample of primary schools is required to further investigate these effects. An a priori power calculation to determine the sample size was not possible as there was no previous research using similar methods that could inform such an analysis. To determine the required sample for future research the mean effect size of all mean differences recorded in experiment 1 ($d = 0.14$) was used to compute an a priori calculation ($\alpha = 0.05, \beta = 0.2$) that revealed that a sample size of 1,102 primary schools is required to achieve an observed power of 0.8. For reference, the analysis described in experiment 1 used a sample of 292 primary schools.

Aside from primary schools, it would be interesting to uncover if such effects manifest in secondary schools or in other cultures where school uniform colour varies between schools. Further findings that demonstrate a negative effect of red uniform on academic KPI's could
be influential in school uniform policy and in general classroom design. Changing uniform colour is a relatively cheap process (the cost of a couple of sweatshirts per student) and so potentially represents a cost effective way in which to improve academic performance – if the RACE manifests in further analyses.

This chapter has shown the possibility of a small, but potentially meaningful, and consistent effect of school uniform colour on academic performance that is congruent with other research on the RACE (i.e. red stimuli impede cognitive performance). However, it is also possible that schools that perform worse in KPI’s are more likely to require red uniform. This study has demonstrated limited evidence that the RACE does manifest in applied settings and this effect could differ as a function of gender. However, further analysis with increased statistical power is required to clarify and add to these findings.

This thesis will now move away from secondary research methods towards primary research methods to further investigate the RACE. Employing a range of research methods allows for a thorough and robust investigation of the RACE across different settings that each vary in terms of ecological validity and precision over examined and extraneous variables. The next chapter describes two experiments set in an online environment that participants access from remote locations. These experiments assess if red stimuli can influence cognitive performance in cognitive measures delivered online. This applied environment represents a transition between the secondary research methods that explicitly assess applied settings with high ecological validity (i.e. experiment 1) and research that assess the RACE in a controlled laboratory setting (i.e. experiment 5).
Chapter 4: The influence of red stimuli on cognitive performance in an online setting

4.1. Experiment 2: Replication of the RACE in an online environment (part I)

4.1.1. Introduction

The results from experiment 1 demonstrated that, in UK primary schools, there is some evidence that suggests the RACE influences measures of educational success. However, the recorded effects are very small and the analysis was underpowered. Further research is required that assesses larger samples of primary schools to clarify this finding. This chapter moves the thesis from an observational study of the RACE in an applied environment to an experimental study that assesses if the RACE can be uncovered in an applied environment; an online setting.

By assessing if the RACE manifests in an online environment this chapter examines aim 1 of this thesis; to assess the RACE in an applied context (see 2.5.). Experiment 1 showed the possibility that the RACE emerges in an applied, ecologically valid environment, (in UK primary schools) and so experiment 2 will further examine this possibility by asking participants to complete measures of cognitive performance in another applied, ecologically valid environment; on their smartphones, tablets and laptops.

This chapter also furthers the current research by attending to aim 3 of the thesis; to provide an independent, direct replication of the RACE as well as providing other conceptual replications where relevant. Although direct replications represent the ‘gold standard’ of scientific replications, sometimes conceptual replications are more appropriate to a line of
research where consistency between a series of experiments is important. The experiment within this chapter is one such example where a conceptual replication is appropriate.

Gnambs, Appel and Batinic (2010) have previously examined the RACE in an online environment. However, it is not appropriate to conduct a direct replication of this experiment for two reasons; (1) the use of general knowledge as the dependent measure and (2) the use of ‘dynamic features’ as the colour stimulus presentation technique. These two reasons are explained in more detail below.

1: The use of a general knowledge measure as a dependent variable in Gnambs, Appel and Batinic (2010)

Gnambs, Appel and Batinic (2010) used a measure of ‘general knowledge’ as the dependent variable, a measure they describe as assessing ‘crystallised intelligence’. Although this thesis does not have the scope to debate the validity and differences of the fluid and crystallised intelligence dichotomy, it is clear that the Gnambs, Appel and Batinic (2010) measure of general knowledge (e.g. history, biology, etc.) relies more on long term memory and previous exposure to relevant material than the verbal and numerical reasoning tasks used in Elliot et al., (2007); Maier, Elliot and Lichtenfeld, (2008); and Lichtenfeld et al., (2009) that rely more on working memory. Within this thesis it is important that a dependent measure is used consistently across experiments so that reliable comparisons between experiments can be made. It is also important that the chosen dependent measure has been used regularly in previous studies of the RACE so that valid comparisons can be made between studies too.
As measures that require working memory have been used more regularly in the RACE literature (see section 2.3. for a description of such work) it was decided that this type of cognitive measure will be used within the thesis. A ‘verbal analogy’ task was chosen as the specific measure of cognitive performance in this thesis as previous work on the RACE by Elliot et al., (2007) indicated that this cognitive measure reveals the largest effect size ($\eta^2_p = 0.3$ and $\eta^2_p = 0.5$), compared to anagram tasks ($\eta^2_p = 0.09$) and numerical sequencing tasks ($\eta^2_p = 0.11$).

2: The use of ‘dynamic features‘ as the colour stimulus presentation technique

To further ensure consistency between experiments within this thesis and across existing studies it is critical to identify a method of colour stimulus presentation that can be used in both online and laboratory settings. The colour presentation method from Gnambs, Appel and Batinic (2010) is the colouring of ‘dynamic features’, i.e. on-screen ‘progress bars’ and ‘forward buttons’. This method of stimulus presentation would not have been possible in future paper based experiments and additionally has only been used once in published literature.

However, the colour presentation method used by Elliot et al., (2007), Maier, Elliot and Lichtenfeld (2009) and Payen et al., (2011). a coloured square presented immediately before the task displayed for 5s, is possible to be implemented in both online and laboratory settings. This colour stimulus presentation technique therefore represents the optimal method of stimulus presentation to be used consistently throughout this thesis. Such consistency minimises any possible unsystematic variation caused by colour presentation methods that could influence results.
4.1.1.1. The present experiment

The present experiment examines the RACE in an online environment and will show participants a coloured square for 5s immediately before they take a verbal analogy task. This experiment seeks to examine three of the four aims of this thesis (see section 2.5). This experiment will assess the RACE in an applied context (aim 1) as participants will complete the tasks on personal web enabled devices (e.g. computer, smartphone, tablet) from remote, ecologically valid locations. In doing so, this will test the robustness of the effect outside of the laboratory. The design of this experiment provides a conceptual replication (aim 2) based on the combination of two existing methods (Gnambs, Appel and Batinic, 2010; Elliot et al., 2007). The experiment will seek to shed light on the gender discrepancy in the literature (aim 4) by running specific statistical analyses to uncover if gender influences the RACE.

4.1.2. Method

4.1.2.1. Participants

Members of the online public ($N = 98$, 47 male) were recruited through advertisements on social media platforms (for examples of these see B). Participation was restricted to native English speakers over the age of 16. In total 14 participants were excluded from the data set, 2 males with a known colour vision deficiency and 12 participants (7 males) who did not complete the task. The remaining 84 participants (36 male, $M_{age}=32.6$, $SD_{age}=9.3$, $range_{age}=18$ - 63) reported normal or corrected-to-normal visual acuity and normal colour vision. This was assessed by asking participants two post-task questions; 'Do you wear glasses?' and
'Have you ever been diagnosed with a colour vision deficiency by a medical professional?'. As the experiment was taking place in an online environment it was unfeasible to assess colour vision deficiencies in a more stringent manner (e.g. ‘City University Color Vision Test’, Birch, 1997)

The minimum sample size required ($n=39$) was determined from an a priori power analysis (calculated by G*Power software: Faul, Erdfelder, Lang & Buchner, 2007) using a desired power value of 0.8 and the pooled effect size ($\eta^2_p = 0.18$) from the two previous studies that are replicated in this experiment. Given that there is a wide-range of previous work showing the RACE, and that these effects are predicted by theory, it was decided to use the standard criterion power value of 0.8 (Cohen, 1992). The pooled effect size ($\eta^2_p = 0.18$) was calculated by combining the effect sizes reported in experiment 3 ($\eta^2_p = 0.3$) of Elliot et al., (2007) and experiment 1 ($\eta^2_p = 0.06$) of Gnambs, Appel and Batinic, (2010). The effect size from experiment 3 in Elliot et al., (2007) was selected as it reported the lowest (and thus most conservative) effect size when assessing the influence of the RACE on verbal analogies. The effect size from experiment 1 in Gnambs, Appel and Batinic (2010) was chosen as this was the only experiment in that paper that reported a significant effect when assessing both genders. Therefore the pooled effect size for this experiment was calculated as: $(0.3 + 0.06) \div 2 = 0.18$. As the pooled effect size was calculated using an estimation of two very different effect sizes it should be cautiously interpreted and assumed to contain a degree of error.

Therefore, the desired sample was set at 117, three times the required minimum sample ($39 \times 3 = 117$) to ensure that the desired power value was achieved. Although recruiting enough participants is critical to ensuring sufficient power exists within an analysis it is also important to note that very large samples can result in the smallest (and therefore
pragmatically meaningless) effect sizes becoming statistically significant (Runkell, 2012; Fritz, Morris and Richler, 2012). To guard against this possible issue the interpretation of statistical analysis will consider both dichotomous analysis (e.g. null hypothesis significance testing - NHST) and estimation analysis (e.g. 95% CI's on effect sizes). The experiment was stopped after 5 consecutive days of inactivity, 98 participants had completed the experiment at this point.

The participants remained anonymous and because of this, free and informed consent was highlighted but not recorded as recommended by the University of Surrey Ethics Committee who granted ethical approval.

4.1.2.2. Materials and measures

The dependent variable was performance in a 10 item verbal analogy (VA) task drawn from the 'Verbal Ability Word Relationships Test' (Newton & Bristoll, 2013). A VA task comprises of a verbal equation with a missing word (e.g. forest and trees = meadow and ?) and five possible answer words (e.g. 'grass, hay, feed, plants, pasture'). The task is to select the answer word which, when included in the equation, produces two word pairs with an analogous relationship. In the example given the answer is 'grass' as trees form a forest and grass forms a meadow. This measure is not the same VA measure used in Elliot et al., (2007) (the Intelligence Structure Test, [IST], Beauducel, Liepman, Horn, Brocke and Amthauer, 2007) but its question structure is identical.

Two other variables previously shown to influence cognitive performance were also measured; age (Craik & Salthouse, 2011) and gender (Miller & Halpern, 2014). Gender is
especially important to measure given the gender discrepancy in the RACE literature previously discussed (see section 2.5.4.). Demographic information was recorded at the start of the experiment and a post-experiment questionnaire was given at the end of the experiment. The post-experiment questions assessed if participants were aware of the coloured background, asked if they had any existing colour vision deficiencies and if they could guess the study's aims. This post-experiment questionnaire was split into two sequential pages, the first asked about the aims of the experiment and the second dealt with colour issues. This was done to ensure participants could not guess the study's aims from the post-experiment questions.

4.1.2.3. Colour stimuli

The independent variable was the colour of a square presented to the participant for 5s immediately before attempting the VA task. The coloured square (960px X 720px in size) was displayed for 5 seconds in the centre of the page. Inside the square the words: 'Test starts in 5 seconds...' were displayed in black Arial font, size 36. Four colour conditions were produced; red, blue, green and grey. The LCh values for these stimuli are given in table 4.1. For details of how the colour stimuli were produced and a description of issues regarding colour categorisation, measurement and production see appendix A.
Table 4.1. Lightness, chroma and hue values for the four colour conditions

<table>
<thead>
<tr>
<th>Colour</th>
<th>Lightness</th>
<th>Chroma</th>
<th>Hue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>46.23</td>
<td>52.01</td>
<td>27.48</td>
</tr>
<tr>
<td>Blue</td>
<td>45.94</td>
<td>53.22</td>
<td>284.07</td>
</tr>
<tr>
<td>Green</td>
<td>48.49</td>
<td>55.07</td>
<td>149.04</td>
</tr>
<tr>
<td>Grey</td>
<td>50.80</td>
<td>8.70*</td>
<td>151.54*</td>
</tr>
</tbody>
</table>

Note: '*' denotes the fact that chroma and hue in grey stimuli are not able to be equated to chromatic colours.

Design of the colour stimuli was implemented on a calibrated display, however it must be noted that these colours, although matched at source on lightness and chroma, would have been displayed differently to each participant owing to individual device display settings. Previous work on the influence of colour on cognition has stated that this could confound the interpretation of findings (Valdez & Mehrabian, 1994; Elliot & Maier, 2012). However, this method of colour presentation (i.e. matching colour attributes at source but allowing variation due to remote access) has been used in several other studies that have uncovered colour effects (e.g. Gnambs, Appel and Batinic, 2010; Lichtenfeld et al., 2012) and is characteristic of colours experienced in applied settings.

4.1.2.4. Design and procedure

The experiments were designed, accessed and administered using Qualtrics Research Suite Software (Qualtrics, 2014). Participants accessed the experiment remotely, in their own time, through a hyperlink distributed on social media platforms. The survey was optimised to run on laptop/desktop computers, tablets and mobile devices. Participants were
randomly allocated to one of four between-subjects experimental colour conditions (red, blue, green and grey). Randomisation was facilitated through participants indicating which day of the month their birthday fell on. Those whose birthday fell on the 1st to the 8th were assigned to the red condition, those born on the 9th to the 16th were assigned to the blue condition, those born on the 17th to the 24th were assigned to the green condition and those born from the 25th to the 31st were assigned to the grey condition. As the grey condition assignment criteria were smaller (owing to differing month lengths) the dates that colour conditions related to were switched every four days. This ensured equal group sizes across conditions. Aside from the colour manipulation there were no chromatic cues in the task; the background was dark grey; black text appeared in white boxes and navigation buttons were light grey. Additionally a grey and white 'Qualtrics' logo was present throughout in the top left hand corner.

After viewing an information page and entering demographic information, participants were shown the instructions for the VA task. On advancement to the next page participants were shown an ostensible loading page (which acted as the colour manipulation), presented for 5s before auto-advancing. This page contained a large coloured square with the words 'Test starts in 5 seconds..' printed in Arial bold, font size 24, in the middle of the square. The next page contained the VA task where participants were given 90s to complete as many items as possible. This 90s time frame was calculated from a pilot study in which participants (N = 11) took a mean time of 105.3s (SD = 9.46s) to complete 10 items. The 90s time allowance was used to reduce the probability of a ceiling effect. Participants were then given feedback for each question, their total score and the mean score of all participants for the VA task. Finally, participants completed the post experiment questionnaire (separated across two
4.1.3. Results

4.1.3.1. Analyses of verbal analogy performance

Preliminary screening of data revealed that the VA task performance variable conformed to parametric assumptions. Further analysis revealed that the specific ANCOVA test assumptions were met.

4.1.3.1.1 Dichotomous analysis

A two-way (colour condition [red vs. blue vs. green vs. grey] gender [male vs. female]) between-subjects analysis of covariance (ANCOVA) was conducted on VA performance with age as a covariate. For mean VA task scores for each colour condition and their 95% confidence intervals, see figure 4.1.
The analysis revealed no significant effect of age, $F(1,75) = .001, p = .974, \eta^2_p > .001$, on VA task performance. Considering gender, the analysis revealed no significant effect of gender on VA task performance, $F(1,75) = .047, p = .829, \eta^2_p = .001$. Regarding the central hypotheses, the analysis revealed no significant effect of colour on VA task performance, $F(3,75) = 1.26, p = .296, \eta^2_p = .048$, and no significant interaction effect between colour and gender on VA task performance, $F(3,75) = .576, p = .632, \eta^2_p = .023$. Overall, this NHST analysis indicates no significant differences between the task performance of the red, blue, green and grey conditions, regardless of whether gender is factored into the analysis. For mean VA task score, SD and SE of each colour condition, see table 4.2.
Table 4.2. Mean VA task score, SD and SE of colour conditions

<table>
<thead>
<tr>
<th>Colour Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>21</td>
<td>6.90</td>
<td>1.97</td>
<td>.430</td>
</tr>
<tr>
<td>Blue</td>
<td>21</td>
<td>6.86</td>
<td>2.37</td>
<td>.518</td>
</tr>
<tr>
<td>Green</td>
<td>21</td>
<td>6.52</td>
<td>2.52</td>
<td>.550</td>
</tr>
<tr>
<td>Grey</td>
<td>21</td>
<td>7.67</td>
<td>2.20</td>
<td>.480</td>
</tr>
</tbody>
</table>

4.1.3.1.2. Estimation analysis

The dichotomous analysis revealed no effect within the data. As the key prediction between the colour conditions is that those in the red condition will perform worse than those in any other condition the estimation analysis examined the differences between scores of the red condition and each of the other three conditions. Table 4.3. contains the mean difference and Cohen’s d effect size of each colour comparison along with the associated 95% CI.

Table 4.3. Mean difference, Cohen’s d effect size and 95% CI’s of the colour comparisons.

<table>
<thead>
<tr>
<th>Colour</th>
<th>M_{diff}</th>
<th>95% CI</th>
<th>Cohen’s d</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>.04</td>
<td>[1.41, -1.31]</td>
<td>0.22</td>
<td>[0.82, -0.39]</td>
</tr>
<tr>
<td>Green</td>
<td>.38</td>
<td>[1.79, -1.03]</td>
<td>0.17</td>
<td>[0.77, -0.44]</td>
</tr>
<tr>
<td>Grey</td>
<td>-.77</td>
<td>[0.54, -2.06]</td>
<td>0.36</td>
<td>[0.97, -0.25]</td>
</tr>
</tbody>
</table>

Note. $M_{diff}$ = mean difference between colour condition and red colour condition. CI = confidence interval.
There is very little difference between the means of the blue and green colour conditions and the red condition. This is reflected in a Cohen's $d$ value that is interpreted as a small to non-existent effect size (Cohen, 1988) and contextually, the raw mean difference of less than half a question also supports the small effect size interpretation. Additionally the direction of the difference is the opposite to what was expected, i.e. those in the red condition outscored those in the blue and green condition. The direction of difference is what was expected in the red and grey comparison and also exhibited the largest mean difference of 0.77 (11% of the mean score) and a Cohen's $d$ effect size ($d = 0.36$) that would be considered a small-to-medium effect (Cohen, 1988). However, the 95% CI's demonstrate that this difference could vary considerably and could even show the effect in an opposite direction.

4.1.3.1.3. Estimation analysis of male participants

As some literature suggests that the RACE only manifests in male participants (see 2.4.4.) the above estimation analysis was replicated but using only male participants. Dichotomous analysis of male participants was achieved through including 'gender' as an independent variable in the NHST. Unfortunately, given the random nature of recruiting participants in this experiment, the group sizes are unequal and the number of male participants in the green and grey condition are particularly small. Therefore these results should be interpreted with great caution. See table 4.4. for sample size, $M$, $SD$ and $SE$ of these conditions.
The mean difference and Cohen's $d$ effect size of each colour comparison along with the associated 95% CI for male participants are presented in table 4.5.

### Table 4.5. Mean difference, Cohen's $d$ effect size and 95% CI's of the colour comparisons.

<table>
<thead>
<tr>
<th>Colour</th>
<th>$M_{diff}$</th>
<th>95% CI</th>
<th>Cohen's $d$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>-.43</td>
<td>[1.68, -2.54]</td>
<td>-.18</td>
<td>[0.99, -0.64]</td>
</tr>
<tr>
<td>Green</td>
<td>.75</td>
<td>[2.98, -1.48]</td>
<td>0.34</td>
<td>[1.27, -0.60]</td>
</tr>
<tr>
<td>Grey</td>
<td>-1.58</td>
<td>[0.49, -3.66]</td>
<td>-0.80</td>
<td>[1.81, -0.44]</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval.

The first note to make is that the difference between means is larger in male participants than in the mixed gender sample. The blue condition has swung from a small difference in the opposite direction expected to a difference of 0.43 in the hypothesised direction, revealing an effect size that is approaching what could be considered a 'small' effect size of $d = 0.18$ (Cohen, 1988). The difference between the green and red condition remains in the opposite direction to what was expected but the effect size has more than doubled. The
findings of the two chromatic conditions indicate that the scores of male participants varied more than the scores of female participants. The effect size between the red and grey conditions has doubled revealing a large effect size (Cohen, 1988). However, the small sample size and relatively large 95% CI's that span zero in each comparison mean no firm conclusions can be drawn from this.

4.1.3.2. Analysis of post-experiment questionnaire

1. Awareness of colour manipulation

This question asked participants to name the colour of the loading page in the online experiment. In total 84.7% of participants answered this question. All participants stated the correct colour indicating that they were aware of the colour being present; however it is possible that the participants who did not answer this question did so because they could not accurately recall the colour.

2. Awareness of study aims

In total 83.2% of participants answered the question 'What do you think we were trying to find out?' and none of these participants correctly guessed the aims of the experiment. In general participants commented on the verbal nature of the dependent measure (e.g. ‘assessing verbal skills’ or stated that they were unsure (e.g. ‘no idea!’)).

3. Colour vision deficiency

This regarded the question; 'Have you ever been diagnosed with having a colour vision deficiency by a medical professional?' In total 87% of participants answered this question.
Two male participants indicated an existing colour vision deficiency and so were excluded from the analysis.

4.1.4. Discussion

4.1.4.1. Summary of results

The combination of analyses from the dichotomous and estimation analysis clearly demonstrate that no effects (in particular the RACE) were uncovered during this experiment. The direction of the effects are not consistent with previous work, the NHST analysis revealed all non-significant differences, the mean difference and Cohen’s $d$ effect size were approaching small or small-to-medium across all conditions and all 95% confidence intervals of the estimation analysis crossed zero.

The estimation analysis of the male participants showed a little more evidence for the RACE but it is not conclusive nor entirely consistent with previous work. Although those in the blue and grey condition outscored those in the red condition (as expected), those in the red condition outscored those in the green condition by 0.75 of a correct answer (as not expected). The comparison of red to the achromatic condition (grey) is in line with the expected direction of the effect but the variation in data indicated that the effect could be much smaller or even in the opposite direction if replicated in a similar sample. These results require even more caution given that they emerge from a very small and uneven distribution of male participants.
4.1.4.2. Experiment 2: Findings, limitations and considerations for further research

One key difference between this experiment and Gnambs, Appel and Batinic (2010) is the duration of colour exposure. Experiment 2 used the 5s exposure employed in the paper-based experiment by Elliot et al., (2007) whereas the original online Gnambs, Appel and Batinic (2010) experiment used a ‘progress bar’ giving constant exposure to the colour stimuli. It could therefore be that a 5s exposure to colour stimuli is sufficient to activate the RACE in a controlled laboratory setting whereas a longer exposure is required in an online environment given the reduced control over other confounding variables within the setting. For example, a participant could have been taking the experiment on a red laptop or on a phone housed in a green case. Such colour cues could impact the potency of the RACE and perhaps the continual exposure to colour that Gnambs, Appel and Batinic (2010) presented to participants limited this. To investigate this possibility, further experiments in this thesis will assess if constant exposure to colour stimuli activates the RACE in an online environment (experiment 3) and also if the 5s colour exposure method activates the RACE in controlled laboratory settings (experiments 4, 5 and 6).

Although care was taken to identify an appropriate dependent measure it may be that the 10 item VA task used was not sensitive enough to detect any differences in performance. Although the question structure was identical, the 20 item VA task used by Elliot et al., (2007) was longer and developed by different authors to the one used in this experiment. Therefore it seems sensible to adjust the VA task in future experiments to ensure the sensitivity of the dependent variable is not masking any effects. Experiments 2, 3, 4 and 5 in
this thesis will employ longer VA tasks to aid in determining if the sensitivity of the
dependent measure is a contributing factor.

The effect sizes previously reported in the literature were not found. However, the
experiment recruited 2.5 times the participants required as calculated in the a priori power
analyses and so power is unlikely to be a contributing issue. Instead it is likely that the effect
size calculated from the combination of Elliot et al., (2007) and Gnambs, Appel and Batinic
(2010) was inaccurate and more cautious effect sizes will be used in further experiments
meaning larger samples will be required. However, it is important to bear in mind that the
further the theorised effect size is reduced the less meaningful the RACE becomes in applied
situations.

4.1.4.3. Experiment 2 summary

In summary, this experiment did not uncover the RACE in an online setting. Some colour
comparisons showed trends in the direction reported in the literature (e.g. the comparison
of red and grey conditions), especially in male participants. However, these possible effects
are too vague and inconsistent with previous work to be considered anything but
unsystematic variation at this point. Possible confounds include the colour stimulus
exposure being too short for an online environment or the dependent measure of cognitive
performance not being sensitive enough. Issues that require further exploration include the
accuracy of the effect size used for a priori analyses, the possibility that the effect is only
present in males and whether the effect occurs outside of the laboratory. It is therefore
appropriate to re-run the experiment with the issues considered in the discussion attended
to. This will aid in clarifying if the replication failed because the effect is not present in our target population or if the experimental methods were not adequate.

4.2. Experiment 3: The influence of red on cognitive performance in an online setting (part II).

4.2.1 Introduction

This chapter details an experiment that will replicate experiment 2 with method adjustments made in accordance with four issues raised in the discussion of experiment 2. These issues are; the 5s colour stimulus exposure was too short to activate the RACE in an online environment (issue 1); the 10 item verbal analogy (VA) task that served as the dependent variable was not sensitive enough to detect changes in performance resulting from the RACE (issue 2); that the estimated effect size that informed the initial power analysis was too large (issue 3); and that too few male participants were recruited to be able to confidently state that the RACE emerged or not (issue 4).

The four issues will be addressed by:

1. Changing the colour stimulus presentation from 5s to constant exposure. The background of the screen that displays the experiment will be coloured (either red, blue or green) from when the participant is randomly allocated to a condition to the end of the dependent measure. Following completion of the dependent measure the colour will revert to default colours (grey and white) whilst the participants complete the post experiment questions to avoid any colour confounds in those answers. This method of presentation is more similar to that used by Gnambs, Appel & Batinic (2010) who used a 'progress bar' which was located in
the top right hand corner of the screen throughout the experiment, thus giving constant colour exposure. The 'whole background' method of colour presentation was selected to ensure that there was no ambiguity regarding whether the participants attended to the colour. As the experiment is accessed from remote locations it is likely that other colour stimuli feature in the participant’s visual field (e.g. a green phone case/red laptop), therefore a more prominent and prolonged colour stimulus within the experiment could aid in minimising any other colour effects. If the RACE is to manifest in an online environment it is logical to assume its occurrence is most likely with an overt and prominent presentation.

2. The dependent variable will be changed to a different 20 item VA task; the VA items from the Alice Heim 4 verbal reasoning tests (Heim, Watts & Simmonds, 1975) and Alice Heim 5 verbal reasoning tests (Heim, 1968). These reasoning tests have a long history of being used in psychological experiments (Shipley, Der, Taylor & Deary, 2012) whereas the task used in experiment 2 (Newton & Bristoll, 2012) was drawn from a psychometric test used by employers. The AH 4/5 tests should be more sensitive to any differences in performance given their historical use in experimental psychology and greater number of items. This measure is not the same VA measure used in Elliot et al., (2007) (IST; Beauducel, et al., 2007) but its question structure is identical and both are intended to measure verbal ability in adult populations.

3. The values used to determine the estimate of the effect size used for a priori power calculations will be reviewed to be more conservative. Previous work suggests that the effect size of the RACE is smaller when measured in an online setting (see the meta-analysis
in chapter 8 for a comparison of effect sizes) and so very conservative criteria will be used to calculate the estimated effect size here.

4. More male participants will be recruited for the experiment to further examine if the RACE is only present in males. This will be achieved through recruiting more participants overall as recruitment advertising that places a special emphasis on male participants could put off potential female participants, especially as recruitment is through social media where broadcast posts have the potential to be viewed by many people. This increase in participant numbers complements issue 3 as a conservative effect size will mean more participants are required to achieve sufficient statistical power.

To summarise, for the second time in this thesis, this experiment attempts to replicate the RACE in an online setting. The method of this experiment reflects differences imposed as a consequence of issues raised from the results of experiment 2. which did not uncover the RACE. Based on previous findings it is predicted that those in the red condition will perform worse than any other colour condition and that there will be no significant differences between the non-red colour conditions.

4.2.2. Method

4.2.2.1. Participants

In total, 158 participants were recruited through advertisements on social media platforms (for examples of these advertisements see appendix D). Of these participants, 26 (16 male) were excluded from the analysis; 2 male participants with a known colour vision deficiency, 4 participants who were not native English speakers and 20 participants who did not
complete the experiment. The remaining 132 participants (71 male, $M_{age} = 34.30$, $SD_{age} = 10.30$, $range_{age} = 17$ to 65) reported normal or corrected-to-normal visual acuity, normal colour vision and were native English speakers.

The minimum sample size required ($N = 125$) was determined from an a priori power analysis ($G^*Power software$; Faul et al., 2007) using a desired power value of 0.8 and an expected effect size of $\eta^2_p = 0.06$. This effect size is the same one reported in Gnambs, Appel & Batinic (2010) and is 69% smaller than the pooled version used in experiment 2 ($\eta^2_p = 0.18$), thus ensuring that the analysis has sufficient power to uncover even a small effect size. The experiment was stopped after 5 consecutive days of inactivity, 158 participants had attempted the experiment at this point.

The participant’s remained anonymous and because of this free and informed consent was highlighted but not recorded as recommended by the University of Surrey Ethics Committee who granted ethical approval.

4.2.2.2. Materials and measures

The dependent variable was performance on a 20 item VA task drawn from the Alice Heim 4/5 verbal reasoning tests (Heim, 1968; Heim, Watts & Simmonds, 1975). This measure is not the same VA measure used in Elliot et al., (2007; i.e. the 'Intelligence Structure Test'; Beauducel, et al., [2001]) but its question structure is identical and both are intended to measure verbal ability in adult populations. Two other variables, age and gender, were measured using the same demographic and post-experiment questions used in experiment 2.
4.2.2.3. Colour stimuli

The independent variable was the colour of the background on which the experiment was displayed in the web browser. Three colour conditions were produced; red, blue & green. A grey condition was also produced but unfortunately a technical error meant that these data were not collected; participants that should have been directed to the grey condition were directed one of the three other conditions.

