DOCTOR OF PHILOSOPHY

Measuring implicit attitudes towards pictures : presenting a modified version of the Extrinsic Affective Simon Task (EAST) and its applications

Greenwell, Felicity Emma

Award date:
2010

Awarding institution:
Bangor University

Measuring implicit attitudes towards pictures: | Presenting a modified version of the Extrinsic Affective Simon Task (EAST), and its applications.

PRIFYSGOL
BANGOR
UNIVERSITY

Felicity Emma Greenwell

BSc (Hons), MSc

ACKNOWLEDGEMENTS

The work carried out in this thesis was made possible by a ESF Objective one grant, in conjunction with sponsorship from local company blah d blah design. I would like to thank Neil Rylance, Managing director for his willingness to sponsor my work.

I would also like to thank those who have helped me to understand the statistical analysis required for this thesis. Katy Thompson and David Wooff from Durham University, Department of Mathematical sciences have been vital to assisting me in my task of sorting through what was on times a very confusing analysis.

On a personal note there are so many people whom I should thank, who have helped me along the course of my studies, too many to mention in fact. If you find yourself reading this because you know me, then it is likely that you are one of those people, so thank you. There are of course those who require individual mention, my supervisor Dr James Intriligator, who helped me to find PhD funding which was so well suited to my goals, for encouraging me to explore new avenues, and to 'think outside of the box'. I would also like to thank Dr Mary Ellen Large who provided me with a wealth of assistance in getting this thesis complete, for which I am very grateful. Most importantly, I would like to thank my family, my husband Chris, my son Joshua, and my daughter Seren, from the bottom of my heart. Without your support and drive to see me conclude my PhD I would not have finished this thesis. It is through your never failing belief that I could do this that has got me where I am today, thank you. I hope now at last we can enjoy family life without me having to sneak off to the office whenever I see the opportunity.

DEDICATION

I am dedicating this thesis to my beautiful niece Keira Lily Davies who was born sleeping on April 12th 2008.

"Perhaps they are not stars, but rather openings in heaven where the love of our lost ones pours through and shines down upon us to let us know they are happy."

- Eskimo Proverb

TABLE OF CONTENTS

PART 2. INTRODUCING A VARIANT OF THE EAST PARADIGM

Chapter 2. Pilot studies and general methods used throughout

Chapter 3. The diffusion of attention away from the stimulus target towards a task descriptor: Experiments 1-3

PART 3. THEORETICAL IMPLICATIONS AND USES OF THE VARIANT OF FRAMED EXTRINSIC AFFECTIVE SIMON TASK

PART 4. GENERAL DISCUSSION

Chapter 7.

ABSTRACT

Since their introduction in the 1980s, interest in implicit measures has increased rapidly. For example, in 1985 fewer than 10 publications that mentioned implicit/automatic attitudes were published, by 2004 this figure rocketed to over 100 publications with the same terms (Psych Info). This flurry of research has launched several implicit measurement techniques, each with its advantages as well as disadvantages. The studies presented in this thesis utilize a modified version of the Extrinsic Affective Simon Task (De Houwer, 2003). These modifications are discussed in relation to a change in processing, away from the stimulus type and towards a task indicator; a change that may result in a task that is better able to measure implicit responses to stimuli.

Each of the chapters presented will initially describe a methodological change before testing its success both theoretically and in an applied sense. Results are discussed in terms of how successful these methodological changes are, and how useful this modified version is in helping researchers understand the processing and consequences of implicit attitudes.

PART 1.

AN INTRODUCTION TO IMPLICIT MEASUREMENT TECHNIQUES, AND THEIR COMMON USES

Chapter 1. Introduction

Chapter 1. Introduction

Introduction

Attitudes, responses, and reactions

Reflect upon the following...

> What makes up (or constitutes) an attitude? How does an attitude influence the choices that you make? What is the 'best' way of measuring this attitude?

Now, let's consider a scenario. Imagine asking your best friend why they hold an attitude about something and how, or when, this attitude was formed. It is likely that they would be able to explain why they have the attitude, for example, maybe it comes from past experience or hearsay. Often they can tell you when the attitude was formed, or if anything changed that attitude. But, would they be telling you the whole truth? If not, this is likely not because they were lying, not intentionally anyway, but because some attitudes are not readily accessible and easy to verbalise: in short, there are some attitudes that are more implicit in nature.

Implicit attitudes can be held about a multitude of topics and on numerous dimensions. Such attitudes might have been formed during childhood or perhaps were formed more recently. Some such attitudes might have been formed through experience or perhaps through hearsay and, some are might be easily influenced and others more deep-seated and difficult to change. This thesis focuses largely upon a range of attitudes that might be relevant in applied settings, specifically those of a more transient nature that can influence overt behaviour – often behaviours that are "consumer behaviours". As a result of the attitude type being studied, the motivation for the work reported in this thesis is also often of an applied nature. However, in all cases, the applied studies presented herein are first tested in experimental conditions that stem from work with a strong theoretical standpoint. In other words, while one of the primary goals of this thesis is to develop more appropriate "applied" measures, each new measure is motivated from a theoretical perspective and is tested on both "applied" stimuli (e.g. brands) and "laboratory" stimuli (e.g. photos from the IAPS).

For a variety of reasons, I chose to use the Extrinsic Affective Simon Task (EAST) as the primary experimental paradigm. The reasons behind this choice shall be discussed in greater detail later in this chapter. Firstly, I will introduce the broad area of attitudes and implicit research, along with the methods and techniques commonly used for their measurement.

Attitudes

The question 'what is an attitude?' is one which has been hotly debated for some time, and is not one that can easily be addressed in a short space. It is worth noting that, for the purpose of this thesis, I am concerned only with the retrieval and behavioural manifestations of an attitude (e.g. choices influenced by attitudes) rather than with the original attitude formation.

At perhaps the most basic level, attitudes are valenced feelings towards salient objects. Attitudes are said to be composed of expectations, beliefs, and emotional responses to a psychological entity which operate at or below the level of awareness to influence attitude-related behaviour (Bagozzi, 1982). There are two main views about attitudes. The first is that attitudes are relatively permanent and that they are automatically activated; the second is that attitudes often change and that both the attitudes themselves and their impact are context-sensitive. Those which support the viewpoint that attitudes remain in memory and persists over time (Sherif and Cantril, 1947) have recently discovered a further component of this view: that these attitudes tend to come to mind automatically, quickly, and with little conscious control; (Fazio et al, 1986). This finding has largely been replicated using a variety of different attitude types. But, what is not agreed upon, is whether this applies to all attitudes or just to strong/deep-seated ones.

For those that support the opposing view – that attitudes are much more context-dependant (Anderson, 1974) – it has been argued that attitudes can be so context dependant that there is no such thing as a 'true' attitude. There is no doubt that attitudes can be affected by many things such as current state of mind, mood, own behaviour and social context, all of which have an empirical standing. Wilson, Hodges and LaFleur (1995) are amongst those who have demonstrated that when people are asked to think about the way that they feel and their reasons for those feelings they often construct an entirely new attitude based upon reasons that are readily accessible. So, it is entirely sensible, based on these findings, to propose that attitudes are best viewed as the current state of activation of a connectionist system (Smith, 1996) rather than as long-standing evaluations that are stored semi-statically in memory. Arieli, Sterkin, Grinvald and Aertesen (1996) conceptualize attitudes as stored evaluations that are also context dependant. Many parallel distributed processing models assume that mental representations are highly sensitive to context, since context influences the pattern of activation that determines and perhaps "embodies" mental representations. Given these two varying view points about what an attitude is, it is important to contemplate the division of attitudes into "implicit" and "explicit" components (or types), and to consider how this distinction has shaped modern attitude research methods.

Implicit versus Explicit attitudes

As the chemist and philosopher Michael Polanyi once wrote, "We know more than we can say" (1967). For decades now, scientists have tried to explore "implicit" knowledge and attitudes: How can we go beyond what people *say* they think or feel, to find out what they actually *do* think or feel? There is a growing movement of scientists who are calling for a more stringent definition of what makes something implicit and not explicit. De Houwer and Moors (2006) are two such advocates of this stance, adopting a definition of an implicit measurement as 'measurement outcomes that reflect the to-be-measured construct by virtue of processes that are uncontrolled, unintentional, goal dependant, purely stimulus driven, autonomous, unconscious, efficient or fast'.

There is currently a raging debate about what makes an attitude implicit or explicit, and indeed, which is the 'true' attitude. To help gain some clarity, it will prove helpful to look at the theoretical literature behind what makes an attitude implicit or explicit. Allport (1935) stated that the function of an attitude is to evaluate, and to provide valanced summaries of our environment. Fazio (1986) conceptualized attitudes as signals for people about what is good and bad in their environment and therefore believed that they are vital for everyday life via, for example, approach/avoidance behaviours. If the primary psychological impact of attitudes is on approach/avoidance behaviours, then one might assume that attitudes have a single valance, e.g. good or bad. In other words, if attitudes link primarily to approach/avoid, then they might well be a 'one-dimensional summary statement' (Thompson, Zanna and Griffin, 1995). Consistent with this belief, consumer and social studies traditionally are measured using unidimensional scales e.g. is the positive or negative?

However, is this unidimensional viewpoint true to life? How often is it the case that you think you hold an attitude only to be proved wrong? For example, love and hate are often reported as easily confused. Dovidio, Kawakami, Johnson and Johnson (1997) discuss models of prejudice and suggest that it is possible to hold an attitude learned in childhood whilst simultaneously also holding a more recent attitude about the same topic. One question relevant to this topic is: can we really have two attitudes at the same time, or when we form a new attitude does it erase the old one? Wilson, Lindsey and Schooler (2000) propose that people have dual attitudes: 'one which is an automatic, implicit attitude and the other of which is an explicit attitude. The attitude that people endorse at any point in time depends on whether they have the cognitive capacity to retrieve the explicit attitude, and whether the explicit attitude overrides the implicit one'.

Theories concerning implicit measures and their relationship with explicit measures

Given that researchers remain divided about the interaction, or lack of interaction, between explicit and implicit attitudes, before any research is undertaken to explore an implicit measurement technique it is prudent to explore the different viewpoints. A large volume of research has found support for the idea that implicit attitudes are better at predicting spontaneous behaviour, and explicit attitudes are better at predicting purposeful behaviour (e.g. Dovidio, Kawakami and Gaertner, 2002). However, others (e.g. Hugenberg and Bodenhausen, 2003) postulate that both explicit and implicit attitudes can predict behaviour, or that they interact in some way to ultimately inform behaviour (McGregor and Marigold, 2003).

A useful model that provides a framework for understanding the benefits of using implicit measures to study behaviour was provided by Fazio (1990). The "MODE model", as it is known, 'proposes that attitudes can exert influence through relatively spontaneous or deliberate processes. The former involve judgments of, or behaviour toward, an object being influenced by one's construal of the object in the immediate situation – perceptions that themselves can be affected by individuals' attitudes having been automatically activated upon encountering the attitude object. In contrast, deliberate processing involves a more effortful, cost-benefit analysis of the utility of a particular behaviour'. In other words, when a person has the time to consider a decision, behaviour is deliberate and best predicted by explicit attitudes, when the person's behaviour is spontaneous, then implicit measures are more suited. There is now evidence to support the idea that implicit measures, for example those that arise from the IAT (Implicit Association Task), are good predictors of spontaneous behaviour (Poehlman, Uhlmann, Greenwald and Banaji, 2009).

Zajonc (1980) postulated that people don't tend to think *through* their attitudes, they just are what they are, a spontaneous and automatic evaluation of the world around them. Explicit measures force attention to be allocated to the attitude, evoking a whole series of cognitive events; and thus, such measures no longer measure the spontaneous and "real" attitudes – instead, they measure the conscious/cognitive construction or interpretation of the attitude. In the case of the experiments reported in this thesis, which are concerned with consumer behaviours, implicit measures are likely well suited. This is because, as was remarked by Fazio and Olsen (2003), such measures are particularly helpful in predicting behaviour that is intrinsically difficult to control, where people are un-motivated or cannot control the impact of automatic behaviour.

Attitude measurement

After much discussion amongst attitude researchers it is often agreed that there are three main methodological approaches towards attitude research. The first uses explicit techniques, using methods which minimize the incentive for a person to report what they see as the socially desirable response. Of course, this assumes that the person is always aware of their attitude in order to report it. An example of such a method is the 'bogus pipeline' (Sigall and Page, 1971), where participants are encouraged to believe that researchers are able to discern their true attitude so lying becomes an embarrassment and pointless. Bradburn, Sudman and Warnsink (2004) used a technique where confidentiality and anonymity are assured by either simple reassurance or randomized response techniques. These two techniques have successfully demonstrated an increase in the frequency of self-reported socially undesirable answers; however, as mentioned earlier, they rely on the participant 'knowing' their attitude. It may be that the participant is not willing to admit the attitude even to themselves (Paulhus, 1984) or that they simply lack any cognitive access to the attitude. This is certainly the case when researching socially sensitive attitudes such as racial attitudes (Crosby, Bromley and Saxe, 1980).

The second and third methodological techniques are concerned with 'implicit' measures of attitudes. That is, those methods which do not involve explicit self report. Implicit measures have become very popular in many sub disciplines of psychology (Fazio and Olson, 2003). De Houwer and Moors (2007) discussed the different definitions of what it means to say that something is an implicit measure, arriving at the conclusion that they can be defined as 'measures that capture attitudes in an automatic manner'. Such implicit measures are often employed as they are believed to be less biased than explicit measures wherein the participant may attempt to conceal their 'real' attitude, or where they may simply be unaware of such attitudes (Dovidio and Fazio, 1992; Greenwald and Banaji, 1995). Broadly speaking there are two different ways in which implicit attitudes can be measured, these are indirect measures and physiological measures.

The first of the two implicit techniques to be presented uses indirect measures to assess attitudes. In fact, proponents of such techniques claim that, in many cases, the participant is ideally unaware of the relationship between the measure and the attitude. An example of such a technique, which will be discussed in more detail below, is the "affective priming paradigm." The underlying theory is that an attitude will exert an influence upon the performance of certain tasks, and the amount of this influence can be used to assess the strength of the attitude tested. The technique aims to assess attitudes by measuring the immediate and automatic responses towards an object (word, image, etc) quickly – often with a few hundred milliseconds from presentation. Such attitudes are conceptualized as being unintentional, resource efficient, and outside of conscious control and awareness (Shiffrin and Schneider, 1977).

The second of the implicit techniques uses methods such as physiological reactions and measurements of brain activity, to avoid the possibility that the participant is responding strategically. The next section of this introduction will focus on the first of these implicit techniques, as it is within this genre of investigations that the experiments reported in this thesis reside.

Implicit techniques

Most of these methods quantify the implicit reaction to a stimulus based on the amount it influences the performance of a primary task, and typically they rely on response time measurements. Examples of such tasks include: priming (Fazio et al, 1986), the Go/No-Go Association Task (GNAT, Nosek, and Banaji, 2001), the Implicit Association Test (IAT, Greenwald et al, 1998), and the Extrinsic Affective Simon Task (EAST, DeHouwer, 2003).

Priming

It is largely accepted amongst cognitive science scholars that when a participant is exposed to a concept such as 'teacher', this will facilitate retrieval of successive related concepts, e.g. 'pupil'. This is often explained via a "spreading activation" type of model: the participant is first exposed to the initial concept (the prime) which in turn activates semantically related concepts, this results in a time reduction for subsequent decisions related to these "primed" terms. There are two types of priming measures used in attitude measurement, concept priming and evaluative priming.

Concept priming

Concept priming, takes advantage of the semantic link between the prime and target. The prime activates internal representations which are associated with that prime in memory. This activation spreads via existing links in the semantic network to associated representations. As a consequence of this spreading of activation, the internal representation of the target receives a level of activation prior to the actual display of that target. This results in reduced time required for the target to be recognized and/or responded to (Posner and Snyder, 1975). The priming just described relies upon facilitatory priming effects, the other type of response time measure relies upon the opposite to this, interference effects. The measurement technique used in the Experiments reported throughout this thesis uses interference effects, three of which are briefly described below (after a description of evaluative priming) beginning with the Implicit Association Task.

Evaluative priming

The main difference between evaluative and concept priming is that evaluative priming studies focus on evaluative associations, whereas concept priming focuses on conceptually driven relationships between the target and the prime. An example of evaluative priming is shown by Fazio et al (1986) who presented target words with a general evaluative meaning (such as 'awful'), and then asked their participants to judge the word's evaluative context, either positive or negative. So, participants were presented with descriptive target words, and a decision task was then presented (evaluation priming) to identify the word. The researcher would then determine whether exposure to the prime facilitated the evaluation of the target word. Fazio (2001) later reflected that studies such as the one just referred to use targets that are equal to 'good' and 'bad'. Other studies, such as DeHouwer et al (2002) use target words with a clear valence, but that are otherwise diverse in meaning. Fazio (2001) suggests that in the former case, (with the targets equaling good and bad), results can be explained by the spreading activation account as with concept priming. When the target words are more diverse this could reflect response competition.

IAT (Implicit Association Task)

The first measurement method using interference techniques discussed here is the Implicit Association Test (IAT) (Greenwald, McGhee and Schwartz, 1998) since it is the most widely known and used group of interference measures. The IAT consists of two discrimination tasks which are combined across five different steps. A common example of an IAT experiment involves using the test to measure a participant's implicit attitude about race. Such an experiment takes place in five different phases or steps. In the first phase names are presented one at a time on a computer screen. The names are common names that are typically used by either white or black people, the participant is asked to classify them, this is done by pressing either a 'white' or a 'black' key. In the second phase, words with a positive or negative connotation are presented and the participant classifies them using the left or right key (the same keys that had previously been used for the name classification). In the third phase, phases one and two are essentially "superimposed": participants now press the right key if they see a "white name" or positive words, and the left key if they see a "black name" or negative word. In the fourth phase, the assignment of the keys as 'white' and 'black' are reversed. Finally in the fifth phase, the two tasks are again superimposed but in reverse, so this time participants press the right key if they see a black name or a positive word and the left key if they see a white name or a negative word. The primary measure of interest is the response times for phases three and five. In the example described above, if the participant responds quicker in step three than five this is indicative of positive attitudes held towards 'whites' and negative towards 'blacks'. Essentially, the underlying logic is that it is easier to respond when similar responses are grouped and thus require the same hand for responding. And, contrary, it will be more difficult (slower) to make a response if the response key is mapped to different/opposing concepts.

Go/No-Go Association Task

The go/no-go association task (GNAT; Nosek and Banaji, 2001) uses a similar logic to the IAT and is a related response competition task. There are theoretical similarities to the IAT as it examines automatic associations between concept (e.g., gender) and attribute (e.g., evaluation) categories. However, despite its similarities, it varies in two main ways. Firstly, in terms of how effects are quantified, the GNAT uses signal detection statistics in its calculations. Secondly, with other implicit measures such as the IAT there is a requirement for an attitude toward one category (e.g. insects) to be assessed relative to a second category (e.g. flowers). However, in the GNAT, insects could be evaluated in the context of a single category (flowers), a superordinate category (animals), a generic category (objects), or with no context at all.

Extrinsic Affective Simon Task

The Extrinsic Affective Simon TASK (EAST) paradigm has its roots in an early paper by De Houwer and Eelen (1998) and, as is the case for the GNAT, it has its roots based in the IAT. In this initial paper the authors showed that it was possible to create an affective (emotionally-valanced) variant of the traditional Simon paradigm (Simon, 1969). In the original Simon task, a participant reports changes to a key feature of a stimulus (for example, its colour) and tries to ignore "irrelevant" features, such as spatial position (Kornblum, 1992). The classic finding is that the irrelevant feature powerfully influences response times; the irrelevant feature affects performance of the primary task.

For their 1998 paper De Houwer and Eelen created a task in which both the relevant and irrelevant features had affective content. As their primary task, participants classified certain emotionally affective target words as either nouns or adjectives. They indicated this classification verbally, using a word label that also had an affective valence (e.g., "flower" to label nouns and "cancer" for adjectives). Investigators instructed participants to ignore the affective meaning of the target words and to identify them simply as parts of speech. The results revealed a Simon effect in the affective domain: reaction times were fastest when the valence of the target word (e.g. "gift") matched the valence of the correct word-classification response (say, "flower" identifying it as a noun). This result suggests that the affective content of the word was processed sufficiently (perhaps subconsciously) to influence task performance even when participants were instructed to ignore the affective meaning entirely.

Investigators developed this affective variant of the Simon paradigm into the EAST paradigm - a technique that could be used to measure attitudes indirectly (DeHouwer, 2003). The primary innovation in EAST is that, rather than evoking pre-existing connotations (e.g., "'flower' is good"), it creates an extrinsic response connotation over the course of the experiment (e.g., "the 'M' key is good", or "the 'D' key is "happy", or "the 'L' key is "sad"). Once the experiment has created this extrinsic connotation, researchers can use it to measure implicit affective attitudes by looking for stimulus/response compatibility effects. For example, imagine we have one key that became extrinsically linked to the concept of "happy" and another became linked to "sad". Then, any subsequent classification task that required the use of these same keys should suffer/benefit from these extrinsic linkages, in direct proportion to how happy/sad the participants find the target stimulus.

On each trial in the original EAST experiment a word was presented on a black background. The colour of the word was either white, green, or blue. When a white word appeared, participants used one of two response keys to classify its valence (positive/negative). This task established the extrinsic key mapping: a certain key for positive words, another for negative. When a coloured word appeared, researchers instructed participants to ignore its meaning and instead to just classify the colour (simply as either green or blue), using the same two keys as in the white (valence) task. Thus, thanks to the extrinsic key mapping, the extrinsic meaning of the keys had an opportunity to influence responding. The authors found that coloured words with a positive valence produced faster correct responses when the correct colour classification response key corresponded to that assigned to positive white words. Therefore, the relative speed/accuracy of colour classification in relation to the extrinsic meaning of the keys could be used to measure implicit reactions to the target stimuli.

