

The Stroop Task Redefined: The Multiple-Item Stroop Task

by

Sarah Uzzaman

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Department of Psychology
University of Toronto

© Copyright by Sarah Uzzaman 2017

The Stroop Task Redefined: The Multiple-Item Stroop Task

Sarah Uzzaman

Doctor of Philosophy

Department of Psychology
University of Toronto

2017

Abstract

A Multiple-Item Stroop Task was created with the goal of providing a sensitive and objective measure of attentional biases. It is a modification of the traditional single-item emotional Stroop task that is intended to assess the extent to which attentional biases distract participants from a colour identification task. The traditional version was viewed as having clinical relevance but that promise has been largely abandoned due to small effect sizes. The goal of this thesis was to investigate whether procedural modifications, informed by theory, might serve to increase the sensitivity of capturing attentional biases using the new Multiple-Item Stroop task.

Experiment 1 demonstrated that the standard Stroop effect was observed in both single-item and multiple-item variants, with the magnitude of effect being similar as is consistent with current theories of the Stroop effect, assuming that a single item is sufficient to induce a maximal effect. The emotional Stroop effect was examined but disappointingly found absent in Experiment 2 wherein a small set of emotional words was employed. Subsequently, using a larger stimulus set and more potent taboo words, a significantly enhanced emotional/taboo Stroop effect using the

new Multiple-Item paradigm was demonstrated (Experiment 3) and replicated (Experiment 4). Additionally, content relations between the observed effects and independent measures that should be related to attentional capture gave rise to a relatively consistent pattern of relation using the new multiple-item paradigm, at least relative to the traditional single-item variant (also Experiment 4). Taken together, with the use of undergraduates, these empirical results support the use of my proposed new paradigm in capturing attentional biases with healthy individuals and potentially within clinical populations. Limitations and future studies are explored.

Acknowledgments

First and foremost, all thanks and praise is to the Almighty God, The Most Gracious and The Most Merciful.

I would like to express my sincere gratitude to my supervisor Dr. Steve Joordens for the continuous support of my graduate studies, for his understanding, motivation and immense knowledge. As a supervisor, he gave me the freedom to pursue an area of research that I was interested in, and for that I would like to thank him – for his trust in me to do so. He guided and supported me throughout the research process as well as helped me to become a better and effective writer. I can honestly say that I would not be where I am today without Dr. Joordens's guidance and understanding. His open door policy allowed me to venture into his office during my second year as an undergraduate, which ultimately led to an honours thesis project together. That opportunity captured my passion for research, and ultimately paved the way for my decision to pursue it as a career. In addition, Dr. Joordens was extremely understanding of the leave of absence I took after the birth of my son and insured that my status as a graduate student and my funding were not affected in any way. I cannot imagine having a better supervisor, mentor and friend.

Besides my supervisor, I would like to thank the rest of my thesis committee, Dr. George Cree and Dr. Jonathan Cant, for their insightful comments and encouragement, but also for the hard questions and critiques which incited me to think critically and widen my perspectives. Dr. Cree was the co-supervisor during my master's degree and I am grateful for his continuous support and guidance throughout my graduate career. I would like to thank Dr. Cant for having faith in me and joining my committee mid-way without having seen any of my work – I thank him for his trust in me.

A special thanks to Mr. Ayaan Chaudhry for all his help and support. He was my lab mate, my programmer and my go-to guy for anything in the lab. We spent countless hours figuring out logistics for the experiments and he was always around whenever I needed help and also for a quick chat about life in general (when I didn't want to think about my experiment anymore!). I

thank him immensely and wish him all the best in his future endeavours. Additionally, I'd like to thank all the faculty and staff at the Psychology Department at University of Toronto Scarborough, who, in the last twelve years, have become a second family to me.

A special thanks to my family. Words cannot express how grateful I am for the encouragement and support I received from them all. Specifically, I would like to thank my parents, who have been supportive of my education and graduate career from the beginning, and helped me every step of the way.

Lastly, yet the most important of all, I would like to thank my one true rock: My beloved husband. My graduate studies would have been impossible to complete without the love, support and help from him. From the beginning, he pushed me to complete my Ph.D. and worked very hard to take care of our son while I was at the lab. His time, insight, sacrifice and patience were invaluable throughout and I am utterly grateful for them. I would be completely lost without his support; he is my pillar of strength and I cannot imagine completing my graduate studies without him.

Table of Contents

Acknowledgments.....	iv
Table of Contents.....	vi
List of Tables	ix
List of Figures	x
List of Appendices	xii
Chapter 1 Introduction	1
1 Attention.....	3
2 Consequences of Emotional Biases in Anxiety Disorders.....	7
3 Stroop.....	15
3.1 The Classic Stroop Task	15
3.2 The Robust Stroop Effect.....	17
3.3 The Relation of the Stroop Task to Attention.....	19
4 The Emotional Stroop Task	23
5 Purpose of the Thesis: Introducing the Multiple-Item Stroop Task.....	29
Chapter 2 Standard and Emotional Stroop Effects	33
6 Experiment 1: Single-item vs. Multiple-item Stroop task	34
6.1 Method	37
6.1.1 Participants.....	37
6.1.2 Materials and Apparatus	37
6.1.3 Procedure	40
6.2 Results.....	43

6.2.1	Reaction Times	44
6.2.2	Accuracy	46
6.3	Discussion	47
7	Experiment 2: Single-item and Multiple-item Emotional Stroop	47
7.1	Method	49
7.1.1	Participants.....	49
7.1.2	Material and Apparatus.....	50
7.1.3	Procedure	51
7.2	Results.....	52
7.2.1	Reaction Times	53
7.2.2	Accuracy	54
7.3	Discussion.....	55
8	General Discussion.....	56
	Chapter 3 Taboo Stroop Effect and Anxiety Measures	58
9	Experiment 3: Single-item vs. Multiple-item Positive, Negative and Taboo Stroop.....	61
9.1	Method	61
9.1.1	Participants.....	61
9.1.2	Materials and Apparatus	62
9.1.3	Procedure	63
9.2	Results.....	64
9.2.1	Reaction Times	65
9.2.2	Accuracy	67
9.3	Discussion.....	68

10 Experiment 4: Single-item vs. Multiple-item Positive, Negative and Taboo Stroop with Anxiety Measures.....	68
10.1 Method	71
10.1.1 Participants.....	71
10.1.2 Material and Apparatus	71
10.1.3 Procedure	72
10.2 Behavioural Data	73
10.2.1 Results.....	73
10.2.2 Reaction Times	74
10.2.3 Accuracy	76
10.2.4 Discussion.....	77
10.3 Relations with Independent Measures of Anxiety	77
10.3.1 Results.....	77
10.3.2 Discussion.....	84
11 General Discussion.....	85
Chapter 4 Implications and Future Endeavors	87
12 There is Stroop, and Then There is Stroop	88
13 Modelling Attentional Biases within an Undergraduate Populations	90
14 Next Steps	92
References.....	94
Appendices.....	105

List of Tables

Table 1. Experiment 1: Descriptive statistics associated with mean accuracy rates for single-item and multiple-item Stroop effect.

Table 2. Experiment 2: Descriptive statistics associated with mean accuracy rates for single-item and multiple-item Stroop effect.

Table 3. Experiment 3: Descriptive statistics associated with mean accuracy rates for single-item and multiple-item Stroop effect in positive, negative and taboo target and control words.

Table 4. Experiment 4: Descriptive statistics associated with mean accuracy rates for single-item and multiple-item Stroop effect in positive, negative and taboo target and control words.

Table 5. Pearson correlations quantifying the relation between the single-item and multiple-item Stroop effects and the various predictors of anxiety as measured in Experiment 4.

List of Figures

Figure 1. Multiple-Item Congruent Trial: The word *GREEN* written in green ink colour on 60% of the screen, and in blue ink colour on 40% of the screen. Correct response would be ‘green’. This is a congruent condition as the word and the dominant colour match.

Figure 2. Multiple-Item Incongruent Primed Trial: The word *BLUE* written in green ink colour on 60% of the screen and in blue ink colour on 40% of the screen. The correct response would be green, as the dominant ink colour is green. This is an incongruent condition as the word and the dominant ink colour do not match, and the participant is ‘primed’ or more inclined toward the incorrect response of ‘blue’ as the subordinate colour matches the word presented.

Figure 3. Multiple-Item Incongruent Neutral Trial: The word *BLUE* written in green ink colour on 60% of the screen and in red ink colour on 40% of the screen. The correct response would be green, as the dominant ink colour is green. This is an incongruent condition as the word and the dominant colour do not match; additionally, it can be said to be neutral in nature as the subordinate colour and the word are different and thus do not sway the participant in responding either correctly or incorrectly.

Figure 4. Multiple-Item Control Trial: The *XXXX* written in green ink colour on 60% of the screen and in blue ink colour on 40% of the screen. The correct response would be green, as the dominant ink colour is green.

Figure 5. Experiment 1: Mean reaction times (milliseconds) and standard error for single-item vs multiple-item Stroop effect.

Figure 6. Experiment 2: Mean reaction times (milliseconds) and standard error for single-item vs multiple-item Stroop effect for negative and neutral words and control non-words.

Figure 7. Experiment 3: Mean reaction times (milliseconds) and standard error for single-item vs multiple-item Stroop effect for positive, negative and taboo target and control words.

Figure 8. Experiment 4: Mean reaction times (milliseconds) and standard error for single-item vs multiple-item Stroop effect for positive, negative and taboo target and control words.

Figure 9. Experiment 4: The taboo Stroop effect for multiple-item and single-item in relation to the State Score in the STAI.

Figure 10. Experiment 4: The taboo Stroop effect for multiple-item and single-item in relation to the Trait Score in the STAI.

Figure 11. Experiment 4: The taboo Stroop effect for multiple-item and single-item in relation to the Taboo Questionnaire score.

List of Appendices

Appendix A – Informed Consent Form for Experiment 1

Appendix B – Informed Consent Form for Experiment 2

Appendix C – Stimulus word list for Experiments 3 and 4

Appendix D– Informed Consent Form for Experiment 3

Appendix E – Informed Consent Form for Experiment 4

Appendix F – Taboo Questionnaire for Experiment 4

Chapter 1

Introduction

Attention is a mechanism that allows certain information to be selected for further or deeper processing, and even healthy adults do not have full conscious control with respect to what they attend to (Galotti, 2004; Goldstein, 2008). In some cases, attention acts as a top-down process, meaning that we direct our attention to gather information for a particular task or goal. However, there are also instances when attention acts bottom-up, meaning that stimuli in the environment can capture our attention independently of our current goals.

The relevance of this selective attention mechanism is demonstrated in anxiety and depressive disorders wherein hyper-vigilance to certain disorder-related concepts and stimuli can lead patients to become hypersensitive to those stimuli, giving rise to a negative feedback loop that further strengthens those concepts (Beck, 2008). The existence of attentional biases of this sort is well documented and best conceptualized within anxiety individuals; although it is important to note that biases occurring early in the information-processing stream feed directly into interpretation, memory, and other higher order cognitive abilities, which can magnify the severity of the particular biases significantly. Thus this problem in attentional bias is extremely serious.

A potentially valuable method that has been used to examine attentional biases in healthy adults and patient populations is the emotional Stroop paradigm (MacLeod, C. M. & MacLeod, C., 2005; Williams, Mathews, & MacLeod, C., 1996). The emotional Stroop effect refers to the increased latency associated with the naming of the ink colour of emotional words compared to neutral words. Both the standard Stroop effect and the emotional Stroop effect reflect failures to fully focus on colour induced by our natural tendency to attend to the carrier word. However, unlike the standard Stroop task that is assumed to arise from response competition, the emotional Stroop effect is assumed to occur due to the interference of the emotionally valent carrier word even though that information is irrelevant to the simple colour-naming task (Williams et al., 1996). The colour responses to the emotional words are usually slower and more error prone than

responses to the neutral control words, possibly indicating the automatic allocation of attention to emotional stimuli (Frings, Englert, Wentura, & Bermeitinger, 2010).

Given this logic, the magnitude of emotional Stroop effect is reflected by the amount of slow down between emotional, or disorder-specific words, and neutral words, which could be used to measure the extent to which participants have particular attentional biases. Thus, the emotional Stroop task might allow one to objectively measure the extent to which specific issues are dominant in a participant's mind. In theory, the promise of the emotional Stroop task is great. However, in practice, that promise has fallen short due to small to non-existent effect sizes, difficulty in replication, context specificity, and reliability issues (Eide, Kemp, Silberstein, Nathan, & Stough, 2002; MacKay, Shafto, Taylor, Marian, Abrams, & Dyers, 2004; MacLeod, C. M. & Hodder, 1998b). The goal of this thesis is to investigate whether procedural modifications, informed by theory, might serve to increase the sensitivity of capturing attentional biases using the new Multiple-Item Stroop task.

The goal of this chapter is to describe some of the existing research and ideas relevant to the central story that I will be exploring in this thesis. As such, I will first discuss the general concept of attention, and the extent to which attention allocation is under our conscious control. I will highlight the manner in which our thoughts and experiences can bias the automatic capture of attention, using anxiety disorders as an example. I will then discuss the Stroop effect generally, and then focus in on the emotional Stroop effect more specifically, and the promise it has for providing an objective measure of attentional biases. Finally, I will discuss why that promise appears not to have been realized, thereby setting the stage for the remainder of my thesis wherein I explore ways in which that promise might be reinstated. Specifically, I have created a modified version of the standard colour-word Stroop experiment, with the hopes of enhancing the sensitivity of capturing attentional biases.

1 Attention

In this section, the mechanism of selective attention is explored: How attention is directed through conscious effort and in what ways it is unintentionally captured by salient stimuli is discussed. Although we consciously select specific information for further processing, the unattended information does not vanish but is actually processed to a certain degree in regard to its perceptual and semantic meaning. Even at that level, the unattended information has the capacity to influence our judgment and the decisions that we consciously make, to the extent where it can even switch our attention to focus on potentially salient stimuli. In the following section, I will discuss selective attention generally, highlighting priming and automaticity in regard to the spread of activation model (Collins & Loftus, 1975). In reference to the central story, as attention can be unconsciously drawn toward stimuli and concepts that are primed or hold a significance in our minds, particular biases can accentuate this process where participants are unable to concentrate on tasks without constantly being pulled toward salient stimulus features in reference to their biases.

Our senses are continuously bombarded by stimulation from both external and internal sources. At any given moment, we encounter a myriad of various sights, sounds, smells, tastes and touch sensations. In addition, we have our own cognitive thoughts, feelings, tasks and goals in mind. The amalgamation of all these sources of information confronting us, sources whose specifics change on a moment to moment basis, is quite enormous.

The brain is incapable of processing all of this incoming information deeply; and instead, it uses processes of selective attention to focus on a subset of the incoming stimuli (Anderson, 2005; Cowan, 2005). The information is usually selected based on what we require, such as to achieve a goal or to complete a task (Broadbent, 1958). Information that is selected then enters our working memory, which forms the basis of what we are 'aware' of. This leads to the subjective experience of having one thing 'in our mind' at any given second, although it changes from moment to moment as different stimuli enter and leave our consciousness. This notion of 'stream

of consciousness', first introduced by William James (1892), is the idea that there are different, ever changing, thoughts and ideas in our minds that continuously flow (hence the 'stream').

Cognitive psychologists often describe attention as a resource-limited capacity, where the brain is described as having a limited supply of mental resources that can be allocated to processing of information (Galotti, 2004; Goldstein, 2008). At any given moment, people have only a certain amount of mental energy to devote to all the possible tasks and all the incoming information confronting them. If they devote some portion of those resources to one task, less is available for others and the more complex and unfamiliar the task, the more mental resources that must be allocated to that task to perform it successfully (Broadbent, 1958; Yiend, Mathews, & Cowan, 2005).

Of particular importance to this thesis is the question of how certain incoming information is identified as being worthy of selection, and generally speaking psychologists have described two distinct ways in which selection occurs. Top-down attention refers to when we have a goal in mind and, using prior knowledge, we seek out information to achieve the goal or complete a task. This is also known as endogenous attention, which is voluntary, conceptually-driven and requires conscious effort (Galotti, 2004; Goldstein, 2008). On the other hand, bottom-up attention is activated in a reflexive manner, as salient stimulus features seemingly demand further processing. This is also known as exogenous attention, which is involuntary, stimulus-driven and occurs in an automatic fashion. There is evidence to support the notion that endogenous and exogenous attention differ in several ways, such as the source and temporal relay of their activation, the degree of control that the observer has, and the depth of processing each receives (Galotti, 2004; Goldstein, 2008).

A second critical question about selective attention centers on the fate of the stimulation that is not selected for deeper processing; is it not processed at all, or does it still receive some shallow level of processing? To explore this issue, a clever paradigm was designed where different

auditory messages were played simultaneously in each ear and participants had to attend to one of the messages. Cherry (1953) asked participants to shadow (i.e., repeat aloud) the attended message to ensure their focus, while another message was playing in the unattended ear. When asked about the message in the unattended ear, participants could only say that there was a message and could perhaps identify the gender of the voice; however they were unable to report the content of the message (Cherry, 1953; Goldstein, 2008). This suggests that the unattended message is being processed enough to allow the listener to be aware of its existence but the processing of that information may be restricted to very low levels of primarily perceptual analysis (e.g., pitch differences reflective of voice gender).

However, subsequent work suggested that despite the lack of awareness for the unattended information, it was perhaps being processed more deeply than initially assumed. For example, Eich (1984) presented homophone words (e.g., mail/male, sail/sale) in the attended ear and simultaneously presented biasing words in the unattended ear (e.g., letter/man, boat/shop). Once again, participants reported no memory of the items presented in the unattended ear. However, when asked to write out the words presented in the attended ear, participants did so with the spelling that corresponded with the biasing words presented in the unattended ear (e.g., wrote 'male' if MAN was presented; but spelled 'mail' when LETTER was presented). Thus collectively, this suggests that the unattended information is being processed to some degree, and can actually bias how we discern the attended information (Treisman, 1964).

Not only is unattended information processed but, in some cases, it is processed in a way that directly affects the focus of our selection, pulling attention towards it, as described previously when discussing the bottom-up or exogenous attention capture. In their seminal experiment, Gray and Wedderburn (1960) asked participants to shadow the message in the left ear, in which 'Dear 7 Jane' was presented at one point in the attended ear and simultaneously '9 Aunt 6' was presented in the unattended ear. Rather than reporting the 'Dear 7 Jane' message that was presented to the attended ear, participants reported hearing 'Dear Aunt Jane'. The switch to the unattended ear to hear 'Aunt' meant that the participant's attention had jumped, signifying that they were taking the semantic meaning of the message into consideration. This is in line with the

attenuation model of attention, where all incoming information is processed at a very shallow level, but due to the limited capacity of attention, only certain stimuli go on to be processed at a deeper level while the rest do not (Treisman, 1964). However, if the unattended information is salient enough, attention can be captured and switched to the salient stimulus.

To understand why some unattended stimuli can capture attention while most do not, we need to consider a concept called priming. One account of priming relates to the cognitive structure of semantic representations, which builds on the spreading activation model (Collins & Loftus, 1975; Henik, Friedrich, & Kellogg, 1983). The basic tenet is that a concept has some resting level of activation, and stimulation from either the external or internal worlds can raise that level. For example, activating the word ‘pilot’ would be assumed to spread more activation to the word ‘plane’ compared to ‘uniform’, suggesting that activation spreads but diminishes in strength having a smaller effect on less related items (Doyen, Klein, Simon, & Cleereman, 2014). If the level of activation associated with some concept ever surpasses some activation threshold, then the concept enters awareness, implying that it has captured the selection process. By this account, priming occurs automatically and passively; as long as the prime is perceived (either with or without awareness), it triggers a cascade of semantic associations, leading to faster processing of related words and concepts (Collins & Loftus, 1975; Doyen et al., 2014; Goldstein, 2008). Thus within this account of priming, for some unattended stimulus to capture attention, the amount of incoming stimulation, when summed with its resting level, must surpass threshold.

With the priming explanation of how exogenous selection occurs in mind, we can now reconsider some of the previously described experiments more clearly. When probed regarding the unattended message, participants in Cherry’s (1953) experiment were only able to report that there was a message and perhaps the gender. In an alternate version, Cherry presented the participants’ name in the unattended message, and some participants were able to detect it. Participant’s attention would automatically, without conscious control, switch to the unattended ear (Galotti, 2004; Goldstein, 2008). One’s own name is a highly salient concept, which means that it starts with a high resting level of activation, so it does not require much activation to reach ‘awareness’ threshold and capture attention. Thus the crucial concept is the factor of relevance or

importance to each individual that is causing the priming to occur. This relevance could be in the form of semantic knowledge (i.e., red means stop, green means go) or personal importance (i.e., name of a loved one). In regard to the Dear Aunt Jane experiment, the unintentional switch from one ear to the other is in the linguistic context, which is creating a ‘feeling’ of words that go together (Gray & Wedderburn, 1960). This reflects directly upon the structure of semantic representation model, where there is faster processing of related words and concepts (Collins & Loftus, 1975). Hence, the semantic context is adding the activation (priming) in addition to what is coming from the external words, and the amalgamation of these two sources is enough to draw the attention to ‘Aunt’ in one ear and then back to ‘Jane’ in the other ear.

Thus, it seems the brain is able to effectively deal with the myriad of incoming stimuli thanks to a flexible selective attention process that is sensitive to current goals but also to specific aspects of the current stimulus stream. Generally, attention is focused on information that is assumed to be goal relevant, but if any stimulus in the stream reaches a sufficiently high level of activation it can draw attention to itself. This is typically a very good thing; it is a strategic method for the mind to speed up processing of words or concepts that have subjective importance (such as one’s own name) or that signal danger (fire! help!). For example, if you are scavenging for berries when you feel your ‘predator sound’ concept become active, it would be a good idea to stop looking for little red things and instead focus on the sounds that triggered the concept. However, as we will see in subsequent sections, sometimes this process can become problematic to the point where it may lead to or even maintain mental disorders.

2 Consequences of Emotional Biases in Anxiety Disorders

This thesis is focused on gaining better measurements of attentional bias which, of course, assumes that such biases exist. Perhaps the context in which this bias has been most clearly documented, and the one the best highlights the critical role they can play in affecting cognition, comes from the clinical literature related to mental disorders. Thus, in this section, I will describe research highlighting emotional biases in mental disorders.

In line with the previous section, it is well documented that anxious individuals' attention tends to be captured unintentionally by stimuli that are disorder-specific such as threat related words, but the impact of these biases carries through to other levels of cognitive processing as well (Beck, 1976, 2008). In this section of my thesis, I will describe several experimental techniques used to demonstrate that anxious individuals display a bias toward threatening stimuli, and have even exhibited the tendency to perceive neutral stimuli as threatening. Subsequently memory is also skewed in line with this bias in the processing of emotional stimuli or events. This creates a vicious cycle, where the tendency to attend to or notice specific stimuli disproportionately feeds into better recall for those stimuli and subsequently other higher-order cognitive functions. In reference to the central story, as attentional biases exists both in healthy individuals as well as in anxious individuals, the advent of a paradigm that is sensitive in capturing attentional biases would be quite beneficial.

By definition, bias is an inclination or a perspective that is consistently chosen over other possibilities. Holding a bias toward specific information is by no means a globally negative process, nor is it exclusive to clinical populations. As elaborated in the previous section, healthy individuals constantly undergo and process information based on selectively attending to specific stimuli, for example stimuli relevant to their safety at any given time. However when there is a systematic pattern of selecting specific stimuli or schemas that results in perceptual distortion or inaccurate judgment, this can cause significant anxiety and impairment in social functioning (Clark & Beck, 2010; Mathews, 1990). This type of bias can be very debilitating and cause reduced quality of life.

Most mental illnesses, especially those related to emotions, encompass some form of bias within their etiology. These biases can be quite general, in the case of generalized anxiety disorder, or quite specific, in the case of post-traumatic stress disorder; biases can be taken from external stimuli, such as phobias, or can be internalized, such as anorexia nervosa (Davidson, Neale, Blankstein, & Flett, 2005). Thus there are several forms of biases that exist in many different

mental disorders. Although the following discussion focuses on the biases common in those with anxiety disorders specifically, the general kinds of cognitive biases that I will highlight occur within the context of many other mood and personality disorders. For the purposes of this thesis, I will focus on anxiety disorders for the remainder of this introduction. The more general relevance of my work will be returned to in the General Discussion.

The term ‘anxiety disorders’ is used to refer to a group of mood disorders characterized by feelings of anxiety, defined as worry about future events, and fear, defined in terms of a reaction to current events (Davidson et al., 2005). These feelings can cause physiological reactions, such as dizziness and sympathetic nervous system activation, and often also include mental apprehension. There are a number of different types of anxiety disorders each with its own set of criteria and symptoms, though all include the feeling of anxiety and fear that can range from simple nervousness to bouts of terror (Davidson et al., 2005).

Patients with anxiety disorder deploy their attention in distinctive ways that can be a factor in maintaining their disorder (Mathews & MacLeod, C., 1994, 2002, 2005; Yiend et al., 2005). Cognitive accounts of various mental disorders suggest that maladaptive cognitive processes can influence how emotional information is encoded and can then lead to later cognitive symptoms (Beck, 1976, 2008; Williams et al., 1996). For example, the early cognitive process of interpretation such as viewing skin pigmentation as cancerous can lead to subsequent cognitive symptoms such as rumination and worry. Thus the information gathered in the early stages of the information-processing stream is crucial and any biases within it can cause significant ramifications for later cognitive processes.

A crucial question for this thesis is to determine the stage at which the bias occurs within the information-processing stream. We know that a bias can help or hinder goal performance, and even has the potential strength to switch our currently active goal in both healthy adults as well as anxious individuals. As selective attention is the mind’s first step in determining relevant

incoming information, any biases that exist early in processing can have strong potential to produce large effects downstream.

Many different types of paradigms have been used to determine whether a bias in selective attention exists within anxious individuals. One of the earliest methods used to investigate this issue is the interference task in which the task-relevant stimulus is neutral and the distractor is either emotional or not; it is reasoned that if attention is unintentionally directed toward the emotional distractors this should show up as slower or more error prone performance on the primary task when emotional distractors are present (Goldstein, 2008). I have already discussed one such example, the dichotomous listening task. Using patients with social phobia, agoraphobia, and healthy controls, Burges, Jones, Robertson, Radcliffe, and Emerson (1981) chose words specific to each patient's phobia (e.g., public speaking) and a neutral word (e.g., parking meter), and randomly inserted them into each passage of a dichotic auditory tape. While no difference was found in detection of target words in the attended passage, phobic patients detected significantly more fear-relevant target words in the unattended passage compared to the controls. In addition, there was no difference between the groups for detection of the neutral word in the unattended passage. The difference in detection of the fear-relevant word to the neutral words displays the heightened sensitivity of anxious individuals to stimuli that are semantically associated with their fears. Thus, studies using the dichotic listening task have found that individuals reporting a high level of anxiety display a high tendency to detect emotional stimuli in the unattended channel (Burges et al., 1981; Foa & McNally, 1986; Powell & Hemsley, 1984).