The LCh values for these stimuli are given in table 4.6. For details of how the colour stimuli were produced and a description of issues regarding colour categorisation, measurement and production see appendix A.

Table 4.6. Lightness, chroma and hue values for the three colour conditions

<table>
<thead>
<tr>
<th>Colour</th>
<th>Lightness</th>
<th>Chroma</th>
<th>Hue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>46.23</td>
<td>52.01</td>
<td>27.48</td>
</tr>
<tr>
<td>Blue</td>
<td>45.94</td>
<td>53.22</td>
<td>284.07</td>
</tr>
<tr>
<td>Green</td>
<td>48.49</td>
<td>55.07</td>
<td>149.04</td>
</tr>
</tbody>
</table>

For details regarding the control of colour stimuli that are accessed remotely, see 4.1.2.3.

4.2.2.4. Design and procedure

The experiment's design and procedure was identical to that in experiment 2 with two exceptions. First, the dependent measure was changed from a 10 item VA task to a 20 item VA task and the allotted time therefore changed to 6 minutes. This 6 minute time frame was calculated from a pilot study in which participants (N = 8) took a mean time of 345s (SD = 27.2s) to complete 20 VA items. Second, the presentation of the colour manipulation took
the form of the web page background as opposed to the 5s loading screen used in experiment 2. The colour background was presented immediately after the participant had been allocated to a group and remained until the post experiment questions were asked. See Appendix E for screenshots of the procedure.

4.2.3. Results

4.2.3.1. Analysis of verbal analogy performance

Preliminary screening of data revealed that the VA task performance variable conformed to parametric assumptions. Further analysis revealed that the specific ANCOVA test assumptions were met.

4.2.3.1.1. Dichotomous analysis

A two-way (colour condition [red vs. blue vs. green] and gender [male vs. female]) between-subjects analysis of covariance (ANCOVA) was conducted on VA performance with age as a covariate. For mean VA task scores for each colour condition and their 95% confidence intervals, see figure 4.2
The analysis revealed no significant effect of age, $F(1,125) = 1.936, p = .167, \eta^2_p = .015$, on VA task performance. Considering gender, the analysis revealed no significant effect of gender on VA task performance, $F(1,125) = 1.684, p = .197, \eta^2_p = .013$. Regarding the central hypotheses, the analysis revealed no significant effect of colour on VA task performance, $F(2,125) = .337, p = .714, \eta^2_p = .005$, and no significant interaction effect between colour and gender on VA task performance, $F(2,125) = .576, p = .210, \eta^2_p = .025$. Overall, this NHST analysis indicates no significant differences between the task performance of the red, blue and green conditions, regardless of whether gender is factored into the analysis. For mean VA task score, SD and SE of each colour condition, see table 4.7.
Table 4.7. Mean VA task score, SD and SE of colour conditions

<table>
<thead>
<tr>
<th>Colour Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>44</td>
<td>14.39</td>
<td>2.20</td>
<td>.332</td>
</tr>
<tr>
<td>Blue</td>
<td>44</td>
<td>14.25</td>
<td>2.89</td>
<td>.435</td>
</tr>
<tr>
<td>Green</td>
<td>44</td>
<td>14.02</td>
<td>2.44</td>
<td>.367</td>
</tr>
</tbody>
</table>

4.2.3.1.2. Estimation analysis

The NHST revealed no effect within the data. As the key prediction between the colour conditions is that those in the red condition will perform worse than those in any other condition the estimation analysis examined the differences between scores of the red condition and each of the other two conditions. This difference will be assessed using both the mean difference between conditions and the Cohen's d effect size of this difference. Table 4.8. contains the mean difference and Cohen's d effect size of each colour comparison along with the associated 95% CI.

Table 4.8. Mean difference, upper limit 95% confidence interval and lower limit 95% confidence interval.

<table>
<thead>
<tr>
<th>Colour</th>
<th>$M_{diff}$</th>
<th>95% CI</th>
<th>Cohen's d</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>0.14</td>
<td>[1.22, -0.95]</td>
<td>0.05</td>
<td>[0.37, -0.47]</td>
</tr>
<tr>
<td>Green</td>
<td>0.37</td>
<td>[1.35, -0.62]</td>
<td>0.16</td>
<td>[0.58, -0.26]</td>
</tr>
</tbody>
</table>

*Note. $M_{diff}$ = mean difference between colour condition and red colour condition. CI = confidence interval.*
The mean difference between each colour condition is not in the hypothesised direction as those in the red condition scored marginally better than those in the other two colour conditions. These mean differences are very small both in relation to the mean score for all conditions \((M = 14.21, SD = 2.51)\) and in relation to the size of the 95% CI placed on each comparison. This is mirrored in the Cohen’s \(d\) effect size analysis, which reveals effect sizes that are too small to be considered an effect (Cohen, 1988). The confidence intervals are not only relatively large but also span zero and this suggests that there is little evidence of any difference between colour conditions.

4.2.3.1.3. Estimation analysis in male participants

Experiment 2 showed some possible influence of colour on male participants. As additional literature suggests that the RACE only manifests in male participants (e.g. Gnambs, Appel & Batinic, 2010) the estimation was replicated but using only male participants. Although more male participants were recruited in experiment 3 relative to experiment 2 the randomisation of participants meant that the split across the three conditions was not even. See table 4.9. for sample size, \(M\), \(SD\) and \(SE\) of these conditions.

<table>
<thead>
<tr>
<th>Colour</th>
<th>(n)</th>
<th>(M)</th>
<th>(SD)</th>
<th>(SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>16</td>
<td>14.13</td>
<td>2.47</td>
<td>.618</td>
</tr>
<tr>
<td>Blue</td>
<td>16</td>
<td>15.19</td>
<td>2.10</td>
<td>.526</td>
</tr>
<tr>
<td>Green</td>
<td>23</td>
<td>14.43</td>
<td>2.35</td>
<td>.490</td>
</tr>
</tbody>
</table>
The mean difference and Cohen's $d$ effect size of each colour comparison along with the associated 95% CI for male participants are presented in Table 4.10.

**Table 4.10.** The mean difference, Cohen’s $d$ effect size and the associated 95% confidence intervals.

<table>
<thead>
<tr>
<th>Colour</th>
<th>$M_{diff}$</th>
<th>95% CI</th>
<th>Cohen’s $d$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>-1.06</td>
<td>[-2.72, 0.60]</td>
<td>-0.46</td>
<td>[-1.16, 0.24]</td>
</tr>
<tr>
<td>Green</td>
<td>-0.30</td>
<td>[-1.92, 1.27]</td>
<td>-0.13</td>
<td>[-0.77, 0.51]</td>
</tr>
</tbody>
</table>

The direction of difference has changed in both the blue and green comparisons and whilst the difference between the red and the green conditions represents a small effect size (Cohen, 1988), the difference between the red and blue conditions is a potentially meaningful small to medium sized effect that represents a 7.5% improvement on the overall mean. However, the 95% CI spans zero on both estimations.

**4.2.3.2. Analysis of post-experiment questionnaire**

1. **Awareness of colour manipulation**

This question asked participants to name the background colour of the web based experiment. In total 68% of participants answered this question. Of these participants 2 stated the incorrect colour and 87% gave the correct colour indicating that they were aware of the colour being present. It is possible that the large number of participants not
answering this question \((N = 43)\) could not give the correct colour and therefore the number of incorrect answers could be higher.

2. Awareness of study aims

In total 79.7\% of participants answered the question 'What do you think we were trying to find out?'. None of these participants correctly guessed the aims of the experiment. In general participants commented on the verbal nature of the dependent measure (e.g. ‘assessing verbal skills’) or stated that they were unsure (e.g. ‘no idea!’).

3. Colour vision deficiency

This regarded the question; 'Have you ever been diagnosed with having a colour vision deficiency by a medical professional?' In total 85.5\% of the completed participants answered this question with two male participants indicating an existing colour vision deficiency and so were excluded from the analysis.

4.2.4. Discussion

4.2.4.1. Summary of results

The combination of analysis from the NHST and estimation analysis clearly demonstrates that no effects (in particular the RACE) were uncovered during this experiment. The direction of both comparisons are not consistent with previous work, the NHST analysis revealed no significant results, the mean difference and Cohen’s \(d\) effect size were small across all conditions and all four 95\% CI’s of the estimation values cross zero.
However, similarly to the results from experiment 2 there is some inconclusive evidence that the RACE manifested in male participants during this experiment. The mean difference between colour conditions swung from red outperforming both colour conditions (by 0.14 [blue condition] and 0.37 [green condition]) in the mixed gender analysis to red being outperformed by both the blue (by 1.06 correct answers) and green (by 0.30 correct answers) colour conditions in the male only analysis. The results of the comparison between the red and blue conditions show the strongest evidence of a gender effect, given that in experiment 2 the blue condition also exhibited a direction change towards the hypothesised direction when females were removed from the analysis. However, firm conclusions cannot be drawn from these data yet given that the 95% CI's of all condition comparisons cross zero and most have a relatively large margin of error. Given that Gnambs, Appel and Batinic (2010) found the RACE was driven by male participants it could be that the number of male participants in a sample is an important factor. However, further analysis reveals that Gnambs, Appel and Batinic (2010) had a similar number of males per condition (20) as the present experiment (18.3) and so it is not immediately obvious that a lack of males would account for the discrepancy in the effect sizes reported.

Regarding the analysis of both mixed-gender and male only participants, this experiment found no effect of colour on cognitive performance as predicted by previous literature (Elliot et al, 2007; Gnambs, Appel & Batinic, 2010). This finding is despite changes made to account for issues uncovered in experiment 2 regarding; the colour presentation duration; the validity and sensitivity of the dependent measure; and the number of participants required to achieve the necessary statistical power to detect a more conservative effect.
size. Such caution means it is unlikely that the effect was not uncovered due to low statistical power.

4.2.4.2. General Discussion of Experiments 2 and 3: Findings, limitations and considerations for further research

Changing the colour presentation made no observable difference to the task performance and in this thesis neither a short pre-task burst of colour (experiment 2) nor a constant background colour (experiment 3) has uncovered the RACE in an online setting. The colours used in this study were based on colours previously used to uncover the RACE (see appendix A) but perhaps the variation in colour on each participant’s device screen contributed to the null effect, even though previous research using similar online methods predict that this is not an issue (Gnambs, Appel & Batinic, 2010, Lichtenfeld et al, 2012).

In both experiment 2 and experiment 3, the participants were drawn from the general population through social media. Therefore the cognitive attributes of participants in the sample may have been heterogeneous meaning that unsystematic variation in cognitive ability could be masking any colour based effects. Attributes that could influence cognitive performance include education (Deary & Johnson, 2010), age (Craik & Salthouse, 2011) and socioeconomic status (Strenze, 2006). However, age can probably be ruled out given that it was not related to performance in the analysis of either experiment (see 5.1.3.3.) Previous work on the RACE (e.g. Elliot et al., 2007; Gnambs, Appel & Batinic, 2010) had used homogenous (mostly university student) samples. The participants in such samples are likely to have similar cognitive attributes given that university admissions are based, at least in part, on academic performance criteria. This represents selection bias as experimental
participants in university settings can only be recruited from a pool that has already been
pre-selected on cognitive ability. Failing to replicate the RACE was also demonstrated with a
heterogeneous, non-student population in Larsson and von Stumm (2015). Larsson and von
Stumm link this finding to the amount of time since participants were in education. They
theorise that the strength of the RACE decays the longer that participants have not had
exposure to red coloured evaluative feedback that is common in formal education (e.g. red
pen marking). Given that the RACE is thought to stem from associations much broader than
just academic feedback (i.e. both biological and other societal associations; Elliot & Maier,
2012) it seems unlikely that such a factor would influence the RACE but it is worth further
consideration. This will be tested later in experiments 4, 5 and 6 where homogenous
student populations will be used to examine the RACE.

4.2.4.3. Experiments 2 and 3 summary

In summary these experiments failed to provide evidence for the RACE in an online setting.
In both experiments there was limited evidence to suggest that the RACE was present in
male participants but, despite the number of males being similar to that of previous
research (e.g. Gnambs, Appel & Batinic, 2010), the effect sizes uncovered were not as large
as in previous research and the 95% CI’s indicate the variability of performance in male
participants is considerable. Therefore the mean differences between colour conditions in
the male participants are more likely to be a product of natural variation in the data as
opposed to the RACE.

To further investigate the RACE more control over extraneous variables is required to aid in
clarifying if the results from experiments 2 and 3 are due to method issues. For example it is
possible that the heterogeneity of the participant sample masked the effect. Another possibility is that the variety of other chromatic stimuli present when the task is taken remotely influenced the onset of the RACE. The degree to which the method requires tweaking to uncover the RACE will be indicative of the robustness of the RACE in applied settings. If the RACE does not exist outside of very specific experimental settings its prevalence in the psychological literature is likely due to a combination of publication bias and a lack of direct replications (see 2.4.2. for a description of how this can occur).

Therefore this thesis will now move away from an applied online setting to an offline ‘bricks-and-mortar’ controlled setting (similar to that used in Elliot et al., 2007) where more control over extraneous variables can be achieved and variability between participant demographics and experimental settings can be minimised.
Chapter 5: The influence of red stimuli on cognitive performance in the classroom

5.1. Experiment 4: The influence of red on verbal ability in an applied classroom setting.

5.1.1. Introduction

Chapter 4 detailed two experiments that did not uncover the RACE in an online setting. Previous work suggests that the RACE manifests only in male participants in an online environment (Gnambs, Appel & Batinic, 2010). However, specific analyses of the findings in chapter 4 as a function of gender do not support this finding, even though male participants in the red condition performed slightly worse than those in other colour conditions. The effect sizes of these differences in male participants were not consistent with previous research and the variability within male participants was considerable. Additionally, the analysis of the male participants emerges from a sample size that is comparable to previous work that did uncover the RACE that varied as a function of gender (e.g. Gnambs, Appel & Batinic, 2010) and so lack of statistical power is unlikely to be an issue.

This chapter moves the experimental setting from an online, remotely accessed setting to an offline 'real-world' setting. Previous research that has demonstrated the RACE in mixed gender and female only samples (see section 2.5.4. for a description of gender statistics in the RACE literature) has taken place in either classrooms or psychology laboratories, but not in an online environment. Additionally, the reported effect sizes in offline settings are larger than those reported in online settings (see chapter 8, for a description of effect sizes in the literature). Experiments 2 and 3 did not uncover the RACE in an online setting and it could be that other chromatic stimuli (e.g. phone case, interior design) masked any red induced
avoidance motivation in the red condition or indeed prompted avoidance motivation in non-red colour conditions. Being able to control peripheral chromatic stimuli in experiment 4 will aid in determining if this is a contributing factor to the null results recorded in experiments 2 and 3. Experiment 4 will also aid in clarifying if the null results in experiments 2 and 3 could be partly attributed to the heterogeneity of the participant sample by using a homogenous college student sample. Heterogeneity of participants such as that in experiments 2 and 3 could mask the RACE through the natural variation of cognitive ability in participants.

Therefore, this thesis will now examine the RACE in a mixed gender sample in a classroom setting to explore if findings similar to those demonstrated in other work set in offline settings are uncovered (e.g. Elliot et al., 2007; Maier, Elliot & Lichtenfeld, 2008; Mehta & Zhu, 2009).

The present experiment seeks to replicate experiment 3 from Elliot et al., (2007). From herein the experiments within the Elliot et al., (2007) paper will be denoted with a superscript number indicating the experiment that is being referred to. In short; Elliot et al., (2007)\(^3\) gave small groups of German college students a verbal analogy (VA) task printed on paper inside a ring binder folder. Immediately before the VA task was a title page that contained a coloured square (red, green or grey) with the words 'Anallogies' printed within. On turning to this title page participants viewed this colour stimulus page for 2s. From herein this method of presenting the independent and dependent variables is termed the 'folder method' and it is used frequently in the RACE literature (Elliot et al., 2007; Maier, Elliot and Lichtenfeld, 2008) and in this thesis (experiments 4, 5 and 6). Results in Elliot et
al., (2007)\textsuperscript{3} showed that those in the red condition scored significantly worse than those in the other two conditions ($\eta^2_p= 0.5$).

The present experiment is considered a close conceptual replication as two minor changes were made to the method. Although these are unlikely to make a difference to the outcome of the experiment it would be disingenuous to refer to the experiment as a 'direct replication', even though in practice it would be generally considered so. The first change regards a measure of mood that was included as literature has shown that emotion can influence cognitive functioning (Chepenik, Cornew & Farah, 2007; Scrimin, Mason & Moscardino, 2014) and so it seemed prudent to measure mood in order to partial out this unsystematic variation. The second change regarded Elliot et al.'s., (2007)\textsuperscript{3} use of 'Grade Point Average' (GPA) as a covariate measure of general ability. This metric is not available for UK participants. Instead, a sub-set of Raven's Advanced Progressive Matrices (RAPM; Raven, 1981) was used to measure general ability. The RAPM represented the optimal measure as it predicts both general ability (Raven, 2000; Engle, Tuholski, Luaghlin & Conway, 1999) and is also is directly correlated with performance in VA tasks (Teasdale, 2009; Jones & Estes, 2015). As the practice VA task from Elliot et al., (2007)\textsuperscript{3} was used to determine a base level of VA ability it was decided that, as the RAPM could fulfil this purpose, the practice VA task would be omitted from the experiment. This also aided in keeping the experiment as short as possible due to fewer measures being administered. This was key to running the experiment as it took place during a lunch break in a working college where time was at a premium.
Aside from these two small changes the method and procedure of the present experiment is very similar to Elliot et al., (2007)\(^3\) with regards to the experimental setting, number of participants, gender mix within the sample, colour stimuli values, colour presentation techniques and measure of cognitive ability. It should be noted that although the red stimulus in experiment 4 is functionally equivalent (i.e. within 5 units of the LCh values) to that in Elliot et al., (2007)\(^3\), the green and grey stimuli used within Elliot et al., (2007)\(^3\) were not equated on lightness and chroma. Therefore a decision was made (based on previous literature by the same authors; see Elliot and Maier, 2012) that the equation of lightness and chroma took precedence over replicating the colour stimuli used in Elliot et al., (2007)\(^3\). As the key prediction of the RACE is that a red stimulus will impede performance compared to all other colour stimuli it seemed sensible to ensure that the red stimulus used was functionally equivalent to that used in previous work, whilst other colour stimuli could vary from Elliot et al., (2007) if this meant best research practices (i.e. equating colour values; see appendix) was possible.

See table 5.1. for a comparison of the experimental properties described above. This experiment can be regarded as a close conceptual replication of Elliot et al., (2007)\(^3\).
Table 5.1. Comparison of Elliot et al., (2007)\(^3\) and experiment 4 from this thesis.

<table>
<thead>
<tr>
<th>Experimental property</th>
<th>Elliot et al., (2007)(^3)</th>
<th>Experiment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>German classroom</td>
<td>UK classroom</td>
</tr>
<tr>
<td>Mean Age</td>
<td>17.03</td>
<td>17.89</td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>59</td>
</tr>
<tr>
<td>n per condition</td>
<td>10</td>
<td>14.75</td>
</tr>
<tr>
<td>n per gender</td>
<td>5M, 25F</td>
<td>13M, 46F</td>
</tr>
<tr>
<td>Comparison conditions</td>
<td>Green, Grey</td>
<td>Blue, Green, Grey</td>
</tr>
<tr>
<td>Red LCh values</td>
<td>LCh: 45.70, 55.30, 24.90</td>
<td>LCh: 46.23, 52.01, 27.48</td>
</tr>
<tr>
<td>Green LCh values</td>
<td>LCh: 40.90, 25.80, 154.80</td>
<td>LCh: 48.49, 55.07, 149.04</td>
</tr>
<tr>
<td>Grey LCh values</td>
<td>LCh: 73.4, N/A, N/A</td>
<td>LCh: 50.80, N/A, N/A</td>
</tr>
<tr>
<td>Colour exposure duration</td>
<td>2s</td>
<td>5s</td>
</tr>
<tr>
<td>Dependent Measure</td>
<td>20 item IST VA task (German language version)</td>
<td>20 item IST VA task (English language version)</td>
</tr>
</tbody>
</table>

Note: M = males, F = females. LCh values refer to properties of colour, see 3.1.2. for a description of colour models.

5.1.1.2. The present experiment

In summary, this experiment seeks to inform three aims of this thesis (see 2.6.). Assessing the RACE in a classroom means suggestions can be made about the RACE in an applied setting (aim 1). Specific analyses will assess if gender interacts with the RACE (aim 2). These aims will be investigated through a close conceptual replication (aim 3) of Elliot et al., (2007)\(^3\).
Additionally, the results of experiments 2 and 3 could have been due to heterogeneity in participants’ characteristics (see 4.2.4.2) and so an additional aim of this experiment is to determine if using a more homogenous participant sample (similar to that used in Elliot et al, 2007) is required to uncover the RACE. If a homogenous sample is required then this limits the applicability of the RACE to the general population, although it could still be relevant in settings where sub-populations are generally homogenous (e.g. schools, universities).

5.1.2. Method

5.1.2.1. Participants

Students from a UK college (n = 69, 21 male) were recruited for the study. Recruitment was achieved through posters and group emails (see appendix F for examples of these). Participants received a piece of confectionary and were informed that they could win one of two £20 gift vouchers. In total, 10 participants (8 male) were excluded from the analysis, 1 participant with a known colour vision deficiency, 2 participants who were not native English speakers and 7 participants who did not complete the experiment. The remaining 59 participants (13 male, \(M_{age} = 17.34, SD_{age} = 1.21, range_{age} = 17\) to 20) reported normal or corrected-to-normal visual acuity and normal colour vision by responding to the post-task questions; 'Do you wear glasses?' and 'Have you ever been diagnosed with having a colour vision deficiency by a medical professional?'. As the experiment was taking place in a functioning college and time was at a premium it was unfeasible to assess colour vision deficiencies in a more stringent manner (e.g. ‘City University Color Vision Test’; Birch, 1997)
The minimum sample size required \( (n = 23) \) was determined from an a priori power analysis (using G*Power software: Faul et al., 2007) using a previous effect size \( \eta^2_p = 0.5 \) from Elliot et al., (2007)\(^3\) and a desired power value of 0.95. The desired power value was increased to 0.95 (from 0.8 in previous experiments in this thesis) as a cautious approach to reducing the probability of a type II error, especially given the previous null results in the thesis and the large expected effect size from Elliot et al., (2007)\(^3\). However, even with a cautious approach a required sample size of 23 is very small across four groups. Therefore, the experiment will aim for 1.5 times the number of participants per condition as that in Elliot et al., (2007)\(^3\). In Elliot et al., (2007) there were 10 participants per condition and so this study recruited 15 participants per condition. Therefore the target sample size for this experiment was set at 60 participants. It is relevant to compare the sample size per condition as the present experiment has four colour conditions whereas Elliot et al., (2007)\(^3\) had three. Although recruiting enough participants is critical to ensuring sufficient power exists within an analysis it is also important to note that very large samples can result in the smallest (and therefore pragmatically meaningless) effect sizes becoming statistically significant (Fritz, Morris and Richler, 2012). To guard against this possible issue the interpretation of statistical analysis will consider both dichotomous analysis (e.g. NHST) and estimation analysis (e.g. 95% CI's on effect sizes).

The participants gave written consent, their identities were kept anonymous and all ethical guidelines were followed as per the recommendations from the University of Surrey Ethics Committee who granted ethical approval.
5.1.2.2. Materials and Measures

The dependent variable was performance in a 20 item VA task drawn from the 'Intelligence Structure Test' (IST; Beauducel et al., 2013). This is the same verbal analogy measure used in Elliot et al., (2007)\(^3\) except the present experiment used the 2007 English language version whereas Elliot et al., (2007)\(^3\) used the 2001 German language version. A demographic questionnaire taken at the beginning of the experiment assessed gender and age. Gender was recorded to examine if any gender differences were uncovered regarding the RACE (see section 2.5.4. for description of gender differences in previous work). Age was recorded but not used in the statistical analysis given the small variability of age in the sample (\(M_{age} = 17.89, SD_{age} = 1.18\)). Mood has previously been shown to influence cognitive performance (Chepenik, Cornew & Farah, 2007; Scrimin, Mason & Moscardino, 2014) and so its inclusion is warranted here, despite not being measured in Elliot et al., (2007)\(^3\). Mood was measured using the 'Pleasure-Arousal-Dominance' (PAD) emotional state model (Mehrabian, 1997). General ability was measured using a sub-set of Raven's Advanced Progressive Matrices (RAPM; Raven, 1981). The RAPM represented the optimal measure as it predicts both general ability (Raven, 2000; Engle, Tuholski, Luaghlin & Conway, 1999) and is also is directly correlated with performance in VA tasks (Teasdale, 2009; Jones & Estes, 2015). The same post-experiment questions used in experiment 2 were given at the end of the experiment. The post-experiment questions assessed if participants were aware of the coloured background, asked if they had any existing colour vision deficiencies and if they could guess the study's aims. This post-experiment questionnaire was split into two sequential pages, the first asked about the aims of the experiment and the second dealt with colour issues.
This was done to ensure participants could not guess the study's aims from the post-experiment questions regarding colour vision deficiencies.

The materials and measures described were given to participants in a white A5 ring binder that included (in order): a consent form, demographic questionnaire, a measure of mood, a measure of general ability, VA task and post experiment questions (see appendix G for an example experiment booklet). Immediately before the VA task was a title page that contained the colour manipulation (an 11cm x 11cm printed coloured square) affixed below the title 'Items' printed in black ink.

5.1.2.3. Colour Stimuli

The independent variable was the colour of a square (11cm x 11cm) that was presented immediately before the VA task for 5s. This square was affixed to the paper using spray adhesive. Four colour conditions were produced; red, blue, green and grey. These colour stimuli were created in-house and printed onto HP 'Heavyweight Coated' paper using a HP Designjet 800PS printer. These colours are identical to the ones used in experiments 2 & 3 except those were presented on a computer monitor whilst the colour stimuli here take the form of printed card in the present experiment For details of how the colour stimuli were produced and a description of issues regarding colour categorisation, measurement and production see appendix A. The LCh values for these stimuli are given in table 5.2.
Table 5.2. Lightness, chroma and hue values for the four colour conditions.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Lightness</th>
<th>Chroma</th>
<th>Hue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>46.23</td>
<td>52.01</td>
<td>27.48</td>
</tr>
<tr>
<td>Blue</td>
<td>45.94</td>
<td>53.22</td>
<td>284.07</td>
</tr>
<tr>
<td>Green</td>
<td>48.49</td>
<td>55.07</td>
<td>149.04</td>
</tr>
<tr>
<td>Grey</td>
<td>50.80</td>
<td>8.70*</td>
<td>151.54*</td>
</tr>
</tbody>
</table>

Note: ‘*’ denotes the fact that chroma and hue in grey stimuli are not able to be equated to chromatic colours.

5.1.2.4. Design and Procedure

Participants were recruited through posters and group emails and were each offered a piece of confectionary upon experiment completion. As an additional incentive participants were given the chance to win one of two £20 Amazon gift vouchers. One voucher was given to the highest scoring participant (performance contingency) and the other to a random participant for taking part (participation contingency). The performance contingency was added to ensure the participants viewed the task within an 'achievement context', a setting where competence is evaluated and both success and failure are possible outcomes (Elliot et al., 2007). By explicitly demonstrating how success (being the highest scoring participant) and failure (not being the highest scoring participant) is achieved within the task participants are placed in an achievement context. For the importance of context setting when assessing the RACE see section 2.2. The performance contingency also ensured participants took the task seriously and applied effort throughout as there was a tangible benefit for success. The participation contingency (a random prize draw) was implemented
to encourage participation and was included as an ethical consideration to ensure not only the most intelligent participants benefited from taking part.

A between-subjects design (red vs. blue vs. green vs. grey) compared performance in the VA task between experimental groups. Participants were tested in groups of 12 in a college classroom (one participant to one table), on a lunch break with all four colours being tested at the same time (3 participants in each colour condition). The same room was used for all 6 sessions and was decorated in mostly achromatic colours (grey floor and walls, white ceiling, grey tables and chairs). The only chromatic colours took the form of two wall displays that showed student's work (these were red, blue, green and yellow). Such variation in interior design colours reflects a typical classroom setting and adds to the ecological validity of the experiment. The experimenter wore black shoes, black formal shirt and a grey suit in each experiment. Prior to the participants entering the room the 12 folders were randomly placed on tables so the researcher would be blind to each participant's condition. In addition to the folders, on each table sat an unbranded black pen and an information sheet (see appendix H).

The researcher read from a script (see appendix I) to ensure the instructions were as similar as possible for each session. Although the researcher was blind to each participant's condition, resource limitations meant that an additional researcher blind to the experiment aims could not be recruited. Participants were first instructed to read the information sheet before being guided by the experimenter through each section in the folder. After the consent form and demographic information was completed the participants were given 5 minutes to complete the measure of general ability before completing the measure of
mood. Following this, the participants received VA instructions before being given 5 minutes to complete the VA task. The colour manipulation was presented on the page immediately before each VA task with the words 'Items' printed above. To ensure participants viewed the colour manipulation they were required to verbally confirm they had read the word 'Items' printed above the colour square and that nothing was written within the colour square. The researcher subtly looked away as the colour manipulation was revealed to remain blind to each participant's condition. The researcher remained in the room during the experiment at all times. Following the VA task, participants were asked to complete the post-experiment questions, were thanked for their time, given the opportunity to ask questions, given a piece of confectionary and were then dismissed.

5.1.3. Results

5.1.3.1. Analysis of Verbal Ability Performance

Preliminary screening of data revealed that the VA task performance variable conformed to parametric assumptions. Further analysis revealed that the specific ANCOVA test assumptions were met.

5.1.3.1.1. Dichotomous analysis

To test the hypotheses a 2 (gender; male or female) x 4 (colour condition red vs. blue vs. green vs. grey) between-subjects analysis of covariance (ANCOVA) was conducted on VA performance with mood and general ability as covariates.
The analysis revealed no effect of emotion-arousal, $F(1,49) < .000, p = .985, \eta^2_p < .001$, or emotion-dominance, $F(1,49) = 2.287, p = .137, \eta^2_p = .003$, on task performance. However, the analysis revealed a small effect of the emotion-pleasure variable on task performance, $F(1,49) = 4.063, p = .050, \eta^2_p = .076$, indicating that those participants in a more pleasurable mood scored higher in the VA task.