Subsequent to the initial experiments that measured valence with the EAST paradigm, strong evidence has emerged suggesting that the technique is a powerful tool for assessing affective attitudes about a wide range of concepts. Examples include food and physical activity (Craeynest et al, 2005), alcohol (DeHouwer et al, 2004), and phobias (Huijding and Jong, 2005).

Huijding and Jong (2005) discuss the use of frames with pictorial stimuli. However, these authors opted to use frame form (instead of colour) as the relevant feature, fearing that if the picture content and picture form are not linked, the picture may be ignored, and thus the interference effects may be reduced. Although this concern is valid, work by Degner, Wentura, Gniewosz and Noack (2005) reported evidence to suggest subliminal affective priming tasks, where the participant is unaware of the attitude object, are effective. In these studies participants were oblivious to the object being recorded, thus, giving further support to the idea that a stimulus need not be directly attended in order to record an implicit attitude towards it. Huijding and Jong (2005) also discussed the possible use of a coloured overlay over pictorial stimuli. But, they reject this approach citing work with an emotional Stroop task (Constantine, McNally and Hornig, 2001) suggesting that an overlay may undermine the strength or validity of the stimulus. It is a modification of the EAST method which is used in each of the experiments reported as part of this thesis.

Physiological measures

Whist the focus of this thesis uses response time and accuracy measures, it is important to acknowledge the growing trend towards using physiological response and brain activity measures. The main measurement techniques used have been: functional Magnetic Resonance Imaging (fMRI); Event Related Potentials (ERP); Facial electromyography (EMG); Heart Rate measures and Galvanic Skin Responses (GSR). Each of these physiological measures are discussed briefly in the following sections.

Magnetic Resonance Imaging (fMRI)

fMRI is one of the most recent techniques used to measure implicit attitudes. The premise is that with an increase in neural activity comes an increase in blood flow, which can be measured using this method. This technique has excellent spatial resolution, however there are issues surrounding poor temporal resolution as the time course of blood flow changes lags behind that of neural changes. fMRI signal does not typically start until 2 seconds after stimulus onset, reaching its peak 2-4 seconds after onset and returning to baseline 10-14 seconds after onset (Menon and Kim, 1999). Research with fMRI has lead to a general understanding that the amygdala and the right inferior Pre Frontal Cortex (PFC) are both involved with both implicit and explicit possessing (Cunningham, Johnson, Raye, et al., 2004). Precisely what this activation reflects remains unclear although there have been many functions attributed to it. These include evaluations of the threatening nature of stimuli, assessments of attentional demands, assessments of biological/personal significance to a stimulus, and binding together information to form one unit. Whilst what the activation reflects remains unclear, what is clear is that amygdala activation is similar to what would be expected with explicit ratings of stimuli, without conscious recognition of the stimulus being required.

Event Related Potentials (ERP)

This is a direct measurement of electrical activity measured through the scalp. ERPs have a much better level of temporal resolution compared to fMRI, allowing for easier measurements of processes such as implicit attitudes. Also, the cost of maintaining and running ERP experiments is considerably less expensive than fMRI studies. However, where there are positives there are invariably negatives; in this case it is the lack of spatial resolution, also, in order for electrical activity to register there needs to be a sufficient signal of activity, with deeper brain structures often registering little activity (Fabiani, Gratton and Coles, 2000). Researchers have used ERPs to discover that larger P1 (Clark and Hillyard, 1996), N100 (Fabianni et al, 2000), P200 (Smith et al, 2003), and the LPP (Late Positive Potential) and the P300 (Cacioppo, Crites, Berntson and Coles, 1993) responses can provide information about the implicit mechanisms that underlie attitudes.

Facial electromyography (EMG)

EMG uses small electrodes placed on the facial muscles of interest, and are relatively easy to acquire. Cacioppo and Petty (1981) were the first to demonstrate that EMG can assess both the cognitive and affective processing which underlies attitudes. In a previous study Cacioppo and Petty (1979) showed attitude valence was reflected in EMG activity over the brow and cheek regions. Support for these finding comes from studies with other affective stimuli such as film clips, words, sounds as well as affective pictures (Cacioppo, Bush and Tassinary, 1992). Since facial muscles can be controlled consciously, studies using EMG to assess implicit processes need to designed accordingly. One such example of such a method is using short presentations and backwards masking techniques (Winkielman and Cacioppo, 2001), in this study contour outlines of common objects were presented for 16 milliseconds, immediately masked by a simple line drawing. Perceptual fluency suggests that perception of the target should be facilitated if both the target and prime are the same.

Furthermore, increased fluency is associated with greater positive affect, and as such greater activity was found in the cheek when the target and prime matched.

Heart Rate measures and Galvanic Skin Responses (GSR).

Both heart rate and GSR are easy to record measures of activity of the autonomic nervous system. These methods are not without their opponents, but their ease of capture makes them remain popular measures. However, their use in implicit attitude study is limited as they are insensitive to response valence (Cacioppo et al, 1986). In other words, both positive and negative reactions can produce a nearly identical response, so some prior knowledge about the stimulus valence is required to correctly interpret any such physiological measure. In studies where there is prior knowledge about attitudes held by individuals these techniques can be used to measure the strength of these attitudes (Sheilds and Harriman, 1984). There is a growing body of evidence that casts doubt on heart rate being used as a measure of intensity, as it can also be influenced by other factors such as the metabolic demands associated with an affective state (Cacioppo, Berntson, Klein and Poehlmann, 1997). GSR measures or Electrodermal activity (EDA) dates as far back as back Fere (1888). It is believed to be a highly sensitive but non specific measure of the sympathetic nervous system (Cacioppo and Tassinary, 1990). As with heart rate, it is often used as a measure of the intensity of a reaction where prior attitude knowledge is already known (or separately acquired). However, there can be problems with this as the measure is so sensitive that a response can be elicited by something other than the materials being tested (Blascovich, 2000).

Why focus on the EAST?

As I hope to have demonstrated there is no one implicit measure that is without its limitations. The IAT, despite being very successful and widely replicated can only reveal relative attitudes. For example, in the racial IAT described above, the results are able to indicate the attitude of the participant towards blacks relative towards whites. It does not indicate where this attitude lies on the scale of positive or negative, so both attitudes towards white and blacks could be positive, both negative or polar opposites. Also, the IAT has been shown to be influenced by factors other than personal attitudes (Rothermund and Wentura, 2004). It is these limitations that have encouraged the growth of other implicit measurement techniques such as the EAST.

The EAST has been shown to overcome some of these limitations found in the IAT. The EAST is able to test both single and multiple attitudes in a non-relative manner. Also, since the EAST is based on a comparison of trials within a single task, rather than between two tasks as is seen with the IAT, it has been postulated that the EAST may be less sensitive to non-attitudinal factors that participants may use to recode the different IAT tasks (De Houwer, 2003).

Despite these benefits of the EAST over the IAT it is important to note that there is evidence that suggests that the EAST compares poorly to the IAT at providing a measure of inter-individual differences of attitudes. Schmukle and Egloff (2006) demonstrated that the IAT was far superior at predicting criterion variables compared to the EAST, also that EAST scores have far lower reliability than IAT scores. The experiments in this thesis are not designed to study individual differences, but are interested in using a modified version of the EAST to divulge group trends and, more specifically, to answer applied questions within the context of consumer psychology.

The modified version involves separating attention between the task indicator (a frame) and the stimulus (either a word or a picture).

Why might a spatial separation of the task indicator and stimulus picture assist in the measure of implicit attitudes?

Some argue that attention is the process that allows a cognitive system to preferentially focus on a given object in order to process information from a world that would otherwise be far too complex. It is logical to assume that whilst items are being attended for such preferential processing, other information from our visual field will still be processed to some extent. For example, when driving a car, the driver is focused and attending to the road, however, they still need to be aware of what is going on in the periphery in order to drive safely. By spatially separating the stimulus from the task indicator, this will ensure that the participant is focused on the stimulus in order to complete the task, whilst the task indicator will still receive some amount of processing without being the sole focus of attention.

This seems particularly important when using emotionally charged pictures since Paulhus and Levitt (1987) found that emotionally charged objects attract and hold our attention longer, and that emotional items displayed in the center of a computer screen will grab an individuals' attention and can even disrupt further cognitive processes.

In the design of our experiments we assumed that an implicit measure remains implicit until the participant considers their specific answer in response to a stimulus. Or, put another way, a measure will remain implicit until a participant attends to that aspect of the stimulus that is being measured. Once attention is wholly focused on the stimulus of interest, cognitive processes will come into play and any response will be a mixture of both implicit and explicit processing. If the task indicator and stimulus are too closely linked, then the experimenter runs the risk of having attention overly focused onto that stimulus, thus creating a certain amount of considered thought about it. By making the stimulus irrelevant to the task, the stimulus receives scant thought and so the effects that are measured could be argued to be of a "more" implicit nature.

Attention is not the only cognitive process linked to implicit attitudes, memory has also been a subject of relevant research. Roskos-Ewoldsen and Fazio (1992) demonstrated that when a participant holds an implicit attitude about an object, that object becomes more accessible from memory. Furthermore, these 'attitude-evoking objects' are more likely to attract attention when included in a visual display along with objects that have less accessible attitudes. Once again, this evidence suggests that, when using a more emotionally charged stimulus, distracting attention away from the object may ensure a more accurate implicit measure.

How could this modified EAST employ the idea of diffused attention?

Whilst there is a lack of relevant research directly applicable to answer this question, work from areas other than attitude research can provide a basis for discussion as to whether an over commitment of attentional resources can detrimentally affect task performance. One such interesting example which examined the effects of diffused attention is a study by Olivers and Nieuwenhuis (2005) which found that participants performed better on a temporal task when they were distracted by music. The authors argue that a diffused attentional state yields 'better' performance on an attentionally demanding task. Another example lies in relaxation research. Smilek, Enns, Eastwood and Merikle (2006) examined the effects of relaxation on an attentionally-demanding visual search task. They found that the simple act of asking a participant to relax influenced performance. These examples if applied to the EAST, argue at least theoretically, that a spatial separation of the task indicator and stimulus type should at the very least match the success of the traditional EAST method, if not exceed its capabilities.

Theories concerning attention and implicit attitudes towards emotional stimuli

There are numerous theories surrounding this area, often conflicting, and specific to the stimulus type, either words or picture based. Furthermore, the influence of affective stimuli upon attention has been examined in several different area of psychology making comparisons difficult. To name but a few, for example, Ohman, Flykt and Esteves (2001) studied the effects of threatening stimuli on attention, Pratto and John (1991) examined at the effects of personality traits on attention, and Williams, Matthews and MacLeod (1996) researched the influence of depression and anxiety disorders on attention towards stimuli.

Schimmack (2004) postulates that any theories based on studies using only words as stimuli are limited, as it is difficult to understand whether the results from words alone are generalisable to other stimuli. This is because pictures are more ecologically valid than words (Kindt and Bersschot, 1999). Also, words are encountered more often than pictures, thus it can be assumed that their meaning is more readily accessible compared to a new picture not encountered before – which will require a greater depth of processing. Support for this viewpoint has been provided by Kindt and Bersschot (1999), Lavy and van de Hout (1993), and Kindt and

Bersschot (1997), who reported stronger effects for word stimuli compared to pictorial stimuli. There are three major theories about the influence of emotion upon attention which are particularly important to this thesis as the methodological alteration allocates attention towards affective pictures. These are categorical negativity theory (Pratto and John, 1991), the evolutionary threat theory (Ohman et al, 2001) and the arousal theory (Anderson , 2003).

Categorical negativity theory is based on two assumptions, the first of these is that all stimuli in the environment are evaluated as either positive or negative, and it is only the negative stimuli that attract attention. The second assumption is that this automatic evaluation does not provide information about *how negative* the stimuli is, so attention is allocated equally for both very and slightly negative stimuli. Support for this theory comes from the emotional stroop task (Williams et al, 1996). Here trait words such as honest and mean were presented in colour and the participant was asked to name the colour. The results showed that negative words produced longer response times, where as the extremity of the word meaning had no effect. Although there is support for this theory, there remains uncertainty as the whether it would 'stand up' to scrutiny with different stimulus types such as arousing (e.g. erotic) stimuli.

The second theory is the evolutionary threat theory. This begins in a similar way to that of the categorical negativity theory, where negative stimuli that threaten survival attract more attention than stimuli with no such threat. It is argued that the influence of affect upon attention is triggered by threat to survival. Visual search experiments provide support for this theory, for example, when searching for a picture, participants find a threatening picture such as a snake in an array of mushrooms quicker than when they attempt the reverse of looking for a mushroom in an array of snakes. Further support comes from work with faces and expressions where participants find an angry face amongst happy faces faster than when asked to do the reverse (Hansen and Hansen, 1988; Ohman et al, 2001).

Constantine, McNally, and Hornig (2001) used an emotional stroop task with pictures including snakes, bunnies, and cows. According to the evolutionary threat theory more attention would be allocated to the snake pictures. However, it was found that longer response latencies were found for both bunnies and snakes compared to cows, with no significant difference between bunnies and snakes. Other work carried out with both phobics and non phobics have found that pictures of spiders and snakes have not produced more attention than neutral pictures (Kindt and Brosschot, 1997; Lavy, van de Hout and Arntz, 1993).

The final theory is that of the arousal hypothesis. The arousal theory predicts that affective influences upon attention can be predicted by arousal ratings of affective pictures. It also predicts that the strongest effects are found for unpleasant pictures as well as those of opposite sex models, the argument for this being that these picture types produce the highest arousal level (Bradley et al, 2001). It has been argued that a flaw in the research supporting both the evolutionary threat theory and negativity bias theory is that they fail to control for arousal, especially since negative or threat images are often arousing. It is possible that any effects found in these studies were actually due to the arousal created by the pictures used, and not based on the other factors that those studies claimed to support.

Areas addressed as part of this thesis

This thesis consists of four main parts, each designed to cover a specific point; the introduction; the beginnings of the new method including pilot work, methodology and initial experiments one through to three; further exploration of the new method and how far reaching it is; the discussion. Chapter 2 begins by introducing the basic methodology that is used throughout all studies in the thesis, it then goes on to address the pilot work which was carried out as part of the process of developing this methodology. Experiments one through three are then presented. Experiment 1 replicates the EAST paradigm set out by De Houwer (2003) using a new set of words and the several small methodological changes as a consequence of the pilot work. The purpose of this study is to ensure that the stimulus set is as effective as that used by De Houwer, before moving on to Experiment 2 where the methodology is changed more radically. Experiment 3 then exploits the methodological changes in Experiments 1 and 2 by using pictorial stimuli. There is prior evidence suggesting that, compared to words, pictorial stimuli can more directly activate associated attitudes in memory (De Houwer and Herman, 1994). Additionally, pictorial stimuli have an advantage within the applied context in that they are not limited by language, age, or educational background. This chapter focuses on how this methodological change may also bring about a cognitive processing change, which yields promising results as well as expanding the use of an already respected existing technique.

Part 3 then presents experiments four to seven and their sub sections through the chapters four to six. Experiment 4 uses stimuli tested on the dimensions of arousal, dominance and valence. By using these three different dimensions for each stimulus picture, the flexibility and sensitivity of this new method are pushed further. Experiment 5 then shows how this pictorial version testing different dimensions of the same stimulus set is not only theoretically interesting but also can be used to assist in the pursuit of answers to 'real world' questions using the country of origin effect

(COO). Chapter 6 takes the pictorial version of the EAST, devised in chapter 3, and expands its repertoire by moving into the field of vision research. So often implicit measures are primarily used to measure attitudes in participant on matters such as individual differences, social and personality, and consumer science, rarely are they used the better understand more traditional scientific questions. In experiment 6 the pictorial EAST is pushed further than before, examining whether it is sensitive enough to measure the effects of gaze and visual field. Experiment 7 demonstrates a 'real' application of gaze and visual field and how it can aid in applied (industrial) choices.

The final chapter in part 4, chapter 7, contains a discussion of each of the experiments, how successful they have been in their goal of demonstrating a new method and pushing its use and flexibility.

PART 2.

INTRODUCING A VARIANT OF THE EAST PARADIGM

The development of the FEAST

The purpose of this chapter is three-fold: firstly, to provide a description of the methodology which is common to each of the studies presented; secondly, to provide a description of the analysis used in all studies; and finally to provide a brief background about the methodological choices made based on pilot work. While these general methods apply to all the studies presented in this thesis, in experiments where methodological changes are utilized, these are presented at the beginning of the relevant experiment.

Method

Participants

Ethical approval was obtained through the correct departmental procedures for each of the experiments reported, and full consent for participation was provided by all participants prior to beginning each experiment laid out in this thesis. In addition anonymity was made clear, as was the right to withdraw at any point without penalty in accordance with departmental guidelines. In the majority of cases the participants were gathered from the departmental participant pool and took part in order to gain course and printer credits. When it was difficult to gather participants from the demographics required, for example with Experiment 5, participants were found through using the university community panel and by internal advertising on the university intranet. Participants received payment (cash, printer-credits, and/or course credits) at the set departmental rate depending on the time required.

All participants used in the pilot data were from the university participant panel and no payment was provided, all ethical procedures were followed as mentioned above.

Materials and measures

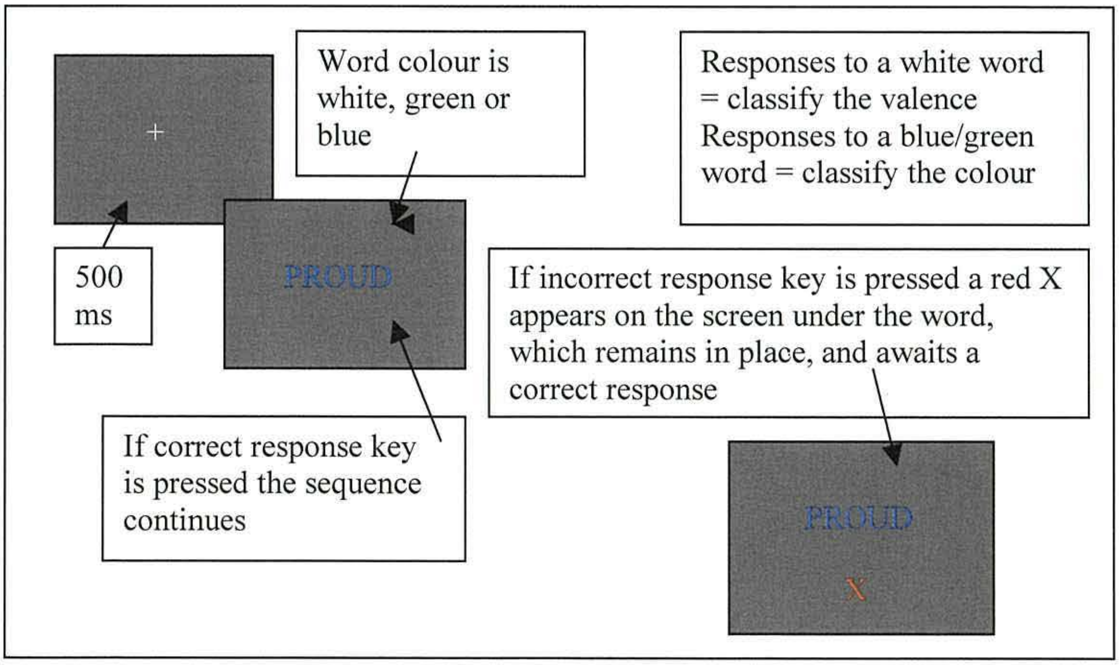
The program was run using e-prime version 1.1. In each of the experiments reported there was either a word or frame which appeared in each green, blue or white. The colour green was created using the RGB values: {90,225,150}; the colour blue was created using the RGB values: {80,210,255}. As was the case in the De Houwer (2003) paper which formed the basis for all experimental work that is reported, the green and blue colours were chosen to be quite similar in appearance (see pilot studies). The letters were centered on the screen, a maximum of 7 mm high and 5mm wide, with a distance from the top and bottom of the screen being approximately 17 cm. Participants were tested in two separate testing rooms, each containing three separate computer workstations, with dividing walls between them (to prevent participants from being distracted by each other). The stimuli were viewed on individual 17 inch monitors placed approximately 40 cms from each participant. Audio feedback was given (via headphones) to indicate whether or not participants had correctly responded. The wearing of headphones also gave more privacy in the room.

Procedure –the experiment

All participants worked individually. Instructions were given in groups both visually via the means of a PowerPoint display and verbally before beginning the experiment and again, during the course of the experiment, on the computer screen. Where participants were to be asked to carry out more than one experiment, the presentation order of the experiments was counterbalanced to prevent learning effects and to allow for a direct comparison when necessary, with each participant in the room following a different run order.

Each trial would follow the same sequence, beginning with a white fixation cross appearing at the center of the screen for 500 milliseconds (ms) on a dark grey background. A word then appeared in the center of the screen on a dark grey background. The letter colour was white, green, or blue. When the word was white, participants were asked to classify it based on its valence alone, that is, whether the concept was positive or negative. When the word was green or blue, participants were to classify it based on colour alone, that is, whether the font colour was green or blue, ignoring the word meaning. The word remained on the screen until a response was made. If the participant made an incorrect response, a tone played, a small red "x" appeared below the target word, and the computer waited for the correct response before continuing. Figure 1 shows the experiment flow.

Figure 1.



Using a QWERTY keyboard, response keys were always 'Q' or 'P', with each of these keys being assigned two different values. For example, for one participant 'Q' indicated both "positive" and "blue", while 'P' indicated both "negative" and "green". For another participant 'Q' indicated both "positive" and "green", while 'P' indicated both "negative" and "blue". Assignment of response keys were counterbalanced across participants both for valence and colour responses. Participants were informed that reaction times were recorded to encourage speedy and accurate responses.