The Stroop task has been at the forefront in attention research since its conception, and as such has been used extensively in attentional bias research (Stroop, 1935). In essence, the standard colour-word Stroop paradigm unfolds as follows: Different colour words (red, green and blue), control words (non-colour words) and/or non-words (series of XXX, or ###) are written in different coloured ink (red, green or blue). On congruent trials, the word and the ink colour are the same (e.g., RED written in red ink), whereas on incongruent trials, the word and the ink colour are different (e.g., RED written in blue ink). In all cases, the participants' task is to name

the colour of the ink as fast as possible. Stroop interference is measured by subtracting the congruent from incongruent trials reaction times. The Stroop effect is said to be present when the reaction times for incongruent trials are longer than the reaction times for congruent trials.

With an increased interest in cognitive accounts of emotions, several investigators proposed variations of the classic Stroop task to examine the cognitive processes associated with emotional disturbance (MacLeod, C. M., 2005; Williams et al., 1996). As those variations are critical to the empirical work of this thesis, I will describe them in detail in a subsequent section of the introduction that will follow shortly. In the remainder of this section, I discuss results from other paradigms that support the notion of an attentional bias.

Another direct method for the assessment of attentional bias is the attentional probe task (Anderson, 2005; Mathews & MacLeod, C., 1994, 2005). An array of stimuli, one emotional and one neutral, is briefly shown at two separate locations, and subsequently participants are required to detect the occurrence of a small dot probe at either one of the locations (Galloti, 2004; Goldstein, 2008). Research with this spatial attentional probe task has revealed that anxious individuals demonstrate faster detection latencies for probes occurring at the spatial location of the previously shown emotional words compared to neutral words, even when exposure was restricted in awareness (Mathews & MacLeod, C., 1994, 2005). The basic tenet for the task is that faster response rates would be seen for the locations that participants were attending to, and typically anxious individuals are faster to respond to the probe when it occurs in the same location as the emotional stimulus, be it a word or a facial expression (Anderson, 2005; Bradley, Mogg, Falla, & Hamilton, 1998; Bradley, Mogg, & Millar, 2000; MacLeod, C., & Mathews, 1988, 1991, 2004; Mathews & MacLeod, C., 1994, 2005).

Thus there is empirical evidence to support the existence of a bias in selective attention within anxious individuals, which leads to the question of whether biases are also present in other early cognitive processes. The perception of emotionality is subjective and as such differs from person

to person; a neutral stimulus to one person may be threatening to another. Thus the perception of neutral stimuli as threatening can also be viewed as a bias. The concept of priming plays a role in developing these notions, where ambiguous stimuli get linked with negative schemas as anxious individuals tend to have a preoccupation with threat-related stimuli and events (Conrad, 1974; Mathews & MacLeod, C., 1994, 2005; Yiend et al., 2005).

There is empirical evidence to support the notion of interpretive bias, a tendency to deduce ambiguous stimuli as threatening or negative leading to higher frequency of anxious reactions to everyday stimuli and events. Much of the research conducted has used homophones, for which there is an emotional and a neutral meaning (e.g. die/dye, pain/pane, guilt/gilt). When these homophone words are presented to be remembered via an audio tape, results demonstrate that anxious individuals subsequently provide a higher number of negative-meaning spellings compared to neutral homophone spellings (Eysenck, MacLeod C., & Mathews, 1987; Eysenck, Mogg, May, Richards, & Mathews, 1991).

Consistent evidence is also provided with the use of the lexical decision task where a homograph, an ambiguous word with a negative and a neutral meaning (e.g., shot, sentence, growth, beat), is used as a prime (Richard & French, 1992). Using delayed recognition memory, Eysenck et al. (1991) had participants listen to a set of ambiguous sentences (e.g., the doctor examined little Emma's growth) and subsequently undergo a recognition test, which included disambiguated versions of the sentence (e.g., either referring to cancer or height). Results indicated that anxious individuals claimed to recognize the threatening version of the sentence to a greater degree, while non-anxious individuals displayed a bias more toward the positive meaning of the word (Eysenck et al., 1991; Hirsch & Mathews, 2000; Mathews & MacLeod, C., 1994). The results show that anxious individuals lean more toward negative interpretations rather than neutral or positive meanings of words.

Wilson, MacLeod C., Mathews, and Rutherford (2006) used training procedures to systematically induce interpretive bias favouring either threatening or non-threatening meanings of ambiguous information. In the training procedure of the threatening condition, a threatening homograph prime was presented and then followed by a word completion task, where the prime was always a clue as to which meaning the word fragment would successfully complete; thus the participant would develop a bias favouring interpretation of ambiguous stimuli as threatening. Wilson et al. (2006) found participants who were trained with a threatening bias had faster response latencies in the assessment task when the target was associated with the threatening meaning. In addition, participants were also shown a series of anxiety inducing videos, and subsequent emotional reactions revealed that those participants trained with the threatening bias showed elevated response, in both anxiety and depression reactivity. Thus Wilson et al. were able to induce interpretive bias, and demonstrate how that bias skewed subsequent information interpretation and analysis.

Taken together, these studies show that anxious individuals not only attend to negative stimuli disproportionately, but also are more likely to perceive ambiguous stimuli in a threatening and negative manner. Thus the initial stages of the information-processing stream show a profound bias in terms of both the selection of stimuli as well as their interpretation. Most, if not all, of the information that is being processed is seen with a distorted lens, and thus by extension their recollection of events could also be skewed in the same manner.

If attention and interpretation can be negatively affected by disorder-related biases, it is perhaps not surprising that memory is also affected. A memory enhancement for threatening stimuli has been well documented in the literature (Morgan, 2010; MacLeod, C., & Mathews, 2004). Studies have shown that anxious individuals retrieved more anxious/worried memories when cued with neutral stimuli, and more fearful memories when cued with disorder-specific stimuli (Richards & Whittaker, 1990; Wenzel & Cochran, 2006). However, as autobiographical memories are implicated with individual differences, word lists with emotional and neutral words have also shown enhanced memory for negative valence words compared to positive words in anxious individuals, while the opposite is true for healthy controls (Denny & Hunt, 1992; Groeger, 1997).

To understand the memory enhancement observed within anxious individuals, we need to understand the distinction between implicit and explicit memory. Generally speaking, implicit memory is stored information that affects subsequent responses or actions without awareness, whereas explicit memory is the conscious recollection of events or facts (Galotti, 2004; Goldstein, 2008). The distinction between implicit and explicit memories has allowed researchers to investigate memory without relying on intentional retrieval alone. Mathews, Mogg, May and Eysenck (1989) found no significant difference in memory using a cued recall task, an explicit memory test, within anxious individuals. However, in a word completion task, a test of implicit memory, anxious individuals produced more words from the threatening lists than normal controls. Thus the two memory systems can work independently of each other, and reveal that anxious individuals unconsciously perceive and remember threatening stimuli disproportionately (Mathews, Mogg, et al., 1989; Ridout, Astell, Reid, Glen, & O'Carroll, 2003).

From the review of the literature, it can be conclusively stated that there is a bias in the information-processing stream in individuals suffering from anxiety disorders. A recent meta-analysis found that attentional bias appears to occur in all anxiety disorders in equal magnitude, suggesting that elevated anxiety is an integral component (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Waechter, Nelson, Wright, Hyatt, & Oakman, 2014). Thus the empirical evidence is in line with the theoretical cognitive accounts of anxiety disorders that emphasize specific forms of bias in the information-processing stream. Because of their disorder, anxious individuals not only selectively attend to and recall threatening stimuli disproportionately, ambiguous stimuli are also perceived as threatening to a higher degree. All of these biases occur unconsciously, without intent, leading to later cognitive symptoms such as rumination and worry. In reference to the central story of my thesis, a measure that is sensitive in its ability to capture attentional biases would be beneficial, as biases can occur at early stages of the information-processing stream.

3 Stroop

Up to this point of the Introduction, I have discussed attention and how it can be unintentionally captured within healthy individuals as well as patient populations; and how, within anxious individuals, these biases cascade to additional levels of information-processing biasing them also toward disorder-related stimuli and schemas (Clark & Beck, 2010). I have discussed several experimental designs including the Stroop paradigm which I discussed only briefly in anticipation of this section of my thesis. Given its critical relevance to the research I will subsequently present, in the following sections, I will discuss the original experiments conducted by J. R. Stroop (1935) in which he revolutionized attention paradigms by combining two dimensions simultaneously. I will then highlight how the Stroop effect has remained quite robust in the face of the many procedural and stimulus modifications. I will then go on to discuss the underlying cognitive processes that are at play that connect the Stroop effect to selective attention. Given that I have already ascertained the seriousness of the attentional biases, I will highlight why the Stroop paradigm is the ideal candidate to capture attentional bias.

3.1 The Classic Stroop Task

The origins of the Stroop task can be traced back to Cattell (1886) who, under the supervision of Wundt, discovered that reading the word blue is faster than saying blue to a blue colour patch. In line with the modern views, Cattell proposed that words and letters harbour associations between ideas, which have taken place “so often that the process has become automatic, whereas in the case of colors and pictures we must by a voluntary effort choose the name” (1886, p. 65). Cattell’s discovery propelled many experiments, some of which further supported the automatic/voluntary view, while others offered alternative explanations (Brown, 1915; Hollingworth, 1915). However, none were as influential and famous as Stroop’s (1935) classic experiments.

In a series of experiments, Stroop (1935) combined words and colours into a single stimulus thereby capturing both dimensions simultaneously. In Experiment 1, Stroop (1935) presented

subjects with colour words, in either a different ink colour (incongruent trials) or in black ink colour (control trial). Stroop determined that incompatible ink colour had no significant effect on word reading speed. In Experiment 2, subjects were presented with incongruent trials as in the first experiment, however Stroop used solid colour squares as his control trials. His dependent measure was the total time it took participants to respond to all 100 items on a stimulus card. He discovered that subjects took significantly longer (average 47 s) to name the colour of the ink of incongruent words than solid colour squares. For the third experiment, he looked at practice effects of naming incongruent ink colours over an eight-day period. He found that interference for incongruent trials did decrease with practice, however an ink-colour naming baseline was not taken each day. Thus, there may be a general practice effect or learning to learn effect (MacLeod, C. M., 1991; Stroop, 1935). Experiment 3 also reported the first account of the reverse Stroop effect, as the practice naming ink colours over the course of eight days created an interference with word reading when pre and post tests were compared. That is, when subjects had significant practice naming colours, those colours began to interfere with word naming.

By combining the two dimensions into a single stimulus, Stroop (1935) was able to examine the interference between two conflicting processes. He suggested that his experiments provided a unique basis “for comparing the effectiveness of the two type of associations” (p. 659). He believed that “these associations are products of training, and since the difference in their strength corresponds roughly to the difference in training in reading words and naming colors, it seems reasonable to conclude that the difference in speed of reading names of colour and in naming colour may be satisfactorily accounted for by the difference in training in the two activities” (p. 660). Stroop created a unique test that was relatively easy to administer, and appeared to have good validity and reliability (MacLeod, C. M., 1991).

The now standard Stroop colour-word task is the procedure used by Stroop (1935) in his Experiment 2. The difference in response times, called the Stroop effect, is said to provide a measure of interference in information processing. Although the focus here is on interference, Stroop facilitation also occurs for congruent stimuli (word red written in red ink colour), although the strength is weaker than the corresponding interference (MacLeod, C. M., 1991).

Many different versions and procedures have been created using the standard Stroop colour-word task as the basis (see review, MacLeod, C. M., 1991). Since its inception, researchers have attempted to manipulate the Stroop effect, either by enhancing or diminishing it, to understand the underlying cognitive processes that are at play.

3.2 The Robust Stroop Effect

Perhaps one of the most fascinating aspects of the Stroop effect is its robustness. In 1991, C. M. MacLeod conducted a literature review of the different versions, modifications, and procedures to date, and found modifications to the Stroop task modestly affected its magnitude, but not its qualitative form. In this section, I will highlight how the Stroop effect has withstood the test of time and procedural modifications, and has shown itself to be a robust phenomenon. As such, the Stroop effect would provide a solid foundation upon which variations could be attempted with the goal of assessing attentional biases.

Early research showed that changes to stimulus dimensions did not adversely affect the Stroop effect. For example, the original Stroop paradigm employed colours and words as visual stimuli, but switching those dimensions for pictures or auditory stimuli does not change the overall Stroop effect. Sorting and matching versions of the colour-word task, as well as the picture-word interference task both observed normal Stroop interference (Chmiel, 1984; MacLeod, C. M., 1991; Tecce & Happ, 1964). The auditory analogy of the Stroop task was also investigated, where subjects were asked to say 'low' to the word low presented at a low pitch (110 Hz) and 'high' to the word high said in a high pitch (175 Hz); in the incompatible condition, there was marked interference present (Hamers & Lambert, 1972; Lutfi-Proctor, Elliot, & Cowan, 2014). The Stroop effect was also intact when colour hue and saturation were varied (MacLeod, C. M., 1991). Thus changing the dimensions of the stimuli does not significantly alter the principle Stroop effect.

As one of the fundamental components of the Stroop paradigm is the presentation of one stimulus that incorporates two different dimensions, researchers have investigated the Stroop effect using integrated and non-integrated stimuli. The published research demonstrates that even when irrelevant words and colours are spatially or sequentially separated (i.e., non-integrated), the Stroop effect remains intact, albeit smaller (Dyer & Severance, 1973; MacLeod, C. M. & Bors, 2002). As the time lapse between presentations of the two stimulus dimensions is increased, or when the difference between the spatial locations is increased, the Stroop effect tends to decrease (MacLeod, C. M. & Bors, 2002; MacLeod, C. M. & Hodder, 1998a). Given that the hallmark of the Stroop task is its ability to tap into two different dimensions simultaneously (word and colour dimension), the integration of stimuli generally gives rise to the most robust effects.

In addition, the use of different response modalities does not alter the magnitude of the Stroop effect drastically. The most robust effects with the Stroop paradigm are seen using vocal response output, although the use of manual response output does elicit the Stroop effect, albeit smaller. In his original works, Nielsen (1975) empirically compared both response modalities within the same experiment. He observed interference in both vocal and manual responses, although the interference was more pronounced with vocal responses (MacLeod, C. M., 1991; Nielsen, 1975). In addition, the Stroop effect has been shown to occur across cultures and languages. Lee and Chan (2000) recruited monolinguals and bilinguals in both English and Chinese, and were able to produce a Stroop effect even when the languages were intermixed. Thus there is evidence that the Stroop effect is observed regardless of the response modality or language utilized.

Generally, the Stroop effect seems to be timeless and ageless, where gender does not have an effect, and where age differences are related more to education level than age (MacLeod, C. M., 1991). Interference has been shown to begin early in school years as reading skill level develops, remaining stable throughout adulthood and waning in later years around 60 years of age (Esposito, Baker-Ward, & Meuller, 2013; Goldstein, 2008). Individual differences may be present, as some might be better at task switching than others, though usually not to a significant

degree. As we will see, the use of the Stroop task in a clinical setting has suggested that the degree of Stroop interference may be able to be used to differentiate between certain populations (see Emotional Stroop section).

From the research presented above, I agree with C. M. MacLeod's (1991) conclusion that the qualitative form of the Stroop task remains quite stable. Gender and age does not change the basic form of the Stroop effect and neither does modality of response nor language. Although the dual dimension of stimulus is a fundamental component of the Stroop stimulus, simultaneous presentation of the non-integrated stimuli does elicit the Stroop effect. Research thus far has shown the Stroop task to be quite resilient to change. In fact, it is so robust that it is often used for live demonstration purposes in noisy settings such as classrooms as the effect is almost always borne out. Given that the goal of this thesis is to create a paradigm that taps into attentional biases, what better paradigm to modify that one than has been shown to be robust, reliable, and resilient to change?

3.3 The Relation of the Stroop Task to Attention

Evidently the Stroop effect reflects a problem with the selective attention process; otherwise participants could completely ignore the incongruent dimension when required. The robustness of the Stroop effects speaks volumes regarding the internal conflict, however it raises the question of why and how this conflict occurs: Is it that attention is always captured by irrelevant dimensions? What is the relation between the dimensions and the primary task? In this section, I highlight negative-priming research which suggests that participants attempt to suppress one dimension over the other when dealing with the internal conflict. I will then discuss how it is 'automaticity' that is at the heart of this internal conflict; and how word reading is a highly practiced process that activates meaning, creating the response conflict between colour naming and word reading. Thus, the Stroop task taps into a highly automatic process resulting in an attentional bias toward word reading, paralleling the attentional biases seen within the clinical populations. Given the similarities, and importance to the current thesis, the Stroop paradigm would be an ideal candidate to objectively measure attentional biases.

Even when participants are explicitly instructed not to read the word, but to identify the colour of the ink, participants seem unable to do so. This gives rise to Stroop interference, demonstrating an evident struggle in the selective-attention process. This attentional bias is perfectly captured in the trial-by-trial effects, which is when the response on one trial affects the response on the subsequent trial (Juvina & Taatgen, 2009; Lowe & Mitterer, 1982). A critical example of this is when the to-be-ignored word on one trial turns out to be the to-be-named colour on the next trial (e.g., red written in blue ink preceding green written red ink). In such cases the interference is larger, theoretically because the response 'red' had just previously been suppressed in an attempt to control its interfering effects. This basic finding is known as negative priming (Juvina & Taatgen, 2009; Lowe, 1985). Although the theoretical models for negative priming are still under debate, the fact that it occurs is consistent with the notion that participants are actively trying to suppress the word and name the colour of the ink. This gives rise to the question of why the words are even being processed when it is detrimental to the primary task of colour naming.

To understand why the words are being processed and subsequently, the underlying cognitive processes that give rise to the Stroop effect, we need to consider the importance of automaticity. Essentially, automaticity assumes that there are associations between specific stimuli and specific responses, developed through prior learning and frequent experience, which are set in motion by a series of unintentional cognitive and behavioural processes that are effortless and difficult to interrupt or stop (Beck, 2008; Cattell, 1886; Price, Beech, Mitchell, & Humphreys, 2012). Greater automaticity implies that the task requires fewer attentional resources, thus the irrelevant word is processed much faster as it draws on fewer attentional resources. Accordingly, more automatic processes interfere with less automatic processes, but not vice versa.

Rather than an all-or-none phenomenon, automaticity is often described as a gradient that varies as a function of practice (MacLeod, C. M., & Dunbar, 1988). For example, a toddler who does not know how to read will state the colour of ink to a word written in red ink; the automatic reading response has not been established in him, but as he ages and learns to read, that response

will eventually change to saying the actual word rather than the colour. C. M. MacLeod and Dunbar (1988) conducted a pivotal study investigating the effects of practice using the Stroop task. By varying the length of practice (2, 5 or 20 hours), they trained participants in naming four novel shapes by using names of the same four colours. Results showed that increased length of practice led to a Stroop interference with shapes and colour naming, which persisted three months after the experiment ended without further training (MacLeod, C. M., & Dunbar, 1988). This same type of evidence is used when describing the reverse Stroop effect, where with practice colour naming interferes with word reading (MacLeod, C. M., 1991, 1996; Martin, 1981; Stroop, 1935). Thus practice is the critical component that leads to automaticity, which is why the irrelevant words are processed so quickly in the standard colour-word Stroop task.

When the irrelevant word is processed at such a fast pace, this activates the meaning of the word, which can conflict with the colour naming. Since it is from our knowledge that we know the letters /r/e/d/ together make up the word “red” and represent the colour red, it has been shown that the meaning of the word is significantly related to the interference it causes. The classic experiment, Klein (1964) found that the more meaningful the irrelevant word, the more interference it caused. In addition to the congruent and incongruent trials, Klein (1964) introduced a third condition where the irrelevant words were colour-related (e.g., blood, grass, sky). The reaction times for the colour-related condition were faster than the incongruent condition, but not as fast as the congruent condition. This meant that although the colour-related words did not interfere as much as the actual colour words (e.g., red, green, blue), they did cause some interference as they were associated with colour in their meaning (e.g., the word blood is semantically associated with the colour red). Thus there is evidence that any stimulus that activates the association of the colour during the colour-naming task causes interference due to the dual dimension of the stimulus (Burt, 2002; Warren, 1972).

Essentially, the Stroop effect arises because word reading is so practiced that it has basically become automatic: the eyes receive the visual stimuli that are automatically processed (i.e., visual stimuli into letters, then translated into sounds and subsequently semantic meaning), and if the resulting meaning conflicts with the colour naming, this issue must be resolved before

providing the correct response. Thus there is a critical role of response competition that arises. This response conflict is accentuated when looking at the differences when using vocal or manual responses. Generally, the interference is smaller for manual response as the link between specific button and meaning of the word is smaller.

There have been some challenges to the notion of automaticity of word reading. Besner, Stolz, and Boutiller (1997) conducted a series of experiments in which they postulated that rather than an automatic phenomenon, the Stroop effect was due to a mental set serving to bias processing at different levels. Besner et al. (1997, 1999) effectively reduced and even eliminated the Stroop interference by colouring a single letter rather than the entire word. Besner and colleagues argue against the automatic view that words are inevitably processed to the lexical/semantic level; although it is the default mechanism, they suggested that it can be overridden by context and task goal (Bibi, Tzelgov, & Henik, 2000; Manwell, Roberts, & Besner, 2004).

However, it seems like the task constraints in Besner et al. (1997, 1999) were quite contrived and might not represent the natural occurring phenomenon of word reading. Some of the task constraints were single letter colouring, pre-cuing the location of the coloured letter, as well as the elimination of congruent trials. In an attempt to replicate Besner and colleagues (1997, 1999), Augustinova and Ferrand (2007) actually found evidence of both the standard and semantic Stroop effects (Augustinova, Flaudias, & Ferrand, 2010). In addition, the stimulus set of Besner et al. (1997) has been called into question as their neutral baseline consisted of words that began with the same letter as the critical words, which has been shown to complicate results. That is, given the neutral words could cause Stroop interference themselves, using them as baseline would underestimate the Stroop effect (Brown, Joneleit, Robinson, & Brown, 2002).

To investigate the many possible variations, Brown et al. (2002) conducted a series of seven experiments with manipulations to response modality, presence of congruent trials, single letter cueing, the role of spatial attention and orthographic processing. The results showed that these

formats reduced but did not eliminate the Stroop effect. In one of the experiments, Brown et al. created a scenario where there was no incentive to word reading at all (the word was spatially separated from the critical colour patch, always incongruent and never validly pre-cued), yet the presence of the word led to interference in naming the color patch. Thus, the presentation of a word is enough for it to be processed to its semantic level, and cause interference when there is a response conflict.

Taken together, the literature shows that the robustness of the Stroop effect lies in its ability to tap into the automatic word-reading process. Regardless of conscious effort, participants cannot help but process the word to its semantic level, which causes conflict when the correct response is not the same. This attentional bias evident within the Stroop effect is quite similar to the attentional biases seen within the healthy individuals as well as anxious individuals. Given the resemblance, it would be quite valuable if we could objectively measure the attentional biases using the Stroop task.

4 The Emotional Stroop Task

As the Stroop task is relatively stable in its capacity to produce an effect, the emotional Stroop task was created in hopes of mirroring that stability. As the standard colour-word Stroop effect is present in both healthy individuals and patient populations, the use of emotional words in the emotional Stroop task is said to gauge the sensitivity to the emotionality of the word. There are findings to support the use of emotional Stroop task in clinical settings, although several methodological limitations have recently come to light (Algom, Chajut, & Lev, 2004; Price et al., 2012; Williams et al., 1996). In the following section, I will clearly define the emotional Stroop task and its difference from the classic Stroop task, while highlighting some early research and results using the emotional Stroop task. I will specifically discuss the many limitations of the current emotional Stroop task, such as the equivocal lexical characteristics of the words, as well as the differences in valence and arousal of those words. In reference to the thesis, although theoretically the emotional Stroop task is a great tool, it has several methodological downfalls and could benefit from modifications.

The core difference between the Stroop task and the emotional Stroop task is the use of emotionally charged and/or disorder-related words instead of colour words. The inclusion of such words is based on the premise that researchers want to determine whether, and to what degree, specific words may cause interference in the information-processing stream beyond the interference caused by colour-word stimuli. Accordingly, there are no congruent and incongruent trials, but rather target trials (e.g., death written in blue ink) and control trials (e.g., desk written in blue ink). Parallel to the Stroop task, participants are asked to name the colour of the ink of the stimuli, which now is provided via either an emotional word or a neutral word. For the successful completion of the task, participants must filter the semantic meaning of the word as it interferes with colour naming. The response latency between the target and control conditions is calculated to determine the emotional Stroop interference, or rather the bias score, which is postulated to be the result of the interference caused by the emotionality of the target word.

The critical change in the carrier word from the Stroop task to the emotional Stroop task has large consequences as to the interpretation of the cause of the interference. The defining feature of the standard Stroop stimulus is that two dimensions are logically converging on the same semantic system, which is not the case in the emotional Stroop stimulus (e.g., the meaning of the word death is distinctly different than the ink colour blue; the two are not related semantically). In the Stroop task, words in the congruent and incongruent trials are the same and thus interference can be calculated directly. However in the emotional Stroop task, words in the control and target condition are never the same, thereby making it a quasi-experimental design. As such, the emotional Stroop task assumes that the only difference between the target and control words is the difference in emotionality and nothing else. This assumption needs to be carefully validated to claim that there is an emotional Stroop effect present.

It is important to highlight that the emotional Stroop effect, when observed, cannot be attributed to semantic or phonological response competition, as is the case with the standard Stroop effect. That is, neither the meaning nor the sound of the word 'death' conflicts with the meaning or

sound of the colour blue. Therefore any slowing we see cannot be due to a need to resolve competition among conflicting viable responses. Instead, the theoretical interpretation of the emotional Stroop effect relates primarily to biases in attentional capture. Specifically, the assumption is that if a person is cognitively preoccupied with some concept (e.g., death) then stimuli connected to that concept will gain the power to pull attention toward them, just as highlighted in detail earlier in the Introduction. As such, when the word 'death' appears in blue ink colour, participants are drawn to the word at very early stages of processing and are thus essentially pulled away from the primary task of colour naming. Thus it is attentional competition, the difficulty of 'staying on task' that is assumed to produce slowing in the emotional Stroop context. Consequently, if an attentional bias is present, performance on the task suffers as a result of selective attention to emotionally relevant stimuli (Price et al., 2012). The absence of an emotional Stroop effect means that participants are able to ignore the emotional and neutral words to the same extent.

Early researchers compared healthy controls to individuals with anxiety disorders to determine the extent to which specific words caused any attentional bias (Williams et al., 1996). For example, Mathews and MacLeod, C. (1985) found anxious participants to be slower to name threat-related words in general, and more specifically those with social phobia were slower when socially relevant threat words were presented. In another study, rape victims with post-traumatic stress disorder (PTSD) displayed interference with rape-related threat words compared to general threat words (physical harm and death), neutral and non-words, whereas no such interference was found in non-PTSD rape victims and healthy controls (Foa, Feske, Murdock, Kozak, & McCarthy, 1991). More recently, Ball, Mitchell, Touyz, Griffiths, and Beumont (2004) conducted an emotional Stroop task pre and post treatment in patients with anorexia nervosa to determine whether the emotional Stroop task could be used as a treatment outcome measure. Previous to treatment, patients with anorexia nervosa displayed increased response latencies to body image and food related words, which subsequently decreased post treatment. Thus there is a degree of sensitivity in information processing where emotional stimuli have the capacity to produce interference (Dugas et al., 2003; Hirsch, MacLeod, C., Mathews, Sandher, Siyani, & Hayes, 2011; Schultz & Heimberg, 2008; Williams et al., 1996).