As expected, the analysis revealed a moderate-to-strong effect of general ability on task performance, $F(1,49) = 30.36, p < .001, \eta^2_p = .378$, indicating that those who scored higher on the general ability task scored higher on the VA task.

Regarding the central hypotheses the analysis revealed no effect of gender on task performance, $F(1,49) = .022, p = .883, \eta^2_p = .006$, no effect of colour on task performance, $F(3,49) = .965, p = .417, \eta^2_p = .056$ and no interaction of colour and gender, $F(3,49) = 1.31, p = .883, \eta^2_p = .026$. Overall, this analysis indicates no significant differences between the task performance of the red, blue, green and grey conditions. It is worth noting that not only are the mean differences not statistically significant, they are also not in the hypothesised direction (i.e. the red condition outscores all other conditions). From hereon the data reported is adjusted for the covariates in the model. For mean VA task scores for each colour condition and their 95% confidence intervals, see figure 5.1. See table 5.3 for the $M$, $SD$ and $SE$ of these conditions.
Table 5.3. Mean VA task score, SD and SE of colour conditions.

<table>
<thead>
<tr>
<th>Colour Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>15</td>
<td>7.60</td>
<td>2.58</td>
<td>.666</td>
</tr>
<tr>
<td>Blue</td>
<td>15</td>
<td>5.94</td>
<td>2.69</td>
<td>.694</td>
</tr>
<tr>
<td>Green</td>
<td>16</td>
<td>6.87</td>
<td>2.57</td>
<td>.643</td>
</tr>
<tr>
<td>Grey</td>
<td>13</td>
<td>6.57</td>
<td>2.58</td>
<td>.713</td>
</tr>
</tbody>
</table>

5.1.3.1.2. Estimation analysis

The dichotomous analysis revealed no effect within the data. As the key prediction between the colour conditions is that those in the red condition will perform worse than those in any other condition the estimation analysis examined the differences between scores of the red
condition and each of the other three conditions. Table 5.4. contains the mean difference and Cohen’s $d$ effect size of each colour comparison with the associated 95% CI.

**Table 5.4.** Mean difference, Cohen’s $d$ effect size and 95% CI’s of the colour comparisons.

<table>
<thead>
<tr>
<th>Colour</th>
<th>$M_{\text{diff}}$</th>
<th>95% CI</th>
<th>Cohen’s $d$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>1.66</td>
<td>[-1.03, 4.34]</td>
<td>0.63</td>
<td>[-0.11, 1.36]</td>
</tr>
<tr>
<td>Green</td>
<td>0.72</td>
<td>[-1.13, 2.57]</td>
<td>0.28</td>
<td>[-0.43, 0.99]</td>
</tr>
<tr>
<td>Grey</td>
<td>1.03</td>
<td>[-0.94, 3.01]</td>
<td>0.39</td>
<td>[-0.35, 1.15]</td>
</tr>
</tbody>
</table>

*Note. $M_{\text{diff}}$ = mean difference between colour condition and red colour condition. CI = confidence interval. All calculations were conducted using means adjusted for the covariates ‘mood’ and ‘general ability’.*

In each comparison the direction of the effect is not as expected, i.e. the red condition outperforms all other conditions. The mean difference between the red condition and each other condition *could* be considered meaningful (especially comparing blue to red) and this is reflected in the Cohen’s $d$ analysis that represent small to medium sized effects. However, the variation in the sample denoted by the 95% CI’s indicate the effects are not due to the experimental manipulation and are likely due to unsystematic variation in the data.

The low number of male participants ($n = 13$) and the uneven distribution of male participants between conditions (a result of randomised group designation) precluded any male participant only estimation analysis.
5.1.3.2. Analysis of post-experiment questionnaire

1. Awareness of colour manipulation

This question asked participants to name the colour of the square presented before the central task. In total 81.4% of participants answered this question. Of these participants 4 stated the incorrect colour and 44 participants gave the correct colour indicating that they were aware of the colour being present; however it is possible that the participants who did not answer this question did so because they could not accurately recall the colour.

2. Awareness of study aims

In total 81.4% of participants answered the question 'What do you think we were trying to find out?' and none of these participants correctly guessed the aims of the experiment.

3. Colour vision deficiency

This regarded the question; 'Have you ever been diagnosed with having a colour vision deficiency by a medical professional?' In total 81.4% of participants answered this question. One male participants indicated an existing colour vision deficiency and so was excluded from the analysis.

5.1.4. Discussion

5.1.4.1. Summary of results

The combination of analyses from the NHST and estimation analysis clearly demonstrate that no effects (in particular the RACE) were uncovered in experiment 4. The direction of the
effects are not consistent with previous work, the NHST analysis revealed all non-significant results, the estimates of the mean difference and Cohen's $d$ effect sizes crossed zero across all comparisons. The findings in this experiment seem to typify natural variation in data collection.

5.1.4.2. Findings, limitations and considerations for further research

The method of experiment 4 was a close replication of Elliot et al., (2007)\textsuperscript{3} but with a small number of differences. The RAPM was used as a measure of general ability because the GPA metric used in Elliot et al., (2007) was not available in the UK. Additionally the RAPM was also expected to predict performance in the VA task and so the 'practice verbal analogies' measure used by Elliot et al., (2007) was omitted given that this measure provided a base level of VA ability without colour stimuli. Although the RAPM did predict VA performance in hindsight it could have been beneficial to administer the practice VA measure as not only would this provide a direct prediction of VA performance, it may have been beneficial for participants to practice verbal analogies in 'safe environment' (i.e. a practice test) to become familiar with the somewhat novel question structure. Interestingly, those in a more pleasurable mood performed better in the VA task. The direction of this effect is consistent with previous work on cognitive ability and mood (Chepenik, Cornew & Farah, 2007; Scrimin, Mason & Moscardino, 2014). Therefore this measure of mood will be retained in further experiments.

One query that emerged from experiments 2 and 3 of this thesis regarded whether or not the heterogeneity of the participant sample contributed toward the null findings. Experiment 4 used a homogenous participant sample but still did not uncover the RACE.
Therefore the heterogeneity of the participant sample in experiments 2 and 3 may not have contributed towards the null finding. Additionally, the sample recruited in experiment 4 were still in formal education and so, according to Larsson and von Stumm (2015), they could be more prone to being influenced by the RACE.

Although the homogeneity of the sample seemed to make no difference to uncovering the RACE, when comparing the mean scores of participants of experiment 4 to that of Elliot et al., (2007) an interesting difference emerges that could partly explain the lack of the RACE therein. Participants in Elliot et al., (2007) scored those in the present experiment by a mean difference of 4.62 answers (≈25% of total number of items). This difference exists even though the same task was used, the same amount of time was given to participants and the sample demographics were similar. See table 5.5 for the mean scores of participants in both samples. The UK college where the present experiment took place is in an area of low socio-economic status and this was reflected in the college's low academic ranking (reference reserved to protect the participants' anonymity). The German college where Elliot et al., (2007) took place is unknown but could be a higher ranking academic institution. Therefore the measure may have been too difficult for the UK participants this difficulty may have influenced the onset of the RACE. This possibility will be examined by using slightly older participants from high ranking UK universities in experiments 5 and 6.
Table 5.5. Mean score of each colour condition in the present experiment (UK College) and Elliot et al., (2007³; German College).

<table>
<thead>
<tr>
<th>Colour</th>
<th>UK College</th>
<th>German College</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>7.60</td>
<td>9.4</td>
</tr>
<tr>
<td>Blue</td>
<td>5.94</td>
<td>N/A</td>
</tr>
<tr>
<td>Green</td>
<td>6.87</td>
<td>12.7</td>
</tr>
<tr>
<td>Grey</td>
<td>6.57</td>
<td>12.0</td>
</tr>
<tr>
<td>Mean score across all colours</td>
<td>6.74</td>
<td>11.36</td>
</tr>
</tbody>
</table>

Note. The values given for Elliot et al., (2007)³ are approximations drawn from a bar chart as the exact values are not given in text.

5.1.4.3. Consultation with Prof. Elliot

Aside from the differences in participants' general ability it is unclear why, given such a close conceptual replication had been carried out, the results recorded in experiment 4 did not replicate those in Elliot et al., (2007)³.

Following the conclusion of this experiment the researchers had the opportunity to meet with Prof. Andrew Elliot, lead author of Elliot et al., (2007) and other relevant publications (e.g. Elliot & Maier, 2012; Meier, Hill, Elliot & Barton, 2015). During this meeting the methods and results from experiments 2, 3 and 4 in this thesis were discussed. Prof. Elliot made two broad recommendations regarding the present research; (1) that further care needs taking in setting an achievement context and; (2) that demographic information should not be asked prior to the central measure being administered, as doing so could instigate 'stereotype threat' that could confound any red induced avoidance motivation.
Prof. Elliot also highlighted the issue of ensuring the red stimulus used within an experiment is appropriate to eliciting the effect (A. Elliot, personal communication, 19/11/13). The experiments in this thesis had already attended to this issue by ensuring that the red stimulus created was functionally equivalent (i.e. within 5 units of the LCh scale) to that used in the Elliot et al., (2007) paper. However, it is worth mentioning here for other researchers who may not have considered ensuring the 'correct' red was implemented.

5.1.4.3.1. Further considerations of achievement context setting

As previously explained (section 2.2.), the theory that underpins the RACE mechanism (i.e. red stimuli prompting avoidance motivation that impedes cognitive performance) relies on the fact that the task is being taken in an achievement context. For the experiments 2, 3 and 4 the definition of achievement context was taken from Elliot et al., (2007, pp. 156):

"Achievement contexts are situations in which competence is evaluated and both positive outcomes (i.e., success) and negative outcomes (i.e., failure) are possible."

Therefore, it was presumed that in experiment 4 the use of an external performance contingency (i.e. a prize voucher) that relied on evaluated competence (i.e. the VA score) set a clear, tangible pass or fail outcome to ensure the achievement context was salient.

However, Prof. Elliot (A. Elliot, personal communication, 19/11/13) suggested that achievement context setting is a little more nuanced than the Elliot et al., (2007) definition would indicate, especially when considering external performance contingencies. Research demonstrates that external performance contingencies can influence cognitive performance through prompting various forms of approach/avoidance motivation (Deci, Koestner &
Ryan, 1999; Elliot, Shell, Henry & Maier, 2005) and that this could interfere with the avoidance motivation instigated by red stimuli. Such work suggests that external performance contingencies promote task strategies that focus on the external performance contingency and not on the process that facilitates task completion. For example, take a foreign language exam; external performance contingencies may promote the strategy of learning how to pass the exam (e.g. learning the minimum amount of common usage words or answering questions based on marking criteria) rather than learning the foreign language and completing the exam through language mastery. Therefore, external performance contingencies may not be appropriate when assessing how colour influences motivational processes.

In addition to omitting external performance contingencies, Prof. Elliot also suggested making it clear that immediate feedback was to be given by the researcher upon the completion of the task as previous work had shown that this can aid in ensuring an achievement context is set (Elliot, Shell, Henry & Maier, 2005). More general suggestions to provide an adequate achievement context centred on ensuring the setting was as professional and 'serious' as possible through: maintaining a serious, professional rapport with participants; the use of a stopwatch to indicate time pressure; and changing the size of the ring binder folder (from A5 to A4) to make the task look more professional (A. Elliot, personal communication, 19/11/13). These changes were therefore adopted for experiments 5 and 6 of this thesis.
5.1.4.3.2. Stereotype Threat

The second consideration concerns when in the experimental procedure participants were asked to give demographic information. In experiments 2, 3 and 4 this was asked for at the start of the experiment, whereas in Elliot et al., (2007)^3 this was done at the end of the experiment. This was overlooked when devising the procedure for the present experiments. Asking participants to indicate demographic information could elicit 'stereotype threat', a concept that means pre-determined stereotypes of performance can actively influence performance (similar to a 'self fulfilling prophecy'). For example, stereotype threat theory states that when women complete a maths test they risk being judged by the negative stereotype that women are worse at maths than men and that these thoughts can actively impede performance (Spencer, Steele & Quinn, 1999). Other proposed examples of stereotype threat influencing ability include African-American college students enacting the stereotype that they have lower ability than their white counterparts (Steele, 1997; Aronson, Fried & Good, 2002) and those from lower socio-economic groups performing worse at school (Good, Aronson & Inzlicht, 2003). Applying this to experiments 2, 3 and 4, it could be that requesting gender at the start of the experiment enforced stereotype threat that resulted in motivational tendencies across colour conditions that masked any red stimulus induced avoidance motivation. To ensure stereotype threat is minimised, experiments 5 and 6 will ask for demographic information at the end of the study.

5.1.4.3.3. The implications of specific experimental settings required to uncover the RACE

One clear conclusion of the consultation with Prof. Elliot is that the RACE may not be a robust effect as it may be obscured by a multitude of seemingly inconsequential factors, e.g.
offering a prize for performance, asking what gender a person is, or not having a serious and professional rapport with participants. If this is accurate then it would mean that the RACE could only be elicited in very particular experimental settings (i.e. those in Elliot et al., 2007; Gnambs, Appel and Batinic, 2010). If the RACE is simply a novel effect that can be uncovered in specific settings then its value as an applicable finding to real-world settings is severely limited.

5.1.4.4. Experiment 4 summary

In summary, experiment 4 did not uncover the RACE in an applied environment (a college classroom) and as a result did not replicate the findings from Elliot et al., (2007)\textsuperscript{3}. Obtaining a more homogenous participant sample (compared to that recruited in experiments 2 and 3) made no observable difference to the findings. However, it should be noted that the participants' general ability in experiment 4 is likely to be lower than that of Elliot et al., (2007)\textsuperscript{3} as their mean scores were much lower. Consultation with Prof. Elliot highlighted achievement context setting and stereotype threat as possible limitations in experiments 2, 3 and 4. Although experiment 4 demonstrated a close replication of Elliot et al., (2007) it was not a direct replication and the consultation with Prof. Elliot highlights how small changes to a method could make a large difference to results, e.g. offering a prize or the temporal position of the demographic measure. Therefore the next experiment will be a direct replication of Elliot et al., (2007)\textsuperscript{2} (that was set in a laboratory, not a classroom) with special consideration given to the points highlighted by Prof. Elliot. Although logically the next step in this research is to enact the changes suggested by Prof. Elliot it is important to keep in mind that even if the RACE manifests in these altered settings, its fragility may make
its existence largely meaningless given that, ultimately, a key endeavour of psychological research is to inform science about how humans behave in applied contexts.
Chapter 6: The influence of red stimuli on cognitive ability in a laboratory setting

6.1. Experiment 5: The influence of red on verbal ability in a laboratory setting (part I)

6.1.1. Introduction

Chapter 4 detailed experiment 4, which was a close conceptual replication of Elliot et al., (2007). Experiment 4 did not replicate the findings of Elliot et al., (2007). Consultation with the lead author of Elliot et al., (2007), Prof. Elliot highlighted a potentially inadequate achievement setting context and possible enacted stereotype threat as potential reasons for the non-replication. However, as a caveat to these recommendations this thesis highlights that if the RACE emerges only in very specific experimental circumstances then this could devalue the validity of the RACE in applied settings.

Experiment 5 is a direct replication of Elliot et al., (2007) whereas experiment 4 was a close conceptual replication of Elliot et al., (2007). The reason for changing the experiment that is being replicated is that Elliot et al., (2007) took place in a classroom whereas Elliot et al., (2007) took place in a psychology laboratory. Given the problems with uncovering the RACE in experiments 2, 3 and 4 it was decided that it was sensible to conduct experiment 5 in a setting where the most control possible over extraneous variables was available.

In addition to being a direct replication of Elliot et al., (2007), experiment 5 also incorporates the recommendations made by Prof. Elliot detailed in section 5.1.4.3. These considerations regard the setting of an appropriate achievement context and minimising stereotype threat. To aid in setting an appropriate achievement context the following research design issues were attended to; no external performance contingency will be
offered; it will be made clear to the participants that they will receive immediate face-to-face feedback of their performance; a professional and serious rapport will be maintained by the researcher; a stopwatch will be used to indicate time pressure; and the size of the ring binder folder that contains the test materials will be changed to an A4 size. Stereotype threat will be minimised by asking for demographic information at the end of the experiment.

In short Elliot et al., (2007)\textsuperscript{2} describes an experiment where a mixed gender sample of participants, drawn from a German University, are given a folder containing a practice VA task, colour manipulation, the central VA task and post experiment questions. Participants completed the VA tasks individually in a psychology laboratory and were given five minutes for each task. Immediately before the central VA task the participants turned the page of the folder to reveal a coloured square for 5s. Participants were explicitly told they would receive instant feedback from the researcher. Following the task the participants completed the demographics and debrief questionnaire.

The present experiment seeks to directly replicate the method described above with the exception that the sample was drawn from a UK university (and thus an English language version of the IST measure was used) and a measure of mood was taken at the start of the experiment before the task instructions. The measure of mood was added as experiment 4 had shown that mood-pleasure predicted task performance and consultation with Prof. Elliot indicated its inclusion would not elicit stereotype threat (A. Elliot, personal communication, 19/11/13). See table 6.1. for a comparison of the present experiment and Elliot et al., (2007)\textsuperscript{2}. 

181
Table 6.1. Comparison of Elliot et al., (2007)$^2$ and experiment 4 from this thesis.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age</td>
<td>21.67</td>
<td>19.72</td>
</tr>
<tr>
<td>N</td>
<td>46</td>
<td>32</td>
</tr>
<tr>
<td>n per condition</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>n per gender</td>
<td>4 male, 42 female</td>
<td>0 male, 32 female</td>
</tr>
<tr>
<td>Comparison conditions</td>
<td>Green, White</td>
<td>White*</td>
</tr>
<tr>
<td>Red LCh values</td>
<td>LCh: 45.7.23, 55.3, 24.9</td>
<td>LCh: 49.40, 48.30, 27.60</td>
</tr>
<tr>
<td>Green LCh values</td>
<td>LCh: 40.90, 25.80, 154.80</td>
<td>LCh: 50.0, 49.95, 160.0</td>
</tr>
<tr>
<td>White LCh values</td>
<td>LCh: not published</td>
<td>LCh: 100.0, 0.00, 0.00</td>
</tr>
<tr>
<td>Colour exposure duration</td>
<td>5s</td>
<td>5s</td>
</tr>
<tr>
<td>Dependent measure</td>
<td>20 item IST VA task</td>
<td>20 item IST VA task</td>
</tr>
<tr>
<td>Covariate measure of verbal ability</td>
<td>20 item IST VA task</td>
<td>20 item IST VA task</td>
</tr>
</tbody>
</table>

Note: A technical error (see 7.1.2.3.) resulted in the green stimulus being unusable on the week of testing and so only white and red could be compared. This did not pose an issue however as Elliot et al., (2007)$^2$ also used white as one of two comparison colours and so the replication remains valid.

Experiment 5 therefore can be considered a direct, independent replication of Elliot et al, (2007)$^2$. This represents what could be considered the first direct, independent replication of any experiment from Elliot et al., (2007). In replicating Elliot et al. (2007)$^2$, experiment 5 seeks to inform aim 3 of this thesis (to provide an independent, direct replication of the RACE). Experiment 5 also informs aim 2 of the thesis (to investigate the gender discrepancy
within existing work) by examining the RACE in a female only sample. It is by chance that no male participants were recruited for the experiment and reflects the population of the psychology department. Although some previous work suggests the RACE has a stronger influence on males more so than in females, such findings tend to only emerge in experiments set in either an online environment (e.g. Gnambs, Appel and Batinic, 2010) or in a sporting context (e.g. Hill and Barton, 2005). Conversely, work that has examined the RACE in a classroom or laboratory setting has uncovered the effect using samples consisting of both genders and in some cases using exclusively female participants (e.g. Eliot et al., 2007; Maier, Elliot and Lichtenfeld, 2009; Lichtenfeld et al., 2011). See section 2.5.3 for a description of gender differences in the RACE literature. Therefore it is of interest to determine if the RACE can be uncovered in a laboratory setting using only female participants.

6.1.2. Method

6.1.2.1. Participants

In total 35 female participants were recruited from a UK university to take part in the study for course required, research tokens. The demographic profile of the sample is very similar to Elliot et al., (2007)², see table 6.1. for a comparison of sample characteristics. In total 3 participants were excluded from the analysis, 1 participant whose first language was not English and 2 participants who did not complete the experiment due to non-experimental related fatigue. The remaining 32 participants (M_age= 19.72, SD_age = 1.98, range_age = 18 to 28) reported normal or corrected-to-normal visual acuity, normal colour vision and were native English speakers.
The minimum sample size required \((n = 31)\) was determined from an a priori power analysis (G*Power software: Faul et al., 2007) using a previous effect size \((\eta^2_p = 0.3)\) from Elliot et al., (2007)\(^2\) and a desired power value of 0.95. The use of this effect size is appropriate given that this experiment directly replicates Elliot et al, (2007)\(^2\). The desired power value was increased to 0.95 as a cautious approach to reducing the probability of a type II error. This is prudent given the previous null results in this thesis. Ethical approval was granted by the University of Surrey Ethics Committee.

6.1.2.2. Materials and Measures

The dependent variable was performance in a 20 item VA task drawn from the IST (Beauducel et al., 2013). This is the same VA measure used in Elliot et al., (2007)\(^2\) except experiment 5 used the 2013 English language version whereas Elliot et al., (2007) used the 2001 German language version. The practice VA task was also a 20 item test drawn from the IST and this was also the same measure used in Elliot et al., (2007)\(^2\).

As in previous experiments in this thesis, age and gender were also measured. However, to reduce the likelihood of stereotype threat influencing cognitive performance these measures will be taken at the end of the experiment, as was done in Elliot et al., (2007)\(^2\).

As in experiment 4 mood will be measured using the 'Pleasure-Arousal-Dominance' (PAD) emotional state model (Mehrabian, 1997) before task instructions are given. Although this measure was not taken pre-task in Elliot et al., (2007)\(^2\) its inclusion was approved in consultation with Prof. Elliot (A. Elliot, personal communication, 19/11/13). The same post-experiment questions used in experiment 2 was given at the end of the experiment. The
post-experiment questions assessed if participants could report the colour of the square presented to them, asked if they had any existing colour vision deficiencies and if they could guess the study's aims. This post-experiment questionnaire was split into two sequential pages, the first asked about the aims of the experiment and demographic information and the second asked about colour vision issues. This was done to ensure participants could not guess the study's aims from the post-experiment questions.

Participants received a white A4 ring binder which contained (in order): a consent form, a measure of mood, the practice VA task, the central VA task and post experiment questions. Immediately before the measure of mood and VA tasks is a title page and the title page before the central VA task contains the colour manipulation (a coloured square). Before the title page for each VA task is an 'achievement context setting' page that reminds participants whether or not they will receive immediate feedback. This was included as providing immediate feedback is central to setting an achievement context (Elliot, Shell, Henry & Maier, 2005). For the practice VA task the achievement context setting page read; 'You will not receive the score of this test' and for the central VA task the page read; 'The experimenter will give you your results and feedback once you have finished the test'. For an example of the materials described above see appendix J.

6.1.2.3. Colour Stimuli

The independent variable was the colour of a square (13cm x 18cm) with the words 'Test Items' printed in black ink in Calibri font, size 28, which was presented immediately before the VA task for 5s. This square was affixed to the paper using spray adhesive. Two colour conditions were produced; red and white. For details of how the colour stimuli were
produced and a description of issues regarding colour categorisation, measurement and production see appendix A. The LCh values for these stimuli are given in table 7.2. A green condition was planned as this was used as a comparison colour in Elliot et al., (2007). However a technical error with a photocopier destroyed the green stimulus, rendering it unusable on the week of testing and so only white and red could be compared. The green stimulus was obtained from a specialist retailer and this had to be placed on back order to be re-purchased. This did not pose an issue however as Elliot et al., (2007) also used white as one of two comparison colours and so the direct replication remains valid, albeit with one less comparison colour.

Table 6.2. Lightness, chroma and hue values for the two colour conditions.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Lightness</th>
<th>Chroma</th>
<th>Hue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>49.43</td>
<td>48.32</td>
<td>27.62</td>
</tr>
<tr>
<td>White</td>
<td>100.0</td>
<td>0.00*</td>
<td>0.00*</td>
</tr>
</tbody>
</table>

Note: '*' denotes the fact that chroma and hue in white stimuli are not able to be equated to chromatic colours.

The room in which the experiment took place was decorated in achromatic colours and contained no chromatic features in the line of site from where the participant completed the task. The researcher dressed in black shoes, grey trousers and a black t-shirt in all experiment sessions. The researcher used a stopwatch during the central VA task that was coloured grey, this stopwatch aided in setting the appropriate achievement context.
6.1.2.4. Design and procedure

A between-subjects design (red vs. white) compared central VA task performance between the two experimental groups. To ensure the researcher was blind to the experimental condition, one of two experimental folders, which differed only on the colour manipulation, was randomly selected before each experiment. This folder was placed on the table with an unbranded black pen and an information sheet (see appendix K). The researcher read from a script (see appendix L) to ensure the instructions were as similar as possible for each participant. The researcher kept a professional rapport with participants and did not engage in any 'small-talk' prior to the completion of the experiment. Participants were first instructed to read the information sheet before being guided by the experimenter through each section in the experiment pack. Importantly, participants were required to verbally confirm that they had read each title page before turning to the next page. This was done to ensure that when the participant confirmed they had seen the central VA task title page (that contained the colour manipulation) it did not seem odd that they were asked to verbally confirm that they had seen it. To aid in setting the achievement context participants viewed an 'achievement context setting' page before each title page of the VA tasks. Before the practice VA task this page stated that participants would not receive performance feedback, but before the central VA task participants were reminded that they would receive performance feedback and that the VA task was a 'good indication of general intelligence'.

After the consent form and measure of mood were completed the participant received VA instructions before being given 5 minutes to complete the practice VA task. The researcher
stepped out of the room during the completion of the practice VA task. The experimenter re-entered the room (now holding a stopwatch) to guide the participant through the colour manipulation and onto the central VA task before retreating out of the room. To ensure participants viewed the colour manipulation they were required to confirm they had read the words 'Test Items' printed within the colour square. The researcher, standing behind and to the side of the participant, looked away as the colour manipulation was revealed to remain blind to the condition. Following the central VA task, participants were asked to complete the post-experiment questions (that included demographic information) and were given their central VA task results if they accepted the offer to do so. Participants were thanked for their time before being given extra credit and dismissed.

6.1.3. Results

6.1.3.1. Analyses of verbal analogy performance

Preliminary screening of data revealed that the central VA task performance variable conformed to parametric assumptions. Further analysis revealed that the specific ANCOVA test assumptions were met.

6.1.3.1.1. Dichotomous analysis

A one-way (colour condition red vs. white) between-subjects analysis of covariance (ANCOVA) was conducted on central VA performance with practice VA performance and mood as covariates. The analysis revealed no effect of emotion-pleasure, $F(1,26) = .766, p = .389, \eta^2_p = .029$, emotion-arousal, $F(1,26) = 2.70, p = .112, \eta^2_p = .094$, or emotion-dominance, $F(1,26) = 2.60, p = .119, \eta^2_p = .091$, on task performance. As expected, the analysis revealed
a small-to-moderate effect (Cohen, 1988) of practice VA performance on central VA task performance, $F(1,26) = 5.07, p = .03, \eta^2_p = .163$, indicating that those who scored higher on the practice VA task scored higher on the VA task.

Regarding the central hypothesis the analysis revealed a moderate sized effect (Cohen 1988) of colour on task performance, $F(3,50) = 11.731, p = .002, \eta^2_p = .311$. This analysis indicates that those in the white condition outperformed those in the red condition.

From here on the data reported are adjusted for the significant covariate in the model (i.e. practice VA performance). For mean VA task scores for each colour condition and their 95% confidence intervals, see figure 6.1. See table 6.3. for the sample size, $M$, $SD$ and $SE$ of these conditions.

![Figure 6.1. Experiment 5 Mean VA task scores for each colour condition. Confidence intervals (95%) are indicated by error bars.](image-url)
Table 6.3. Sample size, mean VA task score, SD and SE of colour conditions

<table>
<thead>
<tr>
<th>Colour Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>16</td>
<td>9.72</td>
<td>1.70</td>
<td>.426</td>
</tr>
<tr>
<td>White</td>
<td>16</td>
<td>11.85</td>
<td>1.70</td>
<td>.426</td>
</tr>
</tbody>
</table>

6.1.3.1.2. Estimation analysis

The dichotomous analysis revealed the RACE as hypothesised. The estimation analysis will now examine the mean difference and Cohen's $d$ effect size regarding the white and red conditions. Table 7.4. contains the mean difference and Cohen's $d$ effect size of each colour comparison along with the associated 95% CI.

Table 6.4. Mean difference, Cohen's $d$ effect size and 95% CI's of the red and white colour comparisons.

<table>
<thead>
<tr>
<th>Colour</th>
<th>$M_{\text{diff}}$</th>
<th>95% CI</th>
<th>Cohen's $d$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>-2.13</td>
<td>[-3.40, -.849]</td>
<td>-1.25</td>
<td>[-.483, -2.01]</td>
</tr>
</tbody>
</table>

Note. $M_{\text{diff}}$ = mean difference between colour condition and red colour condition. CI = confidence interval.

There is a large mean difference of over 2 correct answers between the red and white condition and this is reflected in the large Cohen's $d$ effect size (Cohen, 1988). The estimates
of the 95% CI's indicate that the difference in means is due to the systematic manipulation of colour prior to taking the task.