The experiment had two sets of practice trials; pilot experiments had shown that this level and type of practice produced the most consistent responses across all trials without seeing more errors initially due to inexperience with the task. The first set of practice began with the white words, so that participants became familiar responding via the positive and negative keys. Then, without warning, the coloured words appeared, to familiarize the participant with the green and blue response keys. The second set of practice trials mirrored the experimental trials in that both white and coloured words were intermixed from the outset. After the practice blocks, each participant then completed 320 trials (176 of these trials being 'white' trials, 72 being 'green' and 72 being 'blue'). Breaking this down further, each participant saw each coloured word eight times (four times in green, four times in blue), and each white word eight times, not forgetting that there is a higher number of white words to coloured words so the frequency of seeing a white compared to a coloured frame was higher. As was the case with De Houwer (2003) the same word could not be presented on two or more consecutive trials, and the specific response key required was never the same on four or more consecutive trials.

Procedure - Rating test

After each experiment was completed by the participant they were asked to run a quick rating test consisting of both the white and coloured words or images that they had seen in the study just completed. Participants saw each of the words or images three times presented in a random order. They were asked to rate each of the images or words using a 5 point likert scale using the descriptors relevant to the study that they had just completed, for example, "positive/negative" or "UK/USA". The scale was reversed for 50% of the participants, randomly assigned. It was made clear that there was no right or wrong answer and that they should feel free to respond as they wished. This was an important part of the debriefing procedure allowing participants to discuss any issues that they may have had with any of the images or words, providing very valuable feedback. Whilst data from these explicit ratings were not without problems, as it could be argued that mapping had already taken place throughout the study so explicit rating could be altered, it provided a useful starting point from which to discuss the images and words, something which is often difficult for participants to engage in. There were also a small number of participants who had not responded in a varied manner, for example they had pressed the same number for each and every required response, and in this case all their data were discarded.

The frequency of white words versus colour

In a final deviation from the traditional EAST method (in which the ratio of coloured to white words is equal), the FEAST uses slightly more white words. Remember, that the white words serve to link certain keyboard keys with either positive or negative emotional valences. It was felt that the stronger the participant internalizes this mapping, the stronger the match or mismatch during the colour classification task. So in order to prevent any decay of this internalized mapping, the percentage of white words was increased.

Design and Analysis

Error rates as well as reaction times were recorded. It is important to note that in all analysis for each experiment, only the initial correct reaction times were included, not those that had been corrected following an incorrect response. This is vital since any correct responses following an incorrect key press are not implicit in nature, they have been guided, or in other words the program has told the participant how to respond, it is not as a result of any implicitly held attitude. A further motivation for the collation of error data was that De Houwer and Eelen (1998) reported that affective Simon effects often emerge through error data.

The De Houwer Method versus the D600 method

De Houwer method

The original method of analysis (the "De Houwer method") calculates separate EAST scores for reaction times and error rates. For the RT time scores, only the RTs from correct responses are used. Reaction times are recoded such that any RT shorter than 300 msec is changed to 300 msec and any reaction time greater than 3000 msec is recoded to 3000 msec. For each participant, the mean reaction time for each of the trial types was calculated. For each stimulus picture, the mean reaction time for trials with extrinsic response dimension high/good was subtracted from the mean reaction time for trials with extrinsic response dimension low/bad, this resulted in a RT EAST score for each participant and stimulus picture. The analysis of the reaction time data was performed separately on the simple reaction times and, in a different analysis, on the logarithm of the reaction times. Although statistical analyses were carried out on the scores using the logarithms of the reaction times (to create more normal distributions), for ease of reading, we report the RT EAST scores calculated from the reaction times.

To calculate the accuracy EAST score, firstly for each participant, an error rate was calculated for each of the trial types. For each stimulus picture, the error rate for trials with extrinsic response dimension good/high was subtracted from the error rate for trials with extrinsic response dimension bad/low, this resulted in an accuracy EAST score for each participant and stimulus picture.

T-tests were performed on each of the sets of scores; the results are shown only when there is evidence to suggest the scores were significantly different from zero. Following De Houwer, to assess the split-half reliability of the EAST scores, the RT EAST and accuracy EAST scores are calculated, separately for the first and second half of the experiment, and then correlations were calculated between them. De Houwer questions the reliability of the scores if this correlation is not significant.

D600 method

Subsequent to De Houwer's original publications and analyses, a new method of EAST analysis has appeared. This so called "D600 method" is considered in several of the experiments presented herein, thus we will give a brief description of the procedure here. In the D600 analysis, first any latencies greater than 10000ms are removed from the analysis. Participants who had reaction times less than 300ms for more than 10% of trials are also removed from the analysis. The reaction times for incorrect responses are penalized (e.g. 600 msecs gets added to the reaction time), and D600 scores are then calculated. The D600 method takes into account both the RT scores and the accuracy scores, which can often showed conflicting results, any combination of the two resulting in a weaker overall score. Due to this, each experiment reported in this thesis will detail scores which are created using the 'De Houwer method' (presenting scores separately for reaction-times and accuracy) and statistical analysis carried out using the 'D600' method (presenting the single score that incorporates both reaction times and accuracy). As suggested by Schmukle & Egloff (2006) error trials were included in this D600 analysis by using the mean latency in correct trials along with a 600ms error penalty. For example, for each participant the mean latency of trials with a correct response was calculated for each of the four types of trials. The latency for any incorrect responses in each type of trial was then replaced by the corresponding mean latency plus 600ms.

Furthermore, in accordance with Schmukle & Egloff (2006), the D600 EAST score (which presents a single score that incorporates both reaction times and accuracy) was calculated for each participant, using the latencies which include the error penalty. For example, the score was calculated in such a way that a more positive score signifies a more positive response towards the word/object, and similarly a more negative score signifies a more negative response towards the word/object. The D600 score was calculated by subtracting the mean latency from trials that required an extrinsically positive response from the mean latency from trials that required an extrinsically negative response; this value was then divided by the standard deviation of latencies from both of the trial types. This was done for trials that had a positive stimulus and trials that had a negative stimulus separately.

Following De Houwer (2003), for each stimulus group, correlations are calculated between the D600 scores in the first half of the experiment and the second half of the experiment. D600 EAST scores were calculated for each word/image. Incorrect responses were included in the data by using an error penalty as described above. A mean D600 EAST score for each stimulus was found by averaging the scores over the participants. These scores were then split into two depending upon the valence of the word or image, a two sample t-test was then performed.

Pilot studies

Although there was a base experiment from which to create any experiments using the DeHouwer (2003) method, it was decided that in order to optimise the method for the participant set that would be used, a series of pilot studies should be run. These examined the areas; images/words used; the colour hues; the presentation timings and the ratio of white words to colour words. Here follows a brief description of each of these pilot studies, their rationale and conclusions.

Pilot 1 – Getting the right words/images

Rationale

Experiments 1 and 2 provide a basis for comparison about how successful the modification of the EAST to a FEAST design is. Where Experiment 1 provides a replication of the EAST method with minor changes and Experiment 2 uses the same words as in Experiment 1 but with frames. It was important that the stimulus set used was appropriate to both the study and the participants. Furthermore, previous EAST work often used non-English words and in their translation they became unsuitable for UK participants. For these reasons it was decided that a new stimulus set should be created. This process involved participants carrying out a quick rating study, as part of this they were also asked to complete a further rating study using the same principles but of images, this was in preparation for Experiment three.

Method

Participants

Twenty eight psychology students (2 male, 26 female) participated as part of a course requirement. In exchange for their participation, they received course credits and printer credits. All participants had normal or corrected-to-normal vision. The mean age of the participants was 20 (*SD* = 5.26).

Materials and measures

Participants were tested using the e-prime version as described above. Several words were rated and informally discussed to include: Coward; Depression; Excellence; Killer; Leader; Love; Rape; Rejected; Romantic; Thrill; Funeral, Murderer, Terrorist, Slave, Thief, Victim, Idiot, Beast, Treasure, Rainbow, Beach, Champion, Vacation, Kitten, Jewel, Home, and Family. Descriptions of images (from the IAPS database) rated are: Angry man; Butterfly; Water park; Gang shooting; Happy child; Owl; Pollution; Scared children; Ski holiday; War; Attack; Boat; Car crash; Bride; Flood; Fireworks; Gun; Flowers; Knives; Monkies; Missiles; Playing children; Missiles; Sky; Snake; Sports car; Wires; and Treat.

Procedure

A rating test was carried out using the method described earlier.

Conclusions

The results from the explicit ratings were checked for any responses were it appeared that the participant had not 'properly' considered their response, these were then removed, for example, if they had simply pressed the same key for each stimulus presented to them for rating. The images and words used in Experiments one to three (presented in chapter three) were selected in this way using this method. Where participants had indicated any unhappiness about an image or word these were also removed, an example of this was 'war' which after discussion with participants was deemed inappropriate for use.

Pilot 2 – Using different colour hues

Rationale

Before experimentation began the decision was made that it was important that the two colour hues, blue and green, would be similar as was the case in the DeHouwer (3003) research paper. The participant was required to pay attention to the frame, and in coloured trials (the implicit test trials) the participant would press a key to indicate whether the frame was blue or green with the assumption that some degree of processing of the image would be carried out despite that not being the instruction to attend to the image. In order to hold the attention of the participant the task of distinguishing between green and blue needed to have some degree of difficulty.

Method

Participants

Twenty five psychology students (5 male, 20 female) participated as part of a course requirement. In exchange for their participation, they received course credits and printer credits. All participants had normal or corrected-to-normal vision. The mean age of the participants was 22 (SD = 4.96).

Materials and measures

Participants were tested using e-prime version 1.1 under the same conditions as described above.

Procedure

Using the words agreed from pilot 1, the EAST was carried out using different coloured hues of blue and green varying on the spectrum between very similar to very distinct. The white words remained the same hue throughout as the task as white words were concerned with the explicit judgment of the word itself, not their colour.

Conclusions

Participants reported verbally that the task was much easier when the colours were very distinct and indeed the data showed a quicker response time when they were more distinct, however along with an increased reaction time was an increased error rate. It was felt from both the verbal feedback and from looking at the data that when the hues were more similar the participants were concentrated on the task producing a better 'quality' of data. Making the task easier by more distinct hues did not mean that the task with similar hues was reported as difficult, simply not as easy. Since an analysis of accuracy scores was also to be used it was important that these were not influenced by speed of response alone. Thus, we chose to use the hues that were slightly more difficult, but which also gave high levels of accuracy. The hues used (in {red, green, blue} space) were: blue = {80, 210, 255}, green = {90, 255, 150}.

Pilot 3 – Getting the timings right

Rationale

As the participant would be completing trials which were largely repetitive in nature it was decided that there should be a sufficient break between each trial to make them distinct. Also, there was concern as to whether there should be a time limit to response to prevent the participant from making a more explicit judgment.

Method

Participants

Thirty psychology students (2 male, 28 female) participated as part of a course requirement. In exchange for their participation, they received course credits and printer credits. All participants had normal or corrected-to-normal vision. The mean age of the participants was 21 (SD = 6.43).

Materials and measures

Participants were tested using e-prime version 1.1 under the same conditions as described above.

Procedure

Using the words from pilot 1 and the colour hues from pilot 2 an EAST experiment was carried out. Lags of 250, 500, and 750ms were trialed for the presentation of the fixation cross.

Conclusions

As described above the decision was taken to begin each trial with a white fixation cross appearing at the center of the screen for 500 milliseconds (ms), followed by the word or image. Rather than enforce a quick response by preventing responses after a certain time period, the decision was taken to take into account any lag in response time and those trials were discarded.

Pilot 4 – Finding the best ratio of white and coloured words/frames

Rationale

The reason behind deciding to trial different white to colour ratios is theoretically based. If the purpose of the white words or images is to promote mapping of the keys, in other words to indentify a key with a certain response, it seemed logical that by increasing the ratio of white stimuli compared to coloured stimuli that mapping would remain reinforced. This was the basis for this particular pilot study. Specifically, the question asked was: would increasing the number of white words make any difference to the results or to how difficult the task was perceived to be.

Method

Participants

Thirty one psychology students (6 male, 25 female) participated as part of a course requirement. In exchange for their participation, they received course credits and printer credits. All participants had normal or corrected-to-normal vision. The mean age of the participants was 19 ($SD = 3.02$).

Materials and measures

Participants were tested using e-prime version 1.1 under the same conditions as described earlier.

Procedure

The EAST method already described was carried out employing the various parameters agreed from the previous pilot work (words, colours, and timings). There were three test groups each with a different ratio of white to coloured words. Group one (identical to the original De Houwer study) had equal numbers of each, with 16 coloured and 16 white words. Group two had 22 white words and 18 coloured words. Group three had 22 white words and 14 coloured words. The words used were those selected from Pilot study 1. Each participant completed each of the three test group studies so that they were able to report any noticeable differences between them in terms of difficulty or to provide any other feedback.

Conclusions

After the data was analyzed it was concluded that Group two (22 white words and 18 coloured words) had the least errors with consistently fast reaction times. Furthermore, following verbal discussions with participants, they reported that they were unable to distinguish any differences between the groups, so the decision as to which ratio to choose was made based solely on the reaction times and accuracy data.

Pilot 5 – The use of feedback

Rationale

Participants were encouraged to respond quickly and accurately as it was explained to them that an analysis of both reaction time and accuracy scores would be used to form any conclusions about the data. However, if participants showed an unusually high percentage of errors their data would have to be removed from any analysis, so in an effort to prevent this happening feedback sounds were added to the program. It was unsure whether this addition of auditory feedback would be detrimental and distracting as it was not mentioned in the DeHouwer (2003) method. Due to this a small pilot study was carried out to compare the data between those with and without feedback sounds.

Method

Participants

Thirty psychology students (4 male, 26 female) participated as part of a course requirement. In exchange for their participation, they received course credits and printer credits. All participants had normal or corrected-to-normal vision. The mean age of the participants was 22 ($SD = 7.71$).

Materials and measures

Participants were tested using e-prime version 1.1 under the same conditions as described above. There were two test groups each using headphones as it prevented outside noise from disturbing them, however only one test group would hear feedback sounds.

Procedure

Participants used a version of the EAST study which now included all conclusions from the above pilot work. The participants were split into two groups in a random manner with one group receiving feedback sounds and the other with no sounds. The feedback sounds were a soft 'click' with each key press to reinforce key pressing and a soft 'horn' type sound for an incorrect response which accompanied the presentation of the X.

Conclusions

It was concluded both from looking at the data and from verbal feedback that feedback sounds were beneficial. The data between the two groups did not show any significant changes; however, verbal feedback indicated that those with feedback sounds found the task less laborious. Thus, feedback sounds were included in all subsequent experiments.

Chapter 3. The diffusion of attention away from the stimulus target towards a task descriptor: Experiments 1-3

Chapter Introduction

The three experiments presented in this chapter are based upon the Extrinsic Affective Simon Task (DeHouwer, 2003), using both words and images. Unlike the EAST and other previous implicit measurement techniques, the method presented here spatially separates the stimulus (a word or picture) from the task cue (a visual frame). On each trial, the participant sees the stimulus item surrounded by a coloured frame (white, blue, or green), and the frame colour indicates the required action. When a white frame appears, the participant must classify the stimulus's meaning or emotional valence (e.g., "happy" vs. "sad"). When the frame is blue or green, participants must ignore the meaning and simply classify the frame colour ("blue" or "green"), using the same keys earlier mapped as "happy" or "sad." Implicit attitudes are measured toward words or concepts by analyzing reaction times and error rates in the frame-classification trials. The logic being: although participants have been asked to ignore the word and respond only to the colour of the frame, the meaning of the word may still influence task performance.

Earlier researchers have attempted to modify the EAST to work with pictures. For example, Huijding and Jong (2005) used a variant of the EAST task (colours overlayed on top of the stimuli) to measure implicit liking of images. In the discussion to their paper, they raised specific concerns about the use of pictures in this paradigm. In the design of Experiments 1 - 3, the concerns highlighted by Huijding and Jong (2005) have been considered. Specifically, that an overlay might in some way alter response patterns due to basic (e.g. low-level) stimulus interference. This is the main reason for a spatially separate task indicator. It was hoped that the spatial separation of the stimulus from the task indicator could actually be beneficial in terms of keeping the stimulus pure and, at the same time, proving to be slightly more attentionally engaging and demanding. An overlay may not have been such a problem in the experiments reported in the first two experiments in this chapter (using word stimuli), but it was deemed important that this method could easily be adapted for use with pictures, so the use of an overlay was discounted.

As already mentioned, this chapter centres on three experiments. The first begins by using the standard EAST paradigm and the parameters determined from the pilot work described in chapter 2. The second experiment incorporates a methodological change where a frame is introduced as the task indicator (instead of the font colour). This methodological change was introduced in the preparation for the third experiment in which pictorial stimuli are used. The introduction of pictures is a crucial development for the overall aims of this thesis; to develop and use a pictorial version of the EAST and to test this version on both theoretical and applied questions.

All three studies used the same participants. It was decided that using the same participants would be to an advantage when trying to understand how successful experimental modifications were without any question about results being due to different subject pools. Experimental presentation was carefully randomized and the techniques discussed earlier in the analysis section of chapter two were designed to remove any data which may have resulted from practice effects, or from a participant having difficulty due to experiment repetition with modifications to each.

The purpose of experiment one was largely to establish a baseline to begin the process of change: the results can serve as a benchmark and can demonstrate that the fundamental design is working and that the method is sound. As a result of the pilot work carried out, described in the previous chapter, there was confidence that Experiment one would yield results similar to those reported in the DeHouwer (2003) paper. It was hoped that results found in both Experiments two and three would then result in a similar set of conclusions in order to give confidence about the framed method before moving forward with more detailed testing using the framed version. The results are discussed as a whole after reporting the results from all three experiments.

Experiment rationale

Experiment 1 provides a starting point from which to begin the journey of methodological modification for inclusion with pictorial stimuli. Whereas the pilot studies have set the groundwork for the stimuli to use, and the particulars such as colour use and timings, Experiment 1 carries out an EAST experiment to test the success of the 'whole package' utilizing each of the decisions taken from the pilot work. It is important that this experiment is carried out as a baseline from which to gauge the changes in Experiment 2 where the task indicator is separated from the stimulus.

Hypothesis

It was hoped that we could replicate the results of DeHouwer (2003) using our specific words, timings, colours and feedback. Specifically, we hoped to find better performance on trials which required an extrinsic response where the same valenced key was used as for the intrinsic valence of the coloured word. To put this differently, it was expected that reaction time would be quicker and there would be fewer errors when a participant responded to a positive coloured word using the extrinsically positive key, or to a negative coloured word using the negative key.

Experiment 1

Method

Participants

Thirty psychology students (2 male, 28 female) participated as part of a course requirement. In exchange for their participation, they received course credits and printer credits. All participants had normal or corrected-to-normal vision. The mean age of the participants was 21 (*SD* = 5.46).

Materials and Measures

40 words were selected from the ANEW database, Bradley, M.M., & Lang, P.J. (1999). These words were selected to be roughly equal in terms of their pre-rated pleasure, arousal, and dominance ratings.

White words

To establish the extrinsic key association, we used the following 22 white (base) words (11 positive, followed by 11 negative): Happy, Success, Pleasure, Victory, Friendly, Loved, Proud, Glamour, Brave, Beauty, Confident, Grief, Failure, Tragedy, Terrible, Rejected, Gloom, Cruel, Lonely, Inferior, Clumsy, and Pain.

Coloured words

We used these 18 coloured (test) words: Funeral, Murderer, Terrorist, Slave, Killer, Thief, Victim, Idiot, Beast, Treasure, Rainbow, Beach, Champion, Vacation, Kitten, Jewel, Home, and Family.

Procedure

After the practice blocks, each participant then completed 320 trials (176 of these trials being 'white' trials, 72 being 'green' and 72 being 'blue'). Breaking this down further, each participant saw each coloured word eight times (four times in green, four times in blue), and each white word eight times, not forgetting that there is a higher number of white words to coloured words so the frequency of seeing a white compared to a coloured frame was higher.

Design and Analysis

It is important to note that when the words were selected for the next two experiments reported here, every attempt was made to provide a mix of scores from high to low valenced words as was taken from the ANEW (Bradley, M.M., & Lang, P.J., 1999), this was done to encourage a spread of EAST scores.

Results

Overview

Following the improved scoring algorithm developed by Greenwald et al. (2003), trials with latencies greater than 10,000ms were eliminated from the analysis, and participants for whom more than 10% of trials have latency less than 300ms were also eliminated.

Table 1 shows the mean untransformed reaction times and error rates as a function of extrinsic valence and stimulus valence (and for Table 2 this includes experiment half). These tables show the latencies before any error penalty is introduced. The latencies were analysed using a 2 (Experiment Half: First or Second) x 2 (Stimulus Valence: Positive or Negative) x 2 (Extrinsic Valence: Positive or Negative) repeated measures ANOVA, this was performed on log transformed latencies, although it is worth noting that the tables showing the mean reaction time and error rates show untransformed data for ease of understanding.

Participant 5 did not have any correct responses for the word Beast when the extrinsic response valence was negative; as a result the incorrect latencies were penalized by using the corresponding mean latency of trials with a correct response for all of the other remaining participants plus 600ms. Participant 7 only had data for the third quarter of the experiment so this data was removed from the analysis (although, it is worth noting that when the data from participant 7 was included in the analysis none of the conclusions changed). No latencies were greater than 10,000ms (the largest latency was 2703ms). Participant 22 had latencies less than 300ms 19% of the time, so was eliminated from the analysis. Therefore for Experiment 1, the analysis used only 28 of the 30 participants. The flow of analysis replicates that used by De

Houwer (2003) which was taken to be a sensible decision since the method was also based on the work of De Houwer (2003).

Untransformed RT and accuracy data

Table 1 shows summaries of the mean untransformed reaction times for correct responses and the mean error rate in Experiment 1 as a function of Extrinsic Valence and Stimulus Valence:

Stimulus Valence	**Extrinsic Response Valence**	
	Positive	Negative
Positive		
Reaction Time (ms)	555.5 (100.1)	577.7 (98.3)
Percentage of Errors	8.16 (9.74)	16.9 (13.4)
Negative		
Reaction Time (ms)	571.4 (102.8)	554.7 (87.3)
Percentage of Errors	18.3 (13.6)	8.73 (8.37)

Table 1

Both tables 1 and 2 show that on average, the reaction times were greater when the stimulus valence was not the same as the extrinsic response valence. Participants also seemed to make more errors when the stimulus valence was not the same as the extrinsic response valence. This data pattern was anticipated; furthermore, it is in accordance with De Houwer's (2003) results and is vital towards answering the question about whether reaction time is quicker with fewer errors when participants respond to a positive coloured word using the positive key, or to a negative coloured word using a negative key, the data in the table does confirm this pattern.