Although there have been apparent successes using the emotional Stroop task there have also been worries about various factors that could be influencing the results, including those associated with the lexical status of experimental stimuli. For emotionality to be the main difference between the control and target words in the emotional Stroop task, lexical characteristics of the words need to be carefully matched. Lexical characteristics are word properties that affect the rate of word recognition (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Larsen, Mercer, & Balota, 2006). The faster you are able to recognize a word, the faster you are able to respond to it. From lexical decision tasks, it has been established that the lexical characteristics of the words play a large role in speed of word recognition (Burt, 2002; Larsen et al., 2006). Word frequency (the frequency with which the word is used) and word length (the number of letters in each word) are important lexical characteristics that influence word recognition speed (Estes & Adelman, 2008; Larsen et al., 2006; Price et al., 2012). Anything that could influence word recognition speed can theoretically impact the magnitude of interference within the emotional Stroop task. As such, the control and target condition words need to be carefully matched on their lexical characteristics.

Within the empirical studies using the emotional Stroop task, the importance of lexical equivalence between target and control words seems to be overlooked significantly, a point that has been noted (Burt, 2002; Balota et al., 2004; Algom et al., 2004). For example, in a modified emotional Stroop study, some of the negative words used were ‘terror, cancer, devil, shame’, whereas neutral words were ‘trend, stone, extra, panel’ and the positive words were ‘comedy, mature, cash, lucky’ (Kitayama & Ishii, 2002). Just by looking at those words, one can see the different lexical characteristics that are present and the many ways in which that could result in a difference in interference above and beyond emotionality. To investigate this issue, Larsen et al. (2006) compiled a list of all the stimuli used within the published works and analyzed them for lexical decision latency and naming speed using the norms of Kučera and Francis (1967) as well as Hyperspace Analogue to Language (HAL; Lund & Burgess, 1996). Of the 1033 unique words within the published words, Larsen et al. found that negative words were significantly less frequent, longer in length, and had smaller orthographic neighbourhoods compared to neutral

words. All of those lexical factors led to longer word-recognition latencies which could lead to larger emotional Stroop interference. Thus, matched lexical characteristics between words within an emotional Stroop paradigm is crucial for the true emotional Stroop effect to be observed.

A fundamental assumption of the emotional Stroop effect is that attention is drawn to emotional words because of their emotional impact. For anxious patients for example, words related to threat or fear are associated with the presence of perceived danger and, thus, the need to attend. Emotionality is a subjective concept and occurs on a gradient within two dichotomous dimensions of valence and arousal. Valence describes how positive (smile) or negative (frown) a stimulus is, and arousal defines how exciting (sex) or calming (sleep) a stimulus is (Anderson, 2005). Historically most early researchers used emotional words as synonymous to negative words, while positive words were either clumped together with the neutral words or not used at all (Anderson, 2005; Kensinger & Schacter, 2006). Thus there has been an empirical bias toward investigating and thus finding attentional effects for negative-valence stimuli in the literature (Frewen, Dozois, & Lanius, 2008). Such valence asymmetry comes from evolutionarily adaptive processes and the automatic vigilance hypothesis, which assumes that special attention is paid to the evaluation of stimuli associated with defense or avoidance (e.g., snake, angry face, gun); it also stems from the notion that negative stimuli are associated with higher arousal than positive stimuli (Anderson, 2005; Ohman, 2005; Schimmack, 2005). This is not completely unwarranted as neuroscientific evidence suggests that a core function of the amygdala is to draw attention toward threats. That is, the brain itself puts more weight on the relevance of negative stimuli (Anderson, 2005; Dijksterhuis & Aarts, 2003; Dolan, 2002). However, recent empirical data counter this notion by showing that positive stimuli can capture attention just as easily as negative stimuli depending on arousal level (Buodo, Sarlo, & Palombo, 2002; Kousta, Vinson, & Vigliocco, 2009; Pratto, 1994).

Currently there is theoretical debate as to whether valence or arousal is the decisive factor in the emotional attention and memory enhancement literature. It has been found that, regardless of polarity, both positive and negative valence words displayed a processing advantage over neutral words (Eviater & Zaidel, 1991; Kousta et al., 2009). Thus in that sense, emotional bias does exist

for both positive and negative stimuli. In their analysis of lexical characteristics, Larsen et al. (2006) also discovered that both negative and positive words were significantly longer and had few orthographic neighbours than neutral words, and negative words were less frequent. Holding these factors constant, Larsen et al. conducted a regression analysis and concluded that negative valence words lost their significance, at least for lexical decision accuracy. Using valence and arousal ratings from the ANEW database and lexical decision and word naming data available from the English Lexicon Project, Estes and Adelman (2008) ran a series of regression analyses that revealed a small yet significant delay for negative words in a lexical decision task when all the factors were accounted for (Balota et al., 2004; Bradley & Lang, 1999). Larsen and colleagues (2008) later corroborated these results, but proposed that arousal interacts with word negativity in a counterintuitive manner where lower arousing negative words produce larger lexical decision latencies than higher arousing negative words. However, Kousta et al. (2009) found confounding results showing a significant effect of valence, whereas Dresler, Meriau, Keekeren, and van der Meer (2009) found that arousal plays the determining role in emotional interference. Thus, theories regarding arousal and valence are still being debated, although both factors are seen as important contributors to attentional bias.

Although the early literature using the emotional Stroop paradigm has shown it to be an effective method to determine attentional bias, recent literature has highlighted that the effect sizes are not large and the effect itself has become difficult to replicate (MacLeod, C. M., & Hodder, 1998b). For example, Borella, Ribaupierre, Cornoldi, and Chickerio (2013) found too much variability in the interference results with children with attention-deficit hyperactivity disorder and concluded that the emotional Stroop task was not a reliable and consistent method. Investigating the test-retest reliability of the emotional Stroop task, Eide and colleagues (2002) found the interference indices taken at two different sessions were not comparable across sessions, and the reliability correlation was low ($r = 0.24$). Low reliability issues have become a common issue for the emotional Stroop task, and have been documented in several empirical studies and across different domains of healthy controls and clinical populations (Algom, et al., 2004; Dresler et al., 2012; Kindt, Bierman, & Brosschot, 1996; Strauss, Allen, Jorgensen, & Cramer, 2005). A recent meta-analysis questioned the automaticity of the test, and found reduced interference by mixed design relative to block design alluding to strategic manipulations by participants (Phaf & Kan,

2007). Thus the meta-analysis suggested that participants were able to execute some sort of strategic approach to the task, which counters the notion of early attention capture and automaticity that is in line with the theoretical model of the emotional Stroop task (Phaf & Kan, 2007).

All of these factors, such as interplay between arousal and valence, lexical characteristics and quasi-experimental design cause us to question the use of current versions of the emotional Stroop task. Theoretically the underlying idea is sound: Emotional words create an attention bias, pulling the participant's attention toward them and away from the primary task, thereby resulting in an increase in the reaction time to name the ink colour (Price et al., 2012; Williams et al., 1996). However, due to the several methodological limitations as well as reliability and replicability issues, the current emotional Stroop task is not accurately depicting the difference in attentional bias, and would greatly benefit from modification and recalibration.

5 Purpose of the Thesis: Introducing the Multiple-Item Stroop Task

The driving goal of this thesis is to propose a modified version of the emotional Stroop task, creating a paradigm that produces larger and more reliable measurements of attentional bias. The preceding literature provides a good overview of the existence of attentional biases as well as the potential of the Stroop task as a tool for quantifying attentional bias. The basic assumption of this thesis, supported by all of the described literature, is that certain populations include a tendency to repeatedly activate certain concepts, which results in biases in attention and subsequent cognition.

My primary theoretical consideration with respect to arriving at a modified paradigm was to hopefully magnify the observed effect by creating a version that required participants to repeatedly engage and disengage attention to complete the task (Posner, Rafal, Choate, & Vaughan, 1985). If it is the engagement or disengagement of attention is biased by these

conditions, then requiring this to happen repeatedly should magnify the effect size. The paradigm I ultimately arrived at is one I term the Multiple-Item Stroop task.

The Multiple-Item Stroop task involves the presentation of a critical word multiple times across the screen, where most of the words will be presented in a so-called 'dominant' ink colour, the rest will be in a 'subordinate' ink colour. The task for participants is to indicate, via button press, which ink colour is dominant, while ignoring the multiple-presented word. Given that the two colours are randomly applied to the stimuli, this requires participants to scan through the display to determine which colour is dominant, meaning the critical word must be repeatedly experienced across the different locations on the screen. Thus, each fixation the participant makes on the screen has now become a selective attention task, where the participant has to select the correct dimension to process; if the incorrect dimension (i.e., the word) is selected, it would take the participant longer to disengage their attention from the semantic meaning of the word and focus on the colour naming task (MacLeod, C. M., & Dunbar, 1988; Posner et al., 1985). And in order to determine the dominant ink colour, theoretically the participant has to make multiple fixations across the screen. As such, if the semantic meaning of the word is one which a person is biased towards (e.g., arachnophobia or test anxiety), theoretically it would take them longer to respond to the colour naming task (Cattell, 1886; Posner et al., 1985; Price et al., 2012).

To some extent this paradigm can be viewed as the opposite of that utilized by Besner and colleagues (1997, 1999). By colouring just a single letter, they allowed participants to focus their attention sharply enough to reduce the likelihood of perceiving the entire word, which resulted in their findings of a lack of a Stroop effect. My paradigm requires a multiple deployments of attention across the display which I hope will exaggerate the effect.

In addition to repeated engagement of attention, by having two different ink colours on the same word within the same trial, there is simultaneous integration of dimensions. From the previous

literature, it has been shown that the advantage of the original Stroop stimuli was the logical integration of two dimensions into one stimulus (Dyers & Severance, 1973; MacLeod, C. M., & Bors, 2002). In the new paradigm, a single word will be coloured by two different inks within a single trial thereby providing an index to compare responses. The meaning of the word will be the same and only one correct response will exist (only one dominant colour), but there will be two ink colours to judge from. This logical integration is what is said to be missing in the emotional Stroop task literature, and will be pertinent in demonstrating how the two ink colour responses are given by the participants. Incorrect responses will also provide information regarding the participants, in addition to the latency of responses as well.

The Multiple-Item Stroop task also has another nice feature: One can essentially titrate the difficulty of the task by altering how dominant the dominant colour is. Imagine a condition wherein 90% of the words are the dominant colour as compared to one in which 55% are the dominant colour. One might expect the 55% condition to be associated with slower responses and more errors. This feature is not exploited much in this thesis aside from pilot work to ascertain an apparent optimal level of dominance. However, it is a useful feature of the paradigm, as the task can be made more or less challenging depending on level of dominant colour (55% dominant colour would be much more difficult than 90%).

The subsequent two chapters of this thesis will elaborate on two experiments in each. In Chapter 2, Experiment 1 will utilize standard Stroop stimuli in a mixed design to demonstrate that the Stroop effect is present with these modifications. Experiment 2 will introduce negative and neutral words as stimuli and will be identical to Experiment 1 in every other manner. Chapter 3 will encompass the next two experiments which will employ a larger subset of stimuli including positive, negative and taboo words and their corresponding control words. Experiment 3 will be identical to the previous experiments with the addition of a larger stimulus set. Experiment 4 will investigate anxiety levels, measured by the State-Trait Anxiety Inventory, and physiological factors, measured via a smartwatch for heart rate, body temperature and galvanic skin response, to determine whether there is any correlation between anxiety level and physiological factors and

the taboo Stroop effect found. Subsequently, Chapter 4 will be a General Discussion with theoretical and practical implications of the findings and future prospects.

Chapter 2

Standard and Emotional Stroop Effects

As described at the end of Chapter 1, my goal in this thesis is to assess how a modified Stroop task could be used to quantitatively measure the extent to which a concept within a person's mind is biasing their experiences. Starting from how attention functions, to how it is captured unintentionally, the literature review within Chapter 1 showed the existence of attentional biases, especially within certain clinical populations. The emotional Stroop task has been used to measure this attentional bias, although a recent review has shown that the paradigm being used is unreliable and riddled with methodological limitations (Algom et al., 2004; Dresler et al., 2012; Kindt et al., 1996; Strauss et al., 2005).

The modifications that I propose are meant to behaviourally magnify any effect of attentional bias present, and as such would tap into the attentional biases. To reiterate, the Multiple-Item Stroop is comprised of a critical word being presented many times across the screen in two different ink colours varied in a predetermined proportion; participants have to determine the dominant ink colour via button press as fast as they can. For example, the stimulus 'RED' will be written throughout the screen, 60% of the instances of RED will be in one ink colour (the dominant colour), 40% in another (the subordinate colour).

The basic theoretical consideration for these modifications is as follows. As the objective of the task is to identify the dominant ink colour, participants are forced to scan the entire screen, and as such would inadvertently fixate on the critical word repeatedly. If the participants are drawn to the critical word disproportionately, or if they have difficulty disengaging attention when they are, the task would take that much longer as the participant would have to consciously disengage from the critical word to determine the dominant ink colour (Posner, 1985). Thus the repeated engagement in a goal-oriented task (i.e., to determine the dominant ink colour) requires repeated disengagement from the distractor (i.e., critical word), and this process should be especially sensitive to attentional biases (Gray & Wedderburn, 1960; Posner et al., 1985).

The experiments that follow examine the potential of the Multiple Item Stroop in two ways. First I compare the traditional single-item procedure with my new multiple-item procedure in the context of the traditional Stroop stimuli, that is, using colour words such as RED, GREEN and BLUE. The primary question I examine is whether I continue to see a standard Stroop effect even when my multiple-item variant is used. In Experiment 2, I contrast single and multiple-item variants in the context of the emotional Stroop task. Do I see the typical small emotional Stroop result when single-items are used, and does this effect magnify in the context of my multiple-item variant?

To anticipate, I do indeed see a standard Stroop effect in Experiment 1 in both the single-item trials and multiple-item trials. However, disappointingly, I did not observe an emotional Stroop effect in Experiment 2 in either the single or multiple-item variants. Reasons for these null findings are considered and are used to inspire additional changes to my paradigm, changes that are then empirically assessed in Chapter 3.

6 Experiment 1: Single-item vs. Multiple-item Stroop task

The aim of Experiment 1 was to make sure the Stroop effect was present given the modifications within the Multiple Item Stroop. Recall from Chapter 1 that in the case of the standard Stroop effect, most theoreticians believe that the observed effect is the result of response interference due to processing of the dual dimensions within the stimuli (Collins & Loftus, 1975). Given that I am proposing rather dramatic changes to the stimulus set within the Multiple Item Stroop task, I need to determine whether the Stroop effect is present, its strength relative to the single-item condition, and whether there are any other effects that may arise.

One theoretical accounts of the standard Stroop effect assume that interference arises due to conflict at the response stage of processing (Posner & Snyder, 1975). The two dimensions (word and colour) compete within the response stage and the one that is more automatic (usually the

word reading) arrives first and thus has the potential to interfere when it is not the dimension of relevance. Especially within the incongruent conditions (e.g., RED written in blue ink), the word dimension usually arrives at the response stage earlier, which results in one of two things occurring: either an incorrect response when the word is produced; or a delayed correct response (the Stroop interference) where a conflict arises that needs to be resolved resulting in extra time. Thus the underlying cognitive process that gives rise to the standard Stroop effect is the automatic word reading, which creates the conflict with the task of colour naming within the response output stage of processing (MacLeod, C. M. & Dunbar, 1988).

In light of the modifications I propose, there is always the worry that when one changes the stimulus display rather dramatically then the changes may cause some unintended effects that might prevent the desired effects from being observed. Given that I am proposing rather significant changes, it seemed prudent to ensure that the standard Stroop effect occurs even using my new procedures; this is the primary goal of Experiment 1.

Theoretically, when one considers the standard Stroop effect in light of my procedural modifications, one can imagine two possible outcomes, depending on whether the degree of response interference is affected by the multiple instances of the stimulus (i.e., the irrelevant dimension). It could be the case that a single distracting word already gives rise to the maximal amount of interference. If this is the case, then I should see equivalent degrees of interference across the single versus multiple-item variants of the task. Observing this outcome would let me know that my changes did not wipe out the effect, and it would also provide some new information about the standard Stroop effect in the sense of showing that a single-item is all one needs to produce the maximal response interference (MacLeod, C. M. & Bors, 2002; MacLeod, C. M. & Hodder, 1998a).

However, it could also be the case that scanning the multiple-item display increases the salience of the distracting dimension, thereby strengthening its ability to interfere. If this is true I would

expect to see a larger standard Stroop effect in the multiple-item condition, relative to the single-item variant. This outcome would not only show that the multiple-item variant seems to 'preserve' the effect, it would also suggest that even the response interference aspect of the Stroop context can be further primed via multiple stimuli. There is some research that provides evidence that demonstrates that a prime could enhance the interference effect observed. For example, Warren (1972) placed a semantically related prime prior to the critical colour naming task and found that the higher the association between prime and irrelevant word, the larger the interference (e.g., the word blood as the prime and the irrelevant word as RED written in blue ink). Similarly, Burt (2002) found greater interference when the meaning was activated by a related word or phrase prior to the colour naming trial. Thus priming has been shown to increase interference, which could be the result when employing the multiple-word condition.

To determine how the Stroop effect would react to the modifications made, Experiment 1 was conducted as a way to measure the standard Stroop effect with the modifications in Multiple Item Stroop. As such, Experiment 1 employed standard Stroop stimuli (red, green, blue) and utilized a mixed design, where standard single-item Stroop trials and multiple-item trials were randomly mixed together to do a direct comparison.

In order to determine the optimal level of difficulty, I ran pilot sessions of the experiment within my lab. As the goal of the multiple-item trial was to force the participants to scan the entire screen, I needed the ink colour to be divided in such a way that would not be too easy (such as an 80/20 split) or too difficult (such as a 55/45 split). After running several sessions, I determined that the optimal level for the ink colour proportions was 60/40 as it allowed one to scan the entire screen to determine the correct answer but was not too difficult as it did not take too long to determine the correct answer.

6.1 Method

6.1.1 Participants

A total of 29 undergraduate students from the University of Toronto Scarborough participated in this study on a volunteer basis for course credit. The only exclusion criterion for the experiment was colour blindness as the primary task of the experiment was to identify different colours. All participants gave informed consent. In accordance with ethical guidelines, participants were explicitly told about the workings of the experiment and the aim of the study. Lastly, the participants were informed that they could leave at any time during the experiment without any repercussions if they felt uncomfortable. This experiment was approved by the Research Ethics Board of University of Toronto Scarborough.

6.1.2 Materials and Apparatus

Testing was conducted using Dell microcomputers running Matlab R2015 software (Brainard, 1997). The stimuli were displayed on 17 inch LCD monitors operating at a 60Hz refresh rate. Subjects observed the stimuli at approximately 5 degrees of visual angle. Responses were collected using a standard Dell keyboard.

The single-item trials represented the standard Stroop task, whereby one stimulus (i.e., RED, GREEN, BLUE, XXX, YYYY, or ZZZZZ) was presented in the centre of the screen in some specific ink colour (red, green, or blue). There were three conditions within single-item trials: On congruent trials the word and the ink colour were the same (i.e., RED written in red ink); on the incongruent trials the word and the ink colour were different (i.e., RED written in blue ink); and on control trials, non-words were written in coloured ink (i.e., XXX written in red ink). The font size used was 50 in the Arial font type. The participants' task was to press the button corresponding to the colour of the item as quickly and accurately as possible.

The unique aspect of this experiment was the inclusion of multiple-stimulus trials in which a given stimulus is repeated multiple times across the screen (10 by 10) and there is a random distribution of two ink colours in 60/40 proportion. The participants' task remains to identify the ink colour, but now they have to determine which ink colour is dominant and identify that ink colour by way of button press. Four different trial types occurred across the multiple-item trials: On congruent trials, the dominant colour matched the word presented (i.e., the word GREEN presented in green ink 60% of the time, and in one of the other colours the other 40% of the screen; see Figure 1). For incongruent primed trials, the word and the subordinate colour ink colour are the same (e.g., the word GREEN presented 60% of the time in blue ink and 40% of the time in green ink; see Figure 2). Incongruent neutral trials, the word differed from both the dominant and subordinate ink colour (e.g., the word BLUE presented 60% of the time in green and 40% in red ink; see Figure 3). Finally, on control trials, non-words written in two ink colours (e.g., XXX presented in, say, green 60% of the time and, say, blue 40% of the time; see Figure 4). The font size used was 30 in the Arial font type.

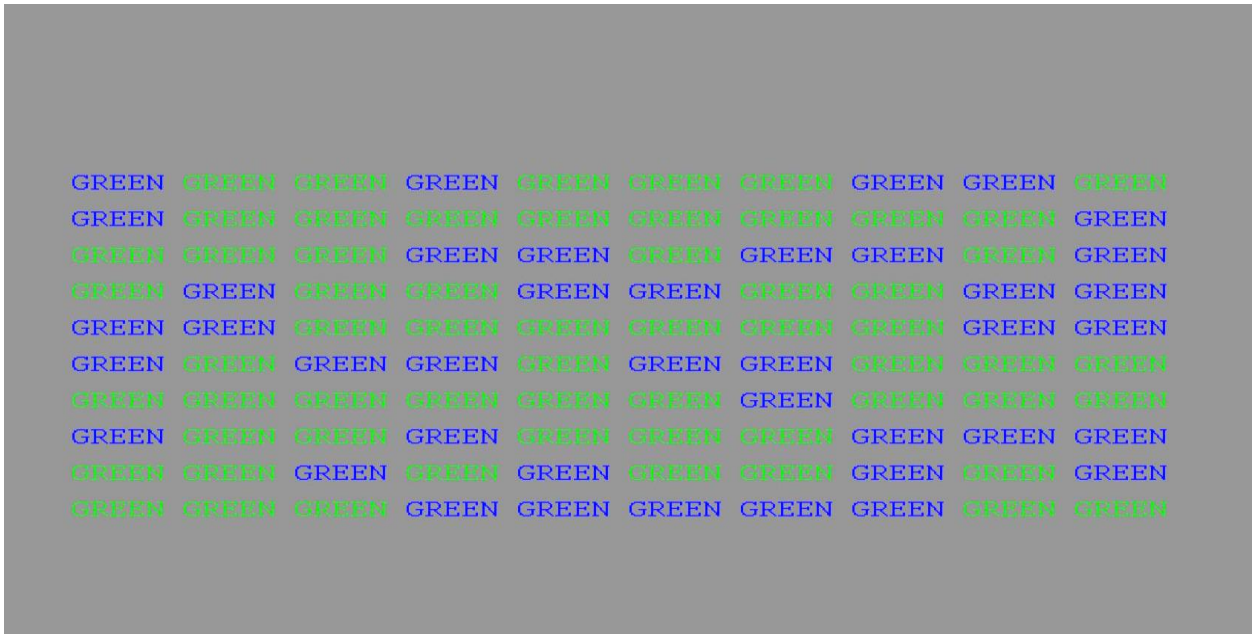


Figure 1. Multiple-Item Congruent Trial: The word *GREEN* written in green ink colour on 60% of the screen, and in blue ink colour on 40% of the screen. Correct response would be 'green'. This is a congruent condition as the word and the dominant colour match.

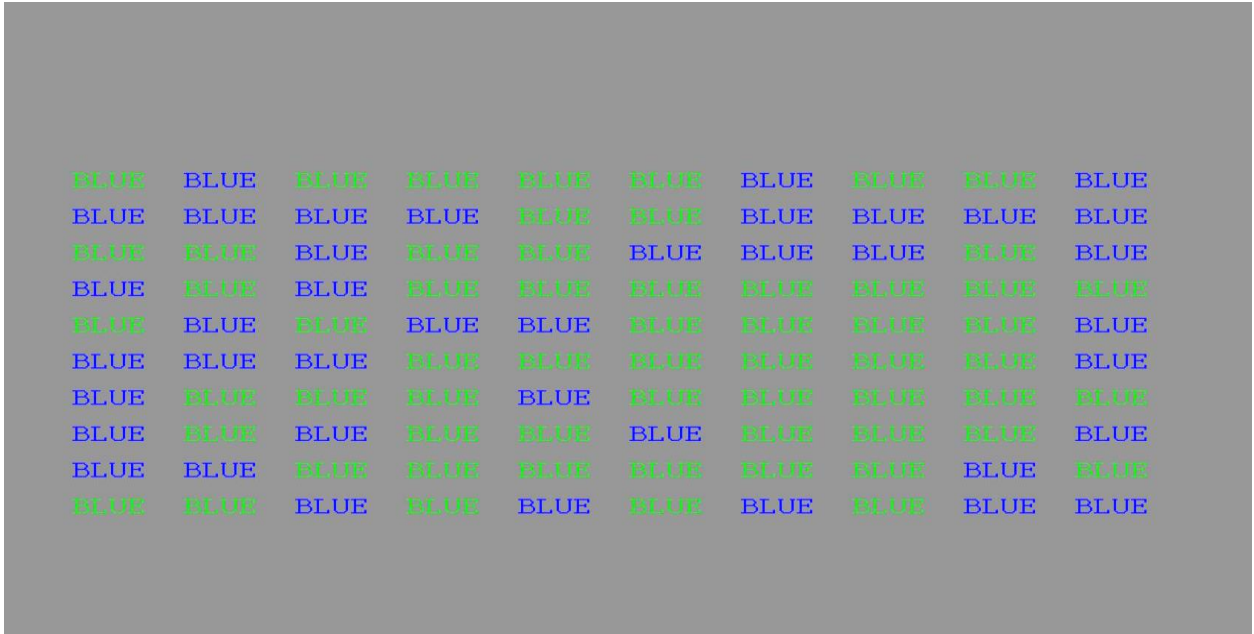


Figure 2. Multiple-Item Incongruent Primed Trial: The word *BLUE* written in green ink colour on 60% of the screen and in blue ink colour on 40% of the screen. The correct response would be green, as the dominant ink colour is green. This is an incongruent condition as the word and the dominant ink colour do not match, and the participant is ‘primed’ or more inclined toward the incorrect response of ‘blue’ as the subordinate colour matches the word presented.