6.1.3.2. Post Experiment Questions

6.1.3.2.1. Awareness of colour manipulation

In total six participants (18.75% of the sample) were unable to correctly recall the colour of the square immediately before the central VA task. In the red condition three participants were unable to name any colour square whereas in the white condition, two participants indicated they had seen a red square and one a black square. Therefore the analysis was re-run with these six participants excluded. This analysis, a one-way (colour condition red vs. white) between-subjects analysis of covariance (ANCOVA) was conducted on central VA performance with practice VA performance and mood as covariates. The analysis revealed no effect of emotion-pleasure, $F(1,20) = .954, \ p = .569, \ η^2_p = .016$, emotion-arousal, $F(1,20) = 2.07, \ p = .404, \ η^2_p = .035$, emotion-dominance, $F(1,20) = 2.14, \ p = .159, \ η^2_p = .096$, or practice VA performance, $F(1,20) = 2.08, \ p = .164, \ η^2_p = .094$ on task performance. This last finding contrasts with the analysis of all participants that found practice VA performance did influence central VA performance. Regarding the central hypothesis the analysis revealed a moderate sized effect (Cohen, 1988) of colour on task performance, $F(3,50) = 11.731, \ p = .002, \ η^2_p = .279$. This analysis indicates that those in the white condition outperformed those in the red condition and the recorded effect size is marginally smaller. See table 6.5. for the $M, SD$ and $SE$ of these conditions.
Table 6.5. Mean VA task score, SD and SE of colour conditions

<table>
<thead>
<tr>
<th>Colour Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>13</td>
<td>9.98</td>
<td>1.70</td>
<td>.494</td>
</tr>
<tr>
<td>White</td>
<td>13</td>
<td>12.02</td>
<td>1.70</td>
<td>.494</td>
</tr>
</tbody>
</table>

The estimation analysis will now re-examine the mean difference and Cohen’s $d$ effect size regarding the white and red conditions. Table 6.6. contains the mean difference and Cohen's $d$ effect size of each colour comparison along with the associated 95% CI.

Table 6.6. Mean difference, Cohen's $d$ effect size and 95% CI's of the red and white colour comparisons.

<table>
<thead>
<tr>
<th>Colour</th>
<th>$M_{diff}$</th>
<th>95% CI</th>
<th>Cohen's $d$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>-2.04</td>
<td>[-3.57, -.510]</td>
<td>-1.03</td>
<td>[-.200, -1.84]</td>
</tr>
</tbody>
</table>

Note. $M_{diff} = mean$ difference between colour condition and red colour condition. CI = confidence interval. All calculations were conducted using means adjusted by the covariates.

As in the original analysis there is a large mean difference of over 2 correct answers between the red and white condition and this is reflected in the large Cohen's $d$ effect size (Cohen, 1988). The lower end of both 95% CI's also indicate that the difference in means is due to the systematic manipulation of colour prior to taking the task.

Overall, the exclusion of participants who could not correctly name the colour presented, or those who named a colour not presented in the task, did not influence the overall interpretation of the data.
6.1.3.2.2. Awareness of study aims

This regarded the question; 'What do you think we were trying to find out?'. In total 100% of the completed participants answered this question. No participant correctly guessed the aims of the experiment. Typical guesses centred on the verbal nature of the task (i.e. "ability to manipulate words") or participants could not guess an aim (e.g. "no idea").

6.1.3.2.3. Colour vision deficiency

This regarded the question; 'Have you ever been diagnosed with having a colour vision deficiency by a medical professional?'. All participants answered this question. No participants had been previously diagnosed with a colour vision deficiency.

6.1.4. Discussion

6.1.4.1. Summary of results

The combination of analysis from the NHST, mean difference and effect size of this data clearly demonstrate that the RACE was uncovered during experiment 5. The direction of the effects are consistent with previous work, the NHST analysis revealed significant differences, the mean difference and Cohen's $d$ effect size were medium sized effects (similar to what was reported in Elliot et al., 2007$^2$) and the 95% CI's indicate that the effect in the expected direction is likely to come from the systematic variation in performance as a result of viewing the red stimuli.
6.1.4.2. Findings, limitations and considerations for further research

In experiment 5 the RACE was uncovered in a sample consisting of female participants only. This contrasts with experiments set in an online environment (Gnambs, Appel & Batinic, 2010) that demonstrated the effect was only found in male participants. The findings from experiment 5 replicate the findings from studies that have found the RACE in samples containing predominantly female participants, including the experiment replicated here (Elliot et al., 2007) and other work by Lichtenfeld et al., (2009). This collection of findings suggests that whilst the RACE may manifest in both genders when examined in controlled offline settings, it manifests only in males in an online setting. Further work is required to examine this gender/setting interaction to determine if there is indeed such an effect.

Experiment 5 represents a successful direct replication of Elliot et al., (2007). Regarding the work on the RACE in this thesis, one direct replication is not sufficient given the previous null replications in this thesis. It is also interesting to investigate if the RACE emerges consistently when the 'perfect' experimental setting is created, such as that in experiment 5. It seems sensible, and good scientific practice, to attempt to replicate the findings of experiment 5.

Experiment 5 represents the first time that the RACE has been uncovered within this thesis. Experiment 5 also represents the first direct, independent replication of an experiment from Elliot et al., 2007. Not only were the direction of the effects consistent with the previous study, so too were the size of the effects recorded. It would seem that the efforts to ensure an appropriate achievement context were set and that stereotype threat were minimised are essential to uncovering the RACE. However, as previously stated in section 5.1.4.3. the
fact that such specific experimental settings are required to elicit the RACE mean its value as an effect to the general population in applied contexts is limited. If the RACE can be negated by offering a prize for doing well or by simply asking an individual their gender or ethnicity (thus prompting stereotype threat), is the RACE an effect that is worth future research? Can the RACE even be considered a valid cognitive effect when its emergence seems to rely so much on a perfect experimental setting? These questions will be considered through both a meta-analysis (see chapter 7) and also a through a discussion of the thesis aims (see chapter 8).

To further examine the fragility of the RACE and to minimise the likelihood of the results of experiment 5 being a type I error, experiment 5 will be replicated again using identical methods. This will also allow the inclusion of the green condition (that should have been included in experiment 5) that will provide a chromatic contrast and provide an 'iron cast' replication of Elliot et al., 2007 that includes all colour comparisons.

6.1.4.3. Experiment 5 summary

In summary, experiment 5 did uncover evidence for the RACE as participants who viewed the red square for 5s performed significantly worse on a VA measure than participants who viewed a white square for 5s prior to the VA measure. The method considerations within this experiment borne from consultation with Prof. Elliot (e.g. setting an appropriate achievement context and asking demographic information at the end of the experiment) call into question not the effect’s existence but the effect’s importance in real world settings given that the effect seems somewhat fragile. Further research examining how formal and rigid the achievement context must be for the effect to manifest is required to investigate
this. To ensure the findings from experiment 5 are not a type I error and to further examine the fragility of the effect (by investigating if the RACE emerges consistently when the 'perfect' experimental setting is created), experiment 5 will be repeated again.

6.2. Experiment 6: The influence of red on verbal ability in a laboratory setting (part II)

6.2.1. Introduction

Section 6.2. describes a replication of experiment 5, which itself is a replication of Elliot et al., (2007)\(^2\). Both experiment 5 and Elliot et al., (2007)\(^2\) demonstrated the RACE and both recorded similar effect sizes (Elliot et al., [2007]\(^2\) recorded $\eta^2_p = 0.3$ and experiment 5 recorded $\eta^2_p = 0.31$). The only difference between experiment 6 and experiment 5 is that experiment 6 includes a third colour condition (green). This replicates the three colour conditions used in Elliot et al., (2007)\(^2\); red, green and white. The green condition could not be implemented in experiment 5 as a technical error rendered the stimuli unusable. Aside from the additional chromatic condition, experiment 5 and 6 are identical including; the population from which the sample was recruited, the experimental setting, the lighting within the setting, the researcher, the clothing the researcher wore, the script that was read to the participants, the materials and measures, the colour stimuli and the design and procedure. The only experimental variable that differed was the time of year. Experiment 5 was undertaken in March and experiment 6 in October. The experiment took place in a windowless room and so this difference in the time of year did not influence the lighting or vary the scenery from the laboratory.
A replication of experiment 5 is justified to ensure that it was not a type I error and to further investigate the fragility of the RACE by assessing if the RACE emerges consistently when the experimental considerations Prof. Elliot described in 5.1.4.3 are implemented into the research design. This represents a thorough and considered approach to research of any type but is especially important here given the series of null results obtained prior to experiment 5 and the lack of direct, independent replications within the RACE literature. It would be somewhat hypocritical of the researcher to not replicate an experimental finding given the importance that was placed upon such a practice in section 2.4.2.

6.2.2. Method

6.2.2.1. Participants

University students ($n = 47$, 9 male) were recruited from a UK psychology department to take part in the study for course required, research tokens. In total 2 participants were excluded from the data set as their first language was not English. The remaining 45 participants ($M_{\text{age}} = 18.64$, $SD_{\text{age}} = 1.03$, range$_{\text{age}} = 17$ to 22) reported normal or corrected-to-normal visual acuity and normal colour vision. This was assessed by asking participants two post-task questions; 'Do you wear glasses?' and 'Have you ever been diagnosed with having a colour vision deficiency by a medical professional?'.

The minimum sample size required ($n = 33$) was determined from an a priori power analysis (G*Power software: Faul et al., 2007) using a previous effect size ($\eta^2_p = 0.31$) from experiment 5 and a desired power value of 0.95. The use of this effect size is appropriate given that this experiment directly replicates experiment 5 (n.b. this effect size is also only
0.01 larger than that recorded in Elliot et al, [2007]², the experiment that experiment 5 replicated). The desired power value was set to 0.95 as a cautious approach to reducing the probability of a type II error. Ethical approval was granted by the University of Surrey Ethics Committee.

6.2.2.2. Materials and Measures

The materials and measures used in this experiment are identical to that in experiment 5 (see 6.1.2.2.).

6.2.2.3. Colour Stimuli

Except for the addition of a third colour condition (green) the colour stimuli were identical to those in experiment 5 (see 6.1.2.3.). For details of how the colour stimuli were produced and a description of issues regarding colour categorisation, measurement and production see appendix A.. The LCh values for these stimuli are given in table 6.7.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Lightness</th>
<th>Chroma</th>
<th>Hue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>49.43</td>
<td>48.32</td>
<td>27.62</td>
</tr>
<tr>
<td>Green</td>
<td>50.00</td>
<td>49.95</td>
<td>160.00</td>
</tr>
<tr>
<td>White</td>
<td>100.00</td>
<td>0.00*</td>
<td>0.00*</td>
</tr>
</tbody>
</table>

Note: ‘*’ denotes the fact that chroma and hue in white stimuli are not able to be equated to chromatic colours.
6.2.2.4. Design and Procedure

Except for the addition of a third colour condition (green) the design and procedure was identical to that in experiment 5 (see 6.1.2.4.).

6.2.3. Results

6.2.3.1. Analyses of verbal analogy performance

Preliminary screening of data revealed that the central VA task performance variable conformed to parametric assumptions. Further analysis revealed that the specific ANCOVA test assumptions were met. Although gender was recorded, only 4 male participants took part in the study and so no gender specific analysis is undertaken.

6.1.3.1.1. Dichotomous analysis

A one-way (colour condition red vs. green vs. white) between-subjects analysis of covariance (ANCOVA) was conducted on central VA performance with gender and practice VA performance as covariates. As expected, the analysis revealed a moderate effect (Cohen, 1988) of practice VA performance on central VA task performance, $F(1,40) = 11.61, p = .002, \eta^2_p = .225$, indicating that those who scored higher on the practice VA task scored higher on the VA task. Regarding the central hypothesis the analysis revealed no effect of colour on task performance, $F(2,40) = .954, p = .394, \eta^2_p = .046$. This analysis indicates no significant differences between the task performance of the red, green and white conditions.
From here on the data reported are adjusted for the significant covariates in the model. For mean VA task scores for each colour condition and their 95% confidence intervals, see figure 6.2. For the sample size, $M$, $SD$ and $SE$ of these conditions see table 6.8.

**Figure 6.2.** Experiment 6 Mean VA task scores for each colour condition. Confidence intervals (95%) are indicated by error bars.

**Table 6.8.** Mean VA task score, SD and SE of colour conditions

<table>
<thead>
<tr>
<th>Colour Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>15</td>
<td>10.65</td>
<td>2.32</td>
<td>.599</td>
</tr>
<tr>
<td>Green</td>
<td>15</td>
<td>10.13</td>
<td>2.31</td>
<td>.596</td>
</tr>
<tr>
<td>White</td>
<td>15</td>
<td>9.48</td>
<td>2.31</td>
<td>.597</td>
</tr>
</tbody>
</table>
6.1.3.1.2. Estimation analysis

The dichotomous analysis revealed no effect within the data. As the key prediction between the colour conditions is that those in the red condition will perform worse than those in any other condition the estimation analysis examined the differences between scores of the red condition and each of the other two conditions. Table 6.9 contains the mean difference and Cohen's \( d \) effect size of each colour comparison along with the associated 95% CI.

**Table 6.9.** Mean difference, Cohen's \( d \) effect size and 95% CI's of the red and white colour comparisons.

<table>
<thead>
<tr>
<th>Colour</th>
<th>( M_{\text{diff}} )</th>
<th>95% CI</th>
<th>Cohen's ( d )</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>0.52</td>
<td>[-1.59, 2.62]</td>
<td>0.23</td>
<td>[-0.50, -0.94]</td>
</tr>
<tr>
<td>White</td>
<td>1.17</td>
<td>[-0.95, 3.29]</td>
<td>0.50</td>
<td>[-0.23, 1.22]</td>
</tr>
</tbody>
</table>

*Note. \( M_{\text{diff}} = \text{mean difference between colour condition and red colour condition. CI = confidence interval. All calculations were conducted using means adjusted by the covariates.} \*

The mean difference in each comparison is in the opposite direction to what was expected (i.e. those in the red condition outperformed those in the white condition). The difference between the red and white represents a medium sized effect but the 95% CI's of the estimation indicates that these differences are due to unsystematic variation in the data and that further samples of data are could show trends in the opposite direction.
6.1.3.2. Post Experiment Questions

1. Awareness of colour manipulation

In total 2 participants (4.44% of the sample) were unable to correctly recall the colour of the square immediately before the central VA task. In the green condition one participant reported seeing a red square whilst in the white condition one participant reported seeing a blue square. Given that this is a very small percentage of participants and that in experiment 5 removing such participants did not influence the interpretation of the data, no further analysis was conducted.

2. Awareness of study aims

This regarded the question; 'What do you think we were trying to find out?'. In total 100% of the completed participants answered this question. No participant correctly guessed the aims of the experiment.

3. Colour vision deficiency

This regarded the question; 'Have you ever been diagnosed with having a colour vision deficiency by a medical professional?'. No participants had been previously diagnosed with a colour vision deficiency.
6.2.4. Discussion

6.2.4.1. Summary of results

The combination of analysis from the NHST, mean difference and effect size of these data clearly demonstrate that no effects (in particular the RACE) were uncovered during experiment 6. The direction of the effects are not consistent with previous work, the NHST analysis revealed all non-significant results, the mean difference and Cohen's d effect size were small across all conditions and all four 95% confidence intervals of the estimation analysis cross zero.

6.2.4.2. Findings, limitations and considerations for further research

The findings of experiment 6 did not replicate the findings of experiment 5 and Elliot et al., (2007)² despite the method directly replicating both experiment 5 and Elliot et al., (2007)² that both did uncover the RACE. It is plausible that the results from experiment 6 are a type II error but this is unlikely given the number of participants recruited that exceeded the a priori power calculation, which itself was based on the effect size recorded in experiment 5. It appears more likely that experiment 5 is the anomaly (being the only experiment to uncover the RACE) in the series of experiments presented in this thesis and could be a type I error.

Another possible explanation regards the fragility of the RACE. This differs from the argument that the RACE does not exist (i.e. it is only a statistical anomaly that can be observed with enough attempts) but instead the RACE exists as an effect that can be uncovered if the experimental setting is manufactured just so. Evidence that supports this
notion comes from the fact that the RACE was not uncovered in experiments 2, 3, 4 and 6 and also from the consultation with Prof. Elliot (see 5.1.4.3.). In this consultation Prof. Elliot highlighted numerous experimental factors that could hinder the onset of the RACE (e.g. the use of external performance contingencies or asking for demographic information at the start of the experiment) and experimental factors that facilitate the onset of the RACE (e.g. ensuring immediate feedback is provided or ensuring an experimenter's rapport is serious and professional). If the RACE is a fragile effect then perhaps it does not emerge consistently in experiments even when methods, settings and populations are identical to that of previous work that did uncover the RACE (e.g. between experiment 5 and experiment 6 in this thesis).

One method by which to further investigate the fragility of the RACE is to conduct a meta-analysis of both published work and grey literature that examines the RACE. A meta-analysis allows the robustness of the RACE to be compared across different experimental settings that incorporate various measures and research practices. Such an analysis allows for an objective interpretation of effect sizes that could provide evidence for the relative fragility of the RACE as posited in this thesis. If the analysis does imply that the RACE is fragile then this calls into question not the existence of the effect, but rather the validity and relative worth of the effect as a recognised cognitive effect. Incorporating grey literature into a meta-analysis aids in minimising the skewing effect that publication bias and the 'file-drawer' problem has on a body of literature (see 2.4.3.).

Not only does a meta-analysis represent the next logical step within this thesis to explore the pattern of results uncovered, a meta-analysis has yet to be published regarding the
RACE and so was set out as a primary aim of the thesis in section 2.5. Given that a meta-analysis is a powerful tool that is becoming more popular in psychology (Cummings, 2013), a meta-analysis fills a clear gap in the current research regarding the RACE that is clearly required. To better understand the RACE as an effect it is imperative research that examines the effect is pooled and analysed within a single analysis.

6.2.4.3. Experiment 6 summary

In summary, experiment 6 did not uncover evidence for the RACE, despite the method being identical to that of experiment 5, which did uncover the RACE. Whilst it is possible that experiment 5 committed a type I error or that experiment 6 a type II error, the evidence presented within the discussion of experiment 6 suggests that the discrepancy in findings may be due to the fact that the RACE is a fragile effect. To investigate the robustness of the RACE a meta-analysis that considers both published and unpublished (i.e. grey literature) research will next be undertaken.
Chapter 7: A meta-analysis of the RACE

Experiment 7: A meta-analysis of the RACE

7.1. Introduction

The experiments in this thesis have uncovered little evidence for the existence of the RACE. Experiment 1 demonstrated a weak, non-significant, but consistent effect of red uniform colour negatively influencing measures of academic success in UK primary schools. However, this finding could in part be attributed to the fact that lower performing Local Education Authorities seem to be more likely to require red uniform and so the causality of the small effects recorded are unknown. Experiments 2, 3 and 4 failed to replicate the RACE in both an online setting and in a college classroom. Following the conclusion of experiment 4, a consultation with the lead author of Elliot et al., (2007) highlighted various experimental practices that were important in research examining the RACE. Following the implementation of these experimental practices, experiment 5 demonstrated that viewing a red (relative to white) coloured square impeded performance in a verbal analogy task. However experiment 6, a direct replication of experiment 5 carried out four months later, did not replicate this finding. This pattern of results is partially reflected in the existing literature of the RACE. Whilst some work supports the RACE (e.g. Elliot, Maier & Lichtenfeld, 2008; Mehta and Zhu, 2009; Thorstenson, 2015) other work only partially supports the RACE (e.g. uncovering the RACE with the caveat that it only manifests in males, [Gnambs, Appel and Batinic, 2010]) or refutes it completely (e.g. Steele, 2014; von Stumm & Larsson, 2015).
When considering the body of work that considers the RACE it is worth remembering that, as discussed in section 2.5.2.1., a body of published literature is likely to overestimate the prevalence of any effect due to systematic publication bias, the 'file-drawer' problem and a tendency for researchers to perform conceptual (relative to direct) replications of existing effects.

Consultation with Prof. Elliot (see 5.1.4.) highlighted the use of external performance contingencies, a lack of immediate feedback and the onset of stereotype threat as possible reasons for why the RACE was not found in the experiments 2, 3 and 4 of this thesis. Although attending to these issues did facilitate a successful replication of the RACE in experiment 5, their implementation in experiment 6, with increased statistical power, did not result in an observation of the RACE.

Further, if factors such as those described above do mask the RACE, then this calls into question the robustness of the RACE and has consequences regarding the validity of the effect in real world settings. To determine the degree to which the RACE is a relevant psychological effect, an examination of both the existence and robustness of the RACE across studies is required. To provide this clarity a thorough investigation of all the published and unpublished material will be undertaken. There are two common methods to interpret such a collection of literature; a systematic review or a meta-analysis. For an area of research such as this, meta-analysis has three main advantages over systematic reviews (Stanley, 2001; Matysiak & Vignoli, 2006). First, a meta-analysis provides an objective, quantitative method of comparing effect sizes that provide less opportunity for researcher bias to surface than the more subjective commentaries of work that are provided within
systematic reviews. Second, a meta-analysis can directly compare effects that are recorded using different measures by converting mean differences into standardised effect size values. For example, a meta-analysis can compare performance between a 10-item verbal fluency task and a 20-item numerical sequencing task by converting the mean differences within each experiment to a standardised effect size (e.g. Cohen’s $d$). Third, a meta-analysis allows a greater number of research papers to be analysed at once, meaning it is possible to include all papers on a given subject that fit preset criteria. To illustrate this point, writing a thorough systematic review of 200 studies would be laborious and protracted to both write and to read and it would be difficult to draw firm conclusions, especially in a field that contains numerous contradicting findings. Conversely, it is much easier to interpret a meta-analysis of 200 papers that has distilled the effect to a single value whilst also weighting the studies by experimental power. Therefore, a meta-analysis provides the optimal method for examining contradicting findings in a body of literature such as the collection of work that describes the RACE.

In brief, a meta-analysis consists of the following three stages; first a systematic search of literature (both published and unpublished) is undertaken to gain a snapshot of all available data at a given time point. Next, mean differences of the effect are recorded and standardised into a common effect size that is weighted based on statistical power. At this stage any other moderating variables that may explain variance are also are specified. Finally, the effect sizes are combined into a single summary effect size that represents the best point estimate and 95% confidence interval (95% CI) for the effect size across all considered studies. This chapter will now conduct a meta-analysis of experiments that have examined the RACE as first described in Elliot et al., (2007).
7.2. Method

7.2.1. Inclusion and exclusion criteria

The meta-analysis will interpret data on the RACE. Therefore, each study in the meta-analysis will be either a conceptual replication or a direct replication of the first paper to formally describe the RACE; Elliot et al., (2007). The definition of the RACE used for the meta-analysis is; *in an achievement context the presence of red stimuli prompts avoidance motivation that results in impaired cognitive performance*. This definition is paraphrased from the original Elliot et al., (2007) paper. The particular criteria for study selection within this meta-analysis are:

- The experiment explicitly examines the RACE, i.e. the study is a conceptual or direct replication of Elliot et al., (2007).
- The experiment is an independent design.
- The independent variable is colour and at least one of those colours is red.
- The dependent variable is an absolute measure of cognitive performance, i.e. not measures of 'speed of completion' or the number of false recalls such as those recorded in Mehta and Zhu (2009), or rates of improvement such as those recorded in Elliot, Payen, Brisswalter, Cury and Thayer, (2011).
- The dependent variable is not a measure of memorisation and recall. The theory that underpins the RACE suggests that red stimuli impede performance through the onset of avoidance motivation mediated through a
narrowing of attentional scope (i.e. local processing). The majority of tasks used within the RACE literature require a broad attentional scope (i.e. global processing) and are therefore impeded by the onset of a narrow attentional scope, e.g. anagram tasks or numerical sequencing tasks. However, tasks that require rote memorisation and recall have been shown to improve using local processing (thought to be due to increased vigilance and attention to detail [Mehta & Zhu, 2009; Jung, Kim & Han, 2011]) and so in these experiments the RACE actually improved performance. Therefore, such tasks will be omitted from the meta-analysis. See section 2.2. for a thorough description of the theoretical mechanism that underpins the RACE.

- The experiment is set in an achievement context, i.e. the participants are asked to perform a task that has a pass or fail element to it. This differs from the evaluation contexts used in similar work that assess the impact of colour on mate evaluations (Elliot & Pazda, 2012; Schwarz & Singer, 2014) or the evaluation of food (Spence, 2015; Rohr, Kamm, Koenigstorfer, Groeppel-Klein & Wentura, 2015).

- That values for the meta-analysis are available. Values required include the sample size of each group in the comparison (i.e. the sample size per condition, not overall N) and a measure of mean difference (e.g. means and standard deviations or a standardised effect size). If these values are not available from the published material then the authors will be contacted to see if they can provide such details.
7.2.2. Moderator and mediator analyses

Two moderator variables were also recorded:

1. Whether or not saturation and brightness are equated across colour stimuli at perception.

This variable indicates if colours were matched at perception on saturation and brightness. Some researchers suggest that the equation of saturation and brightness across colour conditions that vary on hue is vital in colour psychology (e.g. Elliot et al., 2007; Mehta and Zhu, 2009). However, a thorough inspection of published RACE research reveals that this is seldom achieved even in papers that espouse its importance (see section appendix A). The term 'at perception' describes colour stimuli that are equated at the point that the participant's perceive the colour. This differs from methods that equate colour values at source (i.e. as they are being designed) but presents stimuli in a way that varies the colour attributes at perception. For example; Gnambs, Appel and Batinic (2010) match their colour stimuli at source but allow participants to access the online experiment remotely and so 'at perception' the colours appear differently owing to individual device display settings.

2. Whether or not the paper shares authorship overlap with Elliot et al., (2007).

Research on psychology experiment replications (Makel, Plucker & Hegarty, 2012) reveals replications that share no authorship with the original paper are 27.1% less likely to replicate the effect compared to those replications that share some authorship with the original paper. Therefore, whether or not a replication shares authorship with the original Elliot et al., (2007) paper was recorded. This will also aid in highlighting whether or not there
is something 'special' in the research design that co-authors of the Elliot et al., (2007) paper employ that enables the RACE. Possible examples of specific research design factors include the lengths taken to ensure the appropriate achievement context is set and ensuring that stereotype threat is minimised. Although Prof. Elliot highlighted these during a consultation, it was not immediately clear, within the manuscript of Elliot et al., (2007), that such precautions were necessary. Hence, it is possible that other researchers not associated with the authors of Elliot et al., (2007) also overlooked such issues. There is also the possibility that other research design factors not highlighted in Elliot et al., (2007) or by Prof. Elliot could also influence the results. Therefore it is important to determine if the RACE is present when considering only those papers that do not share authorship with Elliot et al., (2007).

7.2.3. Search strategy

This search was undertaken by one researcher (Adam Pedley, PhD candidate, University of Surrey).

7.2.3.1. Published literature

The search strategy used to uncover published literature was adopted from a modified approach taken from Stuck et al., (1993). This approach suggests first thoroughly searching a database of empirical work and selecting a broad range of studies that could be included. Next, the reference lists of these studies are considered before finally acquiring grey literature from other researchers in the field.

Before the search process is explained it is worth noting here that entries to the meta-analysis are made on a per-comparison basis and not a per-experiment basis. For example,
an experiment that contains three conditions (red, blue, green) will yield two comparisons; red vs. blue and red vs. green; not an overall effect and not a comparison of blue vs. green because the latter does not include red. The theory of the RACE states that it is red stimuli that impede cognitive performance through the onset of avoidance motivation - no other predictions or assumptions are made regarding other colours. Thus the number of comparisons within the meta-analysis will be greater than the number of experiments listed.

Applying the Stuck et al., (1999) strategy here; as the RACE was first proposed in Elliot et al., (2007) all subsequent work on the RACE will cite this paper either directly or cite it indirectly through a review paper on the RACE. Therefore, an initial search of the 315 studies that cited Elliot, (2007) on Google Scholar was undertaken. Google Scholar represents the optimum literature search engine available as it is extensive (estimated to contain 90% of all scholarly articles published online, ≈ 100 million articles, [Khabsa & Giles, 2014]) and is freely available to all, meaning that replication of this meta-analysis search is possible for anyone with internet access (although subscriptions are required to obtain most manuscripts over and above the abstracts).

The abstract of each of these studies was read and studies were then excluded on the basis that they did not fit the criteria, rather than being included on the basis that they do fit the criteria. This practice helps minimise the chance of type II errors (i.e. not including a valid study) during a process in which it is inherently difficult to commit a type I error (i.e. including a non-valid study) due to the pre-set criteria. This process resulted in 17 studies for further consideration (see table 7.1 for these studies). These studies were read in full
and 7 studies were excluded, leaving 10 studies remaining. These 10 studies contained 24 experiments providing 34 red-to-another-colour comparisons. The reference sections of these 10 papers were read but no further papers were uncovered.

**Table 7.1.** Studies selected for further consideration from the initial literature search. The reason for exclusion of those studies not retained is also given.

<table>
<thead>
<tr>
<th>Study</th>
<th>Retained?</th>
<th>N of relevant experiments</th>
<th>N of colour comparisons</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elliot et al., (2007).</td>
<td>Yes</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Mehta &amp; Zhu, (2009).</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Maier, Elliot and Lichtenfeld, (2008).</td>
<td>Yes</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Bonnardel, Le Bigot &amp; Piolat, (2008).</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>DV consisted of evaluation and memorisation tasks.</td>
</tr>
<tr>
<td>Lichtenfeld, Elliot &amp; Maier, (2012).</td>
<td>Yes</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Lichtenfeld et al., (2009)</td>
<td>Yes</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Tal, Akers &amp; Hodge, (2008)</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>M and SD of each comparison not available after contacting authors.</td>
</tr>
<tr>
<td>Sorokowski &amp; Szmajke, (2011)</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>DV is not an absolute measure of cognitive performance (arcade game skill).</td>
</tr>
<tr>
<td>Zhang &amp; Han, (2014)</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Shi, Zhang &amp; Jiang, (2015)</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Smajic, Merritt, Banister &amp; Blinebry, (2014)</td>
<td>Yes</td>
<td>3</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Duggan, (2009)</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>IV does not contain a red condition.</td>
</tr>
<tr>
<td>Thorstenson, (2012).</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>Contains admissions of research design error that may bias results.</td>
</tr>
<tr>
<td>Larsson and von Stumm,</td>
<td>Yes</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
During the initial literature search five literature reviews of the RACE were also selected for further consideration (Sburlea, 2012; Elliot & Maier, 2014; Elliot & Maier, 2012; Meier, Hill, Elliot & Barton, 2015; Elliot, 2015). It is possible that replications of Elliot et al., (2007) cited one of these later reviews on the subject and not the original paper. Therefore, any papers that have cited these literature reviews were considered and the reference sections of the literature reviews were also read. This process uncovered four possible papers for inclusion. Two of these studies were ineligible for selection but the other two studies were eligible and yielded two red-to-another-colour comparisons (see table 7.2 for description of these studies).