Table 2 shows summaries of the mean reaction time for correct responses and the mean error rate in Experiment 1 as a function of Extrinsic Valence, Stimulus Valence and Experiment Half:

Stimulus Valence	**Extrinsic Response Valence**			
	Positive		**Negative**	
	Experiment Half		**Experiment Half**	
	1	2	1	2
Positive				
Reaction Time (ms)	552.8 (116.0)	557.6 (91.0)	578.6 (111.0)	576.3 (94.7)
Percentage Errors	9.37 (10.2)	6.94 (10.4)	18.7 (15.2)	15.1 (13.1)
Negative				
Reaction Time (ms)	570.8 (118.0)	572.6 (97.0)	552.8 (92.8)	556.4 (88.3)
Percentage Errors	18.1 (12.2)	18.5 (16.4)	8.73 (9.00)	8.73 (9.49)

Table 2

Table 2 shows that the mean reaction times and error rates do not really change between experiment halves. However in most occasions the variance of the reaction times and error rates reduce in the second half of the experiment. It is important that the data between experiment halves is checked in order to give confidence that the results are not simply due to practice effects. It is expected that there would be no significant difference between experiment halves, as significance indicates the presence of practice effects which the design has attempted to control for by providing a considerable amount of practice trials which are not included in the analysis. A small reduction in reaction times and percentage of errors is anticipated, and was judged to be a consequence of any experiment in which there were repeating trials of a similar nature.

ANOVA on latencies

For the next stage of analysis in Experiment 1 a repeated measures ANOVA was carried out using both the latencies and the logarithms of the latencies, although only the analysis for the logarithms of the latencies are reported as the two data types did not affect the significance levels and the logarithms were more normally distributed. Carrying out an analysis on both the latencies and log transformed latencies was done as an exploratory technique to check for the effects of log transforming the data in terms of resulting significance levels. There were no differences in significance between the two sets of data, and since the convention in previous EAST studies is to analyze using log transformed data this convention was followed for the remaining of the experiments presented in this thesis.

A 2 (Experiment Half) x 2 (Stimulus Valence) x 2 (Extrinsic Valence) repeated measures model was carried out and was found to have one significant term, this was the interaction between stimulus valence and extrinsic response valence: $F(1, 27) = 68.9$, $p<0.001$. The interaction between stimulus valence and extrinsic response is significant at the $p<0.001$ level, the De Houwer (2003) paper indicates that this interaction is the most crucial in terms of which interaction to have as significant. It confirms the interpretation from Table 1 and in turn supports the experiment hypothesis.

In the 4 (Experiment Quarter) x 2 (Stimulus Valence) x 2 (Extrinsic Valence) repeated measures model there was a significant interaction between stimulus valence and extrinsic response valence: $F(1, 27) = 68.7$, $p<0.001$. In addition there was a significant interaction between experiment quarter, stimulus valence and extrinsic response valence: $F(3, 81) = 3.57$, $p=0.02$. This suggests that in one of the quarters, the interaction between stimulus valence and extrinsic response valence is greater.

EAST scores

Given that a significant interaction was found between stimulus valence and extrinsic response valence which supports the interpretation from Table 1, the individual EAST scores are examined both to see the data patterns and whether any words do not follow the expected understanding of a positive EAST scores for a positively valenced word and vice versa. The following two graphs show the EAST scores for the individual target words for both RT and accuracy data, standard error bars are plotted (the standard error does not appear to fluctuate much due to the small values).

Figure 2. Accuracy EAST scores for coloured words

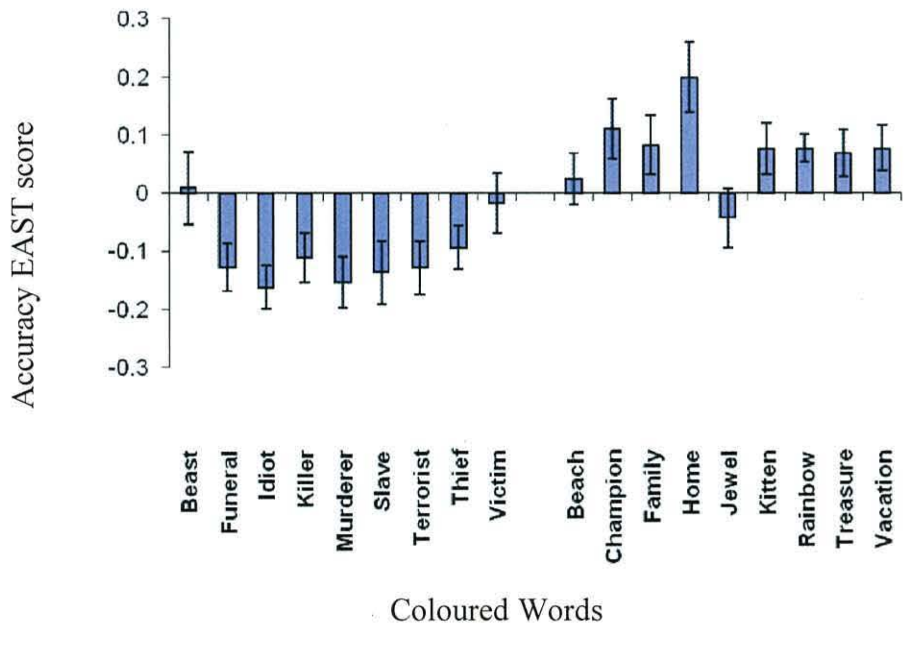
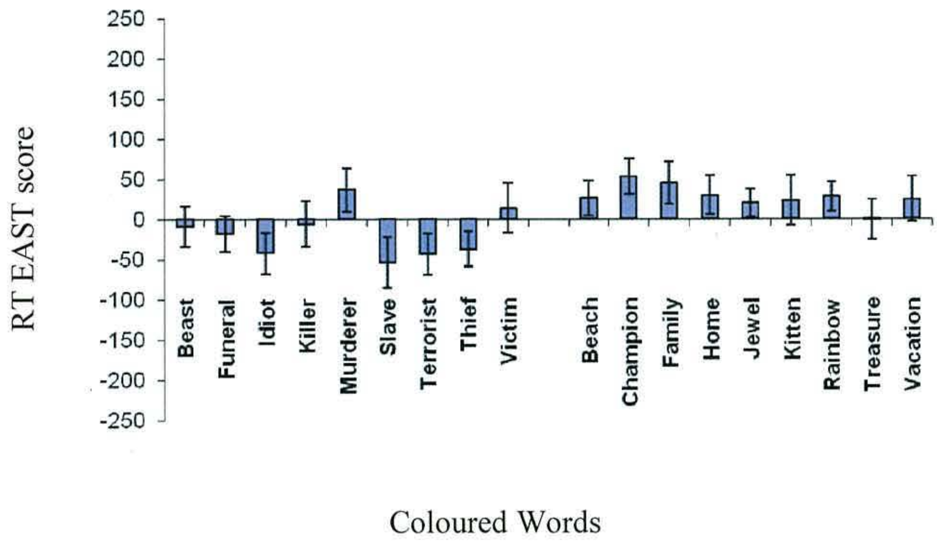


Figure 2 shows that 16 out of 18 of the words scored in the expected valence direction, all except "Jewel" and "Beast." Also, the words "Victim" and "Beach" have scores close to zero although they are in an expected direction on a priori grounds.

Figure 3. RT EAST scores for coloured words

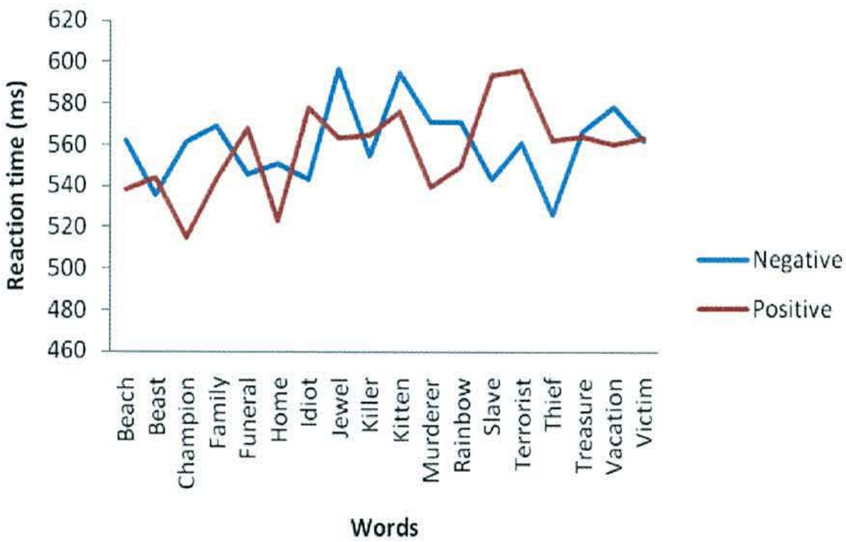


The RT data tells a similar story, in which 15 out of 18 of the words scored in the expected valence direction. Exceptions are "Victim," "Murderer," and "Treasure." The words "Slave" and "Champion" show the largest EAST scores indicating the strongest effect with these two words. When looking at Figures 2 and 3 it is interesting to note that the words which do not yield results in an expected direction differ between the two graphs (accuracy and reaction time).

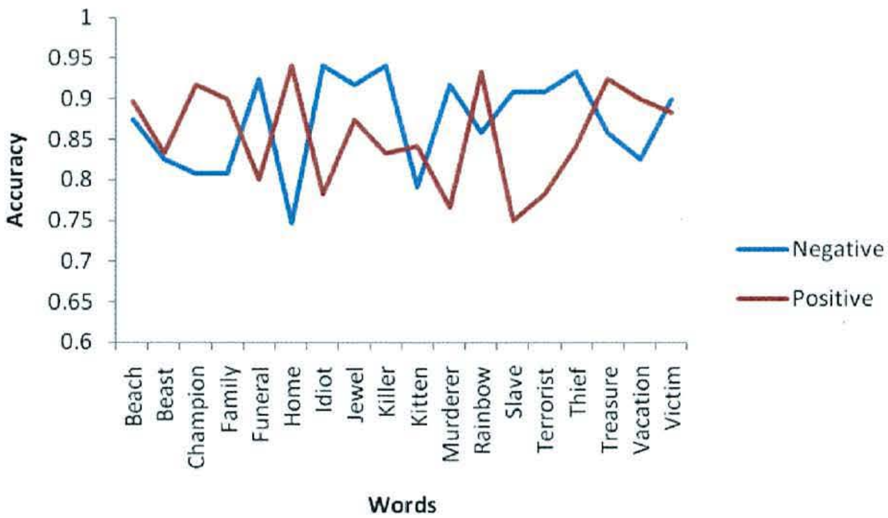
The previous two figures show which direction the words are in terms of EAST scores, either positive or negative. The EAST score is created by subtracting either the accuracy or RT data for trials where the extrinsic response valence matched the intrinsic stimulus valence from those where they did not concur. The next set of analysis examines the RT and accuracy data before it is turned into an EAST score, examining the mean responses for each word when either an extrinsically positive or negatively mapped key was used as the response. This provides information both about what was influencing the significant difference between stimulus valence and extrinsic response key.

Analysis of individual coloured words

T-tests were carried on the individual words in order to interpret whether the results were due to any particular stimulus words, similarly to find out whether any of the stimuli prevented an overall significance. The following figure (figure ??) shows the mean reaction times for each coloured word gathered from data from all participants which were included in the above analysis once the criteria had been met. The figure shows the mean reaction time for each word when either the key pressed was mapped as a negative or a positive key. It is expected that if the word was associated with a positive concept (such as "champion") then reaction times would be quicker when the response key was also positive. It is worth remembering that the participant is responding to the colour of these words not their meaning and that the keys are only positive and negative in nature by the mapping created in response to the previously seen white words. These further t-tests were carried out on both the RT and accuracy data and the words which produced a significant difference. The results are presented initially for the RT data then accuracy data.



Only two words reached significant difference, these being champion t(27) = 2.06, p<.05 and slave t(27) = 2.57, p<.05. These two words are the same words which show the largest EAST scores for individual words.



Many more of the words reached significance for the accuracy data, which does include the two words which reached significance from the RT data (champion and slave). The following words reached significance based on their accuracy data:

Champion t(27) = -2.09. p<.05
Funeral t(27) = 3.04, p<.005
Home t(27) = 3.28, p<.001
Idiot t(27) = 3.90, p<.001
Killer t(27) = 2.54, p<.05
Murderer t(27) = 3.17, p<.005
Rainbow t(27) = -2.19, p<.05
Slave t(27) = 2.79, p<.01
Terrorist t(27) = 2.81, p<.005
Theif t(27) =2.48, p<.01
Vacation t(27) = -2.07, p<.05

Interestingly the words "Jewel" and "Beast" do not reach significance levels. These two words are those which do not show an expected direction of valence when looking at the EAST data for individual words.

There is little consistency between the data provided using RT scores and accuracy scores, in an effort to address this a D600 EAST score was then calculated for each person for both trials with a positive and a negative stimulus, this analysis allows the interpretation of how the word (although it is the colour of the word that is being responded to) has affected the key press speed and accuracy.

D600 analysis

For trials with a positive stimulus, the mean D600 EAST score was D=0.33 indicating that indeed the positive stimuli were implicitly responded to as being positive (i.e. the word meaning was affecting responses even though it was meant to be ignored). For trials with a negative stimulus, the mean D600 EAST score was D=-0.29, the negative score confirming that the negative stimuli were implicitly responded to as being negative. A two sample t-test was carried out between groups when the extrinsic and intrinsic responses were both consistent and when they were not. The two groups were found to be significantly different $t\ (15.8) = 6.28$, $p<0.001$ again adding support to the initial hypothesis although this test only confirms that the two groups were different, not whether a congruent extrinsic and intrinsic response will increase RT and accuracy scores compared to when they are incongruent.

A one sample t-test was performed on the resulting D600 EAST scores for trials that had a positive stimulus and trials that had a negative stimulus separately. T-tests indicated that participants responded to a negative stimulus when the extrinsic response was also negative, $t\ (27) = -4.59$, $p<0.001$, similarly t-tests point towards the explanation that participants tend to respond more quickly and accurately to positive stimuli when the extrinsic response was also positive, $t\ (27) = 5.18$, $p<0.001$.

D600 EAST scores were then calculated for each word. As was the case earlier when looking at the EAST scores, without examining the individual scores for each word no conclusion can be drawn about where the root of the significance lies. An individual D600 score analysis on each word is key to understanding whether the implicit nature of the word/image has affected the speed and accuracy of the key presses (those which are explicitly mapped), and where the significant effects found in the ANOVA reported earlier are from, after taking into account both the RT and accuracy scores.

Figure 4 shows the mean D600 EAST score for each word

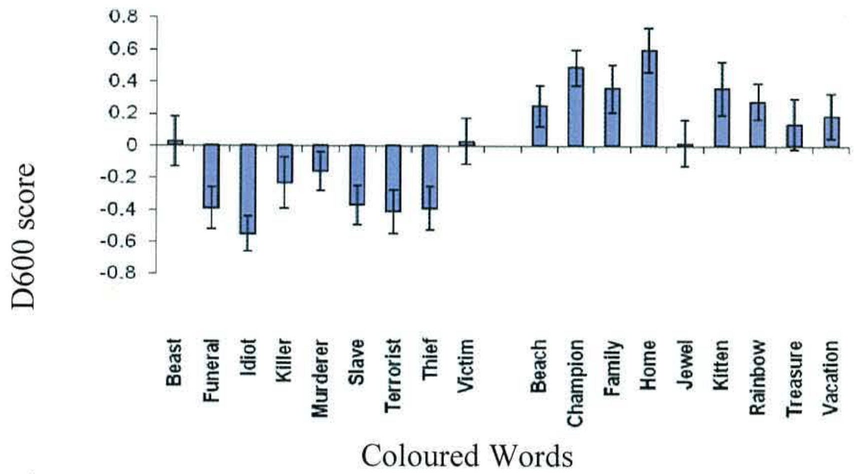


Figure 4

Figure 4 shows that out of 18 words, the D600 EAST score is in the expected direction for 16 of them. We would expect that Beast and Victim would have a negative D600 EAST score; however they both have positive scores, this could be as a result of the word choice which is discussed at the end of Experiments 1, 2 and 3 along with how this issue may have influenced the overall response in each of the first three experiments.

Experiment 2

Experiment rationale

As mentioned earlier, Experiment two uses of new version of the EAST that instead of using a font-colour change to indicate the task, instead uses coloured frames as the task indicator. This version of the task (that incorporates a frame) is what shall hereafter be referred to as the FEAST (Framed Extrinsic Affective Simon Task). It was thought that this spatial separation of the stimulus and target may increase the effect size as it was hoped that by directing attention away from the stimulus any influence that the stimulus may have upon key press will likely be more implicit in nature. In addition, this change was undertaken to create a more flexible version of the EAST paradigm – specifically, one that would be able to use pictorial stimuli.

Hypothesis

The main hypothesis is that the FEAST paradigm would yield the same general results as the traditional EAST paradigm. Thus, in terms of specific hypotheses, they were the same as that in Experiment 1, i.e. that we would find faster responses and fewer errors when the extrinsic (response) valence matched the implicit (stimulus) valence.

Method

Participants

The same participants were used as for Experiment 1.

Materials and Measures

In this study the stimuli used are the same words as used in Experiment 1.

Procedure

Previously, in Experiment 1, participants classified a) white words according to their valence, and b) coloured (blue/green) words according to their colour (using the keys that were also extrinsically linked with a positive or negative valence). Experiment 2 differs from Experiment 1 in one primary way: the word is now always printed in black, but it is surrounded by a frame that is coloured white, blue, or green. The frame served as task-indicator and was presented around the periphery of the screen and with a width of 5 mm.

The frame indicates the required classification task. When the frame is white, participants must classify the word's valence, that is, whether it is positive or negative. When the frame is green or blue, participants must classify its colour, whether the frame is green or blue. Both the frame and word remain on screen until a response is made.

Design and analysis

We analyzed the data using the same procedures as described in Experiment 1.

Results

Overview

All 30 participants were included in this analysis as there were no missing data. Furthermore, no participants showed an unusually high number of latencies less than 300ms, however, there were two latencies greater than 10000ms (59381ms and 14554ms), both of these were eliminated from the analysis.

Untransformed RT and accuracy data

Table 3 shows summaries of the mean reaction time for correct responses and the mean error rate in Experiment 2 as a function of Extrinsic Valence and Stimulus Valence:

Stimulus Valence	**Extrinsic Response Valence** Positive	Negative
Positive		
Reaction Time (ms)	574.8 (106.3)	590.9 (103.7)
Percentage of Errors	10.4 (10.1)	14.8 (11.1)
Negative		
Reaction Time (ms)	587.1 (102.6)	582.7 (106.7)
Percentage of Errors	14.6 (11.5)	8.52 (8.25)

Table 3

Tables 3 and 4 show that on average, the reaction times were greater when the stimulus valence was not the same as the extrinsic response valence. However, the difference in reaction times for trials with a positive extrinsic response and a negative stimulus, and trials with a negative extrinsic response and a negative stimulus is small when compared to trials with a positive stimulus. Furthermore, participants made more errors when the stimulus valence was not the same as the extrinsic response valence. These mean results support the hypothesis stated that RT and accuracy data would increase on trials were the extrinsic response and intrinsic valence concur with each other.

Table 4 shows summaries of the mean reaction time for correct responses and the mean error rate in Experiment 2 as a function of Extrinsic Valence, Stimulus Valence and Experiment Half:

Stimulus Valence	**Extrinsic Response Valence**			
	Positive **Experiment Half**		Negative **Experiment Half**	
	1	2	1	2
Positive				
Reaction Time(ms)	577.2 (100.6)	572.7 (119.3)	587.7 (124.1)	593.8 (93.7)
Percentage Errors	10.6 (10.3)	10.2 (12.6)	13.7 (12.3)	15.9 (12.2)
Negative				
Reaction Time(ms)	589.7 (115.2)	584.1 (97.5)	588.5 (114.5)	576.7 (104.8)
Percentage Errors	15.6 (13.1)	13.6 (12.1)	8.33 (8.48)	8.71 (9.30)

Table 4

Table 5 shows that the mean reaction times and error rates do not really change between experiment halves. However, in most occasions the variance of the reaction times is smaller in the second half of the experiment. Again this result s similar to that found in Experiment one where there was a small reduction in the time taken to respond and in errors made as the experiment progressed. As discussed in Experiment 1 this was expected and not deemed a problem as the differences between the two halves is small.

ANOVA on latencies

In order to test the significance of the interpretations containing the untransformed data two models were fit to the reaction times including the penalised error latencies.