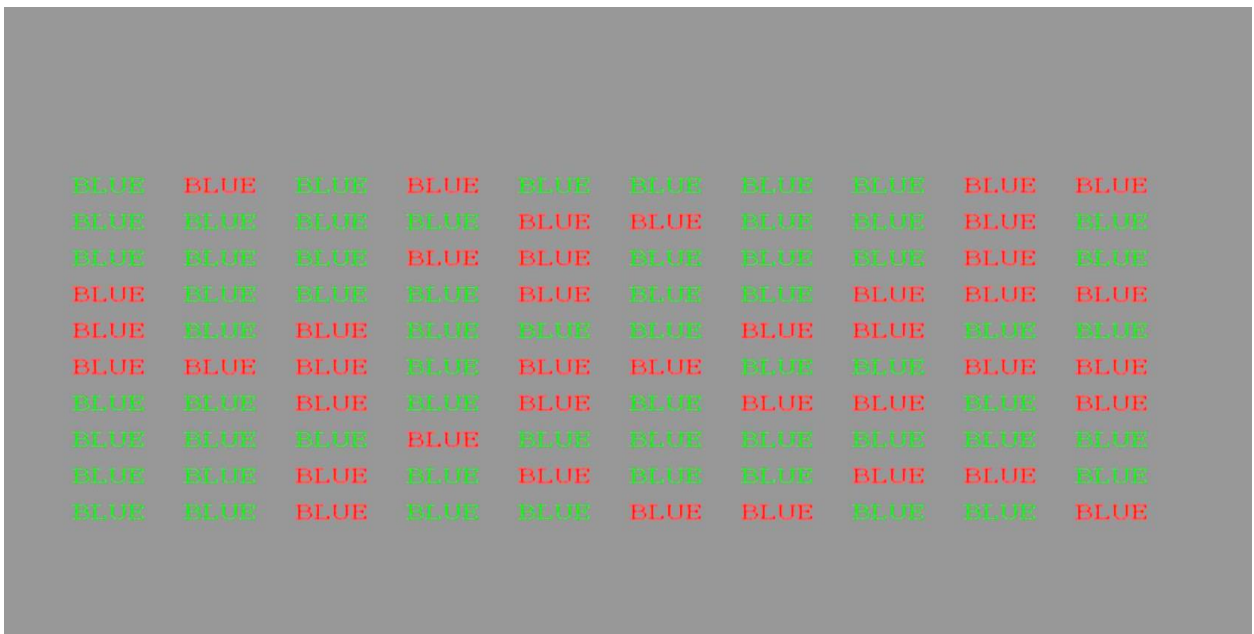


Figure 3. Multiple-Item Incongruent Neutral Trial: The word *BLUE* written in green ink colour on 60% of the screen and in red ink colour on 40% of the screen. The correct response would be green, as the dominant ink colour is green. This is an incongruent condition as the word and the dominant colour do not match; additionally, it can be said to be neutral in nature as the subordinate colour and the word are different and thus do not sway the participant in responding either correctly or incorrectly.

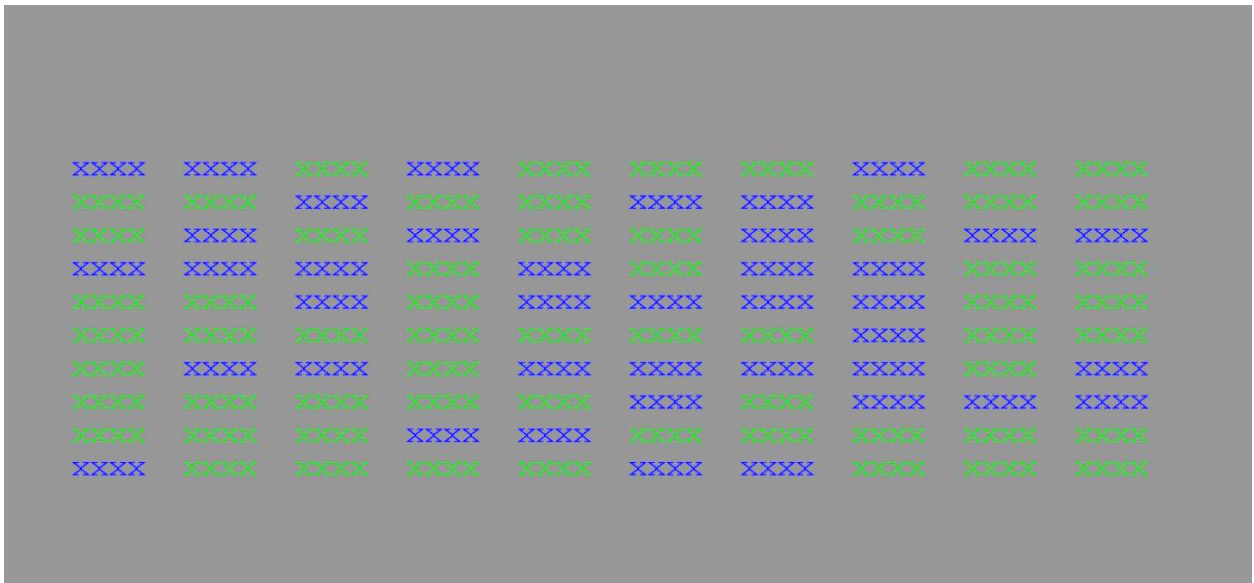


Figure 4. Multiple-Item Control Trial: The XXXX written in green ink colour on 60% of the screen and in blue ink colour on 40% of the screen. The correct response would be green, as the dominant ink colour is green.

6.1.3 Procedure

This experiment took place in a controlled laboratory setting, and consisted of one session of computerized Stroop task, with manual button press responses. The Stroop task was programmed in Matlab v2015a, using the PsychToolBox (PsychToolBox-3; Brainard, 1997). Participants were recruited from the University of Toronto Scarborough student pool, specifically Introduction to Psychology students who participated both for the learning experience and to gain course credit. When the participants arrived in the lab, they were verbally informed

regarding the experiment and their rights as participants, and were given the written Informed Consent form (see Appendix A) to carefully read and sign, if they agreed. The participants were then taken to the computer testing room, where the experiment would take place.

The experiment consisted of a total of 300 trials presented randomly across participants in a mixed fashion, with 150 single-item and 150 multiple-item trials. Of those 150 single-item trials, there were 72 control trials (xxx, yyyy, zzzzz) each appearing 24 times (8 times in each ink colour), while there were 39 congruent and incongruent trials each. Within each of those 39, each of the stimulus items (red, green, blue) appeared 13 times. Of the 150 multiple-item trials, there were a total of 69 controls trials, while there were 27 congruent, incongruent primed and incongruent neutral trials each. Within each of those 27, each of the stimulus items (red, green, blue) appeared 9 times each.

Participants were asked to identify the ink colour as fast as possible without making mistakes. However, there were no penalties if mistakes did occur. The data collected were comprised of gender, age, and the reaction times and accuracy for each trial.

At the start of the computerized session of the experiment, there was an instruction screen followed by a practice block consisting of 24 trials with an equal number of single-item and multiple-item trials. The practice block was the same for all the participants and was given for them to get accustomed to the design of the experiment as well as to remember which button is associated with each colour. The buttons were colour coded, and there was also a colour bar on top of the screen for ease of remembering. Following the practice block, participants had a chance to ask any further questions if required. In addition, the accuracy for the practice trials was also shown, providing the experimenter with the knowledge of the participant's comprehension of the instructions and whether they were performing the task successfully. Subsequently, the instructions for the task were reiterated to the participant before commencement of the experimental block.

In the experimental block, the single-items were always presented at the center of the screen, while the multiple-items were displayed across the entire screen. As such, each trial did not begin with a fixation cross, and participants were allowed to freely move their eyes. There was no delay or blank screen in between trials. The stimulus were kept on the screen until a response was given. In addition, during the experiment there was a brief resting period after every 30 trials, which gave the participants time to rest if they required. The resting screen stated that this was a time for resting, and also reminded participants of the goal of the task as well as the button associated with each colour. It provided a maximum of five seconds of resting, where the participants were given the option to press the spacebar if they wanted to continue the experiment before the end of the five second break. The experiment finished with a screen thanking the participants for their time and a concise debrief of the experiment, as well as instructions to call the experimenter for a more thorough debrief and time for answering further questions.

Stroop interference was measured by subtracting the congruent from incongruent trials reaction times. The Stroop effect is said to be present when the reaction times for the incongruent trials are longer than the reaction times for congruent trials. I opted to use the difference between congruent and incongruent conditions as both have meaning embedded within the stimuli. The control conditions I used did not have any meaning within them and were used to determine baseline colour naming ability.

Once again, the primary purpose of this experiment was to document the existence of the standard Stroop effect using the new Multiple Item Stroop paradigm. Given repeated demonstrations of the robustness of this effect, I certainly expected it to be observed on both single and multiple-item trials, a finding which would allow me to extend the new paradigm using the stimuli and procedures of the emotional Stroop task. It was unclear whether the traditional effect might be magnified by the use of multiple items, and thus this became a secondary question of interest.

6.2 Results

Two dependent variables were measured in this experiment: reaction time (for the correct responses) and accuracy. Each variable was initially analyzed via a 2x3 completely within subject analysis of variance with trial-type (single item vs. multiple item), and condition (congruent vs. incongruent vs. control) as factors. The analysis of variance was then followed by planned comparisons examining the effect of item on each condition. The reaction time and accuracy analyses are presented separately for ease of consideration.

Since the number of trials in each condition was relatively large, I looked at the averages of each condition for the outlier analysis. Thus, the average of each condition within each participant was used to determine any outliers, for both reaction time and accuracy. Each condition was required to meet the criterion (mean of each condition ± 2.5 *standard deviation) for the participants' data to be included. Using this analysis, two participants' data were excluded from the analysis.

One complication of contrasting single and multiple-item trials was the fact that there were two kinds of incongruent trials for the multiple-item trial-type (primed and neutral) but only one for the single-item trial-type. However, a preliminary comparison of the two incongruent conditions within the multiple-item trial type revealed that they did not differ in terms of mean reaction time ($t(26) = 1.249, p = 0.223$) or accuracy ($t(26) = 1.696, p = 0.102$). Thus, for the results that follow, these two conditions are collapsed into a single 'incongruent' trial type.

6.2.1 Reaction Times

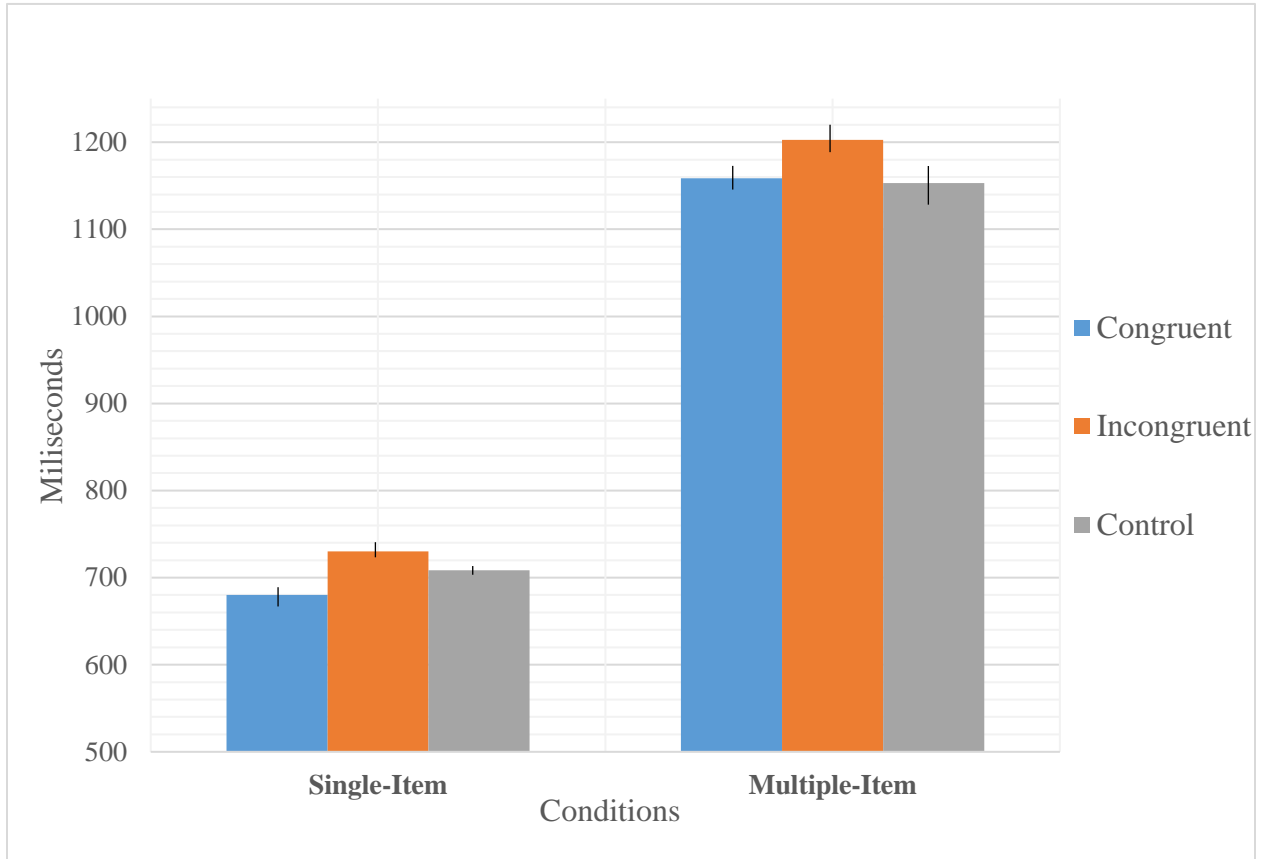


Figure 5. Experiment 1: Mean reaction times (milliseconds) and standard error for single-item vs multiple-item Stroop effect.

As depicted in Figure 5, the pattern of means suggests the following. First, participants are indeed responding faster in the single-item trial-type than the multiple-item trial-type in all conditions. The overall slower latencies in the multiple-item trial-type is theoretically sound as the stimulus display has more information coupled with a decision task thereby increasing the cognitive load. Second, within each of the single-item and multiple-item trial-types, congruent trials are the fastest, followed by control trials and then lastly the incongruent trials. As such there is a Stroop effect present, both when comparing incongruent vs. congruent trials as well as when comparing incongruent vs. control trials both in the single-item and multiple-item trial-types.

To quantify the reliability of these patterns, the data were analyzed via analysis of variance on the mean reaction times, which revealed that both the main effects were significant. That is, participants were faster overall when it was a single-item trial-type compared to multiple-item trial-type, $F(1,26) = 941.11$, $MSe = 9313.44$, $p < 0.001$. The analysis also showed a main effect between the conditions, $F(2,52) = 41.69$, $MSe = 781.42$, $p < 0.001$. In addition, there was also a significant interaction between trial-type and condition, $F(2,52) = 4.92$, $MSe = 906.29$, $p = 0.011$.

In order to explore the data further, I conducted planned comparisons investigating reaction time difference between the conditions within each of the trial-types. I used the Bonferroni correction method to correct for the inflation to the Type 1 error rates as a result of multiple comparisons. There was a significant difference between the single-item congruent vs. single-item incongruent, (50 ms), $t(26) = 4.36$, $p < 0.001$, and between the single-item congruent vs. single-item control, (28 ms), $t(26) = 4.56$, $p < 0.001$.

In addition, there was a significant difference (mean difference = 44 ms) between multiple-item congruent vs. multiple-item incongruent, $t(26) = 4.8$, $p < 0.001$, and a significant difference (50 ms) between multiple-item incongruent vs. multiple-item control ($t(26) = 6.92$, $p < 0.001$), but no significant difference between multiple-item congruent vs. multiple-item control (mean difference = 5.43, $t(26) = 1.13$, $p = 0.265$). The Cohen's d for single-item congruent vs. incongruent was 1.15 while the Cohen's d for the multiple-item congruent vs. incongruent was 0.55. Thus, it can conclusively be said that the Stroop effect was present for both single and multiple item, as congruent trial reaction times were faster than incongruent trial reaction times. However the Stroop effect was not larger in the multiple-item paradigm than in the single-item paradigm; rather the opposite was true with the single-item paradigm having a larger Stroop effect.

6.2.2 Accuracy

Table 1

Experiment 1: Descriptive statistics associated with mean accuracy rates for single-item and multiple-item Stroop effect.

	<u>Single-Item trial-type</u>			<u>Multiple-Item trial-type</u>		
	Congruent	Incongruent	Control	Congruent	Incongruent	Control
Mean	0.975	0.980	0.968	0.919	0.917	0.902
St Dev	0.009	0.009	0.004	0.022	0.029	0.019

The mean accuracy rates are seen in Table 1. The analysis of variance of mean accuracy showed that the main effects for both trialtype, ($F(1, 26) = 223.39$, $MSe = 0.001$, $p < 0.001$), and condition ($F(2, 52) = 25.19$, $MSe = 0.000$, $p < 0.001$) were significant. There was also a significant interaction effect, $F(2, 52) = 3.85$, $MSe = 0.000$, $p = 0.027$.

These effects were then examined via paired t-tests which revealed the following, again using the Bonferroni correction method. There was no difference between single-item congruent and incongruent accuracy, $t(26) = 1.75$, $p = 0.092$, and no difference between multiple-item congruent and incongruent accuracy, $t(26) = 0.71$, $p = 0.48$. Congruent trials had significantly fewer errors than control trials in both single-item trials ($t(26) = 3.80$, $p < 0.001$) and multiple-item trials ($t(26) = 5.09$, $p < 0.001$). Here there is evidence of Stroop facilitation occurring, where the amalgamation of word and colour is leading to fewer errors as oppose to no word at all.

As the direction of error rate differences follows that which was observed in the reaction times results, there was no fear of a speed-accuracy trade-off. In addition, the lack of difference in accuracy between single-item trials and multiple-item trials in terms of congruent vs incongruent demonstrates that participants were not making more errors in any one condition, but the

accuracy is stable across the two trial-types. Although there were more errors made in the control condition compared to congruent conditions for both single-item and multiple-item variants, this was not the case when comparing the congruent vs incongruent conditions.

6.3 Discussion

The aim of this experiment was to establish that the Stroop effect was present using the Multiple Item Stroop paradigm. The multiple-item trials were successfully able to produce a Stroop effect, where the multiple-item congruent trials were faster than multiple-item incongruent trials.

Although the reaction times for the multiple-item trial-type were overall larger than the single-item trial-type, the difference between the congruent and incongruent conditions was significant. Thus, a direct comparison of the standard Stroop task and the Multiple Item Stroop shows that the standard Stroop effect is intact regardless of the modifications.

A key finding from this experiment is that the multiple-item Stroop effect was no larger than the single-item Stroop effect (multiple-item mean difference = 48 ms vs. single-item mean difference = 50 ms). In fact, the single-item Stroop effect Cohen's d can be categorized as a large effect, while the multiple-item Stroop effect Cohen's d is categorized as a medium effect (Cohen, 1992). As previous research suggests, the Stroop effect has been quite resilient to change in its qualitative form, and this experiment exemplifies that a single word is all that is needed to produce the maximal standard Stroop effect (MacLeod, C. M., 1991; Augustinova et al., 2010). Theoretically my intention for using the Multiple Item Stroop was to exaggerate attentional biases, and since the standard Stroop effect is assumed to primarily reflect response competition and not attentional biases, the finding of similar Stroop effect is sensible and theoretically sound.

7 Experiment 2: Single-item and Multiple-item Emotional Stroop

The original emotional Stroop task was based on the traditional Stroop effect, the effect observed using a single colour-related item. My hope is to create a more sensitive version of the

emotional Stroop effect by using multiple items that must be perceptually scanned prior to response. This requires relatively substantial changes to the stimulus display and, thus, it seemed prudent to examine the standard Stroop effect in this new context to ensure that these changes do not fundamentally change how participants approach the experiment. Given that I observed the standard Stroop findings in both the single-item and multiple-item versions of Experiment 1, I now feel comfortable exploring the emotional Stroop effect in Experiment 2.

The ultimate goal of this thesis is to produce a tool that will be able to measure attentional biases. As I am developing this tool within a university context using typical undergraduates as participants, I cannot use disorder-related stimuli as those used with clinical populations. Thus, within this experiment, I opted for using emotional words as my stimuli, with the assumption that any population may show a bias towards negative and threatening stimuli (Bradley, Mogg, et al., 2000; Dijksterhuis & Aarts, 2003; Pratto & John, 1991).

Obviously a negative stimulus presented to a non-clinical undergraduate might not have the same attentional capture ability as a disorder-related word would have for a specific clinical group (e.g., anxiety related words presented to anxious individuals). However, previous research has suggested that undergraduate populations are sensitive to negative stimuli so my hope was that these words would be sufficient to produce an emotional Stroop effect with my sample (Bradley, Mogg, et al., 2000; Pratto & John, 1991). For example, Dijksterhuis and Aarts (2003) demonstrated that negative stimuli were detected faster and more accurately than positive, neutral or no stimuli shown within an undergraduate population. As such, I elected to use negative stimuli as they have the most theoretical evidence, such as the automatic vigilance hypothesis, and seem to be quite salient within an undergraduate population (Ohman, 2005; Pratto & John, 1991; Schimmack, 2005).

Emotionality is a subjective concept and is generally thought to reflect the two dichotomous dimensions of valence and arousal. Stemming from an evolutionary standpoint as well as a

neuroscience basis, negative stimuli are assumed to capture attention as they are associated with defense or avoidance (e.g., snake, gun, angry face) even within a healthy population (Anderson, 2005; Ohman, 2005; Schimmack, 2005). As such, I opted for using negative words with high valence and arousal to capture attention to the highest degree. The neutral words were matched for familiarity, Kucera-Francis written frequency, began with the same letter and had the same number of letters in each word. These constraints were used because lexical characteristics are known to affect word naming latencies (Burt, 2002; Balota et al., 2004; Algom et al., 2004).

As Experiment 1 demonstrated that the modifications within Multiple Item Stroop effectively produced the standard Stroop effect, Experiment 2 was conducted in an attempt to replicate those findings using negative stimuli, specifically words with high valence and high arousal. To keep Experiment 2 consistent with Experiment 1, I opted for negative and neutral stimuli, keeping in line with two conditions along with the one control condition. Thus the goal of this experiment was to directly assess the emotional Stroop effect in both single and multiple-item contexts. The emotional Stroop effect is said to be present when the reaction times are longer for the negative condition than for the neutral condition. We know the effect tends to be unreliable in single-item contexts and thus would not be surprised to find that here. My hope is that the engagement of attention across the multiple-item stimulus will be more sensitive to capture any attentional bias present.

7.1 Method

7.1.1 Participants

A total of 29 undergraduate students from the University of Toronto Scarborough participated in this study on a volunteer basis for course credit. The only exclusion criterion for the experiment was colour blindness as the primary task of the experiment was to identify different colours. All participants gave informed consent. In accordance with ethical guidelines, participants were explicitly told about the workings of the experiment and the aim of the study. Lastly, the participants were informed that they could leave at any time during the experiment without any

repercussions if they felt uncomfortable. This experiment was approved by the Research Ethics Board of University of Toronto Scarborough.

7.1.2 Material and Apparatus

The apparatus used within this experiment was the same as that used in Experiment 1. However, the primary modification was with respect to the stimuli that were presented.

In the present experiment, the stimuli consisted of negative and neutral words and control non-words. The negative words were chosen based on their arousal and valence levels (WAR: arousal level of 7.49 and valence of 2.08; HATE: arousal level of 6.95 and valence of 2.12; ABUSE: arousal level of 6.83 and valence of 1.80) (Bradley & Lang, 1999). These three words were chosen to represent the standard Stroop stimuli of red, green and blue in terms of word length, yet were meant to be distinct with different first letters so as to not replicate a traditional Stroop effect. The neutral words (wet, hold, and allow) were chosen to match the negative words in their familiarity, Kucera-Francis written frequency, first letter, and have the same number of letters. The close similarity between the negative and neutral words would allow for sound comparison between them. The control non-words were kept the same as Experiment 1 (XXX, YYYY, ZZZZZ).

The presentation of the chosen stimuli were similar to Experiment 1. Specifically, the single-item trials represented the traditional emotional Stroop task, whereby one stimulus is presented in the centre of the screen in some specific ink colour (i.e., red, green, or blue). This time, however, the stimulus was either negative, neutral or a control non-word, yet the participants were still required to correctly identify the ink colour. There were three conditions within the single-item trials based on the type of stimuli presented: the negative condition wherein a negative word was presented in coloured ink; the neutral condition wherein a neutral word was presented in coloured ink; and the control condition wherein a non-word was presented in coloured ink. The font size used was 50 in the Arial font type. The participants' task was to press the button corresponding to the colour of the item as quickly and accurately as possible.

Similar to Experiment 1, in the multiple-item trials a given stimulus is repeated multiple times across the screen and there is a random distribution of two ink colours in a 60/40 proportion. The participants' task remains to identify the ink colour, but now they have to determine which ink colour is dominant and identify that ink colour by way of button press. Again, there were three conditions within the multiple-item trials based on the type of stimuli presented: the negative condition wherein a negative word was presented in varied ink colour (60% in one colour and 40% in another colour); the neutral condition wherein a neutral word was presented in proportional coloured ink (60% in one colour and 40% in another colour); and the control condition wherein a non-word was presented in proportional coloured ink (60% in one colour and 40% in another colour). The font size used was 30 in the Arial font type.

7.1.3 Procedure

This experiment was a replication of Experiment 1, with modifications to the stimuli list. As such, the procedures for this experiment were the same as Experiment 1 for the following components: recruitment of participants, verbally informing the participants of their rights, obtaining informed consent (see Appendix B), written instructions within the experiment, practice block ($n = 24$), brief resting screen during the experiment, and thank you/debriefing screen.

The experiment consisted of a total of 378 trials presented in a mixed fashion, with 189 single-item and 189 multiple-item trials. Of those 189 single-item trials, the negative words (war, hate, abuse) were presented 16 times each, and neutral words (wet, hold, allow) were presented 16 times each, while the control items (xxx, yyyy, zzzzz) appeared 31 times each. Of the 189 multiple-item trials, each of the negative words (war, hate, abuse) present 16 times, and neutral words (wet, hold, allow) was presented on 16 times each, while the control items (xxx, yyyy, zzzzz) appeared 31 times each respectively.

My goal for this experiment was to determine whether an emotional Stroop effect would be present in single and multiple-item contexts. Emotional Stroop interference was measured by subtracting the negative from neutral trial reaction times. To explore whether negative words were slower in any form, I also calculated the difference between negative and control trial reaction times. In line with theory, my hypothesis was that negative words would take longer than neutral and controls, as the meaning of those words would be distracting to the participants, capturing their attention and ultimately slowing their responses. My hope was to see a larger difference within the multiple-item trials than single-item trials. As previously mentioned, the emotional Stroop effect has been notably unreliable in the single-item context, as such I would not be surprised if I did not find an effect there. My hope is that the multiple-item condition will be more sensitive in capturing an attentional bias and thus be able to produce the effect more strongly.

7.2 Results

Two dependent variables were measured in this experiment; reaction time (for the correct responses) and accuracy. Each variable was initially analyzed via a 2x3 completely within subject analysis of variance with trial-type (single item vs. multiple item) and condition (negative vs. neutral vs. control) as factors. The analysis of variance was then followed by planned comparisons examining the effect of condition on each trial-type. The reaction times and accuracy analysis are presented separately for ease of consideration. Two participants' data were excluded due to outlier analysis, as described in Experiment 1.