**Table 7.2.** Studies uncovered through searching literature reviews of the RACE.

<table>
<thead>
<tr>
<th>Study</th>
<th>Retained?</th>
<th>N of relevant experiments</th>
<th>N of colour comparisons</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ioan et al., (2007)</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>DV was not absolute performance</td>
</tr>
<tr>
<td>Yamazaki and Eto, (2011)</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>Too few participants in each condition</td>
</tr>
<tr>
<td>Buiks, (2013). Hulshof, (2013)</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>Yes</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
7.2.3.2. Grey literature

To uncover further unpublished work, a call for grey literature was put out detailing the aims of this meta-analysis and the criteria required for studies. See appendix M for an example of this email. An email list was created by listing the authors of each paper uncovered and also any other relevant researcher discussed elsewhere in this paper (see appendix N for this list). The call for grey literature was also sent to a number of relevant research groups to disseminate through their mailing lists; 'Vision Sciences Society', 'European Conference for Visual Perception', 'Colour Group of Great Britain', Progress in Colour Studies'.

The responses from the call for grey literature yielded 11 possible experiments. These experiments were considered in the same manner as the published literature and this left 3 further experiments containing 9 red-to-another-colour comparisons. Details of those studies that did not fit the criteria are not listed as such studies are not freely available in the public domain. However, the majority of these studies were excluded on the basis that their DV did not meet the criteria for the analysis. In addition to grey literature from other researchers, the grey literature from this thesis (experiments 2, 3, 4, 5 and 6) was also included in the meta-analysis. The experiments in this thesis contributed 11 red-to-another-colour comparisons across 5 experiments (see table 7.3. for a description of the grey literature included in the meta-analysis).
Table 7.3. Studies uncovered through a call for grey literature of research that examines the RACE.

<table>
<thead>
<tr>
<th>Grey literature author</th>
<th>N of relevant experiments</th>
<th>N of colour comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thorstenson (2015a, grey lit)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Steele (2015a, grey lit)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Steele (2015b, grey lit)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Steele (2015c, grey lit)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Steele (2015d, grey lit)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Experiment 2 (this thesis)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Experiment 3 (this thesis)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Experiment 4 (this thesis)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Experiment 5 (this thesis)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Experiment 6 (this thesis)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>TOTAL</td>
<td><strong>12</strong></td>
<td><strong>22</strong></td>
</tr>
</tbody>
</table>
7.2.4. Summary of included studies

In total 39 experiments containing 59 red-to-another-colour comparisons were selected for the meta-analysis. Of these:

- 26 experiments (37 comparisons) were drawn from the initial database searches of published materials.
- 2 experiments (2 comparisons) were drawn from reading reference lists of literature reviews of the RACE.
- 11 experiments (20 comparisons) were drawn from an extensive grey literature search.

Table 7.4. lists each colour comparison, the effect size and 95% CI (Cohen’s $d$), the dependent variable used, whether the colour stimuli were matched at perception, and if the paper in question shares authorship with Elliot et al., (2007).

Table 7.4. Each comparison included in the meta-analysis, its effect size (Cohen’s $d$), the dependent variable used, whether the colour stimuli were matched at perception, and if the paper shares authorship with Elliot et al., (2007).

<table>
<thead>
<tr>
<th>Study</th>
<th>Cohen’s d &amp; 95% CI</th>
<th>Shares authorship?</th>
<th>Matched at perception?</th>
<th>DV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buiks (2013)B</td>
<td>-0.50 [-0.92, -0.09]</td>
<td>No</td>
<td>No</td>
<td>Graph interpretation Anagram</td>
</tr>
<tr>
<td>Elliot et al., (2007)1BL</td>
<td>-3.93 [-4.97, -2.88]</td>
<td>Yes</td>
<td>No</td>
<td>Anagram</td>
</tr>
<tr>
<td>Elliot et al., (2007)1G</td>
<td>-3.80 [-4.81, -2.80]</td>
<td>Yes</td>
<td>No</td>
<td>Anagram</td>
</tr>
<tr>
<td>Elliot et al., (2007)2G</td>
<td>-1.05 [-1.82, -0.28]</td>
<td>Yes</td>
<td>No</td>
<td>Verbal Analogy</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Reference</td>
<td>Effect Size</td>
<td>CI Lower</td>
<td>CI Upper</td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Elliot et al. (2007)</td>
<td>2W</td>
<td>-0.89</td>
<td>-1.64</td>
<td>-0.15</td>
</tr>
<tr>
<td>Elliot et al. (2007)</td>
<td>3G</td>
<td>-1.14</td>
<td>-2.05</td>
<td>-0.22</td>
</tr>
<tr>
<td>Elliot et al. (2007)</td>
<td>3GR</td>
<td>-0.86</td>
<td>-1.85</td>
<td>0.12</td>
</tr>
<tr>
<td>Elliot et al. (2007)</td>
<td>4G</td>
<td>-0.51</td>
<td>-1.14</td>
<td>0.13</td>
</tr>
<tr>
<td>Elliot et al. (2007)</td>
<td>4GR</td>
<td>-0.55</td>
<td>-1.23</td>
<td>0.12</td>
</tr>
<tr>
<td>Gnambs, Appel &amp; Batinic, (2010)</td>
<td>1G</td>
<td>-0.50</td>
<td>-0.84</td>
<td>-0.15</td>
</tr>
<tr>
<td>Gnambs, Appel &amp; Batinic, (2010)</td>
<td>2G</td>
<td>-0.75</td>
<td>-1.10</td>
<td>-0.40</td>
</tr>
<tr>
<td>Hulshof, (2013)</td>
<td>1B</td>
<td>-0.36</td>
<td>-0.72</td>
<td>-0.00</td>
</tr>
<tr>
<td>Lichtenfeld, Elliot &amp; Maier, (2012)</td>
<td>G</td>
<td>0.11</td>
<td>-0.73</td>
<td>0.94</td>
</tr>
<tr>
<td>Lichtenfeld, Elliot &amp; Maier, (2012)</td>
<td>GR</td>
<td>-0.75</td>
<td>-1.62</td>
<td>0.12</td>
</tr>
<tr>
<td>Lichtenfeld, Maier, Elliot &amp; Pekrun, (2009)</td>
<td>2GR</td>
<td>-0.72</td>
<td>-1.33</td>
<td>-0.11</td>
</tr>
<tr>
<td>Lichtenfeld, Maier, Elliot &amp; Pekrun, (2009)</td>
<td>3GR</td>
<td>-0.63</td>
<td>-1.26</td>
<td>0.01</td>
</tr>
<tr>
<td>Lichtenfeld, Maier, Elliot &amp; Pekrun, (2009)</td>
<td>4GR</td>
<td>-1.09</td>
<td>-2.04</td>
<td>-0.14</td>
</tr>
<tr>
<td>Maier, Elliot &amp; Lichtenfeld (2008)</td>
<td>2GR</td>
<td>-1.03</td>
<td>-1.96</td>
<td>-0.11</td>
</tr>
<tr>
<td>Maier, Elliot &amp; Lichtenfeld (2008)</td>
<td>1GR</td>
<td>-1.43</td>
<td>-2.44</td>
<td>-0.42</td>
</tr>
<tr>
<td>Mehta &amp; Zhu (2008)</td>
<td>B</td>
<td>-0.67</td>
<td>-1.12</td>
<td>-0.21</td>
</tr>
<tr>
<td>Shi, Zhang &amp; Jiang (2015)</td>
<td>B</td>
<td>-0.57</td>
<td>-1.09</td>
<td>-0.04</td>
</tr>
<tr>
<td>Smajic, Merritt, Banister &amp; Blinebry (2014)</td>
<td>1B</td>
<td>0.03</td>
<td>-0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>Smajic, Merritt, Banister &amp; Blinebry (2014)</td>
<td>1G</td>
<td>-0.02</td>
<td>-0.71</td>
<td>0.68</td>
</tr>
<tr>
<td>Smajic, Merritt, Banister &amp; Blinebry (2014)</td>
<td>1W</td>
<td>0.09</td>
<td>-0.64</td>
<td>0.82</td>
</tr>
<tr>
<td>Smajic, Merritt, Banister &amp; Blinebry</td>
<td>W</td>
<td>0.50</td>
<td>-0.04</td>
<td>1.05</td>
</tr>
<tr>
<td>Reference</td>
<td>Method</td>
<td>Effect Size</td>
<td>95% CI</td>
<td>Verbal Analogy</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------</td>
<td>-------------</td>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Smajic (2014)</td>
<td>Verbal Analogies</td>
<td>0.11 [-0.70, 0.92]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Smajic (2014)</td>
<td>Verbal Analogies</td>
<td>0.22 [-0.74, 1.18]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Smajic (2014)</td>
<td>Verbal Analogies</td>
<td>1.82 [-0.26, 3.90]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Smajic (2014)</td>
<td>Verbal Analogies</td>
<td>1.81 [-0.26, 3.88]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Smajic (2014)</td>
<td>Verbal Analogies</td>
<td>1.79 [-0.27, 3.86]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Steele (2015a)</td>
<td>Anagram</td>
<td>0.07 [-0.20, 0.35]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Steele (2015b)</td>
<td>Anagram</td>
<td>-0.12 [-0.39, 0.16]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Steele (2015b)</td>
<td>Anagram</td>
<td>-0.04 [-0.33, 0.25]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Steele (2015c)</td>
<td>Anagram</td>
<td>-0.06 [-0.37, 0.25]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Steele (2015c)</td>
<td>Anagram</td>
<td>-0.07 [-0.31, 0.17]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Steele (2015d)</td>
<td>Anagram</td>
<td>-0.14 [-0.37, 0.10]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>This thesis (2015d)</td>
<td>Verbal Analogies</td>
<td>0.02 [-0.59, 0.62]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>This thesis (2015d)</td>
<td>Verbal Analogies</td>
<td>0.16 [-0.44, 0.77]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>This thesis (2015d)</td>
<td>Verbal Analogies</td>
<td>-0.36 [-0.97, 0.25]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>This thesis (2015d)</td>
<td>Verbal Analogies</td>
<td>0.05 [-0.36, 0.47]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>This thesis (2015d)</td>
<td>Verbal Analogies</td>
<td>0.16 [-0.26, 0.58]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>This thesis (2015d)</td>
<td>Verbal Analogies</td>
<td>0.61 [-0.12, 1.35]</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>This thesis (2015d)</td>
<td>Verbal Analogies</td>
<td>0.28 [-0.43, 0.98]</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Experiment</td>
<td>Cohen's d</td>
<td>Category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This thesis (Grey Lit)4GR</td>
<td>0.39 [-0.36, 1.14]</td>
<td>No</td>
<td>Yes</td>
<td>Verbal Analogy</td>
</tr>
<tr>
<td>This thesis (Grey Lit)5W</td>
<td>-1.22 [-1.98, -0.46]</td>
<td>No</td>
<td>Yes</td>
<td>Verbal Analogy</td>
</tr>
<tr>
<td>This thesis (Grey Lit)6G</td>
<td>0.22 [-0.49, 0.93]</td>
<td>No</td>
<td>Yes</td>
<td>Verbal Analogy</td>
</tr>
<tr>
<td>This thesis (Grey Lit)6W</td>
<td>0.49 [-0.24, 1.22]</td>
<td>No</td>
<td>Yes</td>
<td>Verbal Analogy</td>
</tr>
<tr>
<td>Thorstenson (2015a)G</td>
<td>0.10 [-0.33, 0.54]</td>
<td>Yes</td>
<td>No</td>
<td>Anagram</td>
</tr>
<tr>
<td>Thorstenson (2015b)G</td>
<td>-0.52 [-0.96, -0.08]</td>
<td>No</td>
<td>No</td>
<td>Anagram</td>
</tr>
<tr>
<td>Thorstenson (2015b, Grey Lit)GR</td>
<td>-0.29 [-0.72, 0.13]</td>
<td>No</td>
<td>No</td>
<td>Anagram</td>
</tr>
<tr>
<td>Larsson and von Stumm (2015)1G</td>
<td>-0.12 [-0.41, 0.17]</td>
<td>No</td>
<td>No</td>
<td>Word beginnings</td>
</tr>
<tr>
<td>Larsson and von Stumm (2015)2G</td>
<td>0.00 [-0.29, 0.30]</td>
<td>No</td>
<td>No</td>
<td>letter sets</td>
</tr>
<tr>
<td>Larsson and von Stumm (2015)3G</td>
<td>-0.04 [-0.33, 0.24]</td>
<td>No</td>
<td>No</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>Larsson and von Stumm (2015)4G</td>
<td>0.03 [-0.26, 0.32]</td>
<td>No</td>
<td>No</td>
<td>Nonsense Syllables</td>
</tr>
<tr>
<td>Larsson and von Stumm (2015)5G</td>
<td>-0.17 [-0.46, 0.12]</td>
<td>No</td>
<td>No</td>
<td>Reasoning</td>
</tr>
<tr>
<td>Larsson and von Stumm (2015)6G</td>
<td>-0.10 [-0.39, 0.19]</td>
<td>No</td>
<td>No</td>
<td>Knowledge</td>
</tr>
<tr>
<td>Zhang &amp; Han, (2014)G</td>
<td>-1.11 [-1.98, -0.24]</td>
<td>No</td>
<td>No</td>
<td>RAPM</td>
</tr>
</tbody>
</table>

Note: Number given after year indicates experiment number within the study. Letter given after the experiment number indicates the comparison colour; 'G' = green; 'B' = blue; 'GR' = grey; 'W' = white; 'BL' = black. 'BIS' = 'Behavioural Inhibition System'. 'RAPM' = Raven's Advanced Progressive Matrices. 'RAT' = Remote Associates Test. 'Grey lit' = grey literature. Minus Cohen's d values indicate that those participants in the red condition scored less than those in the comparison colour condition.
7.2.5. Statistical methods

Effect sizes were calculated as Cohen’s $d$ using Comprehensive Meta-Analysis software (Version 3.0, Biostat, 2014). Cohen’s $d$ is calculated by dividing the mean difference between colour conditions by the pooled standard deviation of the two conditions. Given the heterogeneity between study measures and procedures a random effects model was used to provide pooled point estimates of the RACE. Within this meta-analysis the following interpretations of Cohen’s $d$ effect size will be used from Cohen (1988);

- $<0.2 =$ No effect.
- $0.2 =$ Small effect.
- $0.5 =$ Moderate effect.
- $>0.8 =$ Large effect.

7.3. Results

7.3.1. Model 1: Full analysis.

In total, 59 red-to-another-colour comparisons emerging from 39 experiments (28 published) were included in the overall model. See appendix O for details of these data and a forest plot of these effect sizes. Of the 59 comparisons, 17 recorded participants in the red condition performing significantly worse than the comparison colour and 0 studies recorded participants in the red condition performed significantly better than the comparison colour. Of the 27 other comparisons that did not reach statistical significance (i.e. $p > 0.05$), 21 had a point estimate less than 0 (i.e. the red condition scored worse) and 21 had a point
estimate larger than 0 (i.e. the red condition scored better). The overall effect size of the RACE and its 95% CI is:

\[ d = -0.28, [-0.40, -0.15] \]. Here, Cohen’s \( d \) value indicates a small effect and the 95% CI spans values that range from no effect to a small effect.

7.3.2. Model 2: Analysis split on whether the experiment shares authorship with Elliot et al., (2007).

To determine if the effect sizes reported in papers that share authorship with the original paper (Elliot et al., 2007) differ from those that are independent replications, a meta-analysis, split on this distinction, was run. Within the analysis 16 comparisons shared some authorship with the original Elliot et al., (2007) study (including the original paper that comprised 8 comparisons across 4 experiments). Therefore, 43 comparisons did not share authorship and can be considered independent replications (consisting of both conceptual and direct replications). See appendix P for details of these data and a forest plot of these effect sizes. The overall effect size and 95% CI’s for the studies grouped by shared authorship are given below:

Shared authorship; \( d = -1.08, [-1.56, -0.60] \). Cohen’s \( d \) value indicates a large effect and the 95% CI spans values that range from a moderate effect to a large effect.

Non-shared authorship; \( d = -0.11, [-0.20, -0.02] \). Cohen’s \( d \) value indicates no effect and the 95% CI spans values that sit within the range of no effect to a small effect. It should be noted however that the cut off used for a small effect is 0.2 and so the likelihood of the effect being classified as a 'small effect' is minimal.
7.3.3. Model 3: Meta-analysis split on comparison colour.

To determine if the comparison colour influenced the strength of the RACE a meta-analysis model that was split on colour condition was run. The number of comparisons for each colour comparison is as follows; black (1), blue (11), green (26), grey (15) and white (6). As the black condition was used only once in Elliot et al., (2007) it was omitted from the analysis. See appendix Q for details of these data and a forest plot of these effect sizes. The overall effect size and 95% CI’s for each comparison colour are given below:

Blue; $d = -0.08$, [-0.34, 0.17]. Cohen’s $d$ value indicates no effect and the 95% CI spans values that range from no effect to a small effect.

Green; $d = -0.23$, [-0.40, -0.06]. Cohen’s $d$ value indicates a small effect and the 95% CI spans values that range from no effect to a small effect.

Grey; $d = -0.33$, [-0.50, -0.15]. Cohen’s $d$ value indicates a small effect and the 95% CI spans values that range from a moderate effect to no effect.

White; $d = -0.11$, [-0.80, 0.58]. Cohen’s $d$ value indicates no effect and the 95% CI spans values that range from no effect to a small effect.

7.3.4. Model 4: Analysis split on whether the colour attributes were matched at presentation.

To determine if equating colour stimuli on saturation and brightness properties influenced the RACE a meta-analysis split on this variable was run. Within the analysis 48 experiments used colour stimuli that were not matched at perception, whilst 11 experiments did match
colour stimuli at perception in terms of saturation and brightness. See appendix R for details of these data and a forest plot of these effect sizes. The overall effect size and 95% CI’s for the studies grouped by whether or not colour stimuli was matched at perception are given below:

Colours matched at perception: $d = -0.22$ [-0.64, 0.20]. Cohen’s $d$ value indicates a small effect and the 95% CI spans values that span the range from no effect to a moderate effect. Such a large estimation span indicates a high degree of variability within the analysis.

Colours not matched at perception: $d = -0.26$, [-0.39, -0.13]. Cohen’s $d$ value indicates a small effect and the 95% CI spans values that indicate no effect to a small effect.

7.3.5. Model 5: Analysis split on whether the material emerged from published or grey literature.

To determine if the estimated effect size varied as a function of whether the research was published literature or grey literature, a meta-analysis split on this variable was run. Within the analysis 39 colour comparisons emerged from published material, whilst 20 colour comparisons emerged from grey literature. See appendix S for details of these data and a forest plot of these effect sizes. The overall effect size and 95% CI’s for the studies grouped by whether or not colour stimuli was matched at perception are given below:

Published literature: $d = -0.49$ [-0.68, 0.30]. Cohen’s $d$ value indicates a moderate sized effect and the 95% CI spans values that span the range from a small effect to a moderate effect.
Grey literature: $d = -0.28, [-0.40, -0.15]$. Cohen’s $d$ value indicates a small effect and the 95% CI spans values that indicate no effect to a small effect.

7.4. Discussion

7.4.1. Summary of results

Literature that assessed the RACE through experimentation (both published and unpublished work) was acquired through a systematic search strategy. This literature was filtered and a meta-analysis was run to gain a better understanding of the existence and robustness of the RACE. The discussion regarding the results of the meta-analysis will rely more on the 95% CI’s than the point estimates of the effect size. The analytical approach to results within this thesis (see 5.1.3.1) highlighted the advantages of estimation analysis over that of dichotomous analysis. In brief, it is preferable and more accurate to consider and interpret the range of probable values rather than interpreting the point estimate value solely through the standard canned effect sizes described in 8.2.5. In doing so the interpretation will provide a more measured, and ultimately more accurate, assessment of the meta-analysis.

The five models described in the meta-analysis can be interpreted as follows. When including all colour comparisons in the meta-analysis (model 1) the 95% CI spans an effect size that is interpreted as a continuum that runs from ‘no effect’ to a ‘small effect’. However, when this data is split on whether or not the experiment shares authorship with Elliot et al., (2007) two very different effect sizes emerge (model 2); those who do share authorship with Elliot et al., (2007) reveal an effect size whose 95% CI can be interpreted as
a 'moderate effect' to a 'strong effect', whilst analysis of those who do not share authorship with Elliot et al., (2007) reveal a 95% CI that is interpreted as 'no effect'. Model 3 asked if the comparison colour influenced the size of the RACE. The bias uncovered in model 2 was accounted for in model 3 by not including a colour comparison unless it had been examined by both authors who do share authorship with Elliot et al., (2007) and those who do not (hence black was omitted from the analysis given it was only used in Elliot et al., 2007). The two chromatic colours in this model (blue and green) reveal a 95% CI that indicates either 'no effect' or a 'small effect'. The effect sizes of the two achromatic colours in this model (grey and white) vary more than the chromatic colours and so are less precise in their estimation. The 95% CI of these colours span effect sizes that range from 'no effect' to a 'moderate effect' and the 95% confidence interval of the white comparison is particularly large, spanning effect sizes that indicate the red condition could both impede and improve cognitive performance. Model 4 examined if whether or not colour stimuli had been equated on lightness and chroma at perception influenced the estimate of the effect size. This analysis revealed no differences between those studies that did equate colours values at perception and those that did not; both reveal a 95% CI that is interpreted as either 'no effect' or a 'small effect'. Model 5 asked if the observed effect sizes varied as a function of whether the research was published literature or grey literature. This analysis revealed that the estimation of the effect size of published material ranged from a small effect to a moderate effect, whilst the estimated effect size of grey literature ranged from no effect to a small effect. Interpretation of the 95% CI’s show that both estimations imply the effect could be a small effect. It is possible that the larger effect size of the published material is biased as mean differences in published research generally reach statistical significance (i.e.
publication bias). It is also possible that other grey literature (that may or may not reach statistical significance) was not uncovered in the search for grey literature and this too could influence the findings, although this is an inherent problem in a grey literature search. In communicating with researchers who held grey literature, care was taken to ensure that the research was of similar methodological rigour to that of the published literature and none of the grey literature studies had been previously submitted for publication but rejected. Taking these possible bias factors into consideration it is sensible to assume that no meaningful differences in the effect size estimations of published and grey literature exist.

7.4.2. Findings, limitations and considerations for further research

The most obvious finding to note is that, when all relevant studies are included in a meta-analysis the best estimate of the size of the RACE is that it is either small or does not exist. However, when these data are split on whether or not the paper shares authorship with Elliot et al., (2007) the point estimate of the effect size of those who do share authorship rises from $d = -0.184$ to $d = -0.804$ (interpreted as a 'strong effect'), whilst for those who do not share authorships the point estimate of the effect size drops to $d = -0.121$ (interpreted as 'no effect'). This pattern is typical in psychology research (Makel, Plucker & Hegarty, 2012) although the reasons are not always entirely clear. Such a division could be due to experimenter bias (both intentional and unintentional) but a more likely argument, given the pattern of results so far in this thesis and the consultation with Prof. Elliot, emerges. It seems, as stated previously, that the RACE is a fragile effect. Its presence has been elusive throughout this thesis and the consultation with Prof. Elliot highlighted how small, seemingly inconsequential, changes to the experimental method (see 5.1.4.3.) can have a
large impact over whether or not the RACE is recorded. The RACE emerges from a paper (Elliot et al., 2007) authored by a number of achievement motivation theorists who, between them, have decades of experimental experience in this field. Aspects of research design they may take for granted (e.g. ensuring immediate feedback is given, asking for demographic information at the end of an experiment) may be overlooked by researchers coming from psychology research backgrounds other than achievement motivation. The authors of Elliot et al., (2007) may be adept at ensuring the perfect achievement context environment is created to elicit the RACE but other researchers may not have the necessary experience and knowledge to do so. Whilst the fault in inappropriate achievement setting ultimately lies with the researcher conducting the replication there are some very important points regarding research design (e.g. not including performance contingencies, the importance of immediate feedback, stereotype threat precautions) that are glossed over in the Elliot et al, (2007) paper. For example, at the onset of this thesis the definition of achievement context used (i.e. "Achievement contexts are situations in which competence is evaluated and both positive outcomes (i.e., success) and negative outcomes (i.e., failure) are possible.", pp 156) was taken from Elliot et al., (2007) and so there was no reason to assume that other method issues such as those stated above required consideration. Such oversights are likely to be mirrored by other researchers who have also attempted to replicate the RACE. Therefore, the discrepancy between researches that varies on shared authorship could be due to differences in creating the perfect experimental setting in which the RACE can emerge. Of course, if such a perfect environment is required then the validity of the RACE as a recognised psychological effect is greatly diminished, but that is an argument besides the results of the meta-analysis that will be considered further in the
One other consideration regarding the division in shared authorship is that those authors of Elliot et al., (2007) may be reluctant to publish null findings, especially given that it was them who first identified the RACE. Evidence for this comes from the fact that these authors have not published any null results and also not a single author of Elliot et al., (2007) released any grey literature data for inclusion in the meta-analysis. Rather unintuitively, publishing null effects could actually strengthen their argument (Francis, 2012) as null hypothesis significance testing pre-supposes that (if the typical $\beta$-value of 0.8 is applied) 1 in 5 experiments will result in a type II error. Hence, one could argue that their pattern of significant results every time are 'too good to be true'. This reluctance to disseminate null findings therefore may also be partially responsible for the clear division in effect sizes between those who authored Elliot et al., (2007) and those who did not.

Without wishing to cast aspersions over researchers' integrity it would be remiss to not consider the possibility that a range of research practices used to increase the likelihood of publication (often the key metric in terms of career success) could also partly contribute to the findings regarding shared authorship. Neuroskeptic (2012) describes these practices including 'overselling' (either intentionally or unintentionally exaggerating the importance of their work), 'p-value fishing' (tweaking analysis to reach $p < 0.05$) and the partial publication of data (i.e. intentionally leaving out co-variates that could explain the variation more succinctly). John, Loewenstein and Prelec (2012) measured the prevalence of such questionable research practices through sending anonymous questionnaires to 2,000 research psychologists. Their research found that 63.4% of respondents failed to report all dependent measures, 55.9% of respondents checked whether they needed to collect more data to reach significance (i.e. ‘data peeking’), and 45.8% of respondents selectively
reported experiments that ‘worked’. Whilst this thesis makes no specific assertions about the RACE and the authors of Elliot et al., (2007) it is clear that sub-optimal research practices are prevalent in psychology research and in theory this could explain, to some extent, why those authors from Elliot et al., (2007) are much more likely to perform significant replications of the RACE.

There are several other variables that were planned to be entered into the meta-analysis but were not eligible. The first was to include gender within each analysis. Unfortunately, research papers rarely publish the gender characteristics of the sample per condition (not one instance was uncovered during the literature search) and so this precluded including gender as an additional variable. Gender effects would be interesting to examine within a meta-analysis given the gender discrepancy described in section 2.4.4. The second planned variable to include was the type of dependent variable administered to participants. Unfortunately the variation in the dependent variables used (20 were used in total) was considerable and only one dependent measure (verbal analogies) was used on more than two occasions in both research that did and did not share authorship with Elliot et al., (2007). Model 2 of the analysis highlighted a clear bias in the pattern of results regarding whether or not papers shared authorship with Elliot et al., (2007) and including dependent variables that were used exclusively by these sub-groups could have biased the meta-analysis effect size estimate. The third variable that was planned for the meta-analysis was whether or not an achievement context was set. Unfortunately, the majority of papers obtained for the meta-analysis did not explicitly reference factors that would be needed to determine the degree to which an achievement context was set, for example; if immediate feedback was made salient; if the researchers kept a serious rapport with participants; or if
external performance contingencies were offered. Further research examining gender, the
type of dependent variable and the degree to which an achievement context was set would
be ideal avenues for further research.

7.4.3. Experiment 7 summary

In summary, the five models described in the meta-analysis demonstrated that; when all
available studies are considered, it is likely that the RACE is either non-existent or very small
(model 1); if the study shares authorship with Elliot et al., (2007) it is likely to reveal a
moderate to strong RACE, whilst if the study does not share authorship it is likely to find the
RACE is non-existent (model 2); when considering all available studies there are no
differences between effect sizes depending on the comparison colour (model 3), if colour
stimuli were equated at perception (model 4) or if the research was published or not (model
5). Possible reasons for the results of model 2 include; differences in creating the ideal
experimental achievement context setting; method issues regarding the onset of stereotype
threat; the authors of Elliot et al., (2007) being reluctant to disseminate null findings; and
sub-optimal research practices.

Chapter 7 represents the logical conclusion of primary and secondary research for this
thesis. This thesis began by using secondary research methods in a natural experiment
before undertaking primary research experiments across a range of settings before finally
conducting a thorough meta-analysis of research on the RACE. Chapter 8 will consider the
body of work described in this thesis and identify conclusions, limitations and avenues for
further research.
Chapter 8: General Discussion

8.1. Overview

This thesis examined the influence of red coloured stimuli on cognitive performance in an achievement setting, this effect is termed the red achievement-context cognitive effect (RACE). The first aim of the thesis was to assess the RACE in an applied context, the second aim was to investigate the gender discrepancy within existing work, the third aim was to provide an independent, direct replication of the RACE as well as providing other conceptual replications where relevant and the final aim was to construct and run a meta-analysis of published and grey literature.

8.2. Summary of experimental chapters

8.2.1. Chapter 3: Using secondary research methods to assess the RACE in an applied setting

Using secondary research methods experiment 1 examined if schools that require red colour uniform are more likely to record lower measures of academic achievement. The theory that underpins the RACE (see 2.2.) suggests that in an achievement context, such as that found in school examinations, the presence of red coloured stimuli (e.g. school uniform) could impede performance in examinations through the onset of avoidance motivation. Experiment 1 examined the RACE in an applied context (aim 1) and by assessing gender specific outcome measures of academic achievement also investigated the gender discrepancy within existing work (aim 2). The analysis demonstrated a weak, non-significant, but consistent (27 of 29 outcome measures) effect of red uniform colour negatively influencing measures of academic success in UK primary schools. However, the analysis was
underpowered and so caution should be exercised in the interpretation of the results. Further analysis with much larger samples are required to investigate this.