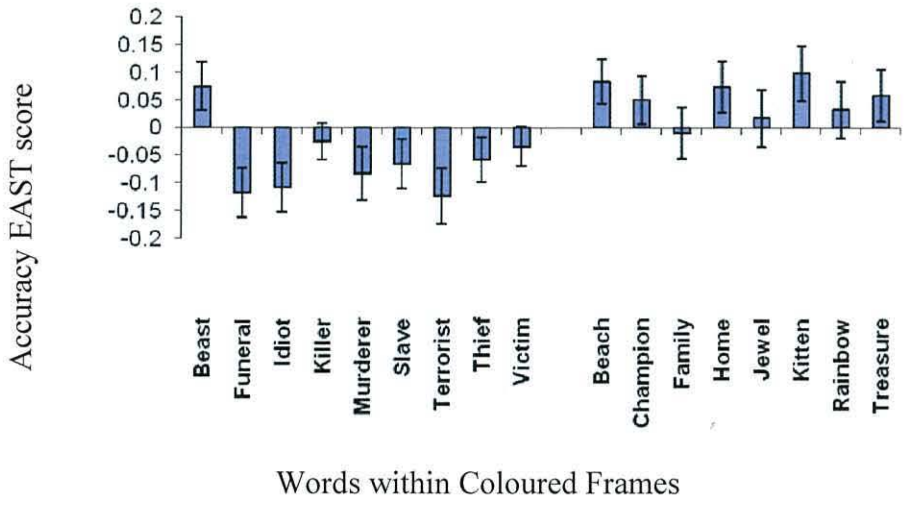
The 2 (Experiment Half) x 2 (Stimulus Valence) x 2 (Extrinsic Valence) repeated measures model only had one significant term, this was the interaction between stimulus valence and extrinsic response valence: $F (1, 29) = 21.5$, $p<0.001$. This interaction is significant as was the case with Experiment 1, not only is this significant interaction important in terms of what De Houwer (2003) stated about the importance of the interaction (explained in Experiment 1), but it is also of particular importance since significance was found in Experiment 1. The results in this, Experiment 2, are hoped to largely mirror those found in Experiment 1 in order to provide confidence that the changed framed method has a strong foundation.

The 4 (Experiment Quarter) x 2 (Stimulus Valence) x 2 (Extrinsic Valence) repeated measures model also had only one significant term. This was the interaction between stimulus valence and extrinsic response valence: $F (1, 27) = 21.6$, $p<0.001$, this provides further weigh in support of the hypothesis. As explained in the previous Experiment whilst these levels of significance are both exciting and important, they do not explain which words are bringing about these significant effects, in order to understand this, the words need to be examined individually.

EAST scores

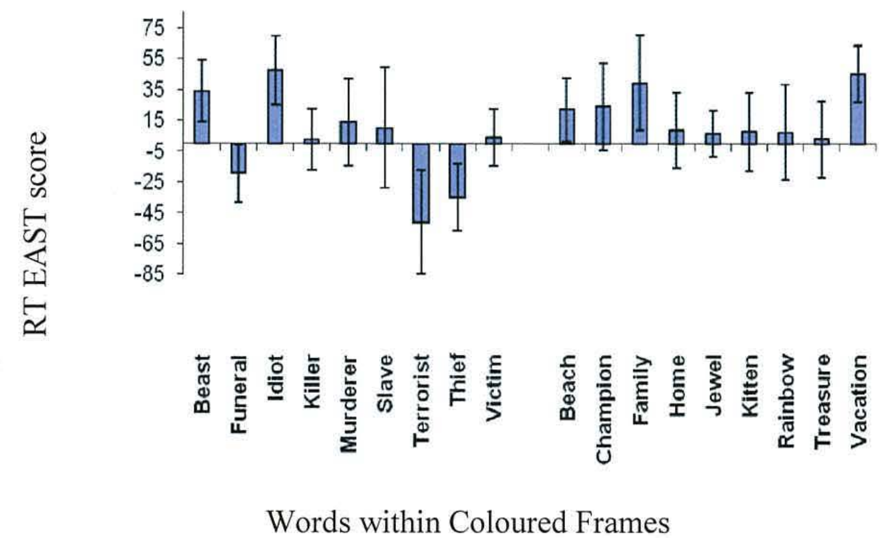
In order to follow up the significant interaction found between stimulus valence and extrinsic response valence the individual EAST scores are examined as was the case in Experiment 1.

Figure 5. Accuracy FEAST scores for coloured words within frames



In terms of accuracy, 16 of the 18 words indicate expected valence direction. Those which do not display the expected direction of positive or negative based on a priori grounds are "Beast" and "Family." The words "Kitten" and "Terrorist" are the words with the strongest EAST accuracy score.

Figure 6. RT FEAST scores for words within coloured frames

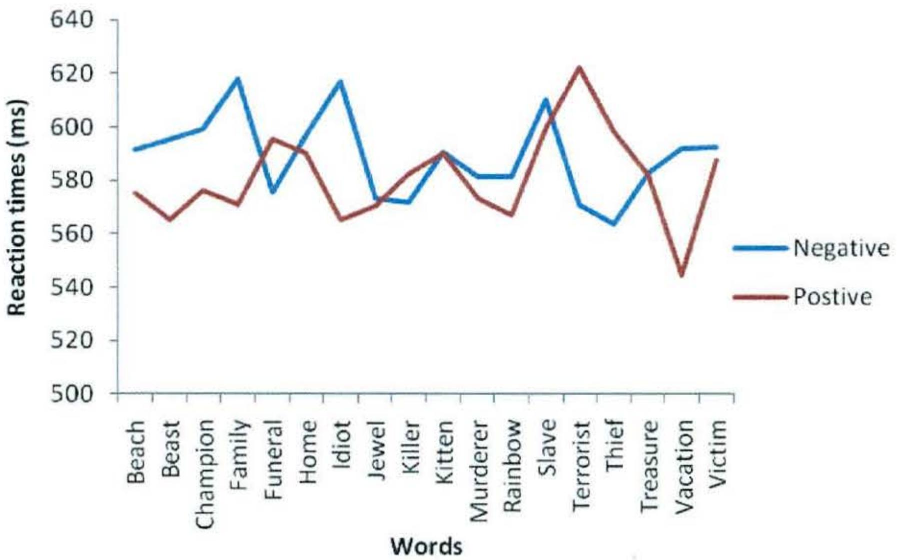


In terms of RT, the story is not as clear as for ACCURACY. Here, 12 out of the 18 words indicate expected valence direction; interestingly all of the positive words are correctly identified by the FEAST score, with only 3 out of the 9 negative words showing a FEAST score in the expected direction. The word "Terrorist" is the word with the strongest EAST score from the negative words as was the case with the accuracy EAST scores; however for the positive words this was different to that produced by the accuracy scores with "Vacation" now being the strongest.

Given that the D600 score is an amalgamation of the RT and accuracy scores, the next analysis uses the RT and accuracy data before it is turned into an EAST score. This analysis is carried out in order to extract where the influence is which brings about the significant difference between stimulus valence and extrinsic response key.

Analysis of individual coloured words

In Experiment 1 an explanation is made regarding how this analysis was carried out, since the same applies for Experiment 2 it shall not be repeated.

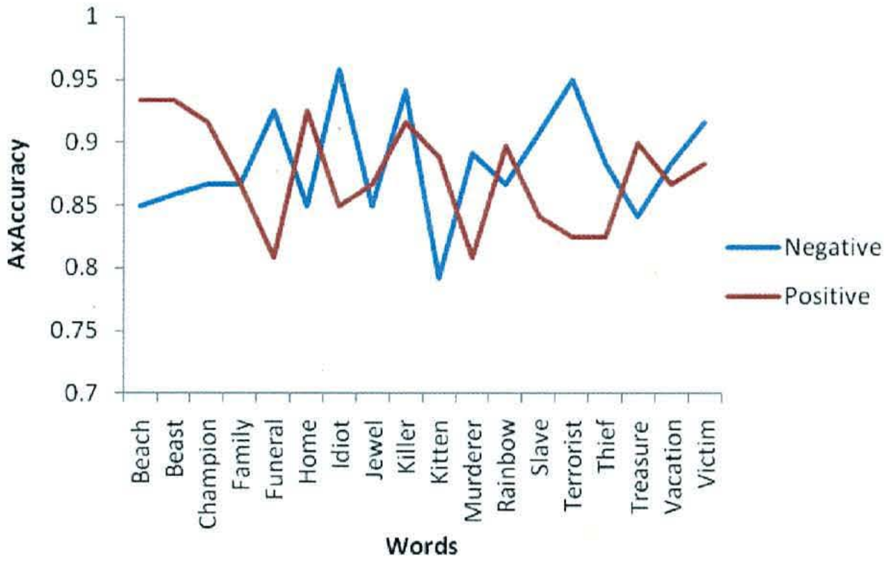


There are four words which reached levels of significant difference, these were:

Family t(29) = 2.85, p<0.01
Idiot t(29) = 2.32, p<0.05
Terrorist t(29) = -.92, p<0.01
Vacation t(29) = 2.56, p<0.05

Interestingly these were no significant values for the words "champion" or "slave".

These two words have significant values in the EAST study (Experiment 1), so it would seem that the framed version is changing how participants repond to the words.



The following four words reached significant levels of difference:

Beach t(29) = -2.07, p<0.05
Funeral t(29) = 2.63, p<0.01
Idiot t(29) = 2.44, p<0.05
Terrorist t(29) = 2.48, p<0.05

Two of the words are repeated in both the accuracy and RT tests with significant levels, these are "Idiot" and "Terrorist". Since the data from the accuracy and RT data are quite inconsistent with each other, D600 scores were calculated to take into account both the RT and accuracy data with one score for each word calculated.

D600 analysis

A D600 EAST score was then calculated for each participant for both trials with a positive and a negative stimulus. For trials with a positive stimulus, the mean D600 EAST score was D=0.18 supporting the view that the stimuli were indeed perceived as positive due to the positive nature of the D600 score. Participants tended to respond more quickly and accurately when the extrinsic response was also positive, $t(29) = 3.01$, $p<0.01$. For trials with a negative stimulus, the mean D600 EAST score was D=-0.16, supporting that the stimuli were perceived as negative indicated by the negative D600 score. Participants tended to respond more quickly and accurately when the extrinsic response was also negative, $t(29) = -2.42$, $p=0.02$. As was the case for Experiment 1 the results from the D600 scores produced faster and more accurate responses when the stimulus valence and the extrinsic response key were either both negative or both positive. A two sample t-test suggests that there is evidence to suggest that the two groups (when the stimulus valence and extrinsic responses are congruent with each other versus incongruent) have different means, $t(9.21) = 3.39$, $p<0.01$. D600 EAST scores were then calculated for each word.

Figure 7 shows the mean D600 EAST score for each word.

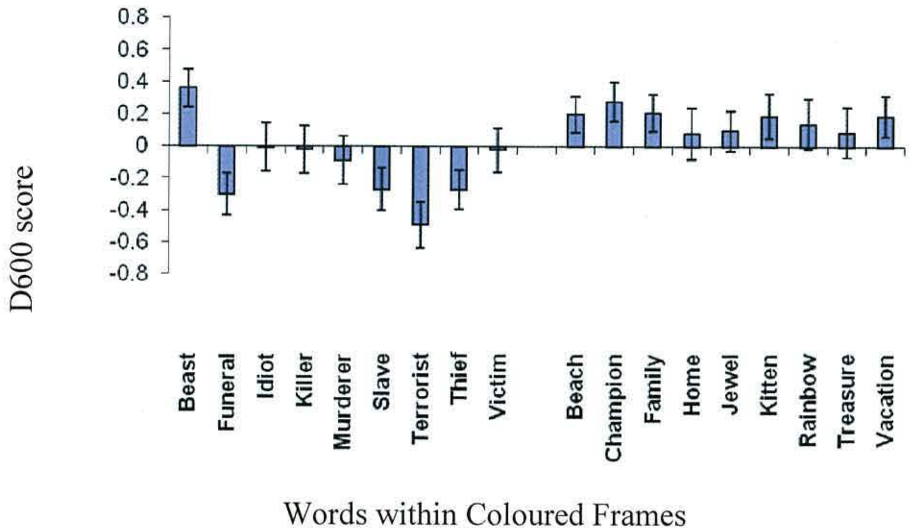


Figure 7

Figure 7 shows that out of 18 words, the D600 EAST score is in the correct direction for 17 of them. We would expect that Beast would have a negative D600 EAST score; however it has a positive score, not only positive but also the largest score when comparing all words. Interestingly "beast" was also one of the two words not in the expected direction for the D600 scores in Experiment 1.

The results of Experiment 2 suggest that this new version of the EAST – that uses a frame as the task indicator (the FEAST) – is capable of measuring implicit reactions to stimuli. Our supposition is that the separation of task-indicator from the stimulus does not substantively influence the basic cognitive processes required to perform the task. Specifically, at an abstract level, on each trial the participant is simply required to know what the basic stimulus is (i.e. what word they are to classify) and the dimension on which it must be classified (valence or colour). As expected, there is no reason why these two pieces of information need to be conveyed by the same stimulus (a word in a colour) and that separating them spatially can yield substantially the same results while also creating the possibility to use more varied stimuli (e.g. pictures). While there were some stimuli that did not have results in the expected direction, the number of such mis-diagnosed stimuli was no greater than that found with the traditional EAST paradigm. And, the important interaction of stimulus valence and extrinsic response was significant in both experiments. In short, overall the paradigm seems to work surprisingly well. Thus, following the analysis of data from Experiments 1 and 2 there was sufficient confidence to go forward and test the FEAST using pictorial stimuli in Experiment 3 to find out whether similar results to these two previous experiments could be replicated.

Experiment 3

Experiment Rationale

One of the major reasons for developing the framed version of the EAST (i.e. the FEAST) was to create the possibility of including pictorial stimuli in this implicit measurement paradigm. Allowing for pictorial stimuli provides much greater flexibility because some concepts are not readily described using words. Furthermore, it would provide a useful method for testing consumer related 'real world' questions which was an important part of this PhD work due to its sponsor. Experiment 3 uses pictorial stimuli by using previously rated pictures from the International Affective Picture System (IAPS) (Lang, Bradely and Cuthbert, 1999).

Method

Participants

The same participants were used as in Experiment 1.

Materials and measures

In this study the stimuli used are pictures selected from the IAPS database, using 18 coloured (blue/green) frame pictures and 22 white frame pictures. Pictures were set to cover 90% of the screen space, with a 15 mm gap between the picture and the 5 mm task-indication frame. Pictures were selected using the pre-ratings from the SAM scores used in the creation of the IAPS database, representing a high to low valence score.

White pictures

As with the traditional EAST (and with the FEAST introduced in Experiment 2), a set of stimuli are required to anchor the explicit response dimension. While traditionally these stimuli (typically words) are used to capture the "good"/"bad" dimension, they really could be selected instead to capture any two sets of concepts. While we get more ambitious and explore other dimensions in later experiments, for the present experiment, we decided to follow the more traditional classification dimension and used pictures that were meant to capture "positive" and "negative". Specifically, the pictures presented within white frames to capture the "positive" dimension have the IAPS descriptions: baby; beach; bunny; country; cub; dolphins; happy child; polar bear. The pictures chosen to capture the "negative" dimension have the IAPS descriptions: angry face; cocaine; fight; flies on pie; insects; roach; scream; shooting; puppy; seal; sunset; skulls; tornado; toxic.

Coloured pictures

The pictures presented in the coloured frames are those to which we wish to measure implicit reactions. In more "applied" settings, these pictures might be of an unknown valence and we would use the FEAST to quantify their valance. However, for this first use of pictures within the FEAST paradigm we chose to use pictures that had already been classified onto known dimensions. Thus, we again used (a different set of) IAPS images. The images were chosen to span the "positive"/"negative" dimension – so that we could assess how well FEAST could measure implicit reactions to a range of valence extremities. Specifically, pictures presented within coloured frames have the IAPS descriptions; attack; boat; carcrash; bride; flood; fireworks; gun; flowers; knives; monkies; missiles; playing children; missiles; sky; snake; sports car; wires; treat (see appendix for images). Table 5 shows the SAM scores for each of the coloured images used. The scores centre around the midpoint of the SAM scale (1-9).

Picture description	Valence	Arousal	Dominance
Snake	3.46	6.87	3.08
Prison	4.04	4.91	3.77
Missiles	4.39	4.88	4
Wires	4.07	4.36	4.05
Flood	3.85	4.83	4.37
Attack	4.09	4.6	4.58
Gun	4.01	4.87	4.81
Sky	7.61	4.51	5
Car Crash	4.51	3.87	5.2
Knives	4.53	3.88	5.29
Fireworks	7.8	5.59	5.56
Monkies	7.62	5.12	6.07
Boat	7.48	4.74	6.18
Flowers	7.36	3.2	6.21
Playing Children	7.57	5.83	6.41
Bride	7.64	5.59	6.53
Sports Car	7.32	4.93	6.54
Treat	7.69	5.14	6.58

Table 5.

Procedure

In a change from Experiment 2, the stimuli are now pictorial, not words. All other elements of the procedure remain unchanged.

Design and analysis

Analysis was carried out using the same procedures as described in Experiments 1 and 2.

Results

Overview

All participants were included in the analysis, furthermore, all data were included as there were no unusually short latencies (less than 300ms) and no latencies greater than 10000ms.

Untransformed RT and accuracy data

It is hoped that the untransformed data will confirm that participants respond faster and more accurately when the stimulus valence in congruent with the extrinsic key press. Remembering that for the coloured framed pictures analysed that participants are responding to colour and the extrinsic valence of the key is created through mapping by responding to the white framed pictures. Table 6 shows summaries of the mean reaction time for correct responses and the mean error rate in Experiment 3 as a function of Extrinsic Valence and Stimulus Valence:

Stimulus Valence	**Extrinsic Response Valence** Positive	Negative
Positive		
Reaction Time (ms)	571.1 (76.9)	592.6 (97.4)
Percentage of Errors	11.4 (13.9)	18.9 (11.9)
Negative		
Reaction Time (ms)	576.9 (101.1)	582.8 (88.8)
Percentage of Errors	20.1 (13.7)	10.7 (10.4)

Table 6

Tables 6 and 7 indeed do show that, on average, mean reaction times for trials with a positive stimulus valence were faster when the extrinsic response valence was also positive. However, unfortunately this is not the case for trials with a negative stimulus valence, where reaction times were slower when the extrinsic response valence was also negative (although the RT data is very close). In the case of accuracy data participants were more accurate when the extrinsic valence matched the stimulus valence.

Table 8 shows summaries of the mean reaction time for correct responses and the mean error rate in Experiment 3 as a function of Extrinsic Valence, Stimulus Valence and how experiment half changed participant responses, if at all:

Stimulus Valence	**Extrinsic Response Valence**			
	Positive		Negative	
	Experiment Half		**Experiment Half**	
	1	2	1	2
Positive				
Reaction Time (ms)	575.8 (85.7)	564.8 (78.8)	588.2 (82.8)	598.0 (117.1)
Percentage of Errors	11.9 (12.7)	10.9 (17.0)	19.1 (11.4)	18.7 (14.0)
Negative				
Reaction Time (ms)	578.8 (106.5)	573.8 (102.3)	584.5 (97.6)	578.7 (94.4)
Percentage of Errors	19.0 (14.1)	21.3 (15.4)	11.1 (13.8)	10.4 (9.08)

Table 7

Table 8 shows that the mean reaction times and error rates did decrease marginally between experiment halves, except the reaction times for trials where the stimulus valence was positive and the extrinsic response valence was negative. The standard deviations of the reaction times also appeared to decrease between experiment halves, again, except for trials where the stimulus valence was positive and the extrinsic response valence was negative. The standard deviation for error rates also changed between experiment halves. As was the case for Experiments 1 and 2 are degree of change between the halves was expected and would be checked for significance as part of the following ANOVA tests.

ANOVA on latencies

Two models were fit to the reaction times including the penalised error latencies. The 2 (Experiment Half) x 2 (Stimulus Valence) x 2 (Extrinsic Valence) repeated measures model only had one significant term, this was the interaction between stimulus valence and extrinsic response valence: $F (1, 29) = 42.7, p<0.001$. This interaction was very important, not only is it deemed vital by DeHouwer (2003), but it was found to be significant in both Experiments 1 and 2, therefore, the interaction was expected to also reach significance if any confidence in a pictorial method was to be achieved.

The 4 (Experiment Quarter) x 2 (Stimulus Valence) x 2 (Extrinsic Valence) repeated measures model also had only one significant term. This was the interaction between stimulus valence and extrinsic response valence: $F (1, 27) = 42.8, p<0.001$. The lack of a significant level of difference being found between the experiment halves was important as it indicated that there had been a sufficient number of practice trials which prevented significant levels of practice effects, and adds weight to the argument that the results are not a product of learning/familiarity since if this were the case it would be expected that significance levels would be found between the halves.

FEAST scores

The significant interaction between stimulus valence and extrinsic response are followed up by examining the FEAST scores. Figure 8 shows accuracy FEAST scores for images presented within coloured frames. Images on the left half of the graph are images that have negative IAPS scores, image on the right have positive scores. Clearly, the FEAST scores agree in all cases with these classifications.