7.2.1 Reaction Times

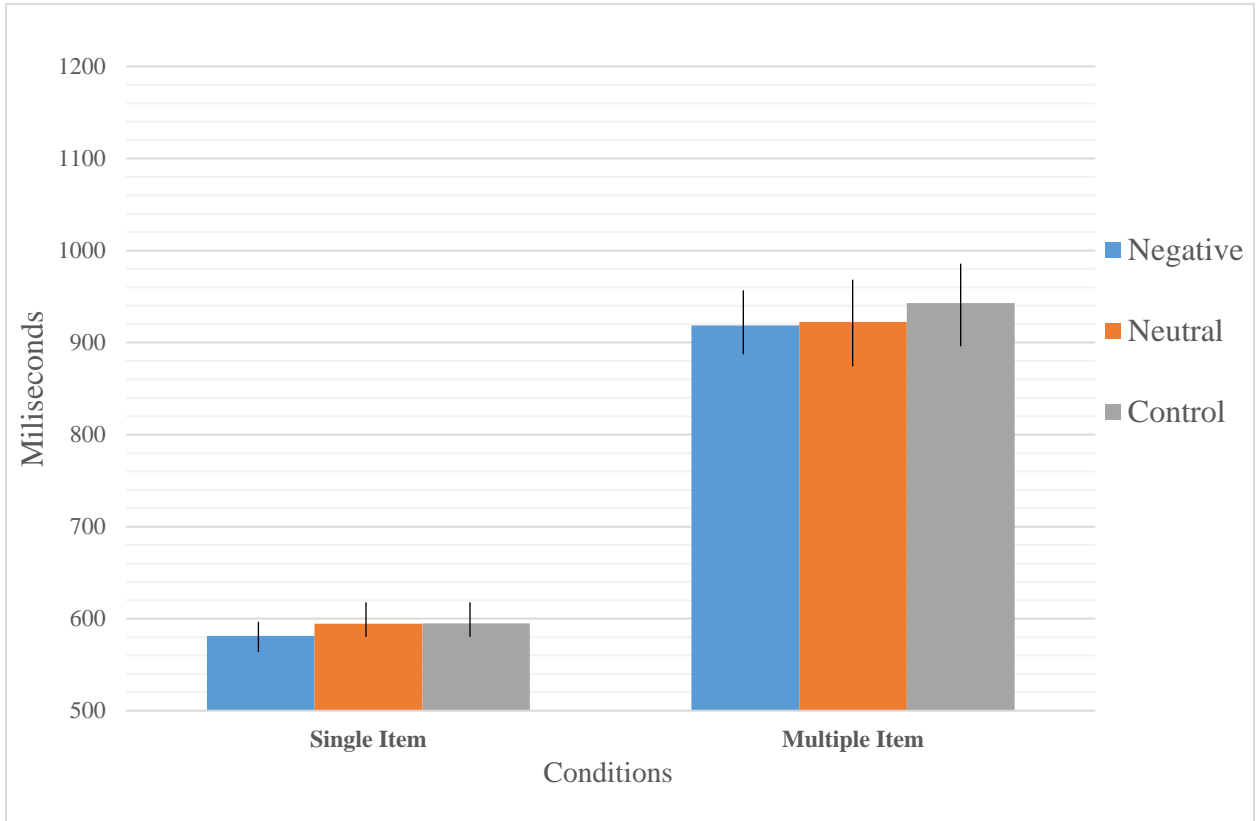


Figure 6. Experiment 2: Mean reaction times (milliseconds) and standard error for single-item vs multiple-item Stroop effect for negative and neutral words and control non-words.

As depicted in Figure 6, the pattern of means suggests the following. First, single-items trial-type are responded to much faster than multiple-item trial-type, replicating findings from Experiment 1. Second, within the single-item trial-type there was not much difference between the conditions at all, suggesting that participants are correctly responding to all negative, neutral and control conditions with equivalent speed. Third, within the multiple-items trial-type, the non-word control condition seems to be taking the longest to respond while the participants were actually faster in responding to the negative condition. Thus, there was no emotional Stroop effect in the single-item trials, which was expected. However, disappointingly, there was also no emotional Stroop effect in the multiple-item trials.

In order to quantify the reliability of these findings, I conducted an analysis of variance on these means, which revealed that only the main effect of trial-type was statistically significant, $F(1, 26) = 137.27$, $MSe = 33645.62$, $p < 0.001$ (all other $F < 2$). That is, participants were overall faster when responding to single-item trials compared to multiple-item trials.

To further explore the findings, pairwise t-tests were conducted, comparing negative vs. neutral and negative vs. control in both single and multiple items, again using the Bonferroni correction method. There were no statistically significant differences found within any of the comparisons.

7.2.2 Accuracy

Table 2

Experiment 2: Descriptive statistics associated with mean accuracy rates for single-item and multiple-item Stroop effect.

	<u>Single-Item trial-type</u>			<u>Multiple-Item trial-type</u>		
	Negative	Neutral	Control	Negative	Neutral	Control
Mean	0.963	0.960	0.965	0.892	0.891	0.882
St Dev	0.028	0.030	0.024	0.062	0.064	0.066

The analysis of variance of the mean accuracy showed a main effect for trial-type, $F(1, 26) = 6688$, $MSe = 0.003$, $p < 0.001$ (see Table 2 for means and standard deviations). No other main effect or interaction was significant (all other $F < 2$). Further analysis using pairwise t-test comparing negative vs. neutral and negative vs. control in both single and multiple item revealed no significant effect (all $t < 2$). The lack of difference in accuracy demonstrates that participants were not making more errors in any one condition; rather the accuracy rate was stable across the conditions.

7.3 Discussion

The hope going into this experiment was that a small emotional Stroop effect would be present in the single-item trials, and that this effect would be magnified in the multiple-item condition. However, I did not observe an emotional Stroop effect in either single-item trials or multiple-item trials. The only significant difference was that single-word trials were overall faster than multiple-word trials, which was also seen in Experiment 1. As the task demands of multiple-item trials are larger than single-item trials, this difference made sense and was expected. However the lack of emotional Stroop effect caused me to question the methodology and stimulus set employed. As such, factors and limitations of the current experimental design were explored.

In retrospect, there are a number of factors that may explain why I did not see the emotional Stroop effect I expected to see. First, in order to keep Experiment 2 as similar to Experiment 1 as possible, there were only three distinct negative stimuli (war, hate, abuse) and three distinct neutral (wet, hold, allow) words that were repeatedly used. This repetition of stimuli may have caused the participants to habituate to the meaning: Initially those words may have caused some discomfort or negative affect, but with repetition the negative effect dissipated. This effect is known as the semantic satiation effect, which is known to make words meaningless after prolonged exposure or massive repetition (Esposito & Pelton, 1971; Shimokido, 2006). Experiments using a lexical decision task have been used to demonstrate the semantic satiation effect in undergraduate students (Shimokido, 2006). This effect fits nicely with the results seen in this experiment.

Second, we know that subjective experience and interpretation have a large effect on how stimuli are viewed and experienced (Eysenck, Mogg, et al., 1991). As such, the word ‘war’ may not be as negative to some who has never lived through a warzone or the word ‘abuse’ may not be traumatic to someone who has not lived through such an experience. Thus, the use of such a limited selection of words with a very large degree of interpretation might have been some of the reasons that I was unable to detect an emotional Stroop effect.

Third, the use of ‘normal’ undergraduate students as participants may have caused some of the effect to dissipate. Related to the previous factor, even if the meaning of the words is getting through with all the repetition, these particular stimuli chosen may not be as negative or jarring enough for this participant pool. As such, these stimuli would not be able to capture the type of attentional biases that would cause one to shift their focus of attention. Thus, I need a set of stimuli that would be jarring even to an undergraduate population, words that should be expected to stand out in an emotional manner. I return to this challenge in Chapter 3.

8 General Discussion

The primary goal of this thesis is to create and assess a modified version of the Stroop task by capitalizing on attentional biases. As such, I created the Multiple Item Stroop paradigm wherein a stimulus is presented multiple times across the screen in proportional ink colours. Experiment 1 demonstrated that the standard Stroop effect was evident with this paradigm, and that there was no drastic difference in the effect compared to the single-item task. In Experiment 2, I used negative and neutral words to produce an emotional Stroop effect; unfortunately no such effect was observed.

Experiment 1 demonstrated that the standard Stroop effect was present using the multiple-item paradigm. Additionally, this effect was not larger in magnitude in the multiple-item trials than in the single-item trials. As the standard Stroop effect is thought to occur due to response competition during the output stage of processing, the presence of the standard Stroop effect using the multiple-item task is theoretically sound. The use of the multiple-item variant was meant to capitalize on attentional biases, which were not present with the use of standard colour stimuli. Thus, a large effect size for the single-item and a medium effect size for the multiple-item was theoretical sensible. Another significant finding was that it took longer for participants to respond to the multiple-item trials than the single-item trials regardless of the condition. Even with more stimuli and a judgment task regarding the dominant colour, the significant Stroop effect speaks volume regarding the strength and endurance of the traditional effect.

Experiment 2 attempted to test the Multiple Item Stroop paradigm with negative and neutral words to achieve an emotional Stroop effect. Unfortunately, no such emotional Stroop effect was obtained. Factors for this null finding were discussed, leading me to question some of the methodology and the stimulus set used. The replication of one finding from Experiment 1 did occur, which was that it took longer for participants to respond to the multiple-item trials than the single-item trials regardless of the condition.

Some of the issues regarding the null finding in Experiment 2 were the use of a limited stimulus set within a 'normal' undergraduate student participant pool. In addition, semantic satiation of the three negative words used may have caused some of the negative effect to dissipate. Given that these potential issues exist, I am not prepared to give up on my new procedure quite yet. In Chapter 3, I will take what I learned from Experiment 2 to create an instantiation of my paradigm that I think has a greater likelihood of supporting and documenting attentional biases even among 'normal undergraduates'.

Chapter 3

Taboo Stroop Effect and Anxiety Measures

As the ultimate aim of this thesis is to assess the magnitude of attentional biases, I created the Multiple Item Stroop in hopes of capturing and magnifying such effects. Although I was able to effectively produce the standard Stroop effect using this new paradigm, unfortunately I was unable to detect the emotional Stroop effect. Consideration of this null finding prompted me to re-evaluate my methods and stimuli, and as such I expanded my stimulus set as well as included a wider range of words. Most critically, I added so-called taboo words to my stimulus set (e.g., TITS, COCK) in hopes they might be more likely to capture attention with an undergraduate population when presented within a laboratory context. With these new parameters, I again set out to test the Multiple Item Stroop task, with much more success.

The theoretical underpinning of the Multiple Item Stroop was to force the participants to experience the critical word repeatedly as they repeatedly engage and disengage attention to determine the dominant ink colour (Posner et al., 1975). As such, if a bias is present toward that critical word, attention will be drawn toward it and thus performance on the task will suffer. Although the modifications were quite drastic, Experiment 1 demonstrated that the standard Stroop effect was present as in the traditional Stroop task. However the same paradigm did not elicit any effect when the stimuli were switched to negative and neutral words, demonstrating no emotional Stroop effect.

A critical element, I believe, regarding the null findings in Experiment 2 was the limited stimulus set, in which the same three critical words were repeated multiple times in an undergraduate participant pool. The first problem with this approach is that the specific words chosen may simply have not been sufficiently arousing, emotionally speaking, to capture attention. Undergraduates today live in a world of horrific terrorist acts and threateningly provocative statements from world leaders. Given this, it is perhaps not surprising that seeing words like WAR, HATE, and ABUSE within a laboratory setting is not all that emotionally arousing.

In addition, the three words I did use were presented repeatedly. As highlighted previously, repeating words across trials can result in semantic satiation; basically the words can stop activating their meaning (Esposito & Pelton, 1971; Shimokido, 2006). Given my use of repetition both within and across trials, any attentional impact these words may have had initially could have been watered down in short order. When one considers the particular words I used and the repetition I allowed together, it is entirely possible that the stimuli I used simply did not pack the punch that is needed to pull attention in the manner that would cause biases to be observed.

As such, the stimulus set used in the current experiment was altered in two ways. First, the number of words was expanded significantly from six to 144 to help reduce the semantic satiation issue. Second, in addition to using neutral, negative and positive items, a new word type was included, one that might hold more impact for students; so-called taboo words. Words that fall in this category are commonly restricted due to social and cultural constraints; these words are typically sexually derogatory and demeaning, and considered swear words. These words are quite negative and most often offensive, creating a perfect stimulus set that is both high in valence and arousal level for use in a ‘normal’ undergraduate participant pool.

Taboo words have been employed as stimuli in experiments previously. As early as 1981, using a dichotomous listening task, Nielsen and Sarason (1981) demonstrated that taboo words were hard to ignore. Geer, Judice, and Jackson (1994) found reading time for erotic and/or sexual material was longer than for neutral material, signifying increased cognitive processing. In addition, it has been shown that taboo words and negative words produce higher interference in native language than non-native language (Eilola & Hevelka, 2011). Physiological arousal has also been documented; Labar and Phelps (1998) found that skin conductance response was higher for taboo words than for neutral words, both when initially encountering the taboo words and when remembering them at a later time. Thus, there is evidence that taboo words disrupt attention, increase cognitive processing, and exhibit unconscious physiological response.

The first instance of taboo words employed in a Stroop task was in 1995 where Siegrist displayed taboo words along with neutral words, incongruent colour words and control non-words. Siegrist found the traditional Stroop effect with the incongruent colour words, but also found significant interference with the taboo words as well. Subsequent research demonstrated that it is the semantic meaning of the taboo words that trigger the emotional reaction and increased interference rather than an orthographic or phonological similarity (Guillet & Arndt, 2009; MacKay & Ahmetzanov, 2005; Siegrist, 1995).

A lot of the work with taboo words has been conducted within memory research. MacKay et al. (2004) conducted a series of five experiments that demonstrated how taboo words disproportionately produce superior memory, both in terms of recognition and recall. Subsequent research has also supported this notion of superior recall and recognition of taboo words compared to neutral, negative and threatening words (Jay, Caldwell-Harris, & King, 2008; Schmidt & Saari, 2007). Additionally, MacKay et al. also found superior memory for font colours of the taboo words compared to neutral words supporting the binding hypothesis theory of how word-specific emotional reactions can trigger better memory for concurrent contextual information. My assumption with the use of taboo words was that they would be anxiety provoking, although other types of emotions are possible (i.e., humour, arousal). I focused on anxiety as the primary emotion as evident from previous research (Foa & McNally, 1986; Geer et al., 1994; Goldstein, 2008; Labar & Phelps, 1998).

The aim of the Multiple Item Stroop task was to capture attentional biases with greater sensitivity. Since I was unable to find an emotional Stroop effect in Experiment 2, I expanded my stimulus set in Experiments 3 and 4 to include taboo words, negative and positive words, and their corresponding neutral words. The primary question was to determine whether I can observe an emotional Stroop effect, using both taboo words as well as emotional words (i.e., positive and negative words). Subsequently, if these effects are present, is the effect magnified in the multiple-item variant compared to the single-item variant. To anticipate, I did indeed see a taboo Stroop effect both in Experiment 3 and 4, and the effect is much more pronounced in the multiple-item trial-type than in the single-item variant. Experiment 4 replicates the finding from

Experiment 3, and extends it by attempting to correlate subjective and objective measures of anxiety.

9 Experiment 3: Single-item vs. Multiple-item Positive, Negative and Taboo Stroop

The aim of Experiment 3 was to determine whether an emotional Stroop effect and a taboo Stroop effect can be observed with the expanded stimulus set within both the single-item and multiple-item variants. The larger stimulus set consisted of positive, negative and taboo words and their corresponding neutral words as well as a larger total number of words used. As the Multiple-Item Stroop task is theoretically supposed to be more sensitive to attentional biases, the hope was to find the effect in both the variations, but that the effect would be magnified in the multiple-item trial-type.

As mentioned earlier, taboo words have been shown to capture attention within experiments (MacKay et al., 2004; Schmidt & Saari, 2007). In a laboratory setting, I believe taboo words would be even more anxiety provoking leading to attentional capture. Specifically from Experiment 2, seeing the word WAR within a laboratory context may not be that alerting; in fact these words might pack a greater emotional bang when they occur in the context of normal conversations or experiences. But the reverse could be true of taboo words. One does not enter a scientific lab within a university setting expecting to be confronted with words like COCK or TITS, as it is a formal context where such items are infrequent. So even if one knows they are going to see words of this sort, there is good reason to think that they will pack an emotional and attentional punch.

9.1 Method

9.1.1 Participants

A total of 48 undergraduate students from the University of Toronto Scarborough participated in this study on a volunteer basis for course credit. The only exclusion criterion for the experiment

was colour blindness as the primary task of the experiment was to identify different colours. All participants gave informed consent. In accordance with ethical guidelines, participants were explicitly told about the workings of the experiment and the aim of the study. Lastly, the participants were informed that they could leave at any time during the experiment without any repercussions if they felt uncomfortable. This experiment was approved by the Research Ethics Board of University of Toronto Scarborough.

9.1.2 Materials and Apparatus

The apparatus used within this experiment was the same as that used in Experiment 1. However, the primary modification was with respect to the stimuli that were presented.

In the present experiment, the stimuli consisted of 24 positive, negative and taboo words each and their 24 corresponding neutral words each. The positive and negative words were chosen based on their arousal and valence level (Bradley & Lang, 1999). The taboo words were harvested from a combination of previous research (e.g., Arnell, Killman, & Fijavz, 2007; Eilola & Hevelka, 2011; MacKay et al., 2004; MacKay & Ahmetzanov, 2005). All the corresponding neutral words were matched for word length and familiarity rating (see Appendix C for the complete list of stimulus set).

The presentation of the chosen stimuli were similar to Experiment 1. Specifically, the single-item trials represented the traditional emotional Stroop task, whereby one stimulus is presented in the centre of the screen in some specific ink colour (i.e., red, green, or blue). This time however, the stimulus was either positive, negative, neutral or a taboo word, yet the participants were still required to correctly identify the dominant ink colour.

There were four conditions within the single-item trials based on the type of stimuli presented: the positive condition wherein a positive word was presented in coloured ink; the negative

condition wherein a negative word was presented in coloured ink; the neutral condition wherein a neutral word was presented in coloured ink; and the taboo condition wherein a taboo word was presented in coloured ink. The font size used was 50 in the Arial font type. The participant's task was to press the button corresponding to the colour of the item as quickly and accurately as possible.

Similar to Experiment 1, in the multiple-item trials, a given stimulus is repeated multiple times across the screen and there is a random distribution of two ink colour in a 60/40 proportion. The participants' task remains to identify the ink colour, but now they have to determine which ink colour is dominant and identify that ink colour by way of button press. Again, there were four conditions within the multiple-item trials based on the type of stimuli presented: the positive condition wherein a positive word was presented in varied ink colour (60% in one colour and 40% in another colour); the negative condition wherein a negative word was presented in varied ink colour (60% in one colour and 40% in another colour); the neutral condition wherein a neutral word was presented in proportional coloured ink (60% in one colour and 40% in another colour); and the taboo condition wherein a taboo word was presented in proportional coloured ink (60% in one colour and 40% in another colour). The font size used was 30 in the Arial font type.

9.1.3 Procedure

This experiment was a replication of Experiment 1, with modifications to the stimulus list. As such, the procedures for this experiment were the same as Experiment 1 for the following components: recruitment of participants, verbally informing the participants to their rights, obtaining informed consent (see Appendix D), written instructions within the experiment, practice block ($n = 24$), brief resting screen during the experiment and thank you/debriefing screen. Participants were informed that they would be viewing taboo words and requested to continue if they were comfortable seeing those words.

The experiment consisted of a total of 576 trials presented in a mixed fashion, with 288 single-item and 288 multiple-item trials. Of those 288 single-item trials, positive, negative and taboo words along with their corresponding neutral words were presented in 48 trials each. Of the 288 multiple-item trials, positive, negative and taboo words along with their corresponding neutral words were presented in 48 trials each. Thus, each given word occurred four times in each of the single and multiple-item contexts.

My goal for this experiment was to determine whether an emotional Stroop effect as well as a taboo Stroop effect would be present in single and multiple-item contexts. Each of the positive, negative and taboo words had their own respective neutral words, thus the effect was calculated by looking at the difference between the reaction times to the target word relative to its associated control word, which was neutral in valence. Thus, the positive Stroop interference was measured by subtracting the positive target word reaction times from positive control word reaction times; the negative Stroop interference was measured by subtracting the negative target word reaction times from negative control word reaction times; and the taboo Stroop interference was measured by subtracting the taboo target word reaction times from taboo control word reaction times.

My hypothesis was that taboo words would have the largest difference in reaction times, followed by negative words and then positive words. Moreover, I hypothesized that this effect would be magnified within the multiple-item trial-type. My hope is that the multiple-item version will be more sensitive in capturing an attentional bias and thus be able to produce the effect more strongly.

9.2 Results

Two dependent variables were measured in this experiment; reaction time (for the correct responses) and accuracy. The reaction time and accuracy analyses are presented separately for

ease of consideration. Two participants' data were excluded due to outlier analysis as described in Experiment 1.

9.2.1 Reaction Times

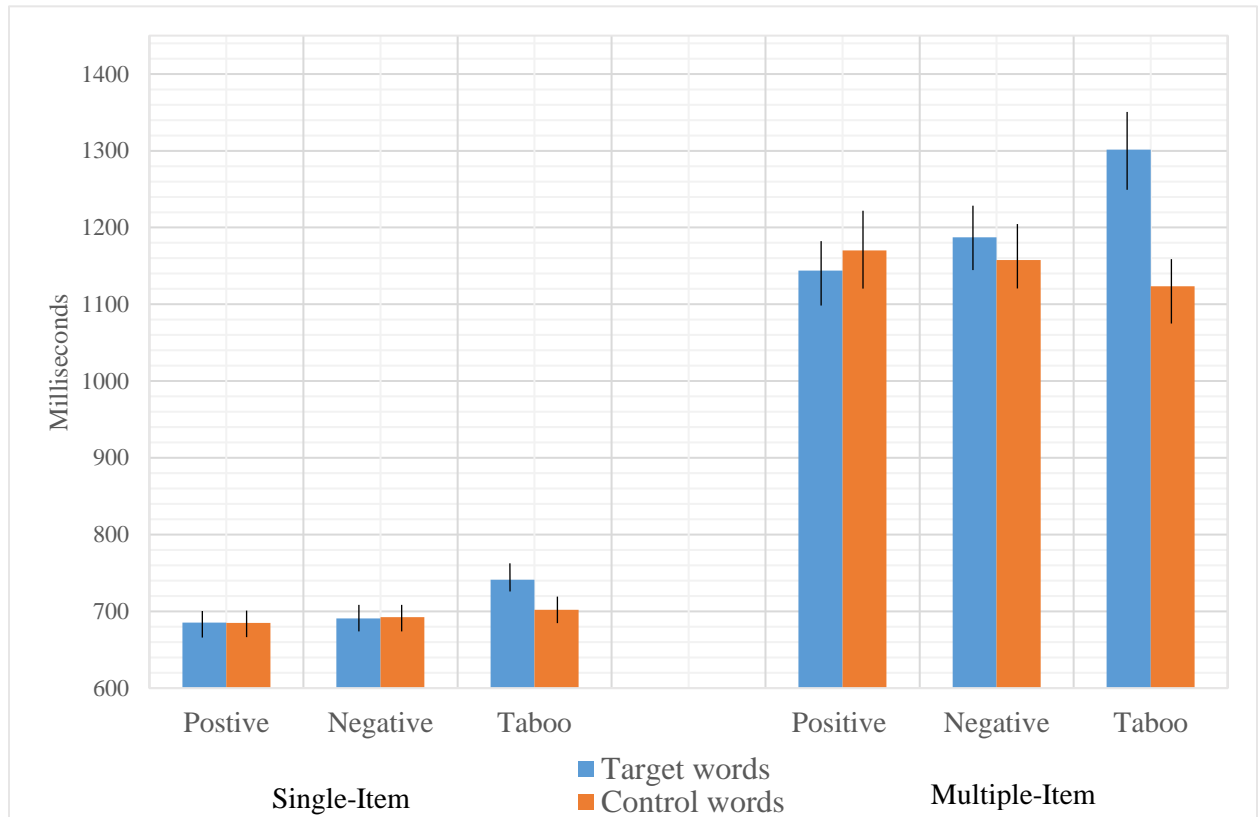


Figure 7. Experiment 3: Mean reaction times (milliseconds) and standard error for single-item vs multiple-item Stroop effect for positive, negative and taboo target and control words.

As depicted in Figure 7, the pattern of mean reaction times suggests the following. First, the single-item trial-type was responded to much faster than the multiple-item trial-type, replicating findings from Experiment 1 and 2. Second, within the single-item trial-type, there was not much difference between positive and negative target words and control words, demonstrating a lack of emotional Stroop effect (for both positive and negative words), a replication of the findings in Experiment 2. However, the taboo target words were slower than the taboo control words, demonstrating a taboo Stroop effect in the single-item trial-type. Third, within the multiple-item trial-type, there was not much difference between positive target and control words, although there was a difference between the negative target and control words signifying a negative emotional Stroop effect in the multiple-item trial-type. Importantly, there was a large difference

between the taboo target and control words within the multiple-item trial-type signifying a taboo Stroop effect. Notably, the taboo Stroop effect was much more pronounced in the multiple-item trial-type than in the single-item trial-type.

To quantify the reliability of these findings, I conducted an analysis of variance on these means via a 2x3x2 completely within subject analysis of variance with trial-type (single item vs. multiple item), valence (positive vs. negative vs. taboo) and word-type (target vs. control) as factors. All three main effects were significant: trial-type $F(1, 45) = 251.40$, $MSe = 127031.11$, $p < 0.001$, valence $F(2, 90) = 8.32$, $MSe = 12759.77$, $p < 0.001$, and word-type $F(1, 45) = 16.27$, $MSe = 11314.14$, $p < 0.001$. Additionally, the two-way interaction of trial-type and word-type was significant, $F(1, 45) = 5.32$, $MSe = 14676.65$, $p = 0.026$, as was the two-way interaction of valence and word-type, $F(2, 90) = 13.03$, $MSe = 14351.18$, $p < 0.001$. Although the interaction between trial-type and valence was not significant, $F(2, 90) = 0.30$, $MSe = 13322.19$, $p = 0.74$, there was a significant 3-way interaction of trial-type by valence by word-type, $F(2, 90) = 6.27$, $MSe = 12945.48$, $p = 0.003$.

In order to disentangle the 3-way interaction, I ran three 2-way analyses of variance with trial-type (single-item vs. multiple-item) and word-type (target vs. control) for each of the valences (positive, negative and taboo). For the positive words, there was only a main effect of trial-type, $F(1, 45) = 195.24$, $MSe = 52463.87$, $p < 0.001$, while the main effect of word-type and the interaction were not significant (all $F < 2$). For the negative words, again there was only a main effect of trial-type, $F(1, 45) = 261.04$, $MSe = 40723.02$, $p < 0.001$, while the main effect of word-type and the interaction were not significant (all $F < 2$). For the taboo words, there was a main effect for both trial-type ($F(1, 45) = 183.01$, $MSe = 60488.65$, $p < 0.001$) and word-type ($F(1, 45) = 53.15$, $MSe = 10192.90$, $p < 0.001$), as well as a significant interaction ($F(1, 45) = 18.18$, $MSe = 12158.67$, $p < 0.001$).

To explore the data further, pairwise t-tests were conducted. I used the Bonferroni correction method to correct for the inflation to the Type 1 error rates as a result of multiple comparisons. Within the single-item trial-types, no statistical difference was found between positive targets vs control words and between negative targets vs control words (all $t < 0.2$). However, there was a difference between taboo targets vs control words (mean difference = 40 ms, $t(45) = 3.04$, $p = 0.004$) demonstrating a taboo Stroop effect. Within the multiple-item trial-types, again there was no statistical difference found between positive target vs control words and between negative targets vs control words (all $t < 1$). However, there was a difference between taboo targets vs control words (mean difference = 178 ms, $t(45) = 6.27$, $p < 0.001$) signifying a taboo Stroop effect. In looking at the effect size, Cohen's d was 0.26 for the single-item trial-type for the taboo Stroop effect, which is classified as a small effect size, whereas Cohen's d was 0.53 for the multiple-item trial-type for the taboo Stroop effect, which is classified as a medium effect size (Cohen, 1992).

9.2.2 Accuracy

Table 3

Experiment 3: Descriptive statistics associated with mean accuracy rates for single-item and multiple-item Stroop effect in positive, negative and taboo target and control words.