8.2.2. Chapter 4: The influence of red stimuli on cognitive performance in an online setting

Experiments 2 and 3 moved towards primary research methods and assessed if red coloured stimuli impeded performance in verbal analogy measure taken in an online setting that was accessed remotely. These experiments examined the RACE in an online (i.e. applied) setting (aim 1) and specific analyses assessed whether or not gender influenced the RACE (aim 2). The methods used to determine these aims represent a conceptual replication of the RACE (aim 3). Both experiments 2 and 3 showed no effect of colour on cognitive performance, regardless of the participant's gender. Possible reasons for not uncovering the RACE include;

1. Extraneous colour stimuli present in the applied setting (e.g. taking the task on a red laptop) could have masked the RACE. However, other previous work had uncovered colour effects (including the RACE) using similar methods (Gnambs, Appel & Batinic, 2010; Lichtenfeld et al., 2012).

2. The possibility that the variability in participants' general ability masked the emergence of the RACE. The majority of previous work that did uncover the RACE (i.e. Elliot et al., 2007; Gnambs, Appel and Batinic, 2010) had used homogenous participant samples (i.e. university students) whereas experiments 2 and 3 recruited from a wider, more general population from social media.
8.2.3. Chapter 5: The influence of red stimuli on cognitive performance in a classroom setting

To investigate if the two reasons detailed above did contribute to not uncovering the RACE, experiment 4 gave participants (recruited from a homogenous student population) a verbal analogy task in a classroom setting. In this setting much more control over extraneous colour stimuli was possible. Experiment 4 examined the RACE in an applied context (i.e. an ecologically valid college classroom; aim 1) and provided a conceptual replication of the RACE (aim 3). Too few male participants were recruited to investigate the gender discrepancy in previous work. No significant or meaningful differences in verbal analogy performance as a function of colour were uncovered.

Following experiment 4, a consultation with Prof. Elliot (lead author of Elliot et al., 2007) highlighted several changes to the experimental method that were implemented in experiment 5. These included further consideration of the achievement context and minimising the possibility of stereotype threat onset.

8.2.4. Chapter 6: The influence of red stimuli on cognitive performance in a laboratory setting

Experiment 5 assessed if a red (relative to white) stimulus impeded performance in a verbal analogy task. Recommendations made by Prof. Elliot were implemented in this experiment. This experiment moved to a controlled laboratory setting to ensure that no extraneous colour stimuli could interfere with the independent variable colour stimuli. Therefore experiment 5 can be considered a direct, independent replication of the second experiment.
by Elliot et al., (2007), thus investigating aim 2 of this thesis. The results of experiment 5 demonstrated that viewing red stimuli did impede performance in the verbal analogy task.

However, experiment 6, a direct replication of experiment 5 with the addition of one more colour condition (green) failed to find any effect of colour on task performance. The inconsistent results between experiments 5 and 6 (that used identical methods) coupled with the admission from Prof. Elliot that very specific methods are required to elicit the RACE, imply that perhaps the RACE is not a robust effect, i.e. its emergence as an effect can be negated through small method changes and even when the correct method is used, it does not emerge consistently.

8.2.5. Chapter 7: Using meta-analytic methods to investigate the RACE

Experiment 7 described a meta-analysis of both published and unpublished research on the RACE. This chapter investigated the fourth aim of this thesis; to construct and run a meta-analysis of published and grey literature. In particular the meta-analysis investigated how robust the RACE is across different experiments. This meta-analysis demonstrated that when considering all research available on the RACE it is likely that the RACE is either a small effect or does not exist. However, when these data are split regarding whether the paper shares authorship with Elliot et al., (2007) or not, two very different effect size estimates emerge. If a paper does share authorship with Elliot et al., (2007) the estimated effect sizes range from a "moderate-to-large" size effect, whilst if the paper does not share authorship with Elliot et al., (2007) the estimated effect sizes indicate there is no effect.
8.3. Main findings

The main findings of this thesis are considered through the lens of the four thesis aims set out in 2.5. The over-arching finding of the thesis (that the RACE is a fragile effect that is likely to be inconsequential in applied contexts) stems from assessing aims 1, 3 and 4 of the thesis. The direct and conceptual replications of the RACE in this thesis (aim 3) demonstrate the fragility of the RACE in both applied contexts (aim 1) and in laboratory settings. The RACE only emerged in one experiment and a replication of this experiment did not uncover the RACE. This finding coupled with the meta-analysis (aim 4) that implies the effects can only be uncovered regularly by those who construct ‘perfect’ experimental settings indicate that the RACE is not a robust effect that emerges in various settings. Consultation with the lead author of Elliot et al., (2007) also referred to the notion that very specific experimental settings are required in which to uncover the RACE. This main finding is discussed at length in 8.3.2. First, a comment will be made regarding aim 3 of the thesis; to investigate the gender discrepancy in the existing literature.

8.3.1. The gender discrepancy in the RACE

In this thesis experiments 1, 2 and 3 examined if the emergence of the RACE varied as a function of gender. Experiment 1 demonstrated that those schools requiring red uniform performed slightly worse than those schools requiring other coloured uniforms. This degradation in ability was marginally larger in male students than female students by 0.95%. Although this difference appears small, a 1% change in performance represents a change of 508 places in the national rankings of primary schools. However, this analysis was underpowered and so further research with a larger sample of schools is required to further
clarify this finding (see 9.4.3. for further details regarding the extension of experiment 1).

Experiments 2 and 3 assessed the RACE in an online environment. No effects of colour were uncovered in either of the mixed gender samples. Considering male participants, the results of experiment 2 showed that both the blue and grey condition performed worse than the red condition but these differences were not statistically significant and the effect sizes were not consistent with previous work. In the green colour condition in experiment 2 male participants in the red condition scored higher, thus the directionality of the difference is not what would be expected. It should be noted that the analysis of male participants in experiment 2 was underpowered and so in experiment 3 a larger number of male participants were recruited and sufficient power was observed. Although male participants in the red condition performed worse than those in the blue or green condition in experiment 3, these effect sizes were small, statistically insignificant and were not consistent with previous work. The observed differences here are more likely to emerge from natural variation in the data opposed to systematic variation caused by the colour stimuli.

In summary, this thesis uncovered no clear evidence that the RACE emerges in male participants or female participants. Although some trends emerged in male participants that suggested those in the red condition performed worse than the other colour conditions, careful interpretation of these data suggest the differences are likely to occur from unsystematic variation in the data.
8.3.2. Fragility of the RACE

8.3.2.1. The evidence for the fragility of the RACE

The experiments in this thesis and the meta-analysis indicate that the emergence of the RACE is inconsistent. Additionally, when the RACE is recorded (i.e. experiment 5) it is not always possible to replicate this finding, even when the methods and participant populations are identical (i.e. experiment 6). Whilst this difference could be attributed to a type II error, the large difference in the effect sizes recorded between experiments 5 and 6 coupled with the results of experiments 1 to 4, and the meta-analysis in this thesis suggest that the failure to replicate is due to the effect being fragile.

In addition to this noted inconsistency, the lead author of Elliot et al., (2007) states that very specific research methods are required to uncover the RACE (see 6.1.4.3). If these are not met, the RACE is unlikely to emerge. Such specific methods include; ensuring no external performance contingencies (i.e. prizes) are provided; ensuring immediate feedback is made salient; maintaining a serious and professional rapport with participants; using a stop watch to indicate time pressure; and asking demographic information at the end of the experiment to ensure that stereotype threat does not mask the RACE. The specificity and importance of these issues are not explicitly stated in Elliot et al., (2007) and so it is likely that many researchers who have attempted to replicate the RACE have not been able to do so as they have fallen foul of these method issues. It is possible that experiments 2, 3 and 4 did not uncover the RACE because the above issues were not attended to. However, although attending to these issues can lead to the emergence of the RACE (i.e. experiment 5) this is not always the case (i.e. experiment 6). It may be that the authors of Elliot et al., (2007) are
adept at creating the ‘perfect’ (albeit with low ecological validity) experimental setting and so are more likely to uncover a fragile effect such as the RACE in such ideal, artificial conditions.

One observation that is incongruous with the idea that the RACE is a fragile effect is the subtlety with which other researchers have managed to elicit the effect. For example, Elliot et al., (2007) showed participants a red square for only 2s; Lichtenfeld et al., (2009) elicited the RACE by having the word ‘red’ printed in the copyright text of the question paper and; Gnambs, Appel and Batinic (2010) evoked the RACE in an online environment (accessed remotely) using a general knowledge dependent measure. Additionally, Tanaka and Tokuno (2011) were able to elicit avoidance motivation (although they did not measure cognitive performance) by simply wearing a red t-shirt under a white lab coat. These findings demonstrate a robustness of the RACE that is not consistent with the observed fragility proposed in this thesis. It is possible that the effect is robust in its onset and then decays quickly if the achievement context is not properly set or it could be that some third variable is influencing the results. Whilst it is not immediately obvious what such a variable could be there are examples in psychological replications that uncover such explanations. Take for example, Doyen’s (2012) direct replication of Bargh’s (1996) finding that priming individuals with words that stereotype the elderly result in participants walking slowly down a corridor. Doyen found that the effect only emerged when experimenters manually timed the duration (using stop watches) and not when using an automated timing system. Hence Doyen concluded that researcher bias, not priming, elicited the effects uncovered in Bargh (1996). It is plausible that something analogous to this is driving the significant results in the published literature. A final explanation is that the three studies described above (Elliot et
al., 2007; Lichtenfeld et al., 2009; Gnambs, Appel & Batinic, 2010) are type I errors that have been published and promoted, due in part to the ‘pressure to publish’ and the tendency for journals to publish papers that uncover effects (i.e. publication bias).

8.3.2.2. The implications of the fragility of the RACE

There are very few (if any) applied contexts where the 'correct' achievement setting is present for the RACE to emerge. In section 2.4.1. several applied contexts were suggested where the RACE may be relevant. However, reflecting on these possibilities now, it seems that none of them would be an appropriate environment in which the RACE can emerge. Take for example educational (e.g. school/university) and professional (e.g. accounting/military) examinations; often there is an external performance contingency (i.e. parental 'prizes' or the lure of increased salary); sometimes feedback is not immediate but takes months to be obtained and often demographic information is asked at the start of the examination. Hence for these three reasons the RACE would probably not emerge in formal examinations. Another possible applied setting suggested in 2.4.1. regards healthcare professionals or business strategists making important decisions about a patient's health or a business venture, again the setting may preclude any influence of the RACE; feedback may not be immediate (it may take years for the correct decision to be known) and in both instances external performance contingencies (i.e. salary, promotion or maintaining a job) could be at play. It is difficult to determine a specific applied setting where the RACE would emerge and this thesis cannot offer a single one. Even if it is possible to determine such an applied setting, experiments 5 and 6 demonstrate that the effect can be inconsistent even when the setting is 'perfect'
If there are no (or very few) instances where the RACE can emerge then its value as an effect is limited and it could be considered an experimental quirk that has no meaningful importance outside of the laboratory.

8.3.2.3. Why do the findings in this thesis contradict those in the published literature on the RACE?

The main finding of this thesis (i.e. that the RACE is a fragile effect that seldom emerges) is not mirrored in the body of published literature that describes the RACE (see 2.3. for a literature review). It is likely that publication bias and other related pressures on researchers have skewed the body of published literature to suggest the RACE is a reliable and meaningful effect, when the research in this thesis would suggest this is not so.

8.3.2.3.1. Publication bias

Section 2.4.2 suggested that further investigation of the RACE (through direct replications) was required as publication bias could be skewing the present literature for three reasons. These three reasons are given and explained below against the backdrop of the main findings in this thesis.

1. Journals are less inclined to publish research that finds no effect.

The traditional funding model of scientific journals requires subscriptions to be commercially successful. This is achieved by publishing 'exciting' and 'interesting' research. To most, research that uncovers psychological effects is inherently more interesting than research that finds no effect. This requirement for exciting and interesting research is forced
upon academic researchers who require publications in order for their careers to progress (van Dalen & Henkens, 2012). Thus, there is a bias towards conducting and publishing research that only uncovers effects, whilst suppressing research that finds no effect. This practice is flawed on two counts. First, it makes it very difficult to disprove an effect once it has been published if research that does not find the effect is not published. As a result, whole fields of research can be built and sustained on a handful of experiments that uncover an effect, even if many more studies (that may never be disseminated) have failed to find the effect. Secondly, the criteria that many journals and academics use to judge if an effect has been found (i.e. NHST and a $p$-value of $<.05$) is flawed and has long been questioned as the sole means to draw meaningful significance regarding if an effect is present in data (Rozeboom, 1960; Gill, 1999; Nickerson, 2000).

A related topic to this is the lack of null replications available from the authors of Elliot et al., (2007). Statistical theory suggests that 1 in 5 attempts to uncover an effect should result in a type II error (Howell, 2013). Work by Francis (2012) suggests that because null results are never found for some psychological effects then they are literally 'too good to be true'. Applying this idea here, the authors of Elliot et al., (2007) have published 11 experiments (Elliot et al., 2007; Maier, Elliot & Lichtenfeld, 2008; Lichtenfeld et al., 2009; Elliot et al., 2011) that all uncover the RACE with zero type II errors. The likelihood of this happening is 8.6%. Of course this is plausible and there may be a multitude of reasons why the null replications are not provided, not least the aforementioned pressure to publish significant results. Null replications were requested from the authors of Elliot et al., (2007) for the meta-analysis but none were provided. The reason given was that further work was either 'preliminary' or 'incomplete'. Whether or not such work was considered preliminary or
incomplete because the methods were not conducive to uncovering the RACE is unclear but remains a possibility without seeing the original data and methods.

2. Journals are less inclined to publish direct replications of existing effects

Another way in which journals are able to be commercially successful is to publish research that is 'novel'. By definition, simply confirming an existing effect is not as novel as proposing a new effect or extending an existing effect. Therefore, researchers rarely perform direct replications of existing work as it is often considered not novel. This is in direct contradiction to the scientific method that governs an objective research process, as a 'true' scientific effect is one that "can be regularly reproduced by anyone who carries out the appropriate experiment in the way prescribed" (Popper, 1959 cited in Novella, 2012). If no one has carried out the experiment in the way prescribed (i.e. directly replicated) then by Popper's definition a scientific effect cannot be described as 'true'. Some researchers argue that conceptual replications are preferable to direct replications as they are able to gauge the generalisability of the effect using different methods (Carpenter, 2012). However, Pashler and Harris (2012) state that this is not the case. They posit that, when successful, conceptual replications are more likely to be published than direct replications as they are novel - thus this presents a bias. But when unsuccessful, conceptual replications are more likely to be published and assimilated into the existing literature through the argument that a confounding variable in the method negated the effect. Thus, conceptual replications will always 'add value' to a field (and in turn a researcher's publication list), through either extending the effect (in the case of a successful conceptual replication) or identifying confounding variables that mask the effect (in the case of an unsuccessful conceptual
replication). Unsuccessful direct replications on the other hand foster controversy, debate and disagreement and often will not be published.

3. The body of literature that describes the RACE is particularly susceptible to issues of publication bias

The final point to make regarding why the findings of this thesis are not consistent with the body of research that describes the RACE is that, as an effect, the RACE is susceptible to publication bias. Pashler and Harris (2012) suggest that psychological effects that are both easy to uncover (i.e. a short study with no expensive equipment) and exciting (i.e. a novel or implausible effect) tempt hundreds of researchers and students to attempt to conceptually replicate the effect. Regardless of whether the effect exists or not some of these studies will produce significant results and these are likely to be published. Additionally, such conditions would promote the use of small, low-powered conceptual replication attempts and these practices have also been found to increase the probability of attaining '$p<0.05$' (and thus have a higher probability of being published), regardless of whether the effect actually exists (Bakker, van Dijk & Wicherts, 2012).

Clearly the RACE meets these criteria. The experiment is short (the experiments in this thesis took around 25 minutes), requires no expensive equipment (only a folder, printed test sheets and colour stimuli are required) and is both novel and implausible (the idea that writing in a red book impedes cognitive performance is both provocative and unusual). Hence it is plausible that hundreds of researchers have attempted to replicate the effect, those that have found the RACE have been published, those that did not have been ignored and overall the body of research seems to indicate that the RACE is a reliable and
meaningful effect. Not only does the work in this thesis suggest that this is not the case, the preceding section clearly demonstrates how the body of published literature may have been skewed.

8.3.2.3.2. Moves towards minimising publication bias

It should be noted that in recent years, there has been a concerted effort within psychological research towards resolving the issues that stem from publication bias through increasing the number of direct replications and also through the onset of online journals.

The pursuit of direct replications to confirm existing findings has recently gained popularity. Take for example, Doyen's (2012) direct replication of Bargh's (1996) study detailed in 8.3.2.1., or a number of direct replications completed in response to Bem's (2011) controversial publication of predicting future events (Robinson, 2011; Ritchie, Wiseman & French, 2012). Another example of direct replications gaining momentum in psychological research regards that of the 'Reproducibility Project; Psychology' (Open Science Collaboration; 2015). This is a collaborative effort of 270 researchers to directly replicate 98 original and influential psychology research papers. Results from the Reproducibility Project showed that the original effects were reproduced in only 39% of the replication attempts.

Regarding the role of journals, the emergence of online only journals (e.g. BMJ Open) minimises the pressure for journals to publish only research that is novel and exciting in two ways. First, an online publication means that there is less competition for space compared to traditional printed journals and so there is less pressure to include only novel and exciting research that attracts readers. Second, several online journals do not rely on subscriptions
for revenue but instead require payment for an article submission and then the research is made available for free (e.g. PLOS one). Not only does this aid in minimising publication bias, it also further democratises research by allowing anyone to obtain it, not only those with access to astronomically priced journal subscriptions.

Whilst these factors may aid in building an accurate picture of psychological effects in the future it may still take decades of research and debate to undo the damage caused by publication bias present since the onset of experimental psychology.

8.4. Future directions for research

8.4.1. Direct replications

Whilst the experiments in this thesis suggest that the RACE is a fragile effect only emerges in specific experimental settings, clearly further clarification regarding this fragility is required. The most obvious way to achieve this is for other researchers to conduct direct replications of the experiments in Elliot et al. (2007) and experiments 5 and 6 in this thesis. If the RACE continues to emerge inconsistently this adds weight to the argument that, even when the experimental conditions are 'perfect', the effect is still fragile.

8.4.2. Extending and clarifying the findings of experiment 1

Experiment 1 assessed the impact of red uniform colour on academic achievement in primary schools. Although the results of this analysis were inconclusive, there was a clear (albeit small and statistically insignificant) trend of red uniform schools performing worse in measures of academic achievement than schools who required other coloured uniforms.
Unlike the other experiments in this paper there were no previous effect sizes that could be implemented in an a priori power calculation and so as a result the analysis was underpowered. Running a further analysis with sufficient power is therefore required. Experiment 1 used data from 292 schools but a post hoc power analysis revealed that a sample of 1,102 primary schools would be required to achieve an observed power value of 0.8. Therefore, to clarify the inconclusive findings of experiment 1 it would be interesting to re-run the experiment with an increased number of primary schools.

8.4.3. Does the fragility of the RACE discredit it as an effect worth researching?

Aside from the issue of whether or not the RACE emerges consistently when the experimental setting is 'perfect' there is a larger issue that requires consideration; if the RACE requires such specific conditions to emerge (evident through both the experiments in this thesis and from the recommendations by Prof. Elliot) then is it worth further research? Such a question would require consideration form a number of angles. First, further research that examines the degree of fragility of the effect is required, i.e. how far can you move from the perfect controlled setting before the RACE consistently fails to emerge? Second, a more philosophical approach that considers the appropriateness of researching an effect that seems to not emerge in any applied contexts is required. Of course, 'blue sky research' (i.e. research where applications to applied contexts are not immediately apparent) has its place in scientific endeavour. But the choice to research one area over another represents an opportunity cost and so it is prudent for researchers to maximise the applied impact of their research, especially given that research grants often use such applied impact criteria within applications. The main finding of this thesis suggests it is
unclear whether further research into the RACE would be considered an efficient use of researcher time to further this aim of understanding.

8.5. Conclusion

The RACE probably does not exist outside of very specific experimental settings. Therefore, for all intents and purposes, it is likely to not exist as a psychological effect that requires consideration in any applied setting. Its recognition as a reliable effect in published literature is likely to stem from publication bias and related pressures. The authors who first proposed the effect acknowledge that the RACE is unlikely to emerge unless a very specific achievement context is set. By proxy, this implies the authors who proposed the RACE acknowledge it is unlikely that the RACE would emerge in applied settings. This thesis is unable to posit a single applied environment where the RACE would emerge but remains open to the possibility of one being suggested. Although further replications are required to clarify these findings, consideration should be taken in deciding whether or not it is sensible to dedicate funding and researcher’s time to an effect that is unlikely to emerge in applied settings and be meaningful in any way.

The research in this thesis suggests there is no meaningful or significant effect of colour on cognitive performance in achievement contexts.
References


Appendix A: Colour categorisation, measurement and production

If colour stimuli are not produced using objective and reliable methods it is difficult to draw firm conclusions regarding colour effects, rendering data virtually ‘uninterpretable’ (Valdez & Mehrabian, 1994). To demonstrate that method issues are attended to empirically, this section will also describe in detail how colour stimuli were produced for the experiments within this thesis.

To understand why control over colour stimuli is important and how this can be achieved, this section begins by detailing how colour is categorised and measured and how these techniques and theories have developed over time. Next this section will examine how issues regarding colour measurement and categorisation apply to the RACE literature. An analysis will then identify specific colours that have been previously used in the RACE literature. Finally, details will be provided regarding how colours were produced to be used within the experiments in this thesis.

A.1 How do we categorise colour?

The categorisation of colour can be split into two distinct fields; the subjective perceptual categorisation of colour, which has arisen naturally throughout history and the more recent objective, scientific description and measurement of colour, termed colorimetry.

A.1.1. Perceptual categorisation of colour

The perceptual categorisation of colour covers common colour terms used in language; 'blue', 'purple', 'pink' for example. These linguistic labels and the range of colours they
describe differ between cultures and for that reason the linguistic categorisation of colour is a key battleground in the debate regarding linguistic relativity (Ozgen & Davies, 2002). Berlin and Kay (1991) conducted a study that assessed cross cultural differences of colour categorisation by investigating the frequency and usage of colour words across cultures. They found that cultures vary in the number of colour categories marked by their language and therefore some cultures use one term to cover a range of colour terms used in other cultures. Take for example the colour terms used in English and Setswana (spoken in Botswana). English contains 11 basic colour terms: black; white; red; green; yellow; blue; brown; purple; pink; orange; and grey. Setswana conversely contains only 5 colour terms that, in English, correspond to: white; black; red; grue (covering English blue and green); and brown (Davies et al., 1992). These observations led to the development of a set of universal rules that explain incremental increases in the number of colour categories used across languages. Those languages with only two colour terms (stage I) would have terms that refer to ‘light’ and ‘dark’ (somewhat analogous to English ‘white’ and ‘black’); those with only three colour terms would also include red (stage II); those with four would include either green or yellow (stage III); up to stage VII, which includes the 11 colour terms in English described above.

More recent experimental work has demonstrated that different colour terms and boundaries can actively influence perception and subsequent categorisation. For example, what in English is termed ‘blue’ is split into two in Russian; ‘goluboy’ (light blue) and ‘siniy’ (dark blue). Winawer et al., (2007) demonstrated that this split is not an arbitrary category distinction that only defines the term used to describe a colour but a concrete distinction that influences cognition, resulting in Russian speakers being able to discriminate between
blues sitting in the *goluboy* and *siniy* categories faster than English speakers. However, research such as this is not definitive and much research using pre-linguistic participants (i.e. infants) suggests that perception drives the linguistic colour terms and not the other way round (Catherwood, Crassini & Freiburg, 1987; Franklin & Davies, 2004). Additionally, Berlin & Kaye’s and Winawer et al.’s methods and conclusions have been debated (Saunders, 2000) and cultural exceptions to the colour hierarchy have been uncovered (Davies et al., 1992). Although the debate continues, the key point that the number of subjective colour categories differ between languages and cultures, regardless of how these differences develop, remains accurate.

A.1.2. Colorimetry

Colorimetry is defined as the science of quantifying and describing human colour perception in an objective manner (Ohno, 2000). As colour perception is an internal representation that differs between individuals the quantification of colour in such a manner is difficult but achievable. Colorimetry as a field is complex and often uses vague terms incorporated from many colour models and spaces that have arisen from several independent applied movements. This thesis will briefly illustrate several colour models from the perspective of three movements:

1. Early attempts at systematically defining colour (≈1758 to ≈1929).
2. Defining colour for the purpose of displaying on televisions and computer monitors (≈1930 to ≈1980).
3. Defining colour to approximate human vision (≈1931 to ≈2002).
A.1.2.1. Early attempts at systematically defining colour (≈1758 to ≈1929)

Early colour models state that colour comprises of three properties, however these properties and their definitions vary between colour models, disciplines and across time. The most agreed upon property is 'hue', which relates to the degree to which we can say a colour is red, blue, green, or yellow (Kuehni, 2001). The next property, known as 'value' by renaissance artists and 'luminosity' by early physicists, is analogous to the modern use of lightness/darkness to differentiate between colours. The final property of colour is one of 'saturation', 'chroma' or 'colourfulness', which relates to the intensity of the colour relative to how 'washed out' it appears (Kuehni, 2001). These terms arose independently to describe similar concepts through practical applications in art and natural philosophy prior to the inception of colour science as a discipline. For example 18th century French artists developed the concept of 'valuer', to define the relative amount of white in paint whilst astronomers independently developed the related concept of 'luminosity' to describe a star's light intensity (Kuehni, 2002).

Of great interest to artists, astronomers and natural philosophers in the 18th century was the mapping of all colours onto a three dimensional space that could be used to systematically define and record colour. These included Mayer's (1758) double triangular pyramid, Runge's (1810) sphere and Wundt's (1896) double cone (all cited in Kuehni, 2001). One of the most successful early models of colour space is the Munsell Colour System (Munsell, 1905 cited in Sebe & Lew, 2003), a colour system based on a sphere. What Munsell did differently to those that preceded him was to not place the 'pure colours' in the centre of the sphere but instead to place the mid-grey at the centre and to determine the
chromatic colours from there, in doing so radically redefining the term 'chroma' in the process (Kuehni, 2001).

The Munsell Colour System uses a form of notation based on three key attributes of colour; hue, value and chroma. Munsell defines hue as the degree to which we distinguish one colour from another, as a red from a yellow, a green, a blue or purple (Kuehni, 2002). Such hues can be elicited from viewing the colour spectrum by dispersing sunlight through a prism. Value was defined as the degree to which we distinguish a light colour from a dark colour and allows a light red (i.e. pink) to be distinguished from a dark red (i.e. maroon). The definition of chroma, Munsell’s great contribution to colorimetry (Kuehni, 2001), relates to the amount of grey in a colour but this somewhat unintuitive concept is best described through Cleland’s ([1921] cited in Kuehni, [2002]; pp. 186) example of emerald stones and green grapes: "We may say that an emerald is green and that it is light, but we can say that certain grapes are green and also light. Yet there is a decided difference between their respective colours if we place them side by side. Both may be green and of the same value of light, but the emerald is strong in colour and the grape is weak in colour or greyer".

A.1.2.2. Defining colour for the purpose of displaying on televisions and computer monitors (=1930 to =1980)

The next advancement in formulating colour spaces developed when attempting to display colour on electronic screens. One such model to achieve this was the RGB (red, green, blue) colour model. The theory behind this model is to mimic the function of cone opsins in the eyes based on the Young-Hemholtz theory of trichromatic colour vision (1850 cited in Hurvich & Jameson, 1957). As there are three cone cells in the eye that are sensitive to red,
green and blue wavelengths it is possible to reproduce colours by displaying combinations of red, green and blue light. Typically each colour is given a value from 0 (no colour) to 255 (full colour) and combinations of RGB give rise to a range of perceptible colours. Whilst this colour space can accurately reproduce colour it is unintuitive for the end user. Consider that to turn a bright red (R = 234, G = 16, B = 16) into pastel pink (R = 249, G = 165, B = 165) a user must slightly increase the red value but greatly increase the green and blue values.

Several researchers attempted to transpose the RGB model to a model more suited to the human perception and categorisation of colour, most notably the Hue, Saturation, Value model (HSV; Smith, 1978) and the Hue, Saturation, Lightness model (HSL; Joblove & Greenberg, 1978). These models returned to the more intuitive three dimensional values that Munsell had used but mapped the RGB model onto complex shapes (hexcones and double hexcones respectively), not simple spheres.

A.1.2.3. Defining colour to approximate human vision (=1931 - ≈2002)

Although both Munsell’s early work and the HSV/HSL colour models made advances in introducing objectivity to colour categorisation their relative simplicity belies the deep complexity of human colour vision processing. Human colour vision cannot be mapped onto a three dimensional shape as the range of true colours that humans perceive is not uniform. For example when comparing a green hue and red hue of equal luminance that are highly saturated, the green hue will appear brighter due to the physical composition of the LMS cones in the retina (Termed the 'Hemholtz-Kohlrausch effect', Nayatani, 1997). Additionally, the composition of the LMS cones differs between individuals meaning that different people will perceive the same colour differently even at this early stage of colour processing. To
counter these issues the International Commission on Illumination (abbreviated to CIE in respect of its French name; Commission Internationale de l’éclairage) developed the ‘CIE standard observer’ (Palmer, 1979), which represents the 'average human chromatic response'. The resulting colour space, the CIE 1931 colour space, represents the gamut of human vision and further refinements of this model: the CIE LAB; CIE LUV; and; CIECAM02 (the most recent adaptation) have become the dominant models in defining human colour perception today. These models define colour using six properties (Fairchild, 2004):

- **Hue.** The degree to which a stimulus can be described as similar to or different from stimuli that are described as red, green, blue, and yellow.
- **Lightness.** The brightness of a stimulus relative to the brightness of a stimulus that appears white under similar viewing situations.
- **Brightness.** The perceived quantity of light emanating from a stimulus.
- **Chroma.** The colorfulness of a stimulus relative to the brightness of a stimulus that appears white under similar viewing conditions.
- **Colourfulness.** The perceived quantity of hue content (difference from grey) in a stimulus. Colorfulness increases with luminance.
- **Saturation.** The colorfulness of a stimulus relative to its own brightness.