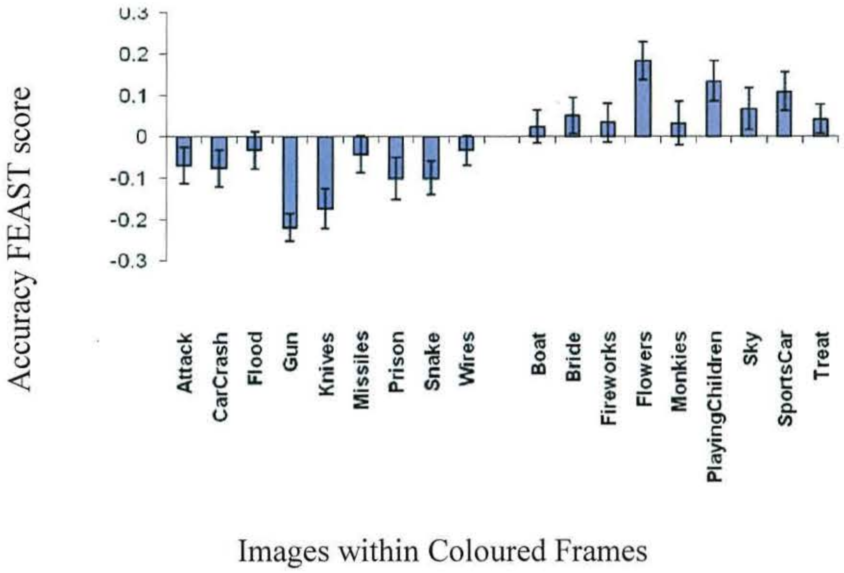
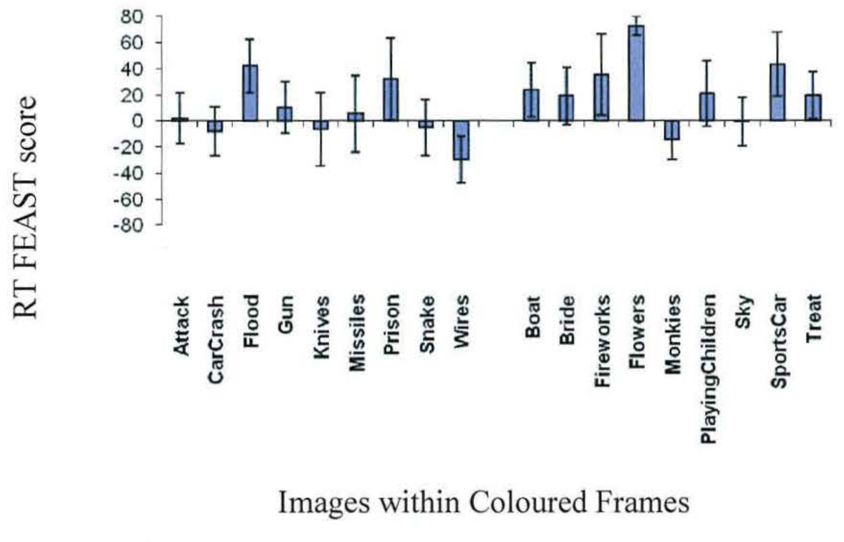


Figure 9 RT FEAST scores for images within frames

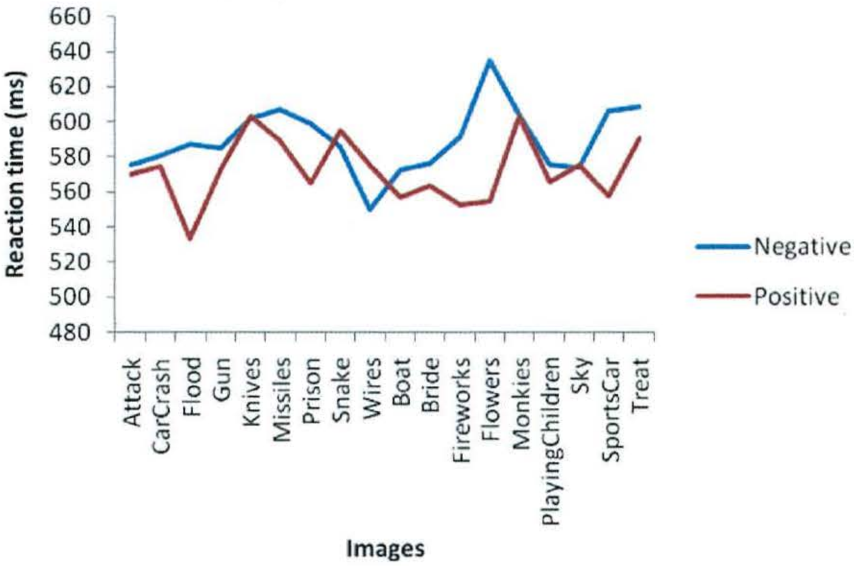


The accuracy data demonstrates that for all images the FEAST was able to capture the valence (as determined on *a priori* grounds based on the SAM scores). However, the RT data only showed 11 out of 18 of the words in the expected valence direction. Once again, as was shown to be the case in Experiments 1 and 2, the interaction between stimulus valence and extrinsic response valence was significant $p<0.001$.

The consistency with which this interaction continues to be significant across each of the Experiments 1, 2 and 3 is very encouraging.

Analysis of individual coloured images

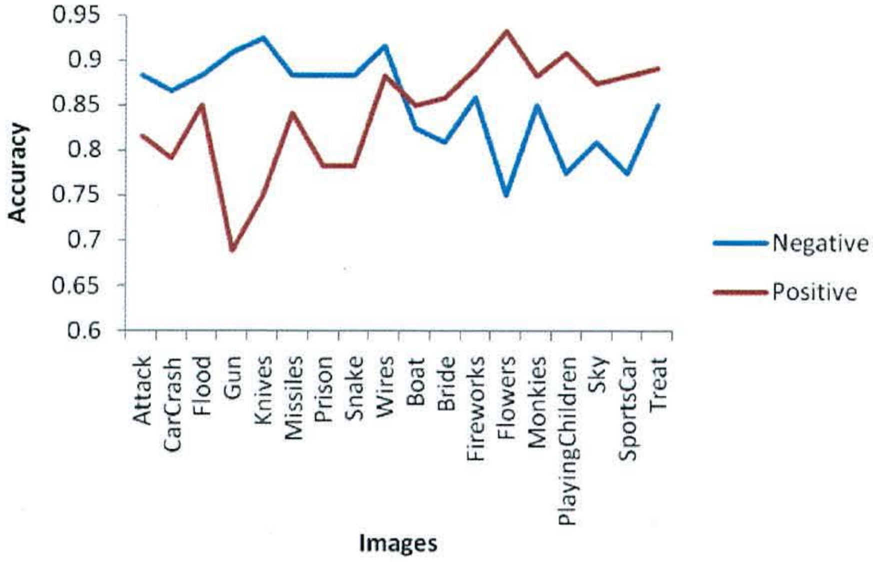
The next analysis used examines the RT and accuracy data before it is turned into a FEAST score. In other words, it looks at the reaction time required when the response is extrinsically positive and when the response is extrinsically negative.



Only two images reached a level of sinificance:

Flood t(29) = -2.43, $p<0.05$
Flowers t(29) = 2.99, $p<0.001$

It is interesting to note that when the FEAST scores for RT are examined then "Flood" indicates a positive valence. The signicance level found when looking the RT scores and the extrinsic responses does not indicate a valence direction mearly that the two key responses types were significantly different from each other.



For accuracy data five of the images reached a level of significance, flowers was the only image repeated in both the accuracy and RT levels of significance.

Gun $t(29) = 3.75$, $p<0.001$
Knives $t(29) = -3.43$, $p<0.001$
Snake $t(29) = 2.56$, $p<0.01$
Flowers $t(29) = -3.34$, $p<0.005$
Sports car $t(29) = -2.28$, $p<0.05$

It is interesting to see where the basis for the significant interaction between stimulus valence and extrinsic response may originate from, and there are certainly indications that the FEAST is correctly indentifying the expected valence of the images. However, often the results are conflicting, the FEAST scores looking at RT in terms of predictive ability of 'correct' valence are not much higher than at chance. Due to this shortfall, the D600 scores pulls together some of the unpredictability into one more meaningful score.

D600 analysis

The FEAST scores as with the previous two experiments are very inconsistent when the RT scores and accuracy scores are examined. The D600 incorporates both the error rates and reaction times within one analysis, the D600 EAST score was then calculated for each person for both trials with a positive and a negative stimulus. For trials with a positive stimulus, the mean D600 EAST score was D=0.25. Participants tended to respond more quickly and accurately when the extrinsic response was also positive, t (29) = 3.21, p<0.005. For trials with a negative stimulus, the mean D600 EAST score was D=-0.19. Participants tended to respond more quickly and accurately when the extrinsic response was also negative, t (29) = -2.99, p<0.01. The positive D600 score for trials with a positive stimulus and a negative D600 score for those with a negative stimulus add support to the assumption that the images used were correctly assigned as positive or negative on a priori grounds. Due to the significance levels when the stimulus valence was congruent with the extrinsic response valence the D600 EAST scores were then calculated for each picture in order to understand better where the significance resulted from.

Figure 10 shows the mean D600 EAST score for each picture.

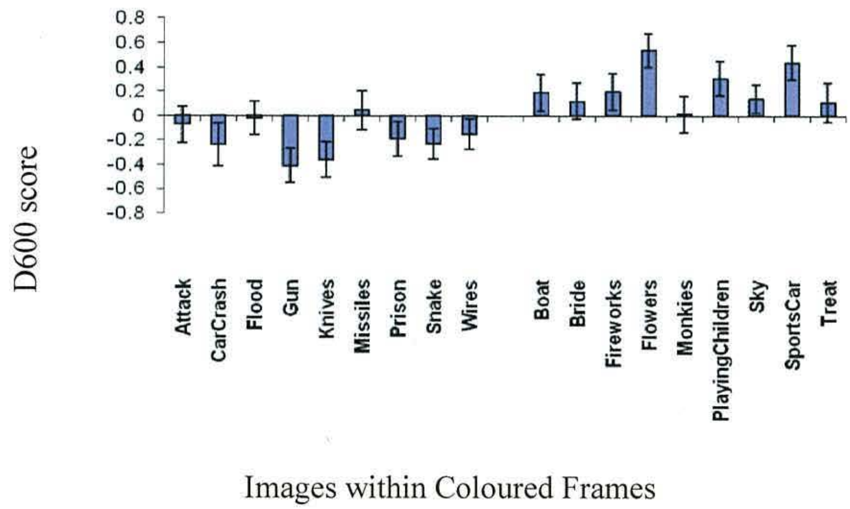


Figure 10

Figure 10 shows the D600 EAST score for each of the IAPS images used in the experiment. Out of 18 pictures, the D600 EAST score is in the correct direction for 17 of them. "Missiles" was expected to have a negative D600 EAST score; however it has a positive score. Interestingly two participants had reported verbally as part of the debriefing process that they had been unaware that the picture was of missiles but had viewed the image as containing space rockets. A two sample t-test suggests that there is evidence that the two groups have different means, $t(15.8) = 5.43$, $p<0.001$.

Discussion of Experiments 1-3

In terms of whether these three Experiments reported agree with the expectations and finding from the DeHouwer (2003) paper on which this work was based, DeHouwer expected that performance would improve when participants were required to respond (based on the colour classification) with the key that extrinsically matched to the intrinsic response. Since here the D600 was chosen as a method of analysis and the stimulus choice varied, the results are not directly comparable. However, it is certainly the case that participants tended to respond faster and more accurately when the extrinsic response key matched the intrinsic response.

It was hoped that by separating the task indicator (the frame) from the stimulus that less attention would be allocated directly to the stimulus and this in turn would lead to larger and perhaps "more implicit" effects. It is extremely difficult if not impossible to say beyond doubt whether this is what was happening, but by examining the D600 results from the EAST experiment 1 and FEAST experiment 2 which used the same words, the D600 scores (either positive or negative) matched the expected direction more often in the FEAST version than in the EAST, although this only differed by one word out of eighteen.

It was often the case that the accuracy scores were in the direction expected whereas the RT scores were not. Much consideration has gone into why this should be the case as it implies that the faster that a participant responded the less likely that they were to respond to the word in an expected manner, when they responded slower and more accurately the results more often reflected what was expected. It was difficult to rectify this since the fast responses helped to ensure a more implicit response to the coloured frames, with a slower response this runs the risk of more consideration being given towards the word or image, when an immediate responses towards the colour reflected less processing of the image or word and a the response being made towards the colour.

As a result of discussions with participants following the experiments it is possible that the choice of words was flawed with words being internalized in a different direction to that intended. For example, the word KILLER could be implicitly attributed as either positive or negative, especially depending on what culture the participant was familiar with. Carrying out ratings on each word by each participant after the FEAST experiments were complete gave some indication of whether a participant was rating a word in a different direction to the majority of the other participants and if this had been the case that participant's data was removed from any analysis.

It was the intention that the same participants would run each of the three experiments to allow for more direct (subject-by-subject) comparisons of the results. This comparison was largely carried out by comparing the D600 scores and their directions for each of the experiments. It was hoped that all three experiments would demonstrate the same implicit directions adding weight to the idea that each of the modifications did not alter the effectiveness of the implicit testing. Each of the participant's data had already been looked at in detail as part of each analysis so this comparison was carried out on the mean D600 scores once all analysis was completed. The only small problem between these comparisons is that data from two participants were discarded from the analysis in Experiment 1, making it not a truly direct comparison. Also of note is that there could be no comparison between the Experiments 1 and 2 and Experiment 3 since the stimulus set varied. When comparing the word D600 scores between Experiments 1 and 2, so essentially looking at what effect moving the task indicator away from the stimulus itself to the periphery frame, it is easy to see that the word score for all positively valenced words have received a positive D600 score. For the negatively valenced words the word BEAST has a positive D600 score which whilst unexpected is welcome in terms of the frame not affecting the overall results. In experiment 1 the word VICTIM also produces a positive score, however the score is very small as is also the case for several of the words.

All the analyses clearly suggest that the FEAST is at least as sensitive to measuring implicit reactions as the EAST (Experiments 1 and 2) and that this can be extended to pictorial stimuli (Experiment 3). Thus, the results from the first three experiments provide a solid basis from which to begin further detailed experiments using the FEAST method. The next set of experiments will move on to look at the sensitivity of the FEAST method using a similar design (pictures and frames) as that used in Experiment 3.

PART 3.

THEORETICAL IMPLICATIONS AND USES OF THE FRAMED EXTRINSIC AFFECTIVE SIMON TASK

Experiment Rationale and introduction

The experiment presented in this chapter is broken into 3 parts and uses the new FEAST method described in Chapter 2 with images. The aim of Experiment 4 is to determine whether this FEAST method is sensitive enough to measure multiple dimensions of the same image. The dimensions used are the same ones used by the IAPS system, namely: valence, arousal and dominance. The experiment was not designed to study valence, arousal, and dominance per se, but instead it was designed to use stimuli from an established image database (IAPS) which was already explicitly rated and to test whether the FEAST was sensitive enough to provide different results for the same stimuli for each different dimension. In other words, the question here is: can the FEAST assess several different dimensions of the same stimulus set? As mentioned earlier, a stimulus can be represented (or rated) on many different dimensions. While this is true of words, it is perhaps particularly important when using pictorial stimuli. It is highly likely that pictorial stimuli can be represented by multiple dimensions and, within the methodology of the FEAST, it should be possible to focus the participant's attention on the relevant dimension by the selection of the proper set of "white words". Ideally, the FEAST should be sensitive enough to distinguish between these different stimulus 'attributes' providing greater flexibility for the researcher. Thus, this chapter is not introducing a new concept, but is instead of exploring the sensitivity of the pictorial FEAST paradigm. Because this is not a new area of research – but instead only an extension to the paradigm testing – the introduction section of chapter 3 covers the same issues as in this chapter.

Hypothesis

It was expected that the FEAST would be sensitive enough to distinguish the three different dimensions, for each stimulus. Thus, we expect that the FEAST scores for each stimulus will vary depending on the dimension that is being judged explicitly (i.e. based on the white-word context set). Furthermore, in order to maintain confidence in the FEAST method, a significant interaction between stimulus valence and extrinsic response was expected between each of the dimensions.

Experiment 4

Method

Participants

Thirty one psychology students (4 male, 27 female) participated as part of a course requirement. All participants had normal or corrected to normal vision. The mean age of the participants was 20.

Materials and measures

The stimuli were gathered from the IAPS database as was the case in Experiment 3 (see Appendix for images). The pictures to be presented within coloured frames (those being implicitly measured) were selected to represent a range of IAPS rated explicit scores on each of the dimensions of valence, dominance, and arousal to reflect those which were highly rated, mid-rated and low-rated (see Table 8). In other words, we selected a set of images that spanned a "chunk" (or cube) of {valence, arousal, dominance} space.

Different images were selected from those used in Experiment 3 because those in Experiment 3 had SAM scores which centred around the midpoint, whereas for Experiment 4 a larger spread of scores was used. It is worth noting that images which were often at the extreme high and low rated ends were not suitable for use in this study due to their distressing or explicit nature, this was especially so for those on the dimensions of valence and arousal.

White framed pictures

16 white frame pictures were selected. These were described in the IAPS database as parrots, gannet, cow, men, spoon, windows, flowers, nature, snake, shark, aimed gun, knife, attack, police, plane crash and sinking ship.

Coloured framed pictures

There were 14 coloured frame pictures, described as aimed gun, attack, babies, basket, beach, earth, kids, lamp, mountain1, mountain2, nature, rabbit, skier and sky. Table 8 shows each of the pictures shown within coloured frames along with the corresponding IAPS slide number and the explicit scores for each dimension. The scores for arousal run from 1 to 9, with 1 representing low arousal, and 9 representing high arousal. The scores for dominance run from 1 to 9, with 1 representing high dominance, and 9 representing low dominance. The scores for valence run from 1 to 9, with 1 representing low valence, and 9 representing high valence.

Description	Slide Number	valence	arousal	dominance
Aimed Gun	6230	2.37	7.35	2.15
Attack	6350	1.9	7.29	2.73
Earth	5890	6.67	4.6	4.15
Mountains	5700	7.61	5.68	4.59
Mountains	5600	7.57	5.19	4.93
Sky	5982	7.61	4.51	5
Skier	8190	8.1	6.28	6.14
Lamp	7175	4.87	1.72	6.47
Kids	2388	7.44	3.77	6.62
Basket	7010	4.94	1.76	6.7
Rabbit	1610	7.82	3.08	6.77
Beach	5833	8.22	5.71	6.97
Babies	2080	8.09	4.7	7.08
Nature	5760	8.05	3.22	7.49

Table 8.

Procedure

The procedure is identical to that of Experiments 3 except for one difference. Each participant completed three versions of the FEAST, each using the exact same stimuli, but the dimension on which they were categorizing white framed pictures differed. In other words, in each of the three experimental blocks, the extrinsic response dimension was different. The three dimensions used (across three blocks) were valence (positive/negative), arousal (high/low), and dominance (high/low). The order of these blocks was varied between participants to control for any practice and/or familiarity effects. To help the participants understand what was meant by each of these three dimensions, they were shown a PowerPoint presentation prior to beginning the experiment. This PowerPoint only contained the images from the white frames, not those intended to be presented in green/blue frame as it was assumed that participants would have no difficulty with the color classification task. The participants were told, for each picture, how to classify the valence, arousal and dominance (either high or low for each of the three dimensions), this desired-classification was in accordance with IAPS rated scores. The "correct" rating was given by the experimenter for each of the "white" images. No further explanation was given about what was meant by these dimensions. Thus, each individual would therefore draw their own conclusions (and form their own interpretations) as to what each dimension represented based upon the Powerpoint they had been shown.

For the purpose of helping the reader: in terms of IAPS ratings, "valence" is on the scale of pleasant to unpleasant, "arousal" is on the scale of calm to excited, and "dominance" ranges from in-control to out of control. It is important to note that no descriptions were given to the participants about what each of the dimensions represented, this was left for the participants to decide.

After each participant performed all the FEAST tasks (for the three dimensions of valance, arousal, and dominance) they then were asked to perform a brief explicit rating task wherein they were presented each stimulus (including all the "white" and the "coloured" stimuli) one at a time and were asked to rate the valence, arousal, or dominance of each image. The different dimensions were rated in separate blocks. And, across participants, the order of these explicit ratings were randomized. This was done to allow for a greater understanding of how each individual participant conceptualized the valence, arousal, and dominance of each individual image used in the FEAST task.

In this first experiment (which was run as one of the three blocks – randomized across participants), participants were asked to rate the arousal level of images appearing within white frames, and to rate the colour of any blue/green framed images that appeared. Thus, over the first set of trials (when purely white-framed images appeared) participants did two things. First, they formed both their own concept of what "arousing" was meant to signify. Recall, that the experiment did not proceed until a "correct" response was given. Thus, the concept that participants formed was actually being shaped over these trials to coincide with the concept of "arousing" that was meant to be captured by the IAPS images. In addition to this concept formation/solidification, the participants (via their responding) were also creating an extrinsic linkage between the concepts "high arousal"/"low arousal" and their respective response keys. Thus, for example, after a certain number of trials, half the participants should have associated the "q" key with "high arousing" and the "p" key with "low arousing". The other half of the participants, of course, should have formed the revers association ("p" = "high arousing"). It is this extrinsic linkage between the key and a concept/dimension that is critical for the success of the FEAST.

Design and analysis

Analysis was initially carried out using the same procedures as described in Experiment 1 including the same number and types of trials.

Results

Arousal

Overview

31 participants took part in the study, however one participant (number 18) had latencies less than 300ms for more than 10% of the trials, and so was removed from the analysis. Thus, the data from 30 participants were included in the final analysis. There was one latency greater than 10000ms (RT=11655ms), which was also removed.

Untransformed RT and accuracy data

In the previous experiments (Experiments 1, 2 and 3) a table was calculated to display both the RT and accuracy scores as a function of the stimulus valence and extrinsic response. As a result of this table, it was easy to see whether participants had responded faster or slower depending on whether the extrinsic response key matched the stimulus valence, and how this congruency/in-congruency affected their error rates. In Experiment 4 it has not been possible to produce such a table since the stimuli selected had a range of explicit SAM scores from high to low, (including several around a mid-point) so a spread of responses was encouraged. However, using arousal as an example, there was no one point at which arousal became high or low. When each participant was asked to explicitly rate each of the pictures used as part of the experiment (after completing the FEAST tasks), we found that each participant had a slightly different mid-point. Due to this individual variability, a generic table (as detailed in the previous experiments) was not produced for each of the dimensions tested in Experiment 4. However, these mean explicit scores were calculated for each individual image for each of the three dimensions covered in this chapter (in Appendix 2).

ANOVA on latencies

A 2 (Experiment Half) x 14 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures model was carried out, this had one significant term, this was the interaction between stimulus picture and extrinsic response valence: F (13,364) = 2.13, p=0.01. As mentioned before this is a very important interaction in terms of measuring the success of the method. This significance does not explain what is happening to achieve a significant result, in order to gain more insight the individual FEAST scores need to be examined.

FEAST scores

The FEAST scores were calculated for each individual image, firstly looking at accuracy scores, followed by RT scores. The chart to the right of each figure shows the explicit SAM scores for each image for the particular relevant dimension (valence, arousal, or dominance). It is important to remember that the SAM scores are explicit in nature, and that the FEAST scores are implicit so a lack of similarity between the two sets of scores is not necessarily a reflection on the FEAST. Instead, measured differences between the two techniques (implicit/explicit) might either be due to measurement error <u>or</u> to differences between the implicit and explicit impact of the images. This is an interesting theoretical question about the nature of attitudes as discussed in the introductory chapter and to which we shall return later in the thesis.