	<u>Single-Item trial-type</u>						<u>Multiple-Item trial-type</u>					
	Positive		Negative		Taboo		Positive		Negative		Taboo	
	Target	Control	Target	Control	Target	Control	Target	Control	Target	Control	Target	Control
Mean	0.962	0.971	0.961	0.958	0.964	0.958	0.930	0.942	0.941	0.945	0.943	0.939
St.Dev	0.047	0.032	0.053	0.045	0.0511	0.070	0.066	0.069	0.085	0.076	0.059	0.093

The analysis of variance of the mean accuracy showed a main effect for trial-type, $F(1, 45) = 10.17$, $MSe = 0.003$, $p = 0.003$ (see Table 3 for mean and standard deviations for accuracy), but

no other main effect or interactions were significant (all other $F < 2$). Further analysis using pairwise t-tests comparing positive targets vs control, negative targets vs control and taboo targets vs control in both single-item trial-type and multiple-item trial-type revealed no significant effects (all $t < 2$). The lack of difference in accuracy demonstrates that participants were not making more errors in any one condition, rather the accuracy was higher for the single-item condition but otherwise was stable across the remaining manipulations.

9.3 Discussion

Given that I used a large stimulus set with a broad range of words, the hope going into this experiment was that I would see a taboo Stroop effect, and that this effect would be much more pronounced in the multiple-item trial-types. Looking at the results, I found exactly that: a taboo Stroop effect was present in both single-item and multiple-item variants, and the effect in the multiple-item trial-type was stronger. Thus the Multiple-Item Stroop task was more sensitive in terms of capturing this difference within the taboo words.

Another significant finding was that the multiple-item trial-types were responded to slower than the single-item trial-types. This finding was a replication of the previous two experiments and was expected, as the cognitive load in the multiple-item variation is larger. No emotional Stroop effect (with both positive and negative words) was detected, again replicating the findings from Experiment 2.

10 Experiment 4: Single-item vs. Multiple-item Positive, Negative and Taboo Stroop with Anxiety Measures

The driving goal of this thesis was to propose a paradigm that is reliable in capturing attentional biases. The previous experiments demonstrated that the Multiple-Item Stroop task produces the traditional Stroop effect as well as a taboo Stroop effect. Of particular importance was the

finding that the taboo Stroop effect was larger in the multiple-item variant than the single-item variant, signifying that the Multiple-Item Stroop task was more sensitive to attentional biases.

A critical aspect in psychological research is the idea of replication of results. Human beings are quite complex, and thus psychological experiments have a great deal of variability to contend with. Therefore it is not surprising that replication can be an issue. However for any field of research to garner the respect it deserves, those working within it must perform replications to better ascertain how reliable a given phenomenon may be. Thus, a primary aim of Experiment 4 is to replicate the results seen in Experiment 3 to ensure the reliability of the effects.

As mentioned in Chapter 1, one of the main pitfalls of the current emotional Stroop task is the lack of replicability and reliability of results. The paradigm does not produce robust results, and thus it is not surprising that the effect often does not replicate (MacLeod, C. M., & Hodder, 1998b; Dresler et al., 2012; Kindt et al., 1996). If the basic empirical effect does not replicate, then it is a waste of time and resources to use that tool, or any type of extension of that tool within a clinical context. As a first step, it seemed prudent to replicate the experiment with least amount of changes to ensure that the same findings are observed with a new sample of participants. Ultimately of course it would be optimal to see others replicate the effect using variations of my paradigm and stimuli.

The previous experiment demonstrates that the Multiple-Item Stroop paradigm can indeed produce a sensitive measure of what I assume to be attentional biases. However my ultimate goal is not to just document attentional biases but, more specifically, to quantify them in a relatively precise manner. That is, if this tool is to have value, it should be sensitive to the range of biases that are present in any given population.

Earlier in this chapter, I described research showing that, on average, undergraduates are sensitive to taboo words, a claim further supported in Experiment 3. Yet it is most certainly the case that seeing these kinds of words results in more anxiety in some participants than in others. Depending on how often one is exposed to such words, or perhaps depending on religious or cultural perceptions of such words, it is reasonable to assume that the power of taboo items to capture attention varies across individuals. If my new paradigm is sufficiently sensitive to such individual differences, and if I could find some independent manner for characterizing the anxiety level of individuals or groups of individuals, then larger taboo Stroop effects should occur for more anxious individuals and this difference should be especially obvious in the multiple-item paradigm.

With this logic in mind, a number of other measures were collected while participants were performing the replication, measures that I think should be related to the subjectively experienced in regards to the emotionality of taboo words and, therefore, measures that may predict the extent to which they capture attention. Higher heart rate, body temperature and galvanic skin response are indicative of higher anxiety and arousal, thus I asked the participants to wear a smartwatch that measured these reactions (Mann & Janelle, 2012). In addition, I administered two separate questionnaires after the Stroop experiment to gauge the participant's comfort level for the taboo words as well as their current and normal anxiety level.

The predictions going into this experiment are, first and foremost, that the effect documented in Experiment 3 replicates. Beyond that I predict that the size of the observed multiple-item taboo Stroop effect will be reflective of the emotional impact of the words as gauged by the additional measures I employed. If the Multiple-Item Stroop task can be demonstrated to be both generally significant and specifically sensitive to the attention-grabbing potential of specific items, then its use with a variety of population can be considered.

10.1 Method

10.1.1 Participants

A total of 63 undergraduate students from the University of Toronto Scarborough participated in this study on a volunteer basis for course credit. The only exclusion criterion for the experiment was colour blindness as the primary task of the experiment was to identify different colours. All participants gave informed consent. In accordance with ethical guidelines, participants were explicitly told about the workings of the experiment and the aim of the study. Lastly, the participants were informed that they could leave at any time during the experiment without any repercussions if they felt uncomfortable. This experiment was approved by the Research Ethics Board of University of Toronto Scarborough.

10.1.2 Material and Apparatus

The apparatus used within this experiment was the same as that used in all previous experiments. There were three additional measures added to determine the level of anxiety that participants were feeling, a smartwatch to record their physiological response, and two separate surveys to obtain their subjective experience of anxiety and comfort level. Each will now be discussed in more detail.

First and foremost, I had participants wear a smartwatch that recorded their heart rate, galvanic skin response and body temperature. These physiological responses are known to increase when one feels anxious or any discomfort (Mann & Janelle, 2015). The smartwatch used was Basis Peak (Basis Peak, 2017), which is equipped with advanced sensor technology with optical heart rate monitor, thermal sensors to measure skin temperature, and galvanic skin response to determine changes in perspiration levels. This was the first attempt to use these devices in our lab, so little was known about their potential to give precise measures ‘on the fly’.

After completing the Stroop component of the experiment, there were two separate questionnaires that were given to the participants. The first was a questionnaire that I created specifically to gauge the use of taboo words as well as participant's comfort level with those taboo words (see Appendix F). This included questions like 'how often do you use these words in front of parents/siblings/friends?' and 'how comfortable will you be reading these words out loud to the experimenter?' This was used to determine how familiar and comfortable the participants were with these words, as we know these can be factors in attentional capture and bias (Yiend et al., 2005). Subsequently, the participants were given the adult form of the State-Trait Anxiety Inventory (STAI; Spielberg, Gorsuch, & Lushene, 1970), which is a two part 20-item scale designed to measure state and trait anxiety (Bados, Gómez-Benito, & Balaguer, 2010). The first part measures state anxiety, which can be described as an in-the-moment transient unpleasant feeling of apprehension, anxiety, nervousness usually accompanied by autonomic nervous system responses as well (Bar-Haim et al., 2007). The second part of the STAI measures trait anxiety, which is more along the lines of a personality disposition where there is a tendency to perceive situations as threatening disproportionately (Spielberg et al., 1970).

10.1.3 Procedure

This experiment was a replication of Experiment 3 with the added subjective and objective measures of anxiety. As such, the procedure for the entire Stroop experiment was exactly the same (see Appendix E for consent form).

Prior to the beginning of the experiment, participants were given the smartwatch and asked to wear it around their non-dominant hand (to decrease interference from any movement). The smartwatch was synced with the lab's iPhone, which recorded real time data from the participants. Once synced, the participants wore the smartwatch for the duration of the entire experiment, including the surveys.

After the completion of the Stroop experiment component, participants were given the surveys to complete. The surveys were programmed right into the experiment, so that the instructions as well as the questions and answers were displayed on the computer screen, where the participants selected (via the mouse) the appropriate options. The Taboo questionnaire was administered first and then the STAI.

The goal of the experiment was to replicate the results of Experiment 3, as well as to determine whether there were any subjective or objective correlations between anxiety level and the corresponding taboo Stroop effect seen. The analysis of the data was conducted in two distinct sections. Initially, the overall Stroop data were analyzed just as they were in Experiment 3. This was done to determine whether in fact there was a replication of the findings. Subsequently, the subjective and objective measures were correlated with the taboo Stroop effect to determine whether there was any relation. Additionally, scatterplots were created to capture the pattern of reaction times for both single-item and multiple-item variants.

10.2 Behavioural Data

10.2.1 Results

Two dependent variables were measured in this experiment; reaction time (for the correct responses) and accuracy. The reaction time and accuracy analyses are presented separately for ease of consideration. Three participants' data were excluded using the outlier analysis mentioned in Experiment 1.

10.2.2 Reaction Times

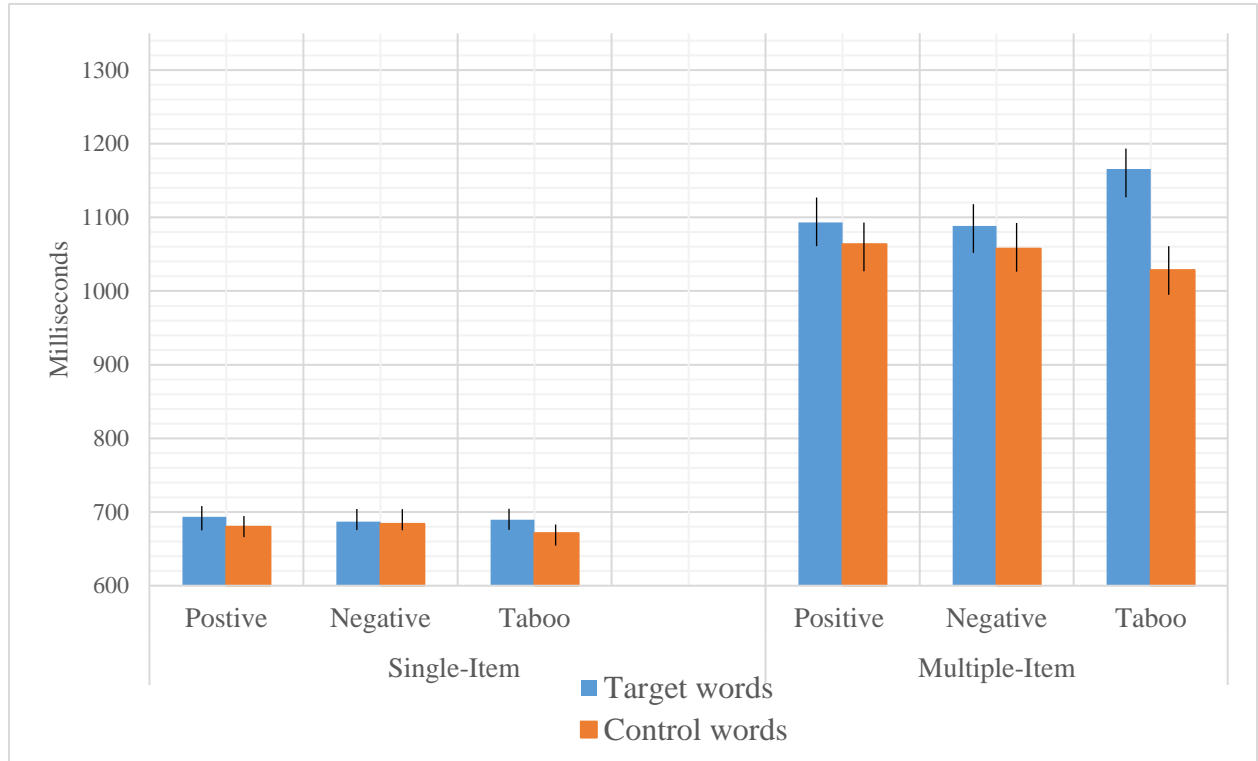


Figure 8. Experiment 4: Mean reaction times (milliseconds) and standard error for single-item vs multiple-item Stroop effect for positive, negative and taboo target and control words.

As depicted in Figure 8, the pattern of mean reaction times suggests the following. First, the single-item trial-type are responded to much faster than the multiple-item trial-type, replicating findings from Experiments 1, 2, and 3. Second, within the single-item trial-type there was not much difference between positive and negative target words and control words, demonstrating a lack of emotional Stroop effect (for both positive and negative words), a replication of the findings in Experiment 2 and 3. However, there was a slight difference between the taboo target and the taboo control words signifying a potential taboo Stroop effect in the single-item trial-type. Third, the target word-types within the multiple-item trial-type were all responded to slower than the control word-types. A difference can be seen in both positive target vs control word-type and negative target vs control word-type, signifying an emotional Stroop effect for both positive and negative target words. Of significant importance to this thesis, there was a large difference between the taboo target and control words within the multiple-item trial-type

signifying a taboo Stroop effect. Notably, the taboo Stroop effect was much more pronounced in the multiple-item trial-type than in the single-item trial-type.

To quantify the reliability of these findings, I conducted an analysis of variance on these means via a 2x3x2 completely within subject analysis of variance with trial-type (single item vs. multiple item), valence (positive vs. negative vs. taboo) and word-type (target vs. control) as factors. Two of the three main effects were significant: trial-type, $F(1, 59) = 400.38$, $MSe = 71432.02$, $p < 0.001$, and word-type, $F(1, 59) = 57.84$, $MSe = 4682.23$, $p < 0.001$, whereas valence was not significant, $F(2, 118) = 1.12$, $MSe = 4919.90$, $p = 0.331$. Additionally, the two-way interaction of trial-type and word-type, $F(1, 59) = 20.94$, $MSe = 6345.69$, $p < 0.001$, was significant as was the two-way interaction of valence and word-type, $F(2, 118) = 19.98$, $MSe = 3452.77$, $p < 0.001$, and the interaction between trial-type and valence, $F(2, 118) = 3.99$, $MSe = 3745.58$, $p = 0.021$. Finally, there was also a significant 3-way interaction of trial-type by valence by condition, $F(2, 118) = 10.80$, $MSe = 4483.09$, $p < 0.001$.

In order to disentangle the 3-way interaction, I ran three 2-way analyses of variance with trial-type (single-item vs. multiple-item) and word-type (target vs. control), separately for each of the valences (positive, negative and taboo). For the positive words, there was a main effect for both trial-type ($F(1, 59) = 325.58$, $MSe = 28256.60$, $p < 0.001$) and word-type ($F(1, 59) = 6.78$, $MSe = 4048.72$, $p = 0.012$), but no significant interaction ($F < 2$). For the negative words, again there was a main effect of both trial-type ($F(1, 59) = 379.21$, $MSe = 23758.95$, $p < 0.001$) and word-type ($F(1, 59) = 5.65$, $MSe = 3102.60$, $p = 0.021$), but no significant interaction ($F < 2$). For the taboo words, there was a main effect for both trial-type ($F(1, 59) = 387.27$, $MSe = 26898.63$, $p < 0.001$) and word-type ($F(1, 59) = 82.00$, $MSe = 4436.44$, $p < 0.001$), as well as a significant interaction ($F(1, 59) = 59.566$, $MSe = 4327.39$, $p < 0.001$).

To explore the data further, pairwise t-tests were conducted, again using the Bonferroni correction method. Within the single-item trial-types, no statistical difference was found between

positive targets vs control words and between negative targets vs control words (all $t < 1$). There was a difference between taboo targets vs control words (mean difference = 18 ms, $t(59) = 2.15$, $p = 0.036$) demonstrating a taboo Stroop effect, but did not reach significance due to the Bonferroni correction. Within the multiple-item trial-types, I found a Stroop effect in negative targets vs. control words (mean difference = 31 ms, $t(59) = 2.89$, $p = 0.005$), and taboo target vs. control words (mean difference = 138 ms, $t(59) = 9.25$, $p < 0.001$). Thus, I found a taboo Stroop effect in both the single-item and multiple-item variants, where the effect was larger in the multiple-item trial-type, replicating the findings from Experiment 3. In looking at the effect size, the Cohen's d was 0.16 for the single-item trial-type taboo Stroop effect which is classified as a small effect size, while the Cohen's d was 0.58 for the multiple-item trial-type taboo Stroop effect which is classified as a medium effect size (Cohen, 1992). These effect sizes are comparable to the ones found in Experiment 3.

10.2.3 Accuracy

Table 4

Experiment 4: Descriptive statistics associated with mean accuracy rates for single-item and multiple-item Stroop effect in positive, negative and taboo target and control words.

	<u>Single-Item trial-type</u>						<u>Multiple-Item trial-type</u>					
	Positive		Negative		Taboo		Positive		Negative		Taboo	
	Target	Control	Target	Control	Target	Control	Target	Control	Target	Control	Target	Control
Mean	0.960	0.965	0.965	0.969	0.974	0.966	0.894	0.900	0.890	0.887	0.881	0.889
St.Dev	0.035	0.033	0.036	0.034	0.024	0.036	0.065	0.055	0.070	0.069	0.074	0.057

The analysis of variance of the mean accuracy showed a main effect for trial-type, $F(1, 59) = 198.88$, $MSe = 0.005$, $p < 0.001$ (see Table 4). No other main effect or interactions were significant (all other $F < 5$). Further analysis using pairwise t-tests comparing positive target vs

control, negative target vs control and taboo target vs control in both single-item trial-type and multiple-item trial-type revealed no significant effects (all $t < 2$). The lack of difference in accuracy demonstrates that participants were not making more errors in any one condition, rather the accuracy was stable across the conditions, again a replication of Experiment 3 results.

10.2.4 Discussion

One of the aims of Experiment 4 was to replicate the results from Experiment 3 to validate the reliability of the findings. I was successfully able to reproduce the taboo Stroop effect seen in Experiment 3. Looking at the results, both the single-item and the multiple-item variant produced a taboo Stroop effect, although the effect was stronger in the multiple-item trial-type. Thus the Multiple-Item Stroop task was more sensitive in capturing this difference within the taboo words.

Additionally, there was a negative emotional Stroop effect observed using the multiple-item trial-type, while no such effect was seen using the single-item variant. This was not observed in Experiment 3, although that might have been due to smaller sample size. An important theoretical consideration is that those same words (both positive and negative) were unable to produce an effect in the single-item condition but were able to reveal a significant effect using the Multiple-Item Stroop task. Once again, these results further show that the Multiple-item Stroop task is more sensitive in capturing even minute attentional biases if present.

10.3 Relations with Independent Measures of Anxiety

10.3.1 Results

As described in the Materials and Apparatus section above, the current experiment also included a number of other measures that could theoretically reflect the extent to which participants in my experiment were anxious about seeing the critical taboo stimuli. If this anxiety is what is feeding the attentional biases, and if the Multiple-Item Stroop task is more sensitive to these biases than is the single-item Stroop task, then I would expect stronger correlations with the size of Stroop

effects in the multiple versus single-item paradigm. This section is intended to assess whether this is indeed the case. To properly frame these analyses I begin by being clear about the theoretical assumptions being made with respect to each of the measures collected. In addition, I focus only on the taboo Stroop effect in both single-item and multiple-item variants. Of the total 63 participants in the entire experiment, measures of anxiety were recorded from only 51 participants due to technical issues.

First, a smartwatch was used to collect three physiological measures as participants completed the experiment; heart rate, body temperature and galvanic skin response. I collected these physiological measures at resting state (prior to the start of the experiment) and while the participants were performing the Stroop experiment; and subsequently calculated the difference between their baseline and their experimental states. The theoretical assumption related to the physiological measures was as follows. To the extent that participants find the taboo words particularly anxiety provoking, their physiological measures should increase relative to the baseline measure. The participants that show the largest increases should also show the largest taboo Stroop effects assuming that anxiety translates into the sort of attentional bias highlighted throughout this thesis. Thus, positive correlations are expected, and if the multiple-item paradigm provides more accurate and precise measures of attentional bias, one would expect the strongest correlations with the Stroop effect observed in that condition.

It is important to note that these technologies are still relatively new, and were designed for recreational rather than research use. It is unclear how precise and time-locked the measurements really are. In fact, to anticipate, I was not able to get an accurate heart rate measure - and thus will not discuss it further - and the measures I was able to get did not succeed in predicting anything. So while the promise of these technologies is high, and while it was certainly worth trying, these measures will not add much to my assessment of my proposed new paradigm.

Second, I administered the State-Trait Anxiety Inventory, an established method of obtaining a subjective measure of anxiety (Spielberg et al., 1970). The two part questionnaire gauges both the state and trait anxiety levels to determine how one is feeling at the current moment and how stable their personality disposition is. The STAI has normative data and has been used reliably for several years (Bados et al., 2010). Theoretically, the presence of taboo words would cause participants to feel anxious, which has been demonstrated in previous studies, and I anticipated that the State component of the STAI would capture that the experiment produced anxiety in participants (MacKay & Ahmetzanov, 2005; Siegrist, 1995). In addition, participants who are more generally anxious (i.e., a higher Trait component of the STAI) may be more prone to anxiety in general and thus should find exposure to taboo words more anxiety producing. As such, my hope was to see a positive correlation between participants who scored high on both components of the STAI and the taboo Stroop effect, particularly on the multiple-item variant.

Third, I created the Taboo Questionnaire to ascertain the use of taboo words as well as the comfort level with taboo words as provided by the participants. This was specifically designed to discover and to gauge how familiar and comfortable the participants were with these words, as we know these can be factors in attentional capture and bias (Yiend et al., 2005). Again, I anticipated a positive correlation between those participants who were uncomfortable with the use of these words and their respective taboo Stroop effect. On the contrary, participants who were relatively comfortable with taboo words should not produce a large taboo Stroop effect in comparison. Both these correlations were predicted to be stronger in the multiple-item variant than in the single-item variant.

With the relevant theoretical assumptions now presented, Table 5 displays the observed correlations between the measures just described, and the magnitude of the taboo Stroop effects observed in the single versus multiple-item Stroop tasks. As is apparent from the table, the physiological markers displayed very weak correlation with the effects seen in the single-item and multiple-item variants of the Stroop task. Both the galvanic skin response (GSR) and the body temperature are less than 0.1, and in the multiple-item Stroop are in the opposite direction of what I predicted. Furthermore, even the two physiological measures (GSR and body

temperature) that should have been strongly inter-correlated displayed a near zero correlation (Pearson coefficient = 0.001). This further suggests a high degree of imprecision in the measurements, and as such, further analysis using a regression analysis was not conducted.

How does the multiple-item Stroop task compare to the single-item variant in terms of being related to the other measures of anxiety I collected? The first thing to note about the data is that only two correlations reached conventional levels of significance, one connected to the multiple-item variant and another connected to the single-item variant. By that method of scoring, it seems to be a tie. However, the fact that only two correlations were significant suggests that this experiment does not have the level of statistical power required to perform this sort of analysis with confidence.

That said, if one looks at the patterns of correlations, the multiple-item Stroop task certainly seems to be providing a more consistent set of data. Although only one correlation, the correlation with Trait anxiety, was significant, all three of the correlations are positive in direction, and the two correlations with the STAI, the most established of our measures of anxiety, are greater than 0.20. Thus the story it tells is sensible overall: The higher the levels of anxiety the larger the observed taboo Stroop effect, and this is especially true for the components of the STAI.

Things are not so consistent when one considers the single-item variant. Yes, the effects observed on the single-item Stroop do positively correlate with the measure of comfort with taboo words derived from the questionnaire designed for this thesis. That makes sense, and it is reliable. However, the correlations with both components of the STAI are negative in direction. Thus the higher one's State or Trait score, the smaller the effect observed on the single-item Stroop. Again, these correlations are not significant so one cannot read too much into the negative pattern, but it certainly does not seem as consistent as the pattern observed in the multiple-item variant of the task.

Table 5

Pearson correlations quantifying the relation between the single and multiple item Stroop effects and the various predictors of anxiety as measured in Experiment 4

	GSR	Temp	State	Trait	Taboo Questionnaire
Single-Item Stroop	.09	.01	-.11	-.12	.30*
Multiple-Item Stroop	-.09	-.04	.22	.29*	.12

* $p < .05$ two-tailed

10.3.1.1 Correlational Analysis of the Survey Data

For the sake of completeness, and to ensure outliers are not colouring the observed statistics too dramatically, I also provide scatterplots of the three relevant predictors to visualize the results. The relation between the State measure of anxiety and the two observed taboo Stroop effects is depicted in Figure 9. As hoped, the relation was stronger in the multiple-item condition than in the single-item condition, even though neither were statistically reliable at this level of power. Nonetheless the direction of the effect within the correlational data is consistent with the notion that the multiple-item version provides a better measure of attentional bias.

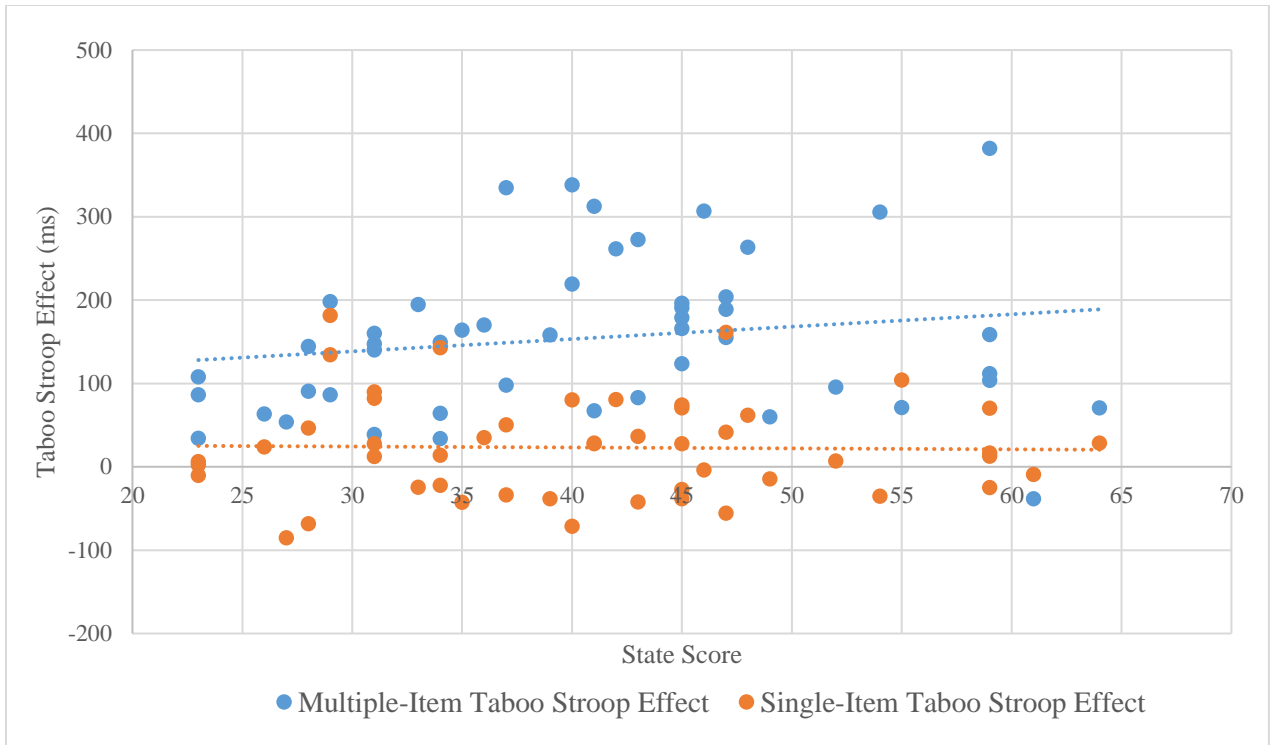


Figure 9. Experiment 4: The taboo Stroop effect for multiple-item and single-item in relation to the State Score in the STAI.