A cylindrical version of this space, the CIE LCh, is much more intuitive than the CIE LAB space for non-colour vision specialists as it lists values that correspond to lightness, chroma and hue. This model combines the depth and accuracy of the CIE LAB colour model with the intuitiveness of the Munsell/HSV-type colour spaces and will be used exclusively throughout the remainder of this thesis. Where colour attributes are given in colour spaces other than
CIE LCh these values have been converted for consistency. For details of how the colours were converted from one space to another see Lindbloom (2013).

A.2. Applying colour measurement and categorisation to the literature on RACE

A.2.1. Equating chroma and lightness when assessing hue

A prevalent issue in colour psychology is ensuring that similar colours are used across different studies. With over 1,000,000 colours discernible to the naked eye (Neitz, Carroll and Neitz, 2001) and the difficulties of defining colours detailed above, it is imperative that researchers properly define and produce the relevant colour. Central to this issue is the notion that colour stimuli should vary only on the property that is intended to be examined. For example, if hue is to be assessed (as it is in the RACE literature) then it is important that only hue varies between the two colour stimuli and that lightness and chroma are kept constant.

This methodological consideration of equating lightness and chroma appears consistently and prominently throughout work that examines the RACE (e.g. Elliot et al., 2007; Maier, Elliot & Lichtenfeld, 2008; Mehta and Zhu, 2009; Gnambs, Appel & Batinic, 2010) and previous work in colour psychology (e.g. Whitfield & Wiltshire, 1990; Valdez & Mehrabian, 1994). It is important to state that this thesis uses the same definition of colour property equality as Elliot and Maier (2012) who state that (in reference to units of the CIE LCh colour model); values within 5 units of each other can be classified as functionally equivalent. The equation of lightness and chroma refers only to the production of chromatic colours (e.g. red, green, blue) and not achromatic colours (i.e. black, white and grey). Achromatic colours
have no chroma and therefore cannot be equated on this dimension. Considering lightness in achromatic colours; black and white cannot be equated on lightness to chromatic colours as they contain 0% and 100% lightness values respectively, however grey can be equated on lightness to chromatic colours.

However, a large number of studies that have uncovered the RACE, including research with high academic impact, did not equate lightness and chroma across stimuli and this finding contradicts many claims in the literature that state that such equation is imperative. Table A1 below details four studies examining the RACE containing eleven experiments. Details are given regarding the colour production technique and whether chromatic and achromatic comparison colours were equated on lightness and chroma. These four studies were selected as they demonstrate strong evidence (i.e. large effect sizes across a rigorous research design) of the RACE in the literature and they also highlight the need for the equation of chroma and lightness in their manuscript. In only four experiments (those in Maier, Elliot & Lichtenfeld, [2008] and Elliot, Payen, Brisswalter, Cury & Thayer, [2011]) are colour attributes equated adequately across chromatic and achromatic colour stimuli. Importantly, in only two of eight instances is a chromatic comparison colour equated on lightness and chroma, suggesting that the equation of these properties is not critical to uncovering the RACE.
Table A.1. A description of comparison colours used in a selection of studies and whether or not these colours were matched on chroma and lightness to the red stimuli.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Elliot et al., (2007) Exp. 1</td>
<td>Green.</td>
<td>Yes</td>
<td>No</td>
<td>Black</td>
<td>N/A</td>
<td>N/A</td>
<td>Pen Ink</td>
</tr>
<tr>
<td>Elliot et al., (2007) Exp. 2 &amp; 3</td>
<td>Green</td>
<td>No</td>
<td>Yes</td>
<td>White</td>
<td>N/A</td>
<td>N/A</td>
<td>Off-the-shelf paper</td>
</tr>
<tr>
<td>Elliot et al., (2007) Exp. 4 &amp; 5</td>
<td>Green</td>
<td>Yes</td>
<td>Yes</td>
<td>Grey</td>
<td>N/A</td>
<td>No</td>
<td>Printed onto paper in-house</td>
</tr>
<tr>
<td>Maier, Elliot &amp; Lichtenfeld, (2008)</td>
<td>None</td>
<td>N/A</td>
<td>N/A</td>
<td>Grey</td>
<td>N/A</td>
<td>Yes</td>
<td>Printed onto paper in-house</td>
</tr>
<tr>
<td>Exp. 1, 2a &amp; 3. Gnambs, Appel &amp; Batinic (2010), experiment 1</td>
<td>Green</td>
<td>No</td>
<td>No</td>
<td>None</td>
<td>N/A</td>
<td>N/A</td>
<td>Computer monitor</td>
</tr>
<tr>
<td>Gnambs, Appel &amp; Batinic (2010), Exp. 2</td>
<td>Blue</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Computer monitor</td>
</tr>
<tr>
<td>Elliot, Payen, Brisswalter, Cury &amp; Thayer, 2011.</td>
<td>Blue</td>
<td>Yes</td>
<td>Yes</td>
<td>Grey</td>
<td>N/A</td>
<td>Yes</td>
<td>Printed onto paper in-house</td>
</tr>
</tbody>
</table>

Note: ‘Exp.’ = Experiment. All chromatic and achromatic colours are matched to a reference red stimulus that differs between studies. The ‘N/A’ (not applicable) designation indicates that it is impossible to match achromatic colours on chroma to chromatic colours and also impossible to match black and white stimuli on lightness values to chromatic colours.

Experiments that present colour stimuli on digital displays (e.g. Gnambs, Appel & Batinic, 2010; Lichtenfeld et al., 2012) have the additional issue of visual display variation; the same colour is presented differently on different displays owing to variation in technology and visual settings. Therefore only experiments that calibrate monitors to accurately display colour (i.e. using an external colorimeter [Mollen, 1999]) would be able to state they had equated colour values effectively. Gnambs, Appel and Batinic (2010) acknowledge this flaw
and state that this represented a trade-off between colour control and ecological validity in the experimental setting.

In summary, the absolute requirement for the equation of colour attributes is unclear given that those who contend it is imperative have published significant results when they have not attended to such issues in the design of their stimuli. However, in the interest of caution and consistency the colours used within this thesis will be equated on lightness and chroma throughout.

A.2.2. Ensuring colours are typical representations

A related problem to ensuring colours are equated on specific values is ensuring colours are perceived as typical representations of the intended hue. As well as ensuring colours are matched on specific values it is also important to consider which colour is being matched to others.

The research that describes the RACE suggests the effect is rooted in the association of red with danger and threat, thought to emerge from both ingrained biological responses (e.g. blood (Gerend and Sias, 2009) and learnt societal responses (e.g. red warning signage, red pen marking [Elliot and Maier, 2012]). These examples consist of vibrant, vivid reds. It may be that darker reds (e.g. maroon) and lighter reds (i.e. those approaching pink) would not elicit the RACE even if they were matched on lightness and chroma to other colours. Thus it is important to ensure a red stimulus is selected that is appropriate to evoking the RACE.

Therefore this thesis will identify a red stimulus that has been previously used to study the RACE. This ensures that a red stimulus, suitable to eliciting the required response is
presented. Comparison colours to use in control conditions will then be produced to be functionally equivalent (i.e. within 5 units of the CIE LCh colour space) to this red. Finally, these colours will be assessed by individual observers to ensure that they are typical representations of the intended hue.

A.3. Defining colour stimuli within the present experiments

A.3.1. Identifying the appropriate red

Given the issues detailed above the colours used in this thesis will be based on colours already used to uncover the RACE. The colour attributes from seven experiments are listed below in table A.2. The studies were selected on the basis that: they explicitly examined the RACE; the colour stimuli were controlled during perception (i.e. either printed stimuli or on-screen stimuli using calibrated monitors); and that evidence for the RACE was shown.

Table A.2. A list of colour values for stimuli used previously in the RACE literature.

<table>
<thead>
<tr>
<th>Study</th>
<th>Experiment</th>
<th>Type of presentation</th>
<th>L</th>
<th>C</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elliot et al., (2007).</td>
<td>2 &amp; 3</td>
<td>Printed on paper</td>
<td>45.7</td>
<td>55.3</td>
<td>24.9</td>
</tr>
<tr>
<td>Elliot et al., (2007).</td>
<td>4 &amp; 5</td>
<td>Printed on paper</td>
<td>44</td>
<td>45</td>
<td>24.3</td>
</tr>
<tr>
<td>Maier, Elliot &amp;Lichtenfeld, (2008).</td>
<td>1 &amp; 2</td>
<td>Printed on paper</td>
<td>48.9</td>
<td>52.2</td>
<td>27</td>
</tr>
<tr>
<td>Elliot et al., (2011).</td>
<td>1</td>
<td>Printed on paper</td>
<td>49.9</td>
<td>50.9</td>
<td>27</td>
</tr>
</tbody>
</table>


The LCh values for the ‘ideal’ red were calculated by taking the mean of each value. Thus, the ideal red stimuli for experiments examining the RACE is: \( L = 47.13, C = 50.85, h = 26.03 \).
The red stimuli used in this thesis will be as close to these values as is feasible, at the very least within 5 units, which is considered functionally equivalent to human vision (Elliot & Maier, 2012).

A.3.2. Producing colours in-house

At the onset of this thesis it was evident that both electronically displayed and printed versions of the colour stimuli would be needed so that consistent colour stimuli could be presented in both online and real-world settings.

Producing a monitor version of the colour was achieved by entering the values identified in table 3.2. into a graphical software package. Stimuli were created on a 20" CRT monitor (Sony GDM-20E21, resolution of 1024x768 at 100Hz). The display was calibrated prior to data collection using a ColorVision Spyder4Pro colour calibration unit and the Psykinematix Visual Psychophysics software.

Obtaining printed versions of the colours was achieved via two methods that were used at different times within the process. The first was to print the colours ‘in-house’ and verify their values using a spectrophotometer, the second was to purchase specialist colour card used in the design industry that had already been measured professionally.

Producing printed colours in house was very difficult. Colour swatches (approx. 1 inch square) of the target colours (red, green, blue and grey) were printed on HP 'Heavyweight Coated' paper using a HP Designjet 800PS printer. These swatches were then measured using a PR-650 SpectraScan Colorimeter in a blacked out room illuminated by a daylight simulation laboratory lamp ('Sol-Source' model by Gretag-Macbeth). These measurements
informed the alterations needed to be made for the next round of colour swatches (e.g. blue needs to be lighter, green needs a higher chroma value). This process was made difficult as changes to a colour value in a graphical software package do not correspond to the same changes of the printed colour swatch, as measured by a spectrophotometer. Additionally, matching four colours at once further compounds the difficulty as moving towards matching a pair of colours may make matching a third colour very difficult. Finding a match for all four colours on all three colour attributes took approximately 300 iterations of printing and measurement and this highlights the difficulty of producing appropriate colours for use in experiments. Further pilot testing ensured all of these colours were perceived as typical representations of their intended hue. Participants were asked to name the colour in a free naming task (100% agreement, \( N = 17 \)). The LCh values of the four colours are given in table 3.3 and these values are functionally equivalent to those used in previous work on the RACE (see section 3.3.1.). These hard copy colours were used only in experiment 4. Equivalent digital versions of these colours were used in experiments 2 and 3, which presented the colours on a computer monitor.

Table A.3. Lightness, Chroma and Hue values for the four colour stimuli produced ‘in-house’.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Lightness</th>
<th>Chroma</th>
<th>Hue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>46.23</td>
<td>52.01</td>
<td>27.48</td>
</tr>
<tr>
<td>Blue</td>
<td>45.94</td>
<td>53.22</td>
<td>284.07</td>
</tr>
<tr>
<td>Green</td>
<td>48.49</td>
<td>55.07</td>
<td>149.04</td>
</tr>
<tr>
<td>Grey</td>
<td>50.80</td>
<td>8.70*</td>
<td>151.54*</td>
</tr>
</tbody>
</table>

Note: ‘*’ denotes the fact that chroma and hue in grey stimuli are not able to be equated to *chromatic colours.*
One major disadvantage to using in-house printed colour stimuli is that it is very difficult for other researchers to reproduce identical colour stimuli given the difficulties described above. The difficult, laborious and time consuming process may deter other researchers from ensuring they match colours across experiments and this could lead to unsystematic variation within the literature. To address this issue, alternative colour stimuli were purchased to be used in experiments 5 and 6.

A.3.3. Purchasing existing colour stimuli

Given the complexities of producing colour stimuli in-house, an alternative is to purchase standardised colour stimuli that are produced to a consistent set of colorimetric values. Not only are such stimuli well controlled in terms of colour values, they are available to other researchers who wish to use the same colour. Consistency in colours used between researchers could aid the progression in colour psychology as it reduces variation in the stimuli. Two colour matching systems, the ‘RAL colour standard’ (2013) and the ‘Natural Colour System’ (NCS: 2012), that are used to provide colour matching in industry design were identified as possible sources of colour stimuli. Both colour models have published, reliable CIE LAB values and it is possible to buy card printed in these colours. Two colours (red and green) were selected that matched the colour values of previous work regarding the RACE. Stimuli from the RAL colour standard was purchased online from e-paint.co.uk and stimuli from the NCS were purchased from ncscolour.co.uk. See table A.4. for CIE LCH values of these colours. Note that these values are functionally equivalent to the other red and green stimuli used within the thesis and used in Elliot et al., (2007).
**Table A.4.** Lightness, chroma and hue colour values for purchased colour stimuli.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Colour System</th>
<th>Specific Colour Code</th>
<th>L</th>
<th>C</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>NCS</td>
<td>S 2060-Y90R</td>
<td>49.43</td>
<td>48.32</td>
<td>27.62</td>
</tr>
<tr>
<td>Green</td>
<td>RAL</td>
<td>160 50 50</td>
<td>50.0</td>
<td>49.95</td>
<td>160.0</td>
</tr>
</tbody>
</table>

*Note: ‘L’ = ‘Lightness’, ‘C’ = ‘Chroma’, ‘H’ = ‘Hue’.*

Further pilot testing ensured all of these colours were perceived as typical representations of their intended hue (100% agreement, \( N = 15 \)). These two colours were used in experiments 5 and 6.

This section served to highlight possible method issues that are relevant to both the general colour research described in chapter 1 and also the more specific RACE literature described in chapter 2. The consideration of the issues described in this section informs the research design within this thesis and helps to ensure research is efficient and effective, reliable and valid. Having described the existing literature and considered the method issues of such work, the next 5 sections describe the empirical research undertaken for this thesis that examine the aims of the thesis set out in section 2.5.
Appendix B. Example of online experiment advertisement for experiment 2. Posted on facebook, twitter and online forums.

“Do you like words?

How good is your verbal reasoning ability?

Try this 4 minute psychology experiment from the University of Surrey to find out: http://goo.gl/0xxx.

This study looks at the ability to complete verbal analogies, which are a good indication of general cognitive ability.”
Appendix C. Screenshots of each webpage of experiment 2

Webpage 1: Information page for participants

Webpage 1 transcript:

Information Page for Participants

Your participant number is (participant number).

Project title: Examining performance in cognitive tasks

To complete this study you must:

Be over the age of 16,

Have English as your first language.
Study summary

This study examines verbal reasoning. You will be asked to complete two tasks. Verbal reasoning has been used in psychology as a measure of general cognitive ability.

The findings from this study will help inform psychological developments regarding how individuals perform verbal reasoning.

This study is run by Adam Pedley under the supervision of Dr. Ally Clifford and Dr. Paul Sowden at the University of Surrey. This study has been reviewed and received a favourable opinion from the University of Surrey Ethics Committee. Any issues, complaints or concerns about any aspects of the study should be addressed to Dr. Ally Clifford, 01483 686907, a.clifford@surrey.ac.uk.

As a participant in this study you have the right to:

**Privacy and confidentiality**

You have been given a participant number at the top of the page which only you know. This will help ensure your anonymity is protected. We will protect your data in accordance with the Data Protection Act (1998). Please do not attempt to take the study multiple times.

**Withdraw from the study at any time**

You are able to withdraw from the study at any time, even after we have finished data collection. If you wish to withdraw from the study during completion please close your browser. To withdraw from the study at any point please email a.pedley@surrey.ac.uk quoting your participant number.
Debriefing

Following completion of the task you will be given your score. If you would like any further information, please email a.pedley@surrey.ac.uk.

Questions

If you have any questions regarding this study please email a.pedley@surrey.ac.uk.

To print this information page press 'ctrl+p' on your keyboard and follow the instructions.
TEST INSTRUCTIONS

In this task you will complete verbal analogies. Verbal analogies consist of three words. The first and second words are connected in some way. Your task is to find a word that has a similar connection to the third word by choosing from a given list of words.

Example 1:

forest : trees = meadow : ?

a) grass  b) hay  c) feed  d) plant  e) pasture

'grass' is the correct answer because trees form a forest and grass forms a meadow.

Example 2:
dark : light = wet : ?

As 'dark' is the opposite of 'light' you know have to find the opposite of 'wet'. Therefore 'dry' is the right answer.

This test consists of 10 verbal analogies and you have a time limit of 90 seconds. After 90 seconds has elapsed the test will stop.

Webpage 4: Colour stimuli presentation (red condition shown)
Webpage 5: Verbal analogy task

Webpage 6: Answer feedback for each question item
Webpage 7: Overall performance feedback with sample mean

Webpage 8: Post experiment question screen 1
Webpage 9: Post experiment question screen 2

Webpage 10: Final screen
Appendix D: Examples of online experiment advertisement for experiment 3. Posted on facebook, twitter and online forums.

Example 1. Used on forums or other groups where the first experiment was not listed:

“Do you like words?

How good is your verbal reasoning ability?

Try this 4 minute psychology experiment from the University of Surrey to find out:
http://goo.gl/0xxx .

This study looks at the ability to complete verbal analogies which are a good indication of general cognitive ability.”

Example 2. Used on platforms where the first experiment was listed:

“Even more verbal ability experiments!

Due to the great response from my last experiment here is another study which looks at the ability to complete verbal puzzles.

Find out how good your verbal reasoning ability is by taking this new 4 minute psychology experiment from the University of Surrey: http://goo.gl/0xxx .

This study looks at the ability to complete verbal analogies which are a good indication of general cognitive ability.”
Appendix E: Screenshots of each webpage of experiment 3 (red condition shown)

Webpage 1: Information page for participants

Webpage 1 transcript:

Information Page for Participants

Your participant number is (participant number).

Project title: Examining performance in cognitive tasks

To complete this study you must:

Be over the age of 16,

Have English as your first language.

Study summary
This study examines verbal reasoning. You will be asked to complete two tasks. Verbal reasoning has been used in psychology as a measure of general cognitive ability.

The findings from this study will help inform psychological developments regarding how individuals perform verbal reasoning.

This study is run by Adam Pedley under the supervision of Dr. Ally Clifford and Dr. Paul Sowden at the University of Surrey. This study has been reviewed and received a favourable opinion from the University of Surrey Ethics Committee. Any issues, complaints or concerns about any aspects of the study should be addressed to Dr. Ally Clifford, 01483 686907, a.clifford@surrey.ac.uk.

As a participant in this study you have the right to:

**Privacy and confidentiality**

You have been given a participant number at the top of the page which only you know. This will help ensure your anonymity is protected. We will protect your data in accordance with the Data Protection Act (1998). Please do not attempt to take the study multiple times.

**Withdraw from the study at any time**

You are able to withdraw from the study at any time, even after we have finished data collection. If you wish to withdraw from the study during completion please close your browser. To withdraw from the study at any point please email a.pedley@surrey.ac.uk quoting your participant number.
Debriefing

Following completion of the task you will be given your score. If you would like any further information, please email a.pedley@surrey.ac.uk.

Questions

If you have any questions regarding this study please email a.pedley@surrey.ac.uk.

To print this information page press 'ctrl+p' on your keyboard and follow the instructions.

Webpage 2: Demographic information
Webpage 3: Task instructions

TEST INSTRUCTIONS
In this task you will complete verbal analogies. Verbal analogies consist of three words. The first and second words are connected in some way.

Your task is to find a word that has a similar connection to the third word by choosing from a given list of words.

Example 1:
Forest: trees = meadow: ?
a) grass b) hay c) field d) plant e) pasture

Grass is the correct answer because trees form a forest and grass forms a meadow.

Example 2:
Dark: light = wet: ?
a) rain b) day c) damp d) wind e) wind

As ‘dark’ is the opposite of light, you have to find the opposite of ‘wet’. Therefore ‘dry’ is the right answer.

This test consists of 20 verbal analogies and you have a time limit of 3 minutes. After 3 minutes have elapsed the test will stop.

Webpage 4: Verbal analogy task

Webpage 4: Verbal analogy task
Webpage 5: Answer feedback for each question item

Webpage 6: Overall performance feedback with sample mean
Webpage 7: Post experiment question screen 1

Webpage 8: Post experiment question screen 2
Thank you for completing this experiment. If you would like further debriefing or have any questions, please email Adam Polkey at apolkey@ualberta.ca as an ongoing part of your participation.

Thank you for your time.

Please close your browser.

[Button: Finish & Submit]
Appendix F: Examples of posters and emails used to recruit participants for experiment 4

COME TAKE PART IN A
PSYCHOLOGY
EXPERIMENT
FREE CHOCOLATE FOR ALL
PLUS YOUR CHANCE TO WIN
A £20 AMAZON VOUCHER

In the Post-16 centre:
14th October at 1.00pm or
15th October at 12.00pm

THE EXPERIMENT WILL TAKE 35
MINS AND EXAMINES VERBAL
REASONING ABILITY

PLEASE DIRECT ALL ENQUIRIES TO: a.pedley@surrey.ac.uk
Example of recruitment email use

'As part of ongoing psychology research at the University of Surrey, this college (name removed to protect anonymity) has been selected to run a psychology experiment. This is your chance to be a participant in a genuine psychology experiment which will help form theory and shape the future of psychological research. Not only that, all participants get free chocolate and will have the chance to win two £20 Amazon vouchers, one for the highest performing participant and one drawn at random for taking part.

The experiment lasts 30 minutes and will take place on the 13th and 14th of May in room XXXX at (XXtime). The experiment looks at verbal reasoning ability in college students. Following analysis of the data the researchers will present a seminar giving details of the findings and how experimental psychology works.

This is an excellent opportunity for you to receive free chocolate, maybe win an Amazon voucher and most importantly be a real-life participant in an experiment which will be published for the world to see!'
Appendix G: Experiment booklet from experiment 4

The experiment packs were printed onto A5 paper and placed in A5 ring binders.

For ease of interpretation a contents of the experiment pack is given below and examples of each page are given on subsequent pages of the appendix. Blank pages (on which the word ‘Blank’ is printed) are inserted to aid in the layout and flow, to ensure colours cannot be seen through translucent paper and to allow for colour stimuli to be re-used. The measures of cognitive ability have not been reproduced as copyright is not held for these.

Blank

Consent Form

Demographic Information

RAPM Instructions

Blank

RAPM task. Three on each page until page 15.

Measure of Mood

Blank

Verbal Analogy Instructions

Blank
Blank

Colour Manipulation

Blank

Verbal Ability task. Six on a page until page 26.

Blank.

28.  Post experiment questions.
Consent Form

Project title: Examining performance in
cognitive tasks

I, the undersigned voluntarily agree to take part in the study on …………………

I have read and understood the Information Sheet provided. I have been given a
full explanation by the investigators of the nature, purpose, location and likely
duration of the study, and of what I will be expected to do. I have been given the
opportunity to ask questions on all aspects of the study and have understood the
advice and information given as a result.

I agree to comply with any instruction given to me during the study and to co-
operate fully with the investigators.

I understand that all personal data relating to volunteers is held and processed in
the strictest confidence, and in accordance with the Data Protection Act (1998). I
agree that I will not seek to restrict the use of the results of the study on the
understanding that my anonymity is preserved.

I understand that I am free to withdraw from the study at any time without needing
to justify my decision and without prejudice.

I understand I am entitled to debriefing of the aims and findings of the study. If I
wish to be debriefed I will provide my email address here(…………………………………………………………………………………………………)
and agree to be contacted to be debriefed.

I confirm that I have read and understood the above and freely consent to
participating in this study. I have been given adequate time to consider my
participation and agree to comply with the instructions and restrictions of the study.

Name of volunteer (BLOCK CAPITALS)

Signed Date

Name of researcher/person taking consent

Signed Date
Demographics
Participant number (printed on your information sheet) :
Age :
Gender :
Measure of General Ability
You will now complete a measure of general ability. An example is shown on the opposite page. You must decide which of the 8 shapes given fit the gap in the main figure.
In this example the correct answer is ‘8’ as the pattern requires a block containing one horizontal line and three vertical lines to complete the pattern.
There are 30 questions in this test.
**Measure of Mood**

Please indicate how you feel right now using the following pictures and scales.

1. The happy-unhappy scale which ranges from a smile to a frown. At the left extreme of the happy vs. unhappy scale, you feel happy, pleased, satisfied, contented, and hopeful. The right extreme end of the scale represents feeling completely unhappy, annoyed and unsatisfied. The middle picture represents feeling completely neutral, neither happy nor unhappy. Please circle which picture best represents your mood right now.

![Happy-Unhappy Scale Diagram]

2. The excited-calm scale. At the left extreme of the scale you feel stimulated, excited, fast, intense, wide awake and agitated. At the right extreme of the scale, you feel completely relaxed, calm, sluggish, dull, sleepy and unaroused. The middle picture represents feeling completely neutral, neither excited nor calm. Please circle which picture best represents your mood right now.

![Excited-Calm Scale Diagram]

3. The controlled-in-control scale. At the left extreme of the scale you feel completely controlled, influenced, cared for, awed, submissive and guided. At the right extreme of the scale you feel completely controlling, influential, in control, important and dominant. The middle picture represents feeling completely neutral, neither controlled or in-control. Please circle which picture best represents your mood right now.

![Controlled-In-Control Scale Diagram]
TEST INSTRUCTIONS

In this task you are given three words. The first and second words are connected in some way. Your task is to find a word that has a similar connection to the third word by choosing from a given list of words.

**Example 1:**
forest : trees = meadow : ?

a) grass b) hay c) feed d) field e) pasture

'Grass' is the correct answer because trees form a forest and grass forms a meadow. Therefore 'a)' is marked as your answer by circling it.

**Example 2:**
dark : light = wet : ?

a) rain b) day c) damp d) wind e) dry

As 'dark' is the opposite of 'light', you now have to find the opposite of 'wet'. Therefore 'e) dry' is the right answer.
ITEMS
Post experiment questions

Please do not refer back to the previous pages in this booklet

What do you think we were trying to find out?
Can you recall the words printed on the front of the experiment booklet?
Can you recall the colour of the square found twice in the booklet?
Can you recall how many questions were in the test?
Do you wear glasses?
Do you have any history of being colour blind?
Appendix H: Information sheet from experiment 4

**Project title:** Examining performance in cognitive tasks  
**Participant number:** .............

**Study summary**
This study examines verbal reasoning. You will be asked to complete a ten minute task. Before that you will be given a task to assess your current mood and a task which will assess general intelligence. You are in no way required to take part in this task. This task will not only be a good test of your own verbal reasoning it will also help form part of a research article which will be submitted for publication. This in turn will be read by psychologists the world over and will help push forward psychological research. We are also offering two prizes (£20 Amazon voucher) for participants. One prize will be given for the highest scoring participant, the other will be given to a participant drawn at random. If you would prefer to leave the room and get back to your lunch, please do so now.

The findings from this study will help inform psychological developments regarding how individuals perform verbal reasoning.

This study is run by Adam Pedley under the supervision of Dr. Ally Clifford and Dr. Paul Sowden at the University of Surrey. This study has been reviewed and received a favourable opinion from the University of Surrey Ethics Committee. Any issues, complaints or concerns about any aspects of the study should be addressed to Dr. Ally Clifford, 01483 686907, a.clifford@surrey.ac.uk.

As a participant in this study you have the right to:
- **Privacy and confidentiality.**
  You will be assigned a participant number which only you know. This will help ensure your anonymity is protected. Your participant number is printed on your information sheet, please note it down. We will protect your data in accordance with the Data Protection Act (1998).
- **Withdraw from the study at any time.**
  You are able to withdraw from the study at any time, even after we have finished data collection. To withdraw from the study at any point please email a.pedley@surrey.ac.uk quoting your participant number.
- **Protection from physical and mental harm.**
  None of the tasks will put you at risk of physical or mental harm.

**Debriefing**
Following data analysis you will be sent a short email (if requested) detailing the results and conclusions of the research. If you would like debriefing, please enter an email address on the consent form. We will not be giving group feedback, not individual scores. You will also have the opportunity to attend a seminar at your college explaining the findings of the research at a later date.

**Questions**
If you have any questions regarding this study please raise your hand and an investigator will come over to provide an answer. If you have any questions regarding the study after the task has been completed please email a.pedley@surrey.ac.uk.

Appendix I: Script used in experiment 4

**INSTRUCTIONS FOR EXPERIMENTER**

Instructions in italics are to be spoken to the participants verbatim. Instructions in bold are actions for you to take.
Before participants enter the room place a folder, a pen and an information sheet on each table. As participants enter the room dictate where they sit so that friend groups are not clustered together.

We will now take part in a psychology experiment that will take approximately 30 minutes. If you do not wish to participate you are free to leave now or once you have read the information sheets which give more detail about the experiment. There are two £20 Amazon vouchers to be won, one for the highest scoring participant and one given to a participant at random. The experiment is split into three parts. First you will take a test of general intelligence, then we will measure your current mood before completing the main verbal ability task.

It is important you follow the instructions from the experimenter so please do not turn through the pages in the folder until instructed.

Any questions?

Please read the information sheet in front of you. These sheets are yours to keep.

Wait for sheets to be read.

If you are happy to continue please open your folders to page 2. Read and sign the consent form on page 2 and fill in the demographic information on page 3. Please do not turn to page 4 until instructed. Please enter today’s date on the first line. Once you have done this please put down your pen so the experimenter knows you are ready.