Description	arousal
Aimed Gun	7.35
Attack	7.29
Earth	4.6
Mountains	5.68
Mountains	5.19
Sky	4.51

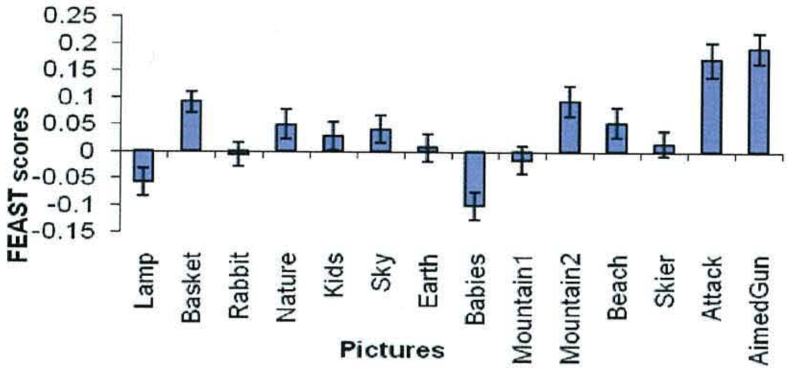
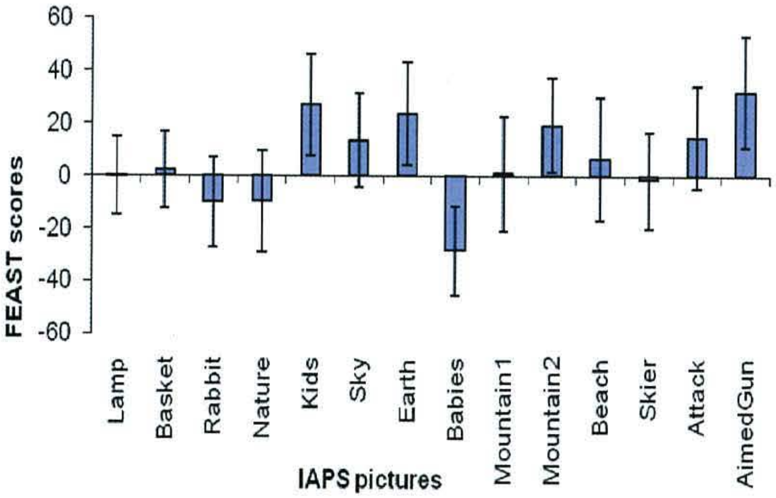


Figure 17

For accuracy FEAST scores (Figure 17), in accordance with the SAM scores, it would seem that "Aimed gun" and "Attack" produce the highest scores. However, for low arousal images there would seem to be no agreement between the two sets of scores with the lowest SAM scores for "Lamp" and "Basket" producing moderate FEAST scores, and the lowest FEAST scores having moderate SAM scores.



Description	arousal
Aimed Gun	7.35
Attack	7.29
Earth	4.6
Mountains	5.68
Mountains	5.19
Sky	4.51
Skier	6.28
Lamp	1.72
Kids	3.77
Basket	1.76
Rabbit	3.08
Beach	5.71
Babies	4.7
Nature	3.22

Figure 18

For RT FEAST scores (Figure 18) "Aimed gun" is again in agreement with the SAM scores are the highest score. "Rabbit" and "Babies" yield a negative (low arousal)

FEAST score with both the accuracy and RT data yet score moderately on the SAM scores. It would seem unwise to draw any substantive conclusions from the FEAST scores since there is so little agreement between the RT and accuracy FEAST scores. Furthermore, although the aim is not to compare implicit and explicit data, the lack of any marriage between the two scores supports the argument that the FEAST scores alone in this example are too varied to draw any one conclusion. It is however, worth noting that the images were chosen to represent a spread of scores and this is indeed reflected in the FEAST scores for both RT and accuracy. And, additionally, that the items which have the highest explicit ratings ("aimed gun" and "attack") also have very high FEAST scores based on both RT and accuracy data. Thus, while the implicit measures of all the individual images might be called into question, it would appear that the FEAST is able to capture implicitly some aspect of the to-be-measured dimension.

Analysis of individual coloured images

Since the FEAST scores do not appear to reveal a consistent message, the next two figures show the RT and then accuracy data before they were 'used' to calculate the FEAST scores. Specifically, the data show RT and accuracy scores when the extrinsically mapped Positive (high arousal) or Negative (low arousal) keys were pressed in response to a coloured frame while displaying the named image.

Description	arousal
Aimed Gun	7.35
Attack	7.29
Earth	4.6
Mountains	5.68
Mountains	5.19
Sky	4.51
Skier	6.28
Lamp	1.72
Kids	3.77
Basket	1.76
Rabbit	3.08
Beach	5.71
Babies	4.7
Nature	3.22

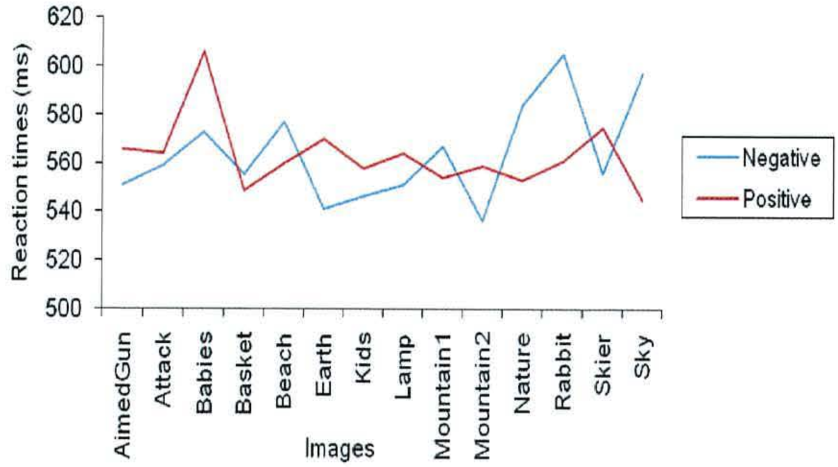
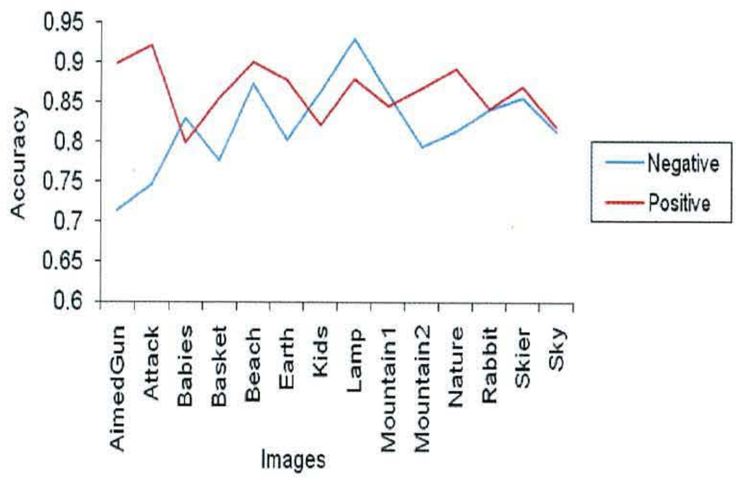


Figure 19

T tests were carried out between the two data sets for each image, and there were no significant differences produced from the RT data. Although from looking at the figure it would appear that the images, Babies, Rabbit and Sky have large differences between the two extrinsic response keys, they did not reach levels of significance.



Description	arousal
Aimed Gun	7.35
Attack	7.29
Earth	4.6
Mountains	5.68
Mountains	5.19
Sky	4.51
Skier	6.28
Lamp	1.72
Kids	3.77
Basket	1.76
Rabbit	3.08
Beach	5.71
Babies	4.7
Nature	3.22

Figure 20

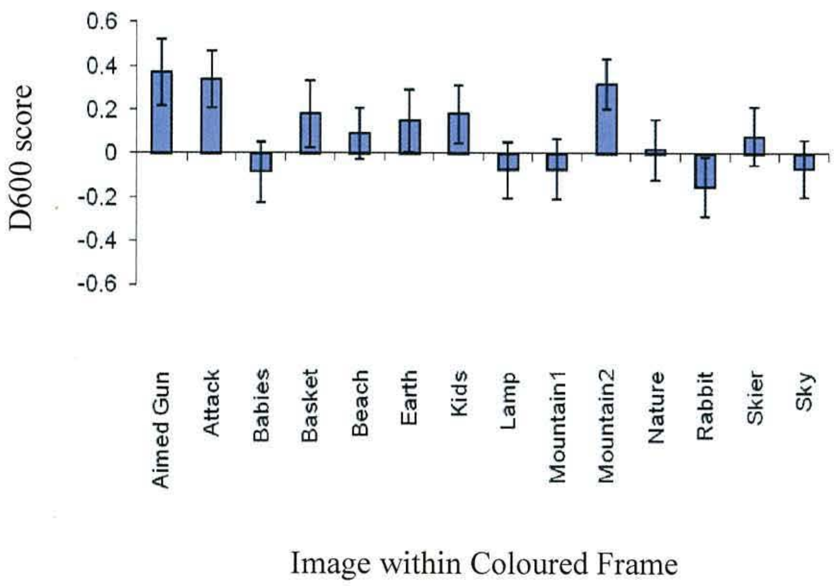
For the accuracy data however, there were three images which did produce a significant level of difference, these are:

Aimed gun t(29) = -2.82, $p<0.01$
Attack t(29) = -3.71, $p<0.001$
Earth t(29) = -0.38, $p<0.05$

Once again, looking at the RT and accuracy data with the two different types of extrinsic responses fails to provide a consistent picture. The next analysis was carried out using the D600 scores where both the RT and accuracy data are combined to produce one score.

D600 analysis

A D600 EAST score for each picture was calculated using the same methods as in Experiment 1.



Description	arousal
Aimed Gun	7.35
Attack	7.29
Earth	4.6
Mountains	5.68
Mountains	5.19
Sky	4.51
Skier	6.28
Lamp	1.72
Kids	3.77
Basket	1.76
Rabbit	3.08
Beach	5.71
Babies	4.7
Nature	3.22

Figure 21

As was the case with the untransformed tables, the three parts of Experiment 4 (valence, arousal, and dominance) are not suitable for a detailed comparison between images scoring high and low on each dimension. This is due to there being no definite point at which the images can be defined as either high arousal or low arousal across participants. By examining the D600 figures, the images "Aimed Gun", "Attack", "Basket", "Beach", "Earth", "Kids", "Mountain2", "Nature" and "Skier" appear to show high arousal, where as the images "Babies", "Lamp", "Mountain1", "Rabbit" and "Sky" low arousal. It seems clear that when the D600 score is compared to the explicit SAM score that there is not a clear and direct relationship between the two. This is not necessarily a problem since it was never expected that the explicit and implicit scores would resemble each other.

Correlations

A Pearson correlation coefficient between the D600 EAST scores and the IAPS explicit scores is 0.421. Although approaching it, the correlation coefficient was not significant (p=0.13).

The following figure displays the D600 scores and the explicit SAM scores.

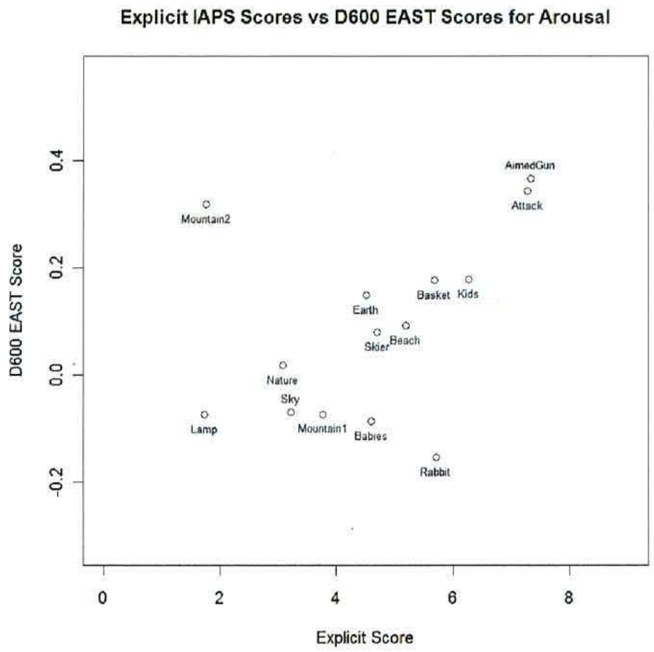


Figure 22

Discussion of Arousal Results

While it was not significant, there did appear to be fairly robust relationships between the implicit (FEAST) and explicit (SAM) measures of arousal. This relationship was particularly clear and strong at the more extreme ends of the scale. It is worth discussing why "arousal" might have been a somewhat difficult concept to capture in the FEAST paradigm. Recall that the logic behind the FEAST is that there should be an increased readiness/likelihood to respond with the extrinsic key if it matches the implicit meaning of the presented stimulus. So, for example, when the participant sees an "Aimed Gun" and must press the (explicitly) "arousing" key to correctly classify the colour, then there will be facilitation. While this logic seems solid, and should hold in most cases, the very nature of "arousing images" will create something of a confound. Specifically, one might predict that the more arousing images (explicitly OR implicitly) will globally facilitate responding via a general arousal/preparedness network. Furthermore, since such flight-or-fight facilitation is likely somewhat long-lasting, one might predict that the overall differences across all trials might be compressed. Thus, it might be that (for this particular dimension and image set) all the differences are being artificially reduced. Given this additional factor (i.e. overall response facilitation via flight-or-fight), it is all the more surprising that we were still able to discern statistically significant differences in FEAST scores for a number of the images.

We next turn to an analysis of the results for the "dominance" dimension. Again, to keep things in perspective, it is important to note that the purpose of these three experiments is to demonstrate that the FEAST paradigm is sensitive enough to measure the same images (taken from the IAPS database) on numerous different dimensions (arousal, dominance, and valence). Keep in mind that all the stimuli being presented to the participants are identical in all three experiments. The "white" images and the "coloured" images are always the same image sets. The only difference is how the participants are being asked to focus their attention (and rate) the different images. Thus, while the relationship between the purported implicit measures (as measured by the FEAST) and the explicit measures (as measured by the SAM) is somewhat important, it is perhaps more important for the purposes of this experiment that the images get different FEAST scores across the three different experiments. This would demonstrate that the FEAST is not based on low-level (e.g. luminance) aspects of the images themselves, but that it is instead measuring dimension-specific attitudinal responses to the images.

Dominance

Introduction

The arousal results (presented above) were promising in that they suggested that the FEAST can measure implicit attitudes towards a set of standardized (IAPS) images. We next try to use the same stimulus set but to ask the participants to focus on a different aspect of the images – namely, the "dominance" of the images.

As a reminder, the concept of "dominance" is meant to capture the idea of whether one is "in control" or is "being controlled by" the situation. Thus, an image of a gun aimed directly at the viewer should evoke feelings of being dominated (and thus receive a very low dominance score). On the other hand, an image of an innocent bunny should make the viewer feel like they are "in control" (dominating) the situation and thus should receive a very high dominance score.

In this experiment (which was run as one of three blocks – randomized across participants), participants were asked to rate the dominance of images appearing within white frames, and to rate the colour of any blue/green frames that appear. Thus, over the first set of trials (when purely white-framed images appeared) participants did two things. They formed both their own concept of what "dominance" was meant to signify and they also created an extrinsic linkage between the concepts "high dominance"/"low dominance" and their respective response keys.

Overview

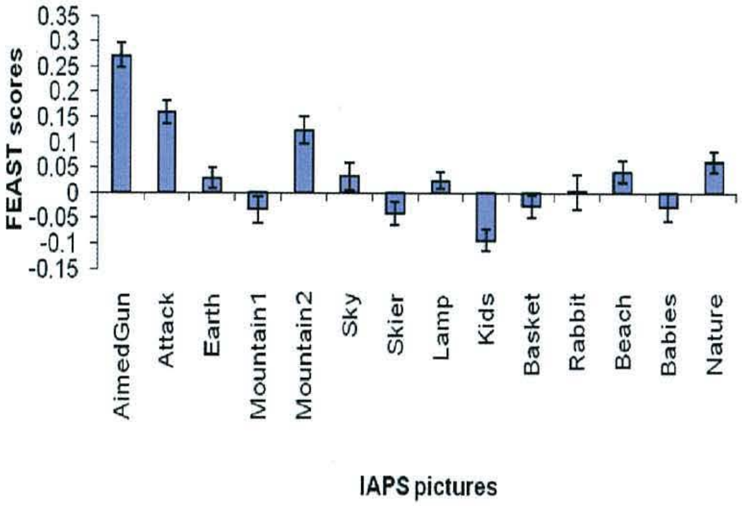
No participants had a large number of very brief latencies, and there were no latencies greater than 10000ms, so all 31 participants were included in the analysis.

ANOVA on latencies

A 2 (Experiment Half) x 14 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures model had a significant interaction between stimulus picture and extrinsic response valence: $F(13,377) = 2.99$, $p<0.001$, again, as in all previous studies reported here, this is an important interaction in terms of fulfilling the expectation of the design.

FEAST scores

In an effort to understand the origins of the significant ANOVA result just described, FEAST scores were calculated for each image, firstly looking at accuracy scores then RT scores.

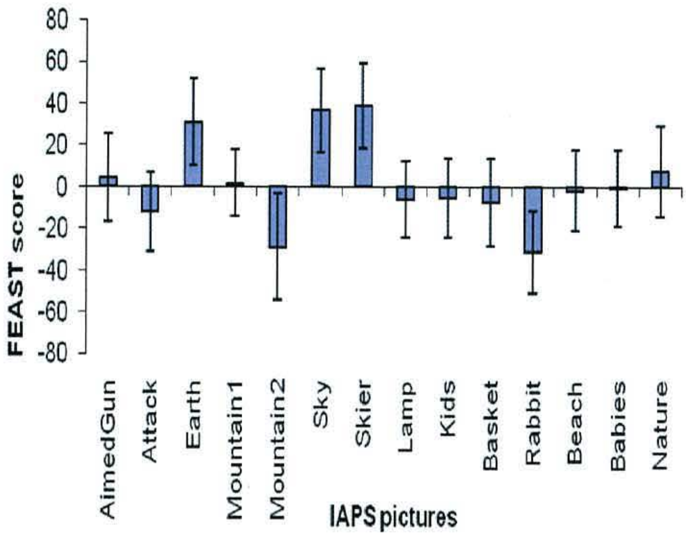


Description	dominance
Aimed Gun	2.15
Attack	2.73
Earth	4.15
Mountains	4.59
Mountains	4.93
Sky	5
Skier	6.14
Lamp	6.47
Kids	6.62
Basket	6.7
Rabbit	6.77
Beach	6.97
Babies	7.08
Nature	7.49

Figure 23

Figure 23 shows the FEAST accuracy scores for the dominance dimension along with the SAM (explicit) dominance ratings for the images. It is difficult to find any direct or obvious pattern of correspondence between the accuracy FEAST scores and the explicit SAM scores (in the chart to the right of the figure). The FEAST scores for

"Aimed gun" and "Attack" are the two strongest scores, yet display the lowest SAM scores. This could be a matter of interpretation, where the lowest SAM scores represented high dominance and stronger FEAST scores the same. In other words, there is a possibility that the participants mis-undestrood (or mis-coded) the direction of the dominance dimension.



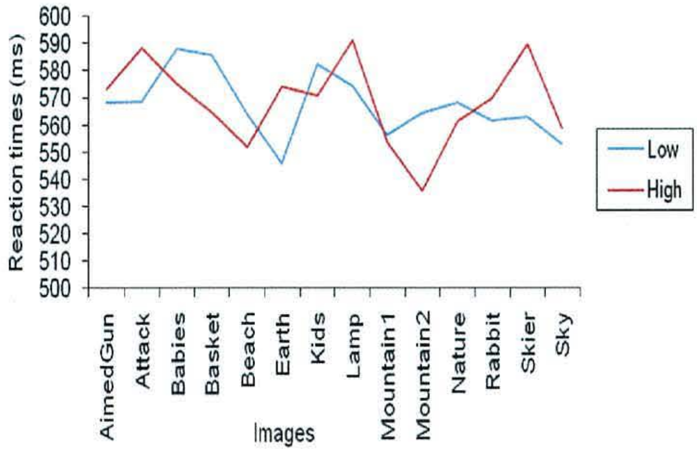
Description	dominance
Aimed Gun	2.15
Attack	2.73
Earth	4.15
Mountains	4.59
Mountains	4.93
Sky	5
Skier	6.14
Lamp	6.47
Kids	6.62
Basket	6.7
Rabbit	6.77
Beach	6.97
Babies	7.08
Nature	7.49

Figure 24

As with the accuracy data, when looking at the RT FEAST scores (Figure 24) there does not seem to be any particular pattern between the SAM and FEAST scores. As mentioned earlier, for a variety of reasons there need not be a direct or simple relationship between explicit and implicit measures. In addition to the reasons mentioned above for this discrepancy, here another possible explanation arises: the concept of "dominance" was the most difficult for participants to clearly conceptualize. This was seen both in terms of the higher variance in explicit ratings (at the end of the experiment) and was also mentioned in the debriefing (exit interviews) with participants.