In terms of the trait anxiety data, a similar scatterplot was created. Looking at Figure 10 below, as the trait anxiety increases, there is a larger taboo Stroop effect within the multiple-item trial-type than when trait anxiety is low. This correlation reached significance, with a Pearson correlation coefficient of 0.29 (see Table 5). In the single-item variant however, the data are a lot more diffuse and again there were some negative values, alluding to more fluctuations in the data.

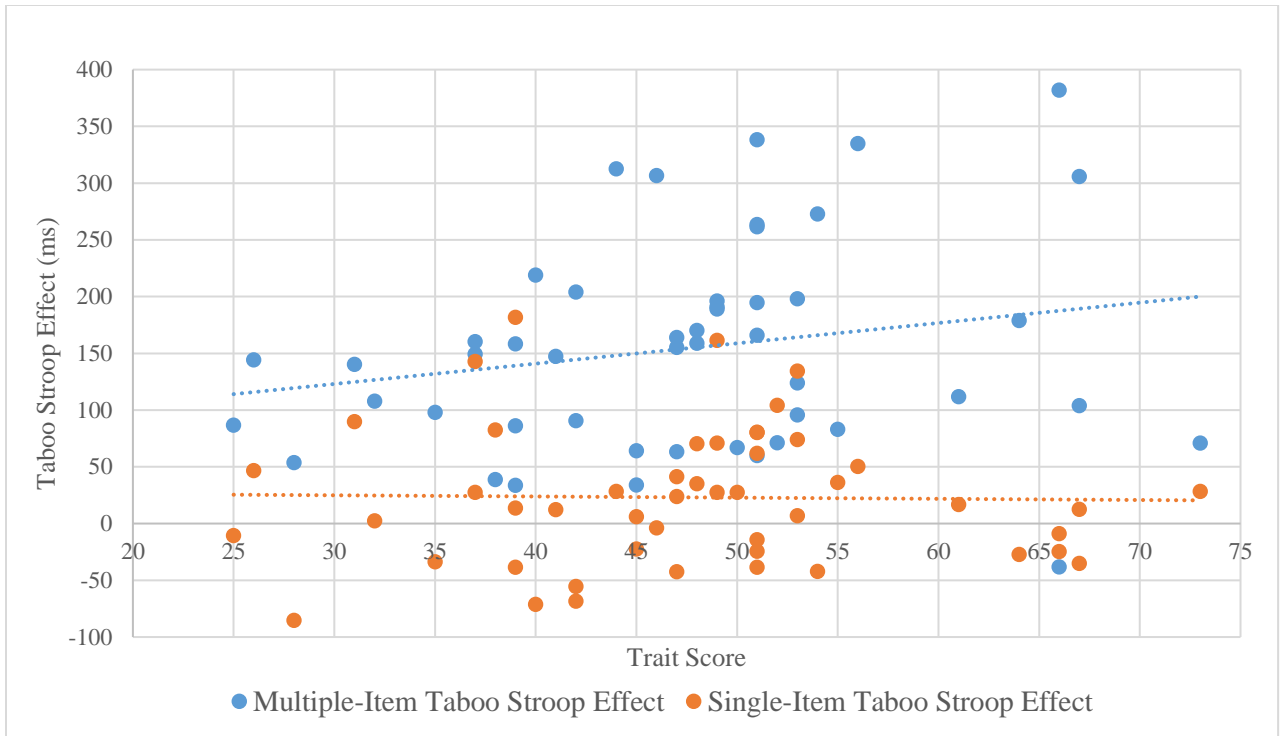


Figure 10. Experiment 4: The taboo Stroop effect for multiple-item and single-item in relation to the Trait Score in the STAI.

The higher scores on the Taboo Questionnaire, the more uncomfortable the participant was with the use of the taboo words. Looking at Figure 11 below, as the score on the Taboo Questionnaire increases (meaning participants who are quite uncomfortable with taboo words), there is also an increase in the taboo Stroop effect compared with those who have a low taboo questionnaire scores in the multiple-item variant. A similar trend is visible in the single-item variant, where the correlation reached significance ($r = 0.30$).

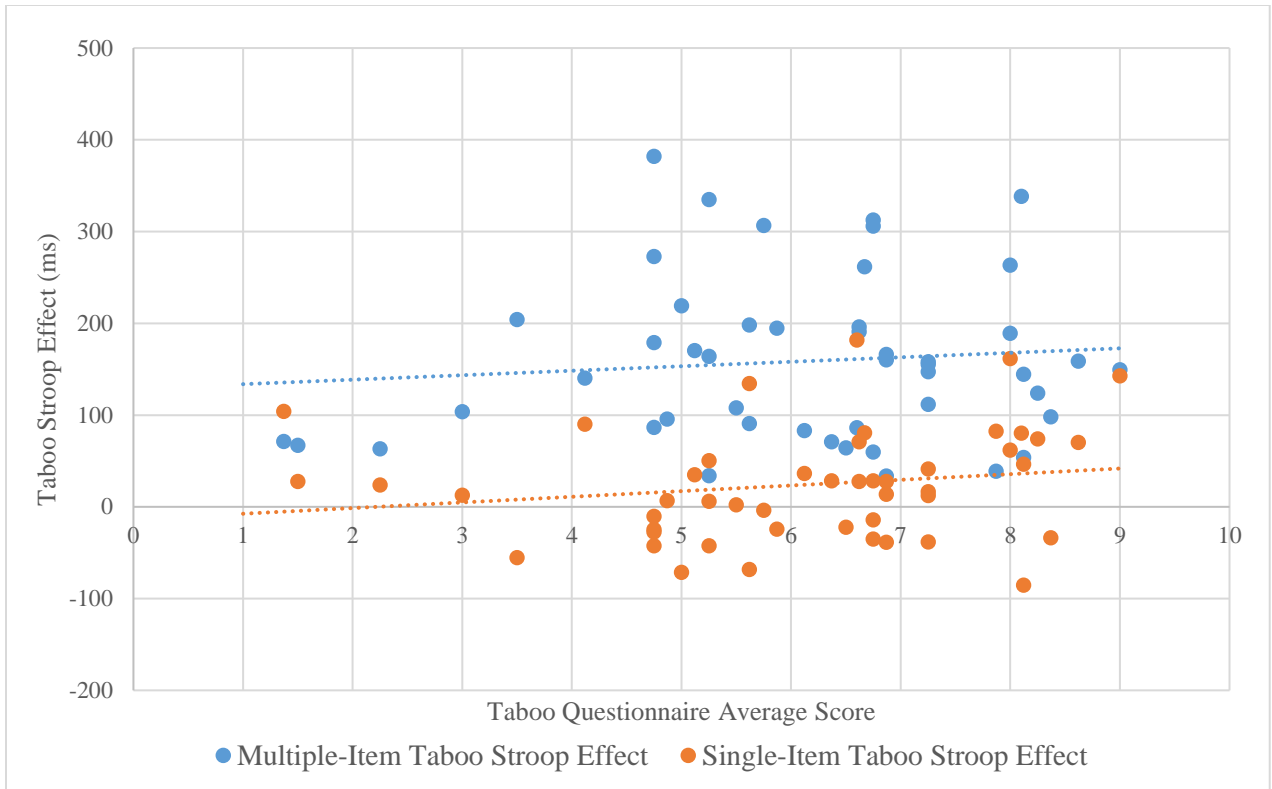


Figure 11. Experiment 4: The taboo Stroop effect for multiple-item and single-item in relation to the Taboo Questionnaire score.

10.3.2 Discussion

In addition to replication, Experiment 4 attempted to collect independent measures of anxiety to potentially correlate them with the taboo Stroop effect. Although not all the correlations reached significance, there was an adequate trend that demonstrated positive correlations with the measures of anxiety within the Multiple-Item Stroop task. It could have been that low power was responsible for the lack of significance. There were some correlations present in the single-item task, however there were far more fluctuations in the data that created an equivocal account. In addition, the physiological measures did not provide any further information.

11 General Discussion

This chapter demonstrated and replicated the Multiple-Item Stroop task's ability to capture the taboo Stroop effect, and also provided some evidence of its ability to capture attentional biases as suggested by independent measures. Specifically, the use of taboo words provided us with a special set of stimuli that would grab attention in an undergraduate participant pool. Experiment 3 revealed the taboo Stroop effect existed within this population and that the multiple-item trial-type provides a powerful tool to capture it. In addition, Experiment 4 replicated the results of Experiment 3 along with obtaining subjective and objective measures of anxiety that were then correlated with the occurrence of the taboo Stroop effect.

Using a large number of stimuli as well as a broad range of stimulus sets, Experiment 3 successfully demonstrated the taboo Stroop effect, both in single-item and multiple-item variants. Considerably though, the multiple-item taboo Stroop effect was superior as evidenced by the large effect size calculated. Although the words utilized were exactly the same, the effect was much more prominent in the multiple-item variant.

Experiment 4 was conducted to ensure the reliability of the results of Experiment 3. As well, physiological measures as well as subjective measures were recorded to determine whether they might predict the magnitude of the taboo Stroop effect observed. The pattern of results displayed that participants who had high state and trait anxiety and/or were relatively uncomfortable with the taboo words generally did produce a larger taboo Stroop effect within the multiple-item trial-type. Even though subjectively participants felt no different regarding the taboo words, the Multiple-Item Stroop task was able to produce large changes in response time. This trend was not seen in the single-item variant, even though the same stimuli were used on the same participants. This precision is what set the Multiple-Item Stroop paradigm apart.

The use of taboo words played two separate roles. First, taboo words provided us with stimuli that had a shock factor associated with them, to be able to use and obtain results within a normal

undergraduate population. The second and the more powerful role taboo words played was to provide a very fitting parallel to disorder-related words.

Taboo words have been known to elicit many of the same the physiological reactions as well as the emotional responses that have been observed using disorder-related words within the patient population (Foa & McNally, 1986; Geer et al., 1994; Goldstein, 2008; Labar & Phelps, 1998). In one study, healthy participants were unable to ignore the distracting taboo word in a rapid serial visual presentation (RSVP) stream when the task was to report a single target (Arnell et al., 2007). When the critical distractor in the RSVP stream was a sexual word, target accuracy decreased substantially, signifying that the taboo words were capturing attention even though it was at the expense of task performance. Furthermore, Barnard, Ramponi, and Battye (2005) conducted a similar study using threat-related words in high state anxiety population in the RSVP stream. Again, similar results were found showing that depending on the level of state anxiety, the threat-related distracting words were more disruptive. Thus these two studies can allow us to create a parallel between the effect that taboo words have on healthy undergraduates and that threat-related words have on highly anxious individuals.

Chapter 4 Implications and Future Endeavors

Throughout this thesis, I have emphasized the critical nature of attentional biases within both healthy individuals as well as clinical populations. Certain populations have a predisposition to have their attention captured by specific stimuli, which then can further reinforce or prime problematic cognitive states, creating a vicious circle. A great method would be to determine ways to detect and quantify the extent to which the problematic biases is occurring. The Multiple-Item Stroop task, the focus of this thesis, was created with this goal in mind.

Selective attention plays a very critical role in what is chosen for further cognitive processing, and a bias within it has tremendous repercussions on how information is selected, interpreted, as well as recalled. The Stroop task has played a vital role in selective attention research as the effect has been shown to be quite resilient to change. Basically, the Stroop effect is said to be due to response competition at the output stage of the information-processing stream.

As an extension, the emotional Stroop task was devised to determine attentional biases for emotional and/or disorder-related words. Any effect seen is not due to response competition as with the traditional effect, but rather is due to the meaning of the emotional or disorder-related word causing interference with colour naming. Previous researchers had attempted to adapt the emotional Stroop effect for this purpose, but the resulting paradigm has been shown to produce small effects and to generally be unreliable (Eide et al., 2002; MacKay et al., 2004; MacLeod, C. M. & Hodder, 1998b).

The Multiple-Item Stroop task provides a new paradigm that is based on the logic of the emotional Stroop effect, but one whose features were designed to amplify any effects that might be present by requiring the repeated engagement and disengagement of attention across multiple stimulus items. Given the changes to the presentation of the stimuli, I first confirmed that the traditional Stroop effect was present within the Multiple-Item Stroop task (Experiment 1). Subsequently, I tested the new paradigm with emotional words, and although it was not

successful when I used a small set of emotional-related items (Experiment 2), when I expanded my stimulus set and included the more potent taboo words I was able to demonstrate (Experiment 3) and replicate (Experiment 4) the finding of a more sensitive taboo Stroop effect within the new multiple-item Stroop paradigm.

In addition, when I assessed the relation between the effects I observed and independent measures that should be related to attentional capture (also Experiment 4), an acceptable pattern of relationship was observed in my new multiple-item paradigm, at least relative to the traditional single-item variant. Taken together, these empirical results arrived at via research with non-clinical undergraduate students support the notion that attentional biases exist and do disrupt selective attention tasks.

In the remainder of this chapter, I raise and discuss a range of issues that this research informs or provokes. First I elaborate on the distinction between the theoretical underpinning of the traditional Stroop effect and the emotional/taboo Stroop effect, a contrast that has been supported by my findings. Subsequently, I highlight the somewhat unique manner in which I used nonclinical undergraduate population as this may be an approach that has more general usage. Finally, I lay the groundwork for the next steps that should follow this thesis to eventually maximize the potential of the Multiple-Item Stroop task.

12 There is Stroop, and Then There is Stroop

The classic Stroop effect has been relatively unchanged since its inception back in 1935. Its strength and endurance was one of the main reasons I chose to use this task for modifications. However, before using the Multiple-Item Stroop task with different stimuli and populations, I needed to determine whether the effect was still present, which was the primary aim of Experiment 1. If the Stroop effect was not present with the modifications, then it would have been futile to move forward with different stimuli or population.

The results from Experiment 1 demonstrated two important findings. First, the Stroop effect was present with the modifications, thus I was able to move forward as hoped. And second, the effect was similar in magnitude to the original Stroop effect, a finding which, in and of itself, has relevant theoretical implications. Since the traditional effect is assumed to arise as a result of response competition at the output stage, this suggests that exposure to a single colour-related word is sufficient to produce the maximal effect. This finding is consistent with the notion that word reading is automatic and thus, can be triggered to the point of generating a complete phonological ‘competitor’ consistently with a single exposure (Beck, 2008; Cattell, 1886; Price et al., 2012).

In contrast, when taboo words were employed there was a stark difference in the Stroop effect seen, where the effect did get larger with the use of multiple stimuli. This finding highlights the theoretical distinction between traditional and emotional Stroop task. Of course, I believe that difference is with respect to what causes the effect itself. It is not the result of response competition but, rather, is the result of attention becoming hijacked from the primary task via the presence of emotionally powerful words. This is consistent with my findings.

Even though the standard Stroop effect was evident using both the single-item and multiple-item variants, the taboo Stroop was observable only using the Multiple-Item Stroop task. The question arises as to why this would be the case: Why is it that the standard Stroop effect does not benefit from the multiple stimulus array whereas the taboo Stroop effect does? The standard Stroop effect is a result of response competition at the output stage, and as such increasing the stimuli does not affect its strength. Whereas for the Multiple-Item Stroop task, I believe it is attentional capture due to salient and powerful emotional stimuli, where repeated engagement and disengagement of attention is occurring. To elaborate, the Multiple-Item Stroop task forces the participants to scan the entire screen to determine the dominant ink colour, but in doing so participants are also forced to repeatedly encounter the critical stimulus. If that critical stimulus is a word that is deeply emotionally arousing and/or anxiety-provoking, it creates a significant

distraction to the primary task, thus causing the increased latencies to the critical stimulus and producing the interference.

Although I discuss the emotional Stroop and the taboo Stroop separately, they are in fact likely related with respect to what causes them: attentional capture. The distinction is likely more related to just how emotional or arousing they are, with taboo words essentially representing the most emotional words possible, at least from the perspective of an undergraduate population. While we use the label ‘Stroop effect’ for both the traditional and the emotional/taboo Stroop effects, they do in fact arise from different underlying causes (Algom et al., 2004). Thus one contribution of this work is that it both confirms and adds to current views of both the traditional and emotional Stroop effects. They do seem to arise from very different causes, and in the case of the traditional Stroop, a single presentation is sufficient to produce the full blown effect.

13 Modelling Attentional Biases within an Undergraduate Populations

Obtaining genuine attentional biases within a laboratory setting is rather difficult, especially within an undergraduate population. As discussed within the Introduction, attentional biases are most clearly documented within clinical populations. Although I did not conduct my experiments within clinical populations, I wanted to model the attentional biases observed in mental disorders with undergraduates, to best emulate the extent of attentional capture due to biases present. Thus, an interesting and unique aspect of this thesis was the way in which I used healthy undergraduate participants to modelled attentional biases. In most cases in the literature, mental disorders are studied in the traditional sense by finding a clinical population to test. Or a subclinical group is sought within the undergraduate community for a particular disorder. However, the method I employed was different from both these regular forms.

In clinical research, the most common form within the literature is to find a population that is currently experiencing a particular disorder. For example, Burgess et al. (1981) used phobic

patients and non-phobic individuals to determine the extent to which specific words caused higher interference in terms of response latencies. Similarly, randomized control trials are used to determine the effectiveness of particular interventions in patient populations (Lam et al., 2003). Thus, one main method of studying a specific population is to use the actual patient populations that one is attempting to target.

However, using an actual patient population in a clinical setting is rather non-optimal with respect to attempting to develop and support a new tool. That is, it can a long duration to develop a tool, where different modifications and versions may not work out immediately. Patients/clients come to a therapist looking for assistance, and sometimes are already experiencing a lot of stress and emotional unrest as a result of their disorders. Subjecting them to research that may or may not work could amount to adding an additional stressor to an already stressful life. It would be far better to perform initial research validating the potential of a tool outside of the clinic.

Another manner in which populations are emulated is to segregate undergraduate participants that show some specific symptoms in a subclinical manner. For example, one could give undergraduates the Beck's Depression Inventory to determine whether some individuals display depressive symptoms, but do not reach clinical threshold levels (Beck, 2008). In other cases, experimenters administer surveys or questionnaires to divide undergraduates into a 'high' or 'low' group in terms of symptoms (MacLeod, C. & Mathews, 1988; Richard & French, 1992). In addition, experimenters also try to induce specific states in individuals, such as happiness or anxiety, to replicate patient response patterns (see Wilson et al., 2006, in Chapter 1 for detailed description).

The approach taken here diverged from both of the above. Attentional biases exist within healthy individuals, but are observed more readily within clinical population, due to their symptomology. Given that the goal was to measure the extent of bias in a general sense, I considered contexts in

which normally functioning undergraduates might display the same kind of bias. That is, rather than searching for some non-typical person (i.e., someone with sub-clinical symptoms), I searched for a non-typical situation in which a normally functioning person might show the same sort of heightened problematic cognitive process. The first attempt at such a context, showing emotional words in a laboratory setting, was not sufficient. However, the move to taboo words seemed to work well. This approach is consistent with the general notion that the cognitive issues often associated with disorders are not in and of themselves “odd” but rather reflect what is normal cognitive functioning going awry. Subsequently, I compared the effectiveness of my new paradigm to the traditional one to measure that issue.

14 Next Steps

I have highlighted and used anxiety disorders throughout this thesis because I am genuinely interested in them. Additionally, they fit nicely with the paradigm as they have similar theoretical underpinnings. However, the Multiple-Item Stroop task is a paradigm that can be applied to almost any type of bias. As the stimuli used within it are malleable, it can be tailored to fit different types of attentional biases (e.g., test anxiety, food aversion/compulsion).

There is precedence for the development of psychological tests in this manner. Many of the standard neuropsychological tests employed within patient population have origins within the undergraduate population or at least within research labs. For example, the use of eye-tracking was originally developed to investigate the movement of eyes during reading and picture viewing to infer the underlying cognitive processes (Rayner, 1998). As such, the use of eye-tracking has moved into the clinical realm, where it has been used to investigate the cognitive processes, or the disruption of specific cognitive processes, within specific disorders. Thus, as attentional biases clearly play a role in some mental disorders, the use of the Multiple-Item Stroop task could be beneficial within the clinical realm, as an objective diagnostic tool or outcome measure.

In terms of my thesis, the next steps that I would consider to follow these experiments would be to ensure the reliability and validity of this new paradigm. Validity and reliability are critical component of any new measure, which need to be properly documented and achieved (Eignor, 2013; Joint American Educational Research Association, 1999). In Experiment 4, I attempted to validate the Multiple-Item Stroop task by correlating some subjective measures of anxiety. Unfortunately, due to lack of statistical power I was unable to add to the assessment of my paradigm. Additionally, reliability is crucial for any paradigm to be used in research and applied settings. As such, I would like to ensure the precision the Multiple-Item Stroop task has in capturing attentional biases.

Additionally, I would be greatly interested to test within clinical populations, as attentional biases are most clearly observed within them. Specifically, I would carry forward the use of anxiety disorder and would attempt to replicate Experiment 3 in an anxiety patient population, with threat-related words specific to the patients. I would want to confirm my belief that there would be much higher interference within the multiple-item trials than the single-item trials. Subsequently, to validate it further, I would also want to do a study where the participants are a mix of clinically diagnosed patients and healthy controls, to which I would be blind. I would want to see if the results of the Multiple-Item Stroop design would be able to differentiate between the two populations in a significant manner. Eventually, I would want to create norms for different population, such that there would be a distribution of reaction times that would be appropriate for a variety of populations and age groups.

References

- Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional stroop phenomenon: A generic slowdown, not a stroop effect. *Journal of Experimental Psychology: General*, *133*(3), 323-338.
- Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology: General*, *134*(5), 258-281.
- Arnell, K. M., Killman, K. V., & Fijavz, D. (2007). Blinded by emotion: Target misses follow attention capture by arousing distractors in RSVP. *Emotion*, *7*(3), 465.
- Augustinova, M., & Ferrand, L. (2007). Influence de la présentation bicolore des mots sur l'effet stroop [First-letter coloring and the Stroop effect]. *L'Année Psychologique*, *107*(2), 163-179.
- Augustinova, M., Flaudias, V., & Ferrand, L. (2010). Single-letter coloring and spatial cuing do not eliminate or reduce a semantic contribution to the Stroop effect. *Psychonomic Bulletin & Review*, *17*(6), 827-833.
- Bados, A., Gómez-Benito, J., & Balaguer, G. (2010). The state-trait anxiety inventory, trait version: Does it really measure anxiety? *Journal of personality assessment*, *92*(6), 560-567.
- Ball, R. J., Mitchell, P. B., Touyz, S. W., Griffiths, R. A., & Beumont, P. J. V. (2004). Clinical utility of the modified Stroop task as a treatment outcome measure: Questions raised. *Clinical Psychologist*, *8*(2), 76-80.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283.
- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1-24.
- Barnard, P., Ramponi, C., Battye, G., & Mackintosh, B. (2005). Anxiety and the deployment of visual attention over time. *Visual Cognition*, *12*(1), 181-211.
- Basis Peak. (2017). An Intel Company. Smartwatch technology. Retrieved from <https://www.mybasis.com/>

- Beck, A. T. (1976). *Cognitive therapy and the emotional disorder*. New York: International Universities Press.
- Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *The American Journal of Psychiatry*, *165*(8), 969-977.
- Besner, D., & Stolz, J. (1999). Context dependency in Stroop's Paradigm: When are words treated as nonlinguistic objects? *Canadian Journal of Experimental Psychology*, *53* (4), 374-380.
- Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, *4*(2), 221-225.
- Bibi, U., Tzelgov, J., & Henik, A. (2000). Stroop effect in words that differ from color words in one letter only. *Psychonomic Bulletin & Review*, *7*(4), 678-683.
- Borella, E., De Ribaupierre, A., Cornoldi, C., & Chicherio, C. (2013). Beyond interference control impairment in ADHD: Evidence from increased intraindividual variability in the color-stroop test. *Child Neuropsychology*, *19*(5), 495-515.
- Bradley, B. P., Mogg, K., & Millar, N. H. (2000). Covert and overt orienting of attention to emotional faces in anxiety. *Cognition and Emotion*, *14*(6), 789-808.
- Bradley, B. P., Mogg, K., Falla, S. J., & Hamilton, L. R. (1998). Attentional bias for threatening facial expressions in anxiety: Manipulation of stimulus duration. *Cognition and Emotion*, *12*(6), 737-753.
- Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW)*. Gainesville, FL: University of Florida, NIMH Center for the Study of Emotion and Attention.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433-436.
- Broadbent, D. (1958). *Perception and Communication*. London: Pergamon Press.
- Brown, T. L., Joneleit, K., Robinson, C. S., & Brown, C. R. (2002). Automaticity in reading and the Stroop task: Testing the limits of involuntary word processing. *The American Journal of Psychology*, *115*(4), 515-543.
- Brown, W. (1915). Practice in associating color names with colors. *Psychological Review*, *22*, 45-55.
- Buodo, G., Sarlo, M., & Palomba, D. (2002). Attentional resources measured by reaction times highlight differences within pleasant and unpleasant, high arousing stimuli. *Motivation and Emotion*, *26*, 123-138.

- Burgess, I. S., Jones, L. M., Robertson, S. A., Radcliffe, W. N., & Emerson, E. (1981). The degree of control exerted by phobic and non-phobic verbal stimuli over the recognition behaviour of phobic and non-phobic subjects. *Behaviour Research and Therapy*, *19*(3), 233-243.
- Burt, J. S. (2002). Why do non-color words interfere with color naming? *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1019-1038.
- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63-65.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975-979.
- Chmiel, N. (1984). Phonological receding for reading: The effect of concurrent articulation in a Stroop task. *British Journal of Psychology*, *75*, 213-220.
- Clark, D. A., & Beck, A. T. (2010). Cognitive theory and therapy of anxiety and depression: Convergence with neurobiological findings. *Trends in Cognitive Sciences*, *14*(9), 418-424.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing.
- Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory and Cognition*, *2*, 130-138.
- Cowan, N. (2005). Selective attention tasks in cognitive research. In A. Wenzel, & D. C. Rubin (Eds.), *Cognitive Methods and their Application to Clinical Research* (pp. 73-97). Washington, D. C.: American Psychological Association.
- Davison, G. C., Neale, J. M., Blankstein, K. R., & Flett, G. L. (2005). *Abnormal Psychology, Second Canadian Edition*. Ontario, Canada: John Wiley & Sons Canada, Ltd.
- Denny, E. B., & Hunt, R. R. (1992). Affective valence and memory in depression: Dissociation of recall and fragment completion. *Journal of Abnormal Psychology*, *101*(3), 575-580.
- Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science*, *14*(1), 14-18.
- Dolan, R. J. (2002). Emotion, cognition, and Behavior. *Science*, *298*, 1191-1194.
- Doyen, S., Klein, O., Simons, D. J., & Cleeremans, A. (2014). On the other side of the mirror: Priming in cognitive and social psychology. *Social Cognition*, *32*, 12-32.