Wait for all instructions to be completed.

Please turn to page 4 of your experiment pack. This page contains the instructions for the measure of general intelligence; I will now read it to you:
You will now complete a measure of general ability. An example is shown on the opposite page. You must decide which of the 8 shapes given fit the gap in the main figure. In this example the correct answer is ‘8’ as the pattern requires a block containing one horizontal line and three vertical lines to complete the pattern.

There are 30 questions in this test.

Any questions?

You have 5 minutes to complete the task; please do not turn past page 15. You may begin.

Time 5 minutes.

Please turn to page 6.

This is a measure of your current mood. I will read the instructions given (read instructions).

Experiment 4 researcher script (continued).

Measure of Mood

Please indicate how you feel right now using the following pictures and scales.

1. The happy-unhappy scale which ranges from a smile to a frown. At the left extreme of the happy vs. unhappy scale, you feel happy, pleased, satisfied, contented, and hopeful. The right extreme end of the scale represents feeling completely unhappy, annoyed and unsatisfied. The middle picture represents feeling completely neutral, neither happy nor unhappy. Please circle which picture best represents your mood right now. There are no right and wrong answers so please just follow your instinct.

(wait for response)

2. The excited-calm scale. At the left extreme of the scale you feel stimulated, excited, frenzied, jittery, wide-awake and aroused. At the right extreme of the scale, you feel completely relaxed, calm, sluggish, dull, sleepy and unaroused. The middle picture
represents feeling completely neutral, neither excited or calm. Please circle which picture
best represents your mood right now. There are no right and wrong answers so please just
follow your instinct.

(wait for response)

3. The controlled-in-control scale. At the left extreme of the scale you feel completely
controlled, influenced, cared for, awed, submissive and guided. At the right extreme of the
scale you feel completely controlling, influential, in control, important and dominant. The
middle picture represents feeling completely neutral, neither controlled or in-control.
Please circle which picture best represents your mood right now. There are no right and
wrong answers so please just follow your instinct.

Wait for all pens to be placed down.

Please turn to page 18. These are instructions for the verbal analogy task. I will now read
them to you.

TEST INSTRUCTIONS

These tasks are called 'verbal analogies'.

In this task you are given three words.

The first and second words are connected in some way. Your task is to find a word that has a
similar connection to the third word by choosing from a given list of words.

Example 1:

forest : trees = meadow : ?

a) grass  b) hay  c) feed  d) field  e) pasture

'Grass' is the correct answer because trees form a forest and grass forms a meadow.
Therefore 'a)' is marked as your answer by circling it.
Example 2:

dark : light = wet : ?

As 'dark' is the opposite of 'light' you know have to find the opposite of 'wet'. Therefore 'dry' is the right answer.

Now you will take two sets of verbal analogies. You will have 4 minutes for each set and a 2 minute break between sets.

Do you have any questions?

Please turn the page and ensure the word 'Items' is printed above the square and nothing is printed within the square. If the word 'Items' is not present or there is other writing within the square please raise your hand (silently count 5s). At this point the participants view the colour manipulation for 5s. Avert your gaze so you remain blind to each participant's condition. Please do not turn past page 26. You may begin (time 5 minutes).

Time up, please turn to page 28. These are the post experiment questions. Please answer all the questions you can and then close your folder.

Wait for post experiment questions to be completed.

The experiment is now complete. You may leave. Thank you for your time and effort. If you have any questions please see Mr. Pedley or email the address given on your information sheet. Please remember the information sheets are yours to keep. You will have the opportunity to find out more about the experiment in a seminar which Mr. Pedley will give at a later date.

Collect equipment. END

Appendix J: Experiment pack used in experiments 5 and 6

The experiment packs were printed onto A4 paper and placed in A4 ringbinders.
For ease of interpretation a contents is given below and examples of each page are given on subsequent pages of the appendix. Blank pages (on which the word ‘Blank’ is printed) are inserted to aid in the layout and flow, to ensure colours cannot be seen through translucent paper and to allow for colour manipulations to be re-used. The measures of cognitive ability are not reproduced because copyright is not held for these.

Page 1: Consent Form

Page 2: Blank

Page 3: Measure of Mood Title Page

Page 4: Measure of Mood

Page 5: Blank

Page 6: VA Test Instructions

Page 7: Practice VA Task Context Setting

Page 8: Blank

Page 9: Practice Items Title Page

Page 10 to 12: Practice VA Task

Page 13: Blank

Page 14: Central VA Task Context Setting
Consent Form

Project title: Examining performance in cognitive tasks

I the undersigned voluntarily agree to take part in the study on ………………..
I have read and understood the Information Sheet provided. I have been given a full explanation by the investigators of the nature, purpose, location and likely duration of the study, and of what I will be expected to do. I have been given the opportunity to ask questions on all aspects of the study and have understood the advice and information given as a result.
I agree to comply with any instruction given to me during the study and to cooperate fully with the investigators.
I understand that all personal data relating to volunteers is held and processed in the strictest confidence, and in accordance with the Data Protection Act (1998). I agree that I will not seek to restrict the use of the results of the study on the understanding that my anonymity is preserved.
I understand that I am free to withdraw from the study at any time without needing to justify my decision and without prejudice.
I understand I am entitled to debriefing of the aims and findings of the study. If I wish to be debriefed I will provide my email address here(……………………………………………………………………………………………) and agree to be contacted to be debriefed.
I confirm that I have read and understood the above and freely consent to participating in this study. I have been given adequate time to consider my participation and agree to comply with the instructions and restrictions of the study.

Name of volunteer (BLOCK CAPITALS)
Signed Date

Name of researcher/person taking consent
Signed Date
Measure of Mood
Measure of Mood

Please indicate how you feel right now using the following pictures and scales.

1. The happy-unhappy scale which ranges from a smile to a frown. At the left extreme of the happy vs. unhappy scale, you feel happy, pleased, satisfied, contented, and hopeful. The right extreme end of the scale represents feeling completely unhappy, annoyed and unsatisfied. The middle picture represents feeling completely neutral, neither happy nor unhappy. Please circle which picture best represents your mood right now.

2. The excited-calm scale. At the left extreme of the scale you feel stimulated, excited, frenzied, wide awake and aroused. At the right extreme of the scale, you feel completely relaxed, calm, sluggish, dull, sleepy and unaroused. The middle picture represents feeling completely neutral, neither excited or calm. Please circle which picture best represents your mood right now.

3. The controlled-in-control scale. At the left extreme of the scale you feel completely controlled, influenced, cared for, awed, submissive and guided. At the right extreme of the scale you feel completely controlling, influential, in control, important and dominant. The middle picture represents feeling completely neutral, neither controlled or in-control. Please circle which picture best represents your mood right now.
TEST INSTRUCTIONS

These tasks are called ‘verbal analogies’.
In this task you are given three words.
The first and second words are connected in some way.
Your task is to find a word that has a similar connection to the third word by choosing from a given list of words.

Example 1:
forest : trees = meadow : ?
a) grass  b) hay  c) feed  d) green  e) pasture
'Grass' is the correct answer because trees form a forest and grass forms a meadow. Therefore ‘a)’ is marked as your answer by circling it.

Example 2:
dark : light = wet : ?
a) rain  b) day  c) damp  d) wind  e) dry
As ‘dark’ is the opposite of ‘light’, you now have to find the opposite of ‘wet’. Therefore ‘e) dry’ is the right answer.
First you will take a practice test of verbal analogies. This will give you the chance to become familiar with verbal analogies before the real test. You will not receive the scores of this test.
Practice Items
Next you will take the real verbal analogy test. This test is a good indication of general intelligence. The experimenter will give you your results and feedback once you have finished the task.
Test Items
Post experiment questions

Please do not refer back to the previous pages in this booklet

What do you think we were trying to find out?

Can you recall the words printed on the front of the experiment booklet?

Can you recall the colour of the square found in the booklet?

Can you recall how many questions were in the test?

Do you wear glasses?

Do you have any history of being colour blind?
Participant number (printed on your information sheet):

Age:

Gender:
Please indicate your A-level results:
A-Level 1:
A-Level 2:
A-Level 3:
A-Level 4:
Appendix K: Information sheet for participants in experiments 5 and 6

---

**Project title:** Examining performance in cognitive tasks  
**Participant number:** ............  

**Study summary**  
This study examines verbal reasoning. You will be asked to complete a ten minute task. Before that you will be given a task to assess your current mood and a task which will assess general intelligence. You are in no way required to take part in this task. This task will not only be a good test of your own verbal reasoning it will also help form part of a research article which will be submitted for publication. This in turn will be read by psychologists the world over and will help push forward psychological research. We are also offering two prizes (£20 Amazon voucher) for participants. One prize will be given for the highest scoring participant, the other will be given to a participant drawn at random. If you would prefer to leave the room and get back to your lunch, please do so now.  

The findings from this study will help inform psychological developments regarding how individuals perform verbal reasoning.  
This study is run by Adam Pedley under the supervision of Dr. Ally Clifford and Dr. Paul Sowden at the University of Surrey. This study has been reviewed and received a favourable opinion from the University of Surrey Ethics Committee. Any issues, complaints or concerns about any aspects of the study should be addressed to Dr. Ally Clifford, 01483 686907, a.clifford@surrey.ac.uk.  
As a participant in this study you have the right to:  

**Privacy and confidentiality.**  
You will be assigned a participant number which only you know. This will help ensure your anonymity is protected. Your participant number is printed on your information sheet, please note it down. We will protect your data in accordance with the Data Protection Act (1998).  

**Withdraw from the study at any time.**  
You are able to withdraw from the study at any time, even after we have finished data collection. To withdraw from the study at any point please email a.pedley@surrey.ac.uk quoting your participant number.  

**Protection from physical and mental harm.**  
None of the tasks will put you at risk of physical or mental harm.  

**Debriefing**  
Following data analysis you will be sent a short email (if requested) detailing the results and conclusions of the research. If you would like debriefing, please enter an email address on the consent form. We will not be giving group feedback, not individual scores. You will also have the opportunity to attend a seminar at your college explaining the findings of the research at a later date.  

**Questions**  
If you have any questions regarding this study please raise your hand and an investigator will come over to provide an answer. If you have any questions regarding the study after the task has been completed please email a.pedley@surrey.ac.uk.  

---

Appendix L: Script used in experiments 5 and 6:  

**INSTRUCTIONS FOR EXPERIMENTER**

*Instructions in italics* are to be spoken to the participants verbatim. *Instructions in bold* are actions for you to take.
Place on the table an information sheet, experiment pack and a black ink pen. Fill out the date prior the participant arriving.

Greet the participant in a formal manner. Explain that you will be reading from a script throughout the experiment and that they can stop you at anytime if they are unsure about anything.

We will now take part in a psychology experiment which will take approximately 20 minutes. If you do not wish to participate you are free to leave now or once you have read the information sheets. The experiment is split into 3 parts. First you will complete a measure of your current mood. You will then complete a practice test before going on to complete an intelligence test. You will receive feedback of your performance on the intelligence test as soon as you have finished from the experimenter.

It is important you follow the instructions given so please do not turn through the pages in the folder until instructed.

Please read the information sheet in front of you. These sheets are yours to keep.

Wait for sheets to be read.

Please open your folder to page 1 and read and sign the consent form if you are happy to participate. Once you have done this please inform the experimenter and please do not turn the page until instructed.

Counter sign consent form when participant has finished.
Please turn to page 3 of the folder.

Do you see the words 'Measure of Mood'?, (wait for 3 seconds), please turn the page.

This is a measure of your current mood.

Please indicate how you feel right now using the following pictures and scales.

1. The happy-unhappy scale which ranges from a smile to a frown. At the left extreme of the happy vs. unhappy scale, you feel happy, pleased, satisfied, contented, and hopeful. The right extreme end of the scale represents feeling completely unhappy, annoyed and unsatisfied. The middle picture represents feeling completely neutral, neither happy nor unhappy. Please circle which picture best represents your mood right now. There are no right and wrong answers so please just follow your instinct.

(wait for response)

2. The excited-calm scale. At the left extreme of the scale you feel stimulated, excited, frenzied, jittery, wide-awake and aroused. At the right extreme of the scale, you feel completely relaxed, calm, sluggish, dull, sleepy and unaroused. The middle picture represents feeling completely neutral, neither excited or calm. Please circle which picture best represents your mood right now. There are no right and wrong answers so please just follow your instinct.

(wait for response)
3. The controlled-in-control scale. At the left extreme of the scale you feel completely controlled, influenced, cared for, awed, submissive and guided. At the right extreme of the scale you feel completely controlling, influential, in control, important and dominant. The middle picture represents feeling completely neutral, neither controlled or in-control. Please circle which picture best represents your mood right now. There are no right and wrong answers so please just follow your instinct.

(wait for response)

Please turn the page.

TEST INSTRUCTIONS

These tasks are called 'verbal analogies'. In this task you are given three words.

The first and second words are connected in some way.

Your task is to find a word that has a similar connection to the third word by choosing from a given list of words.

Example 1:

forest : trees = meadow : ?

a) grass       b) hay       c) feed       d) field       e) pasture
'Grass' is the correct answer because trees form a forest and grass forms a meadow.
Therefore 'a)' is marked as your answer by circling it.

Example 2:

dark : light = wet : ?

a) rain   b) day   c) damp   d) wind   e) dry

As 'dark' is the opposite of 'light', you now have to find the opposite of 'wet'. Therefore 'e) dry' is the right answer.

Now you will take a practice test of verbal analogies. This will give you chance to become familiar with verbal analogies before the real test. You will not receive the scores of this practice test.

Do you have any questions?

You will have four minutes to complete the practice test and the experimenter will leave the room whilst you complete the task. If you finish the practice questions before four minutes is up please inform the experimenter who will be able to hear you through the door.

Indicate the correct answer by circling the appropriate word.

Please turn the page.

Do you see the words 'practice items'?, (wait for 3 seconds), you may begin.

Leave room, time four minutes.
Enter room now holding stopwatch in a prominent position.

Time is up. Please turn the page. Next you will take the main verbal analogy test. This test is a good indication of general intelligence. The experimenter will immediately give you your results and feedback once you have finished the task. You will have 4 minutes to complete the test.

Please turn the page.

_Do you see the words ‘test items’?._ The researcher subtly averts gaze by looking at the script/stopwatch to avoid seeing the colour manipulation and, upon confirmation, silently counts 5 seconds.

You may begin.

Leave room, time four minutes.

Time is up. Please turn to page 22 and complete all the post experiment questions including giving details of your A-level performance on page 23. When you have finished please close your folder.

Wait for participant to close folders.

The experiment is now complete. You may leave. Thank you for your time and effort. If you have any questions please see the researcher or email the address given on your information sheet. Please remember the information sheets are yours to keep.

Give performance feedback if wanted. Give research credit.
Collect folder.

END
Appendix M: Example of email sent to researchers to request grey literature

Hello,

I am a PhD student at the University of Surrey currently collating a meta-analysis of work regarding the influence of red stimuli on cognitive performance.

In particular, I am looking to locate experiments that relate to the effects first formalised in Elliot, Maier, Moller, Friedman & Meinhardt's (2007) paper - Color and Psychological Functioning: The Effect of Red on Performance Attainment. In short, this paper demonstrated that, in an achievement context, the presence of red stimuli prompts avoidance motivation, which results in impaired cognitive performance.

I have been investigating this effect throughout my PhD and am now compiling the published and unpublished literature to run a meta-analysis on the subject. I am contacting you as you may have some relevant studies that you can send me.

In particular, I would be incredibly grateful to receive copies of your published or unpublished research that meet the following specific criteria:

1) An independent variable of colour, comparing red to at least one other chromatic or achromatic colour. The colour stimuli can be either colour in its normal visual form or lexical representations of colour.

2) A dependent variable that comprises some form of cognitive ability including, but not limited to, tests of verbal reasoning, numerical sequencing, working memory and general knowledge.

All experimental settings (e.g. lab based, online, classroom) and populations are welcome. As are student projects or details of pilot studies.

I am very interested in papers that replicate (both directly and conceptually) the red effect set out in Elliot, Maier, Moller, Friedman & Meinhardt (2007) but would also like to include other related effects that may originate from a different theoretical perspective.

Any help will be most appreciated.

Thank you,

Adam Pedley

PhD candidate at the University of Surrey, UK.

Supervised by Dr. Alexandra Clifford and Dr. Paul Sowden.
Appendix N: List of authors the email requesting grey literature was sent to.

Dr. Paul Sowden, University of Surrey, UK.

Dr. Alexandra Grandison, University of Surrey, UK.

Prof. Andrew Elliot, University of Rochester, USA.

Dr. Stephanie Lichtenfeld, University of Munich, Germany.

Prof. Markus Maier, Ludwig-Maximilians University of Munich, Germany.

Dr. Russell Hill, University of Durham, UK.

Prof. Robert Barton, University of Durham, UK.

Dr. Timo Gnambs, University of Osnabruck, UK.

Dr. Markus Appel, University of Koblenz-Landau, Germany.

Prof. Bernard Batinic, Johannes Kepler University Linz, Germany.

Dr. Stephanie von Stumm, Goldsmith’s University of London, UK.

Dr. Andrei Ilie, University of Oxford, UK.

Dr. Ravi Mehta, University of Illinois, USA.

Dr. Juliet Zhu, University of Minnesota, USA.

Dr. Christopher Thorneston, University of Rochester, USA.

Dr. Adnan Smajic, University of Missouri-St Louis, USA.

Dr. Stephanie Merritt, University of Missouri-St Louis, USA.

Dr. Piotr Sorokowski, University of Wroclaw, Poland.
Appendix O: Data and forest plot for model 1 of meta-analysis.

344


### Appendix P: Data and forest plot for model 2 of meta-analysis.

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Red</th>
<th>Non-red</th>
<th>Std. Mean Difference</th>
<th>Std. Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td>Mean</td>
</tr>
<tr>
<td>Elliot et al. (2007)L1</td>
<td>4.4</td>
<td>0.3</td>
<td>19</td>
<td>5.6</td>
</tr>
<tr>
<td>Elliot et al. (2007)L2</td>
<td>4.4</td>
<td>0.3</td>
<td>19</td>
<td>5.6</td>
</tr>
<tr>
<td>Elliot et al. (2007)L3</td>
<td>10.4</td>
<td>2.6</td>
<td>15</td>
<td>13.1</td>
</tr>
<tr>
<td>Elliot et al. (2007)L4</td>
<td>10.4</td>
<td>2.6</td>
<td>15</td>
<td>12.7</td>
</tr>
<tr>
<td>Elliot et al. (2007)L5</td>
<td>9.4</td>
<td>2.6</td>
<td>10</td>
<td>12.4</td>
</tr>
<tr>
<td>Elliot et al. (2007)L6</td>
<td>9.4</td>
<td>2.6</td>
<td>10</td>
<td>12.4</td>
</tr>
<tr>
<td>Elliot et al. (2007)L7</td>
<td>6.0</td>
<td>2.6</td>
<td>19</td>
<td>7.7</td>
</tr>
<tr>
<td>Elliot et al. (2007)L8</td>
<td>6.0</td>
<td>2.6</td>
<td>19</td>
<td>7.7</td>
</tr>
<tr>
<td>Lichtenthal, Elliot &amp; Malter (2012)G</td>
<td>6.5</td>
<td>1.8</td>
<td>11</td>
<td>5.3</td>
</tr>
<tr>
<td>Lichtenthal, Elliot &amp; Malter (2012)R</td>
<td>6.5</td>
<td>1.8</td>
<td>11</td>
<td>8.1</td>
</tr>
<tr>
<td>Lichtenthal, Elliot &amp; Malter, Elliot &amp; Pakhan (2002)G</td>
<td>11.13</td>
<td>2.65</td>
<td>22</td>
<td>13.24</td>
</tr>
<tr>
<td>Lichtenthal, Elliot &amp; Malter, Elliot &amp; Pakhan (2002)R</td>
<td>8.25</td>
<td>3.3</td>
<td>20</td>
<td>11.7</td>
</tr>
<tr>
<td>Lichtenthal, Elliot &amp; Malter, Elliot &amp; Pakhan (2002)OR</td>
<td>10.2</td>
<td>3.5</td>
<td>19</td>
<td>14.7</td>
</tr>
<tr>
<td>Miner, Elliot &amp; Lichtenthal (2003)OR</td>
<td>1.26</td>
<td>1.4</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Miner, Elliot &amp; Lichtenthal, Elliot &amp; Pakhan (2002)OR</td>
<td>1.0</td>
<td>1</td>
<td>10</td>
<td>11.8</td>
</tr>
<tr>
<td>Theronseeran &amp; Greyst (2015)A</td>
<td>3.58</td>
<td>1.93</td>
<td>40</td>
<td>3.37</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>258</td>
<td>277</td>
<td>20.8</td>
<td>-1.88 (0.55, 3.03)</td>
</tr>
</tbody>
</table>

**Heterogeneity Test:** Test for overall effect: Z = 4.14 (P < 0.001)

**P** = 0.48

**Test for overall effect:** Z = 2.30 (P = 0.02)

**Test for overall effect:** Z = 4.35 (P < 0.001)

**Test for subgroup differences:** Chisq = 15.16, df = 1 (P < 0.0001), P = 0.04
Appendix Q: Data and forest plot for model 3 of meta-analysis.

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Fixed</th>
<th>Random</th>
<th>Std. Mean Difference</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.1.1**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Fixed</th>
<th>Random</th>
<th>Std. Mean Difference</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: I² = 0.13, CH² = 91.21, df = 25 (P = 0.0001), P = 0.76%
Test for overall effect Z = 2.61 (P = 0.009)

**Table 3.1.2**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Fixed</th>
<th>Random</th>
<th>Std. Mean Difference</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: I² = 0.11, CH² = 50.04, df = 19 (P = 0.0008), P = 67%
Test for overall effect Z = 0.64 (P = 0.52)

**Table 3.1.3**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Fixed</th>
<th>Random</th>
<th>Std. Mean Difference</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: I² = 0.05, CH² = 26.22, df = 14 (P = 0.02), P = 47%
Test for overall effect Z = 3.66 (P = 0.0003)

**Table 3.1.4**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Fixed</th>
<th>Random</th>
<th>Std. Mean Difference</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: I² = 0.51, CH² = 18.58, df = 5 (P = 0.002), P = 73%
Test for overall effect Z = 0.02 (P = 0.98)

**Table 3.1.5**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Fixed</th>
<th>Random</th>
<th>Std. Mean Difference</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: not applicable
Test for overall effect Z = 3.78 (P = 0.0001)

Total (95% CI) 2.434

Heterogeneity: I² = 0.15, CH² = 22.81, df = 5 (P = 0.0001), P = 75%
Test for overall effect Z = 4.35 (P = 0.0001)

Test for substantial differences: CH² = 43.74, df = 4 (P = 0.0001), P = 92.0%

346
Appendix R: Data and forest plot for model 4 of meta-analysis.

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Red</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Weight</th>
<th>Mean Difference</th>
<th>IV, Random, 95% CI</th>
<th>IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elliott et al. (2002)</td>
<td>6.6</td>
<td>2.12</td>
<td>18</td>
<td>7.7</td>
<td>2.12</td>
<td>22</td>
<td>1.5%</td>
<td>0.51</td>
<td>-1.14</td>
<td>[0.61, 0.63]</td>
<td></td>
</tr>
<tr>
<td>Lichtenfeld, Elliott &amp; Feinman (2002)</td>
<td>5.5</td>
<td>1.9</td>
<td>11</td>
<td>5.2</td>
<td>1.9</td>
<td>11</td>
<td>1.2%</td>
<td>0.10</td>
<td>[0.72, 0.64]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lichtenfeld, Elliott &amp; Feinman (2002)</td>
<td>5.5</td>
<td>1.9</td>
<td>11</td>
<td>5.2</td>
<td>1.9</td>
<td>11</td>
<td>1.2%</td>
<td>0.10</td>
<td>[0.72, 0.64]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water, Elliott &amp; Luczynski (2002/03)</td>
<td>12.6</td>
<td>1.4</td>
<td>16</td>
<td>14.1</td>
<td>1.4</td>
<td>16</td>
<td>1.1%</td>
<td>1.00</td>
<td>[1.00, 0.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water, Elliott &amp; Luczynski (2002/03)</td>
<td>10.0</td>
<td>1.4</td>
<td>16</td>
<td>11.88</td>
<td>1.1</td>
<td>10</td>
<td>1.0%</td>
<td>1.00</td>
<td>[1.00, 0.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This Thesis (Gray)</td>
<td>7.5</td>
<td>2.59</td>
<td>15</td>
<td>8.84</td>
<td>2.59</td>
<td>15</td>
<td>1.0%</td>
<td>0.50</td>
<td>[0.12, 0.34]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This Thesis (Gray)</td>
<td>7.5</td>
<td>2.59</td>
<td>15</td>
<td>8.84</td>
<td>2.59</td>
<td>15</td>
<td>1.0%</td>
<td>0.50</td>
<td>[0.12, 0.34]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This Thesis (Gray)</td>
<td>7.5</td>
<td>2.59</td>
<td>15</td>
<td>8.84</td>
<td>2.59</td>
<td>15</td>
<td>1.0%</td>
<td>0.50</td>
<td>[0.12, 0.34]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This Thesis (Gray)</td>
<td>7.5</td>
<td>2.59</td>
<td>15</td>
<td>8.84</td>
<td>2.59</td>
<td>15</td>
<td>1.0%</td>
<td>0.50</td>
<td>[0.12, 0.34]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This Thesis (Gray)</td>
<td>7.5</td>
<td>2.59</td>
<td>15</td>
<td>8.84</td>
<td>2.59</td>
<td>15</td>
<td>1.0%</td>
<td>0.50</td>
<td>[0.12, 0.34]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This Thesis (Gray)</td>
<td>7.5</td>
<td>2.59</td>
<td>15</td>
<td>8.84</td>
<td>2.59</td>
<td>15</td>
<td>1.0%</td>
<td>0.50</td>
<td>[0.12, 0.34]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>151</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>151</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.34, Chi² = 32.23, df = 10 (P = 0.0043); P = 0.89
Test for overall effect: Z = 1.94 (P = 0.33)

3.3.2 Stimulus included at perception

Bucks (2013) | 0.66 | 0.4 | 47 | 0.94 | 0.3 | 47 | 2.1% | 0.56 | [0.12, 0.36] |
| Elliott et al. (2002) | 4.4 | 0.3 | 19 | 5.6 | 0.3 | 25 | 6.8% | 3.83 | [2.87, 2.60] |
| Elliott et al. (2002) | 4.4 | 0.3 | 19 | 5.6 | 0.3 | 25 | 6.8% | 3.83 | [2.87, 2.60] |
| Elliott et al. (2002) | 10.4 | 2.5 | 15 | 13.1 | 2.5 | 15 | 1.3% | 1.05 | [1.02, 1.08] |
| Elliott et al. (2002) | 10.4 | 2.5 | 15 | 13.1 | 2.5 | 15 | 1.3% | 1.05 | [1.02, 1.08] |
| Elliott et al. (2002) | 10.4 | 2.5 | 15 | 13.1 | 2.5 | 15 | 1.3% | 1.05 | [1.02, 1.08] |
| Elliott et al. (2002) | 10.4 | 2.5 | 15 | 13.1 | 2.5 | 15 | 1.3% | 1.05 | [1.02, 1.08] |
| Subtotal (95% CI) | 151 | | | | | | | 151 | | | |

Heterogeneity: Tau² = 0.31, Chi² = 29.62, df = 9 (P = 0.0008); P = 0.89
Test for overall effect: Z = 1.94 (P = 0.33)

3.3.3 Not applicable

Lichtenfeld, Water, Elliott & Feinman (2002/03) | 11.1 | 2.65 | 22 | 13.14 | 2.65 | 22 | 1.7% | -0.72 | [1.13, 0.11] |
| Lichtenfeld, Water, Elliott & Feinman (2002/03) | 9.20 | 3.3 | 20 | 11.72 | 3.3 | 20 | 1.0% | -0.37 | [1.20, 0.61] |
| Lichtenfeld, Water, Elliott & Feinman (2002/03) | 10.2 | 3.53 | 16 | 14.7 | 3.53 | 16 | 1.0% | -1.36 | [0.64, 0.64] |
| Subtotal (95% CI) | 52 | | | | | | | 52 | | | |

Heterogeneity: Tau² = 0.00, Chi² = 0.04, df = 2 (P = 0.93); * = 0
Test for overall effect: Z = 3.86 (P = 0.00)

Total (95% CI) | 2434 | | | 2766 | 100.0% | -0.29 | [0.40, 0.15] |

Heterogeneity: Tau² = 0.15, Chi² = 229.01, df = 56 (P = 0.0001); * = 70
Test for overall effect: Z = 4.56 (P = 0.00)
## Appendix 5: Data and forest plot for model 5 of meta-analysis.

### 3.5.1 Published Literature

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Literature</th>
<th>Std. Mean Difference IV</th>
<th>Random, 95% Cl</th>
<th>Study or Subgroup</th>
<th>Literature</th>
<th>Std. Mean Difference IV</th>
<th>Random, 95% Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Subgroup 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Study Subgroup 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Subgroup 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Study Subgroup 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Subgroup 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3.5.2 Grey Literature

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Literature</th>
<th>Std. Mean Difference IV</th>
<th>Random, 95% Cl</th>
<th>Study or Subgroup</th>
<th>Literature</th>
<th>Std. Mean Difference IV</th>
<th>Random, 95% Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Subgroup 1</td>
<td></td>
<td></td>
<td></td>
<td>Study Subgroup 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Subgroup 3</td>
<td></td>
<td></td>
<td></td>
<td>Study Subgroup 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Subgroup 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau^2 = 0.24; Chi^2 = 177.12, df = 38 (P = 0.00001); I^2 = 78%
Total effect overall: Z = 5.08 (P = 0.00001)

Heterogeneity: Tau^2 = 0.19; Chi^2 = 220.90, df = 56 (P = 0.00001); I^2 = 75%
Total effect overall: Z = 4.26 (P = 0.0001)

Test for overall effect: Z = 0.14 (P = 0.89)

Test for between-subject differences: Chi^2 = 20.91, df = 1 (P = 0.00001); I^2 = 65%