- Dresler, T., Ehlis, A., Hindi Attar, C., Ernst, L. H., Tupak, S. V., Hahn, T., . . . Fallgatter, A. J. (2012). Reliability of the emotional Stroop task: An investigation of patients with panic disorder. *Journal of Psychiatric Research, 46*(9), 1243-1248.
- Dresler, T., Meriau, K., Heekeren, H. R., & van der Meer, E. (2009). Emotional Stroop task: Effect of word arousal and subject anxiety on emotional interference. *Psychological Research, 73*, 364-371.
- Dugas, M. J., Ladouceur, R., Léger, E., Freeston, M. H., Langolis, F., Provencher, M. D., & Boisvert, J. (2003). Group cognitive-behavioral therapy for generalized anxiety disorder: Treatment outcome and long-term follow-up. *Journal of Consulting and Clinical Psychology, 71*(4), 821-825.
- Dyer, F. N., & Severance, L. J. (1973). Stroop interference with successive presentation of separate incongruent words and colors. *Journal of Experimental Psychology, 98*(2), 438-439.
- Eich, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory & Cognition, 12*(2), 105-111.
- Eide, P., Kemp, A., Silberstein, R. B., Nathan, P. J., & Stough, C. (2002). Test-retest reliability of the emotional Stroop task: Examining the paradox of measurement change. *The Journal of Psychology, 136*(5), 514-520.
- Eignor, D. R. (2013). The standards for educational and psychological testing.
- Eilola, T. M., & Havelka, J. (2011). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism, 15*(3), 353-369.
- Esposito, A. G., Baker-Ward, L., & Mueller, S. T. (2013). Interference suppression vs. response inhibition: An explanation for the absence of a bilingual advantage in preschooler's Stroop task performance. *Cognitive Development, 28*, 354-363.
- Esposito, N. J., & Pelton, L. H. (1971). Review of the measurement of semantic satiation. *Psychological Bulletin, 75*(5), 330-346.
- Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words is categorical and general. *Emotion, 8*(4), 453-457.
- Eysenck, M. W., MacLeod, C., & Mathews, A. (1987). Cognitive function and anxiety. *Psychological Research, 49*, 189-195.

- Eysenck, M. W., Mogg, K., May, J., Richards, A., & Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of Abnormal Psychology, 100*(2), 144-150.
- Foa, E. B., & McNally, R. J. (1986). Sensitivity to feared stimuli in obsessive-compulsives: A dichotic listening analysis. *Cognitive Therapy Research, 10*, 477-486.
- Foa, E. B., Feske, U., Murdock, T. B., Kozak, M. J., & McCarthy, P. R. (1991). Processing of threat-related information in rape victims. *Journal of Abnormal Psychology, 100*(2), 156-162.
- Frewen, P. A., Dozois, D. J. A., & Lanius, R. A. (2008). Neuroimaging studies of psychological interventions for mood and anxiety disorders: Empirical and methodological review. *Clinical Psychology Review, 28*(2), 228-246.
- Frings, C., Englert, J., Wentura, D., & Bermeitinger, C. (2010). Decomposing the emotional Stroop effect. *The Quarterly Journal of Experimental Psychology, 63*(1), 42-49.
- Galotti, K. M. (2004). *Cognitive Psychology: In and out of the laboratory, third edition*. California, USA: Thomson Wadsworth.
- Geer, J. H., Judice, S., & Jackson, S. (1994). Reading times for erotic material: The pause to reflect. *The Journal of General Psychology, 121*(4), 345-352.
- Goldstein, E. B. (2008). *Cognitive Psychology: Connecting minds, research, and everyday experience, second edition*. California, USA: Thomson Wadsworth.
- Gray, J. A., & Wedderburn, A. I. (1960). Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 12*, 180-184.
- Groeger, J. A. (1997). *Memory and remembering: Everyday memory in context*. Harlow, UK: Longman.
- Guillet, R., & Arndt, J. (2009). Taboo words: The effect of emotion on memory for peripheral information. *Memory & Cognition, 37*(6), 866-879.
- Hamers, J. F., & Lambert, W. E. (1972). Bilingual interdependences in auditory perception. *Journal of Verbal Learning and Verbal Behavior, 11*, 303-310.
- Henik, A., Friedrich, F. J., & Kellogg, W. A. (1983). The dependence of semantic relatedness effect upon prime processing. *Memory & Cognition, 11*(4), 366-373.
- Hirsch, C. R., & Mathews, A. (2000). Impaired positive inferential bias in social phobia. *Journal of Abnormal Psychology, 109*(4), 705-712.

- Hirsch, C. R., MacLeod, C., Mathews, A., Sander, O., Siyani, A., & Hayes, A. (2011). The contribution of attentional bias to worry: Distinguishing the role of selective engagement and disengagement. *Journal of Anxiety Disorders, 25*, 272-277.
- Hollingworth, H. L. (1915). Articulation and association. *Journal of Educational Psychology, 6*, 99-105.
- James, W. (1892). *The stream of consciousness*. Psychology.
- Jay, T., Caldwell-Harris, C., & King, K. (2008). Recalling taboo and nontaboo words. *The American journal of psychology, 121*(1) 83-103.
- Joint American Educational Research Association. (1999). *The standards for educational and psychological testing*. American Educational Research Association, Washington, DC.
- Jovina, I., & Taatgen, N. A. (2009). A repetition-suppression account of between-trial effects in a modified Stroop paradigm. *Acta Psychologica, 131*, 72-84.
- Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: Effect of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience, 6*(2), 110-126.
- Kindt, M., Bierman, D., & Brosschot, J. F. (1996). Stroop versus Stroop: Comparison of a card format and a single-trial format of the standard color-word Stroop task and the emotional Stroop task. *Personality and Individual Differences, 21*(5), 653-661.
- Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition & Emotion, 16*(1), 29-59.
- Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology, 77*, 576-588.
- Kousta, S., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*, 473-481.
- Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science, 9*(6), 490-493.
- Lam, D. H., Watkins, E. R., Hayward, P., Bright, J., Wright, K., Kerr, N., . . . Sham, P. (2003). A randomized controlled study of cognitive therapy for relapse prevention for bipolar affective disorder: Outcome of the first year. *Archives of General Psychiatry, 60*(2), 145-152.

- Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion, 6*(1), 62-72.
- Larsen, R. J., Mercer, K. A., & Balota, D. A., Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion, 8*(4), 445-452.
- Lee, T. M., & Chan, C. C. (2000). Stroop interference in Chinese and English. *Journal of Clinical and Experimental Neuropsychology, 22*(4), 465-471.
- Lowe, D. G. (1985). Further investigations of inhibitory mechanisms in attention. *Memory and Cognition, 13*, 74-80.
- Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology, 36*, 684-700.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*(2), 203-208.
- Lutfi-Proctor, D. A., Elliot, E. M., & Cowan, N. (2014). The role of visual stimuli in cross-modal Stroop interference. *PsyCh Journal, 3*, 17-29.
- MacKay, D. G., & Ahmetzanov, M. V. (2005). Emotion, memory, and attention in the taboo Stroop paradigm. *Emotion, Memory and Attention, 16*(1), 25-32.
- MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyers, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory task. *Memory & Cognition, 32*(3), 474-488.
- MacLeod, C. (2005) The Stroop task in clinical research. In A. Wenzel, & D. C. Rubin (Eds.), *Cognitive methods and their application to clinical research* (pp. 41-62). Washington, D. C.: American Psychological Association.
- MacLeod, C., & Mathews, A. (1988). Anxiety and the allocation of attention to threat. *The Quarterly Journal of Experimental Psychology, Section A, 40*(4), 653-670.
- MacLeod, C., & Mathews, A. (1991). Biased cognitive operations in anxiety: Accessibility of information or assignment of processing priorities? *Behavior Research Therapy, 29*(6), 599-610.
- MacLeod, C., & Mathews, A. (2004). Selective memory effect in anxiety disorders: An overview of research findings and their implications. In D. Reisberg, & P. Hertel (Eds.), *Memory and Emotion*, Oxford Scholarship Online.

- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrated review. *Psychological Bulletin, 109*(2), 163-203.
- MacLeod, C. M. (1996). How priming affects two speeded implicit tests of remembering: Naming colors versus reading words. *Consciousness and Cognition: An International Journal, 5*(1-2), 73-90.
- MacLeod, C. M. (2005). The Stroop task in cognitive research. In A. Wenzel, & D. C. Rubin (Eds.), *Cognitive methods and their application to clinical research* (pp. 17-40). Washington, D. C.: American Psychological Association.
- MacLeod, C. M., & Bors, D. A. (2002). Presenting two color words on a single Stroop trial: Evidence for joint influence, not capture. *Memory & Cognition, 30*(5), 789-797.
- MacLeod, C. M., & Dunbar, K. (1988). Training and stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 126-135.
- MacLeod, C. M., & Hodder, S. L. (1998a). Presenting two incongruent color words on a single trial does not alter Stroop interference. *Memory & Cognition, 26*(2), 212-219.
- MacLeod, C. M., & Hodder, S. L. (1998b). On the difficulty in obtaining priming effects in the Stroop task. *Abstracts of the Psychonomic Society, 3*, 37.
- MacLeod, C. M., & MacLeod, C. (2005). The Stroop task: Indirectly measuring concept activation. In A. Wenzel & D. C. Rubin (Eds.), *Cognitive Methods and their Application to Clinical Research* (pp. 13-16). Washington, D. C.: American Psychological Association.
- Mann, D. T. Y., & Janelle, C. M. (2012). Psychophysiology: Equipment in research and practice. *Case studies in applied psychophysiology: Neurofeedback and biofeedback treatments for advances in human performance* (pp. 257-274). Wiley-Blackwell.
- Manwell, L. A., Roberts, M. A., & Besner, D. (2004). Single letter coloring and spatial cuing eliminates a semantic contribution to the Stroop effect. *Psychonomic Bulletin & Review, 11*, 458-462.
- Martin, M. (1981). Reverse Stroop effect with concurrent tasks. *Bulletin of the Psychonomic Society, 17*(1), 8-9.
- Mathew, A., & MacLeod, C. (1985). Selective processing of threat cue in anxiety states. *Behavioral Research Therapy, 23*(5), 563-569.

- Mathew, A., & MacLeod, C. (2002). Induced processing biases have causal effect on anxiety. *Cognition and Emotion, 16*(3), 331-354.
- Mathews, A. (1990). Why worry? The cognitive function of anxiety. *Behaviour Research and Therapy, 28*, 455-468.
- Mathews, A., & MacLeod, C. (1994). Cognitive approaches to emotion and emotional disorders. *Annual Review of Psychology, 45*, 25-50.
- Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology, 1*, 167-195.
- Mathews, A., Mogg, K., May, J., & Eysenck, M. (1989). Implicit and explicit memory bias in anxiety. *Journal of Abnormal Psychology, 98*(3), 236-240.
- Morgan, J. (2010). Autobiographical memory biases in social anxiety. *Clinical Psychology Review, 30*(3), 288-297.
- Nielsen, G. D. (1975). The locus and mechanism of the Stroop color word effect (Doctoral dissertation, University of Wisconsin—Madison, 1974). *Dissertation Abstracts International, 35*, 5672-B.
- Nielsen, S. L., & Sarason, I. G. (1981). Emotion, personality, and selective attention. *Journal of Personality and Social Psychology, 41*(5), 945.
- Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology, 30*, 953-958.
- Phaf, R. H., & Kan, K. (2007). The automaticity of emotional Stroop: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 184-199.
- Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. *Attention and performance V*, 669-682.
- Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive neuropsychology, 2*(3), 211-228.
- Powell, M., & Hemsley, D. R. (1984). Depression: A breakdown of perceptual defense? *The British Journal of Psychiatry, 145*, 358-362.
- Pratto, F. (1994). Consciousness and automatic evaluation. In P. M. Niedenthal & S. Kitayama (Eds.), *The heart's eye: Emotional influences in perception and attention* (pp. 115-143). San Diego, CA: Academic Press.
- Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380-391.

- Price, S. A., Beech, A. R., Mitchell, I. J., Humphreys, G. W. (2012). The promise and perils of the emotional Stroop task: A general review and considerations for use with forensic sample. *Journal of Sexual Aggression, 18*(3), 253-268
- Psychological Review, 82*(6), 407-428.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372-422.
- Richards, A., & French, C. C. (1992). An anxiety-related bias in semantic activation when processing threat/neutral homographs. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 45A*(3), 503-525.
- Richards, A., & Whittaker, T. M. (1990). Effects of anxiety and mood manipulation in autobiographical memory. *British Journal of Clinical Psychology, 29*(2), 145-153.
- Ridout, N., Astell, A. J., Reid, I. C., Glen, T., & O'Carroll, R. E. (2003). Memory bias for emotional facial expressions in major depression. *Cognition and Emotion, 17*(1), 101-122.
- Schimmack, U. (2005). Attentional interference effects of emotional pictures: Threat, negativity, or arousal? *Emotion, 5*(1), 55-66.
- Schmidt, S. R., & Saari, B. (2007). The emotional memory effect: Differential processing or item distinctiveness? *Memory & Cognition, 35*(8), 1905-1916.
- Schultz, L. T., & Heimberg, R. G. (2008). Attentional focus in social anxiety disorder: Potential for interactive processes. *Clinical Psychology Review, 28*(7), 1206-1221.
- Shimokido, T. (2006). An investigation of visual target degradation effect on semantic satiation. *The Japanese Journal of Cognitive Psychology, 4*(1), 25-32.
- Siegrist, M. (1995). Effects of taboo words on color-naming performance on a Stroop tests. *Perceptual and Motor Skills, 85*, 1119-1122.
- Spielberg, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *STAI manual*.
- Strauss, G. P., Allen, D. N., Jorgensen, M. L., & Cramer, S. L. (2005). Test-retest reliability of standard and emotional Stroop tasks: An investigation of color-word and picture-word versions. *Assessment, 12*(3), 330-337.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, IS*, 643-662.
- Tecce, J. J., & Happ, S. J. (1964). Effects of shock-arousal on a card-sorting test of color-word interference. *Perceptual and Motor Skills, 19*, 905-906.

- Treisman, A. (1964). Selective attention in man. *British Medical Bulletin*, 20, 12-16.
- Waechter, S., Nelson, A. L., Wright, C., Hyatt, A., & Oakman, J. (2014). Measuring attentional bias to threat: Reliability of dot probe and eye movement indices. *Cognitive Therapy and Research*, 38(3), 313-333.
- Warren, R. E. (1972). Stimulus encoding and memory. *Journal of Experimental Psychology*, 94(1), 90-100.
- Wenzel, A., & Cochran, C. K. (2006). Autobiographical memories prompted by automatic thoughts in panic disorder and social phobia. *Cognitive Behaviour Therapy*, 35(3), 129-137.
- Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, 120(1), 3-24.
- Wilson, E. J., MacLeod, C., Mathews, A., & Rutherford, E. M. (2006). The causal role of interpretive bias in anxiety reactivity. *Journal of Abnormal Child Psychology*, 115(1), 103-111.
- Yiend, J., Mathews, A., & Cowan, N. (2005). Selective attention in tasks in clinical and cognitive research. In A. Wenzel, & D. C. Rubin (Eds.), *Cognitive Methods and their Application to Clinical Research* (pp. 65-71). Washington, D. C.: American Psychological Association.

Appendices

Appendix A



UNIVERSITY of TORONTO
SCARBOROUGH

DEPARTMENT OF PSYCHOLOGY

Consent Form

Thank you for agreeing to participate in this experiment. This experiment is being conducted by Sarah Uzzaman under the supervision of Professor Joordens. This study is a variation of the classic Colour-Word Stroop tests, where different words, written in different colour ink (red, green, blue) are presented on a computer screen and your task is to identify the ink colour of the word (by way of button press). At times, you will see one stimulus on the screen, and other times you will see a stimulus, repeated all over the screen. Regardless, your task is to identify the ink the stimuli are written in. You will be asked to do this as fast as you can, without making mistakes.

Initially, there will be a practice session in order for you to get accustomed to the design of the experiment, and to remember which button is associated with each colour. The buttons will be colour coded, and there is also a colored bar on top of the screen for your ease of remembering. Again, we are looking for the speed-accuracy trade off, thus we would like you to go as fast as possible without making mistakes. The experiment will take about 30 minutes.

By signing this form you are acknowledging your understanding of the following points:

- Should you ever feel uncomfortable while doing this experiment you are free to quit at any point and without any penalty; your participation is completely voluntary.
- From this point onward your data will only be associated with a participant number thereby assuring your anonymity. Only the experimenter, supervisor and computer programmer will have access to your data.
- There are no risk associated with this experiment. Your benefit is the experience you gain in understanding the process of psychological research method.

- At the end of the experiment the basic purpose of the experiment will be explained to you, and also to answer any questions you may have.
- Your course credit will be given to you by the end of the day.

If you are interested in the outcomes of the experiment, or if you simply have any questions please do not hesitate to contact the experimenter. In addition, if you would like to get a copy of the manuscript once it is completed, please provide the experimenter with your email.

If you have any questions about your rights as a participant, you can contact the Ethics Review Office at ethics.review@utoronto.ca, or (416) 946-3273.

If this is all clear, then please print the date, your name, student number and sign below:

Date: _____

Name: _____

Student number: _____

Signature:

Experimenter:

Sarah Uzzaman,
Graduate Student,
Department of Psychology,
University of Toronto Scarborough,
sarah.uzzaman@utoronto.ca

Supervisor:

Prof. Steve Joordens,
Department of Psychology,
University of Toronto Scarborough,
joordens@utsc.utoronto.ca



UNIVERSITY of TORONTO

SCARBOROUGH

DEPARTMENT OF PSYCHOLOGY

Consent Form

Thank you for agreeing to participate in this experiment. This experiment is being conducted by Sarah Uzzaman under the supervision of Professor Joordens. This study is a variation of the classic Colour-Word Stroop tests, where different words, written in different colour ink (red, green, blue) are presented on a computer screen and your task is to identify the ink colour of the word (by way of button press). At times, you will see one stimulus on the screen, and other times you will see a stimulus, repeated all over the screen. Regardless, your task is to identify the ink the stimuli are written in. You will be asked to do this as fast as you can, without making mistakes.

Initially, there will be a practice session in order for you to get accustomed to the design of the experiment, and to remember which button is associated with each colour. The buttons will be colour coded, and there is also a colored bar on top of the screen for your ease of remembering. Again, we are looking for the speed-accuracy trade off, thus we would like you to go as fast as possible without making mistakes. There will be different types of words, specifically negative and neutral words as well as control words. The experiment will take about 30 minutes.

By signing this form you are acknowledging your understanding of the following points:

- Should you ever feel very uncomfortable while doing this experiment you are free to quit at any point and without any penalty; your participation is completely voluntary.
- From this point onward your data will only be associated with a participant number thereby assuring your anonymity. Only the experimenter, supervisor and computer programmer will have access to your data.
- There are no risks associated with this experiment. Your benefit is the experience you gain in understanding the process of psychological research methods.
- At the end of the experiment the basic purpose of the experiment will be explained further, and also to answer any questions you may have.
- Your course credit will be given to you by the end of the day.

If you are interested in the outcomes of the experiment, or if you simply have any questions please do not hesitate to contact the experimenter. In addition, if you would like to get a copy of the manuscript once it is completed, please provide the experimenter with your email.

If you have any questions about your rights as a participant, you can contact the Ethics Review Office at ethics.review@utoronto.ca, or (416) 946-3273.

If this is all clear, then please print the date, your name, student number and sign below:

Date: _____

Name: _____

Student number: _____

Signature:

Experimenter:

Sarah Uzzaman,
Graduate Student,
Department of Psychology,
University of Toronto Scarborough,
sarah.uzzaman@utoronto.ca

Supervisor:

Prof. Steve Joordens,
Department of Psychology,
University of Toronto Scarborough,
joordens@utsc.utoronto.ca

Stimulus Words List for Experiments 3 & 4

Positive Valence		Negative Valence		Taboo	
Critical	Control	Critical	Control	Critical	Control
good	here	evil	seek	fuck	type
happy	glass	sad	ear	shit	west
beautiful	knowledge	ugly	neat	bitch	candy
kiss	bell	nasty	untie	cock	dash
great	house	terrible	football	asshole	embargo
joy	gem	fear	test	fag	ape
smile	judge	frown	hasty	cum	ant
hug	sag	hit	fig	slut	comb
laugh	paint	cry	add	piss	swan
pleasant	location	anger	catch	pussy	quote
love	past	haunt	wreck	sperm	paper
peace	leave	war	say	ejaculate	chemistry
help	feet	torment	edible	tits	corn
friend	effort	enemy	touch	crotch	patio
kind	held	curse	razor	prick	banjo
excellent	hierarchy	repulsive	carpenter	boobs	dozen
honesty	founder	horrible	compound	nipple	oceans
funny	knock	cruel	hound	penis	spice
handsome	parallel	cheat	dingy	erection	nutrient
perfect	handle	deceive	contour	vagina	pigeon
impressive	sufficient	awful	dandy	clitoris	notation

superb	intact	disaster	autonomy	pubic	cabin
charming	luncheon	sin	hat	dyke	fawn
lucky	cheek	slaughter	staircase	whore	sable



UNIVERSITY of TORONTO

SCARBOROUGH

DEPARTMENT OF PSYCHOLOGY

Consent Form

Thank you for agreeing to participate in this experiment. Your participation is greatly appreciated and valued. This experiment is being conducted by Sarah Uzzaman under the supervision of Professor Joordens. The experiment will take about 30 minutes.

The first part of this study is a variation of the classic Colour-Word Stroop tests, where different words, written in different colour ink (red, green, blue) are presented on a computer screen and your task is to identify the ink colour of the word (by way of button press). At times, you will see one stimulus on the screen, and other times you will see a stimulus, repeated all over the screen. Regardless, your task is to identify the ink the stimuli are written in. You will be asked to do this as fast as you can, without making mistakes. Initially, there will be a practice session in order for you to get accustomed to the design of the experiment, and to remember which button is associated with each colour. The buttons will be colour coded, and there is also a colored bar on top of the screen for your ease of remembering. Again, we are looking for the speed-accuracy trade off, thus we would like you to go as fast as possible without making mistakes.

There will be different types of words, specifically positive, negative, taboo and control words. Taboo words are words that are derogative and sexually charged words that are viewed as highly offensive in nature and by society. Some of the words may make you anxious or uncomfortable, which is part of the experiment.

By signing this form you are acknowledging your understanding of the following points:

- Should you ever feel very uncomfortable while doing this experiment you are free to quit at any point and without any penalty; your participation is completely voluntary.

- From this point onward your data will only be associated with a participant number thereby assuring your anonymity. Only the experimenter, supervisor and computer programmer will have access to your data.
- There are no risk associated with this experiment. Your benefit is the experience you gain in understanding the process of psychological research method.
- At the end of the experiment the basic purpose of the experiment will be explained further, and also to answer any questions you may have.
- Your course credit will be given to you by the end of the day.

If you are interested in the outcomes of the experiment, or if you simply have any questions please do not hesitate to contact the experimenter. In addition, if you would like to get a copy of the manuscript once it is completed, please provide the experimenter with your email.

If you have any questions about your rights as a participant, you can contact the Ethics Review Office at ethics.review@utoronto.ca, or (416) 946-3273.

If this is all clear, then please print the date, your name, student number and sign below:

Date: _____

Name: _____

Student number: _____

Signature: _____

Experimenter:

Sarah Uzzaman,
Graduate Student,
Department of Psychology,
University of Toronto Scarborough,
sarah.uzzaman@utoronto.ca

Supervisor:

Prof. Steve Joordens,
Department of Psychology,
University of Toronto Scarborough,
joordens@utsc.utoronto.ca



UNIVERSITY of TORONTO

SCARBOROUGH

DEPARTMENT OF PSYCHOLOGY

Consent Form

Thank you for agreeing to participate in this experiment. Your participation is greatly appreciated and valued. This experiment is being conducted by Sarah Uzzaman under the supervision of Professor Joordens. The experiment is divided into two parts and will take about 1 hour.

The first part of this study is a variation of the classic Colour-Word Stroop tests, where different words, written in different colour ink (red, green, blue) are presented on a computer screen and your task is to identify the ink colour of the word (by way of button press). At times, you will see one stimulus on the screen, and other times you will see a stimulus, repeated all over the screen. Regardless, your task is to identify the ink the stimuli are written in. You will be asked to do this as fast as you can, without making mistakes. Initially, there will be a practice session in order for you to get accustomed to the design of the experiment, and to remember which button is associated with each colour. The buttons will be colour coded, and there is also a colored bar on top of the screen for your ease of remembering. Again, we are looking for the speed-accuracy trade off, thus we would like you to go as fast as possible without making mistakes.

There will be different types of words, specifically positive, negative, taboo and control words. Taboo words are words that are derogative and sexually charged words that are viewed as highly offensive in nature and by society. Some of the words may make you anxious or uncomfortable, which is part of the experiment. During the duration of the experiment, we will ask you to wear a smart watch in order to measure your heart rate, body temperature and perspiration to objectively capture any physiological responses to the words. After completing the taboo Stroop experiment, the second part of the study will be questionnaires to assess and get your opinion on how those words made you feel.

By signing this form you are acknowledging your understanding of the following points:

- Should you ever feel very uncomfortable while doing this experiment you are free to quit at any point and without any penalty; your participation is completely voluntary.
- From this point onward your data will only be associated with a participant number thereby assuring your anonymity. Only the experimenter, supervisor and computer programmer will have access to your data.
- There are no risk associated with this experiment. Your benefit is the experience you gain in understanding the process of psychological research method.
- At the end of the experiment the basic purpose of the experiment will be explained further, and also to answer any questions you may have.
- Your course credit will be given to you by the end of the day.

If you are interested in the outcomes of the experiment, or if you simply have any questions please do not hesitate to contact the experimenter. In addition, if you would like to get a copy of the manuscript once it is completed, please provide the experimenter with your email.

If you have any questions about your rights as a participant, you can contact the Ethics Review Office at ethics.review@utoronto.ca, or (416) 946-3273.

If this is all clear, then please print the date, your name, student number and sign below:

Date: _____

Name: _____

Student number: _____

Signature:

Experimenter:

Sarah Uzzaman,
Graduate Student,
Department of Psychology,
University of Toronto Scarborough,
sarah.uzzaman@utoronto.ca

Supervisor:

Prof. Steve Joordens,
Department of Psychology,
University of Toronto Scarborough,
joordens@utsc.utoronto.ca

Taboo Questionnaire

1	2	3	4	5	6	7	8	9
Completely Comfortable (use it all the time; would not think about it)	Relatively comfortable (Would use it often)		Somewhat comfortable (use it every so often; would think about it, but still use it)			Not really comfortable (use it only when angry & provoked)		Completely Uncomfortable (would never use it; would be offended)

Please answer the following questions using the scale above.

1. Would you ever use these taboo words with your parents?

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

2. Would you ever use these taboo words with your siblings?

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

3. Would you ever use these taboo words with your friends?

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

4. How comfortable are you hearing these taboo words?

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

5. How comfortable are you using these taboo words verbally?

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

6. How comfortable are you using these taboo words in writing (social media; texting)?

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

7. How comfortable will you be if asked to read this list of taboo words out loud words to the experimenter?

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

8. How comfortable will you be if the experimenter read this list of taboo words out loud to you?

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---