Analysis of individual coloured images

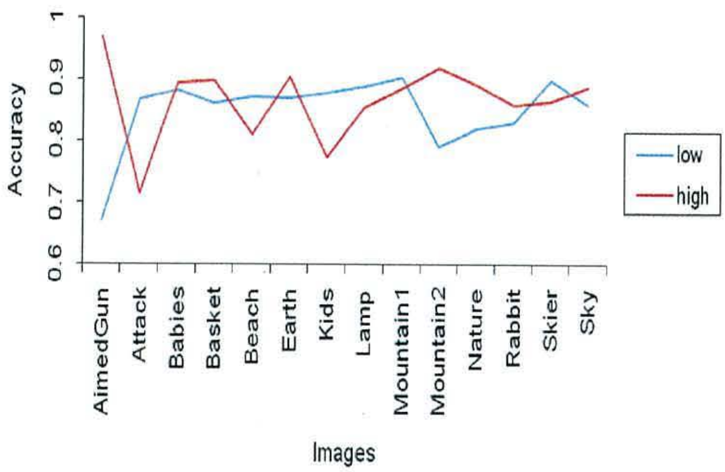
Since the FEAST scores amalgamate both the low and high explicit response keys, and the results in the above two figures were not clear, the next two figures again look at accuracy and RT data separately, but this time it is before a FEAST score is calculated. T tests were run on each image to find out whether there were any significant differences between the responses using the two different mapped keys.



Description	dominance
Aimed Gun	2.15
Attack	2.73
Earth	4.15
Mountains	4.59
Mountains	4.93
Sky	5
Skier	6.14
Lamp	6.47
Kids	6.62
Basket	6.7
Rabbit	6.77
Beach	6.97
Babies	7.08
Nature	7.49

Figure 25

There were no images which produced significant levels of difference for the RT data. Earth, Mountain2, and Skier do appear to have large differences when looking at the figure, these however do not produce a level of significance.



Description	dominance
Aimed Gun	2.15
Attack	2.73
Earth	4.15
Mountains	4.59
Mountains	4.93
Sky	5
Skier	6.14
Lamp	6.47
Kids	6.62
Basket	6.7
Rabbit	6.77
Beach	6.97
Babies	7.08
Nature	7.49

Figure 26

When looking at accuracy data, three of the images had levels of significant differences, these were:

Aimed Gun $t(30) = 2.05$, $p<0.001$
Attack $t(30) = -2.65$, $p<0.01$
Mountain 2 $t(30) = -3.01$, $p<0.01$

This is not terribly surprising, as it could be seen in the accuracy FEAST figure that these three images produced the strongest FEAST scores.

D600 analysis

After examining both the FEAST scores and the data for the parsed explicit response keys there is still no clear understanding about the origins of the significant ANOVA between stimulus valence and extrinsic response. A D600 EAST score for each picture was calculated using the same methods as in experiment 1 in order to hopefully bring together the data in a more meaningful fashion. Values were imputed on two occasions due to missing data.

Description	dominance
Aimed Gun	2.15
Attack	2.73
Earth	4.15
Mountains	4.59
Mountains	4.93
Sky	5

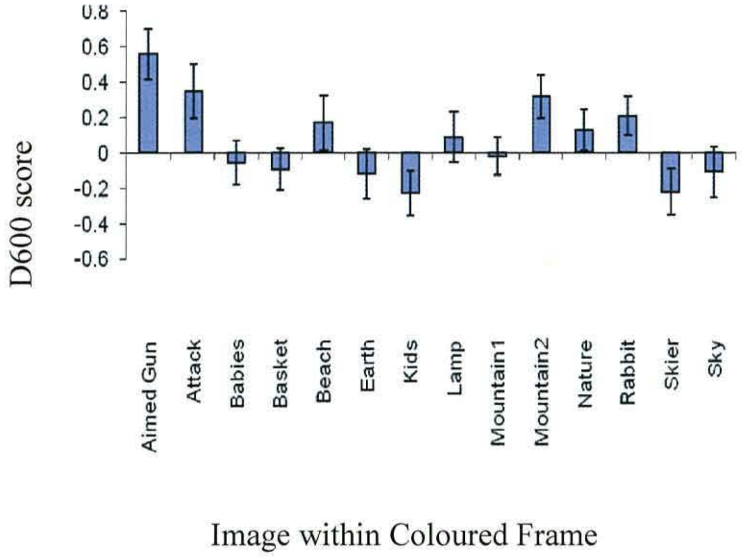


Figure 27

Participants showed a stronger response to the pictures “Aimed Gun”, “Attack”, “Beach”, “Lamp”, “Mountain2”, “Nature” and “Rabbit”. A weaker response was shown towards the remaining seven pictures. Again, just from a simple inspection of the data, it does not appear that the D600 scores (implicit scores) correlate with the SAM scores, however this is checked in the following analysis.

Correlations

The Pearson correlation coefficient between the D600 EAST scores, and the IAPS explicit scores is -0.507. The fact that the correlation is negative is likely due to the reversed interpretation of our participants that was alluded to earlier (i.e. “dominate” Vs “is dominated” is an easy confusion to make). The robust value of this correlation (over 0.5) suggests a strong relationship between the implicit (FEAST) and explicit (SAM) measures. However, although approaching significance, the correlation coefficient was not significant (p=0.064).

Figure 28 shows a scatterplot of explicit (SAM) and implicit (D600) scores.

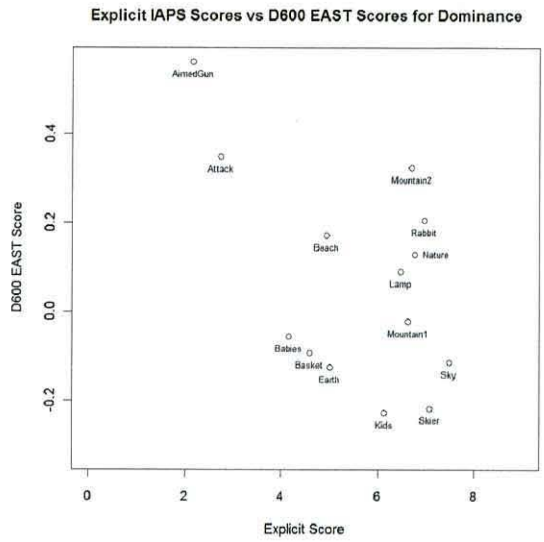


Figure 28

Discussion of Dominance Results

As mentioned earlier, the concept of "dominance" was the least intuitive to our participants. This was evident in their explicit ratings and in their self report. Given the ambiguity of the concept itself, we were actually quite pleased to find the fairly robust and consistent (though not statistically significant) relations between explicit and implicit measures. Again, as mentioned earlier, one would not predict implicit and explicit attitudes (or reactions) to always be highly correlated. And, based on intuition, one might predict that these two would diverge most strongly when social or personal pressures are present. While "dominance" is not as heated an area as racial stereotypes, there is still perhaps some stigma associated with "feeling dominated by" other things in the world. Thus, here it might not be too surprising to find some level of discrepancy between the implicit and explicit measures. In future work it would be interesting to see if such personality traits as "need to conform" correlate with the level of discrepancy between one's implicit and explicit attitudes towards dominance.

Let us turn now to the results of the "valence" blocks.

Valence

Introduction

In this third experiment (which was run as one of the three blocks – randomized across participants), participants were asked to rate the valence of images appearing within white frames, and to rate the colour of any blue/green framed images that appeared. Thus, over the first set of trials (when purely white-framed images appeared) participants formed their own concept of what "valence" was meant to signify and via their responding they also created an extrinsic linkage between the concepts "positive valence"/"negative valence" and their respective response keys. It is this extrinsic linkage between the key and a concept/dimension that is critical for the success of the FEAST.

Overview

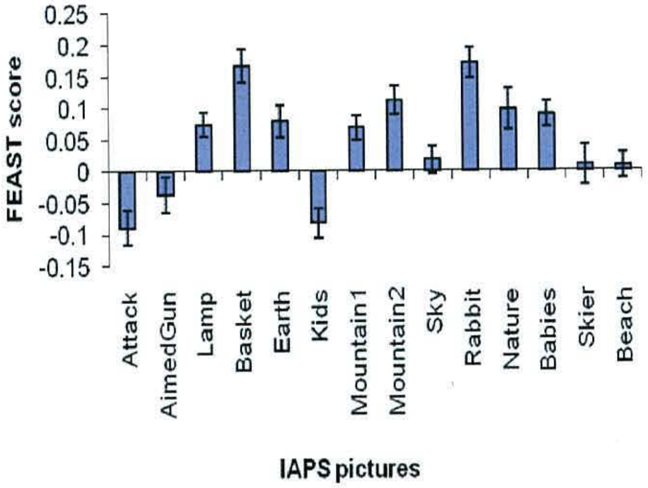
Participants 17, 18 and 22 had latencies faster than 300ms in more than 10% of the trials, so were removed from the analysis. Thus, the overall analysis included 28 participants. There were no latencies greater than 10000ms.

ANOVA on latencies

A 2 (Experiment Half) x 14 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures ANOVA was carried out and was found to have a significant interaction between stimulus picture and extrinsic response valence: $F(13,338) = 3.53$, $p<0.001$, this is, once again, the interaction which is most desired in terms of fulfilling the experiment expectations. Interestingly, the model also included a significant term of Stimulus Picture: $F(13,338) = 2.39$, $p<0.01$, suggesting that there were difference across the pictures. This was not found to be significant in the previous two dimensions of arousal and valence (nor in Experiments 1, 2 and 3).

FEAST scores

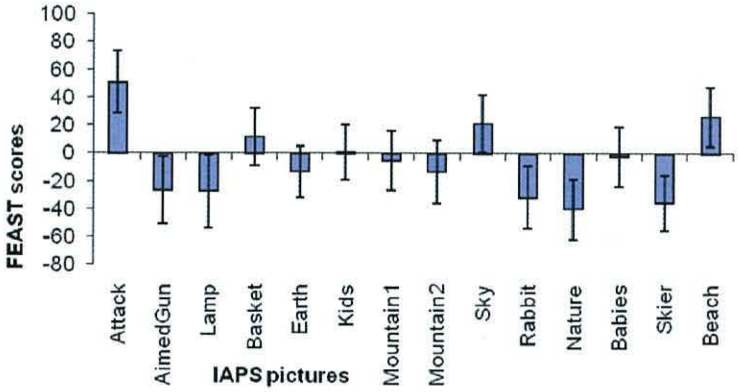
The FEAST scores for the individual images were calculated from the mean RT and accuracy data for each image. This was done in order to gain further understanding into the influence of the significant ANOVA result.



Description	valence
Aimed Gun	2.37
Attack	1.9
Earth	6.67
Mountains	7.61
Mountains	7.57
Sky	7.61
Skier	8.1
Lamp	4.87
Kids	7.44
Basket	4.94
Rabbit	7.82
Beach	8.22
Babies	8.09
Nature	8.05

Figure 29

The FEAST scores based on accuracy data are shown in Figure 29 along with the explicit (SAM) scores for point of reference. Perhaps surprisingly, the accuracy FEAST scores bore only a weak relation to the explicit SAM scores. For example, images such as "Kids" which depicts two children happily going to the beach holding a bucket and spade was certainly not expected to produce a negative FEAST score in the same region as "Attack" where a knife is being held to a throat. That being said, the pictures were being responded to in isolation and not as a comparison with each other, also a young student population who are largely assumed to be without children may be more driven by career and self-gratification (e.g. mountain climbing – which is immensely popular at Bangor where the studies were conducted) rather than images of family. A high positive response to "Basket" was also not anticipated.



Description	valence
Aimed Gun	2.37
Attack	1.9
Earth	6.67
Mountains	7.61
Mountains	7.57
Sky	7.61
Skier	8.1
Lamp	4.87
Kids	7.44
Basket	4.94
Rabbit	7.82
Beach	8.22
Babies	8.09
Nature	8.05

Figure 30

Figure 30 shows the FEAST scores based on the reaction time data. The RT FEAST scores showed a far greater trend towards negative valences. Thus, whereas only three images produced a negative FEAST score using accuracy data, the RT FEAST scores showed nine of the images producing a negative score. Whilst this shows more of a balance and spread of scores, the FEAST scores are again surprising, for instance whilst "Attack" has a large FEAST score it is positive, and "Nature" has the largest negative FEAST score, both of which are difficult to interpret. The definite lack of consistency between the RT and accuracy scores, and between the implicit FEAST and explicit SAM scores, leads to the next part of the analysis where the images are examined before the mean data are converted into a FEAST score.

Analysis of individual coloured images

As was the case for the previous two dimensions studied with these same images, t tests were carried out in order to determine whether any of the images had significant differences when responding with the differently mapped keys.

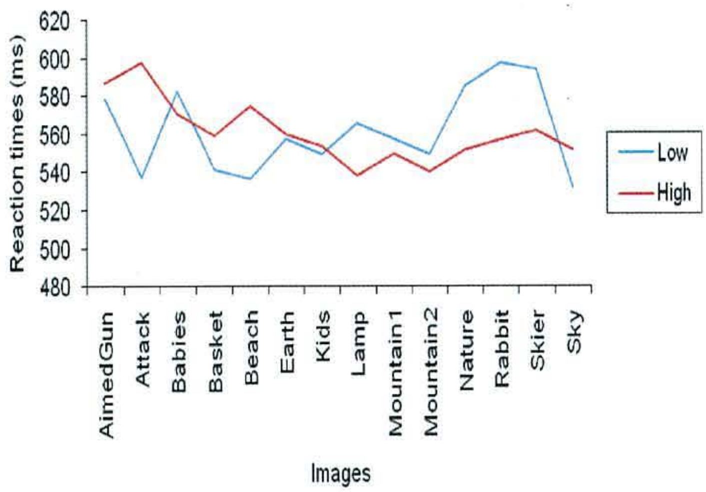


Figure 31

Description	valence
Aimed Gun	2.37
Attack	1.9
Earth	6.67
Mountains	7.61
Mountains	7.57
Sky	7.61
Skier	8.1
Lamp	4.87
Kids	7.44
Basket	4.94
Rabbit	7.82
Beach	8.22
Babies	8.09
Nature	8.05

Out of the fourteen images, nine had significant levels of difference between the differently mapped key responses:

Attack $t(27) = -2.37, p<0.05$
Beach $t(27) = -2.54, p<0.01$
Babies $t(27) = 2.06, p<0.05$
Basket $t(27) = -2.86, p<0.05$
Earth ($t(27) = -2.06, p<0.05$
Kids $t(27) = 2.04, p<0.05$
Mountain 2 $t(27) = -2.51, p<0.01$
Nature $t(27) = -2.89, p<0.05$
Rabbit $t(27) = -3.54, p<0.001$

Description	valence
Aimed Gun	2.37
Attack	1.9
Earth	6.67
Mountains	7.61
Mountains	7.57
Sky	7.61
Skier	8.1
Lamp	4.87
Kids	7.44
Basket	4.94
Rabbit	7.82

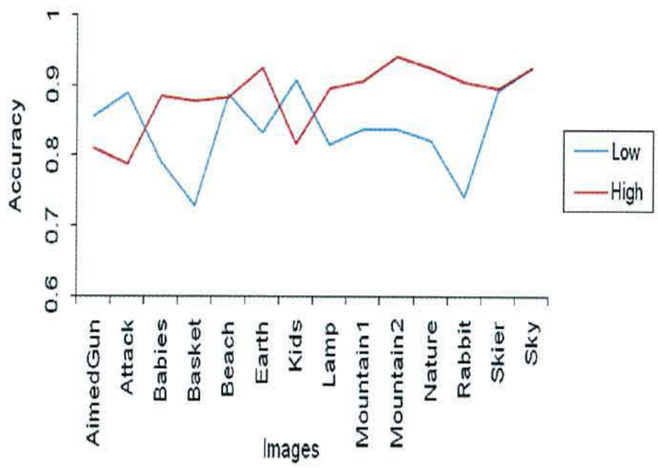


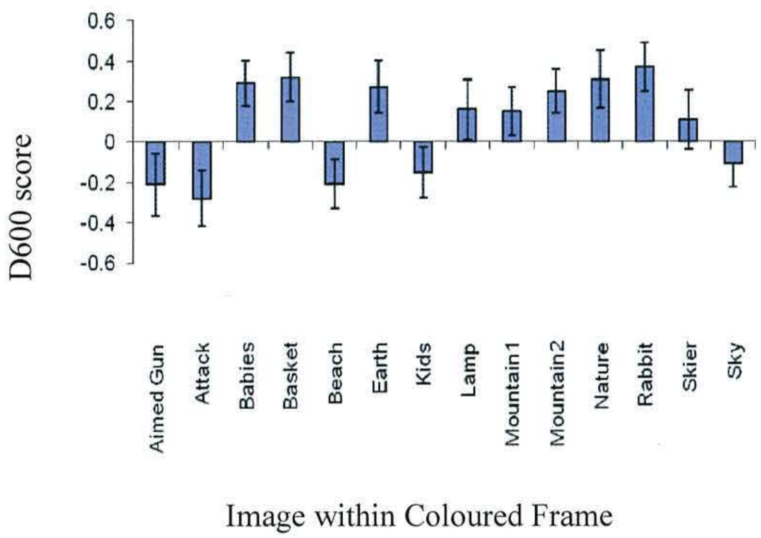
Figure 32

For accuracy data, six of the images had significant differences. Interestingly each of these six images also yielded significant values for the RT data.

Babies $t(27) = -2.06$, $p<0.05$
Basket $t(27) = -2.86$, $p<0.01$
Earth $t(27) = -2.06$, $p<0.05$
Mountain 2 $t(27) = 2.51$, $p<0.05$
Nature $t(27) = -2.09$, $p<0.05$
Rabbit $t(27) = -3.54$, $p<0.01$

D600 analysis

A D600 FEAST score for each picture was calculated using the same methods as has been used throughout the thesis. The purpose of this analysis is to create a single measure that incorporates both the reaction time and accuracy data and thus to help further explore where the significant differences between response and image arise. The FEAST scores alone are confusing and often conflicting, a separation of the key responses which create the FEAST scores do display significance for some images, but there are still differences between the two types of data. The D600 scores for valence are shown in Figure 33. There are clearly some key similarities between the implicit (D600) scores and the explicit SAM scores. For example, the low scores for "Attack" and "Aimed gun", but largely there is a great degree of variability between the implicit and explicit measures.



Description	valence
Aimed Gun	2.37
Attack	1.9
Earth	6.67
Mountains	7.61
Mountains	7.57
Sky	7.61
Skier	8.1
Lamp	4.87
Kids	7.44
Basket	4.94
Rabbit	7.82
Beach	8.22
Babies	8.09
Nature	8.05

Figure 33

Correlations

The Pearson correlation coefficient between the D600 EAST scores and the IAPS explicit scores is 0.447. Although this correlation coefficient is fairly sizeable it was not statistically significant ($p=0.11$).

Figure 34 shows a scatter plot of the implicit (D600) and explicit (SAM) measures of the IAPS images. An examination of the plot suggests that there was a good correspondence between the implicit and explicit measures for many of the images – with the notable exception of Sky, Kids, and Beach.

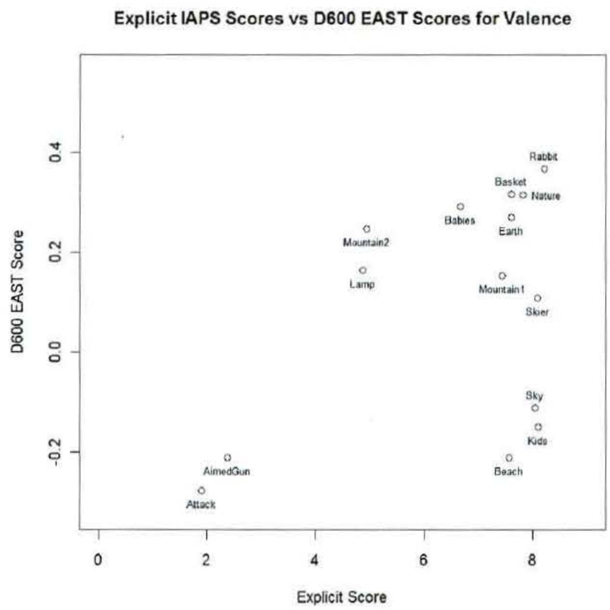


Figure 34

Experiment 4 discussion

The significant interaction between stimulus picture and extrinsic response for each of the dimensions was achieved. This indicates that for each of the three dimensions of interest (valence, arousal and dominance) the FEAST method was sensitive enough to record significant reactions to the various stimuli. Perhaps more surprising, this was true – and also gave different results -- even when the stimulus images remained the same for each dimension. Thus, for example, consider the D600 FEAST scores for the image "rabbit" across the three different measurement blocks/conditions. In all conditions it was the same image appearing within a coloured (blue or green) frame. Thus, in some sense, the error rates and reaction times to judge the colour of the frame should have been the same in all three conditions (it was, very literally, the same stimulus and the same task that the participant was being asked to perform). However, the results for this "same image and same task" are vastly different across the three blocks. Considering the D600 scores, the image "rabbit" scored very high for valence (i.e. was a highly positive image), it scored moderately high for dominance (i.e. the power relationship was unclear), and it scored very low for the D600 FEAST implicit measure of arousal (i.e. the image was not very arousing). This difference across the measurement blocks (dimensions) suggests that the measurements yielded by the FEAST are sensitive to the specific tasks and contexts (dimensions) at hand. This result is exciting as it implies that the FEAST could be used as a very flexible tool. It suggests that, in theory, if one could choose the right "white images" to set the desired dimensional context of interest, then one could implicitly measure attitudes on any number of dimensions on the same set of stimuli.

Although it does not immediately follow, but this result also suggests that it might also be possible to test stimuli not only using paired variables (e.g. good/bad), but also using non paired variables (e.g. exciting/jolly).

Interestingly for each of the dimensions (valence, arousal, and dominance) there were no significant correlations between the explicit SAM scores taken from the IAPS and the implicit (FEAST D600) scores. It should be noted, however, that while there were no statistically significant correlations, the correlations were all fairly sizeable and tended to be weakened (in all cases) by one or two images (such as mountain) which might have been inappropriate for the demographic on which the experiment was tested (i.e. Bangor University undergraduate students). Additionally, the relatively small sample size of this study could have also contributed to the lack of significance. And thus the lack of statistical significance may be more of a reflection of lack of power rather than one of sensitivity of the measure.

Finally, as mentioned previously, there are several ways of looking at this strength of correlation between implicit and explicit measures. One is that the result is exactly what would be *hoped for* since one technique (SAM) is measuring explicit scores and the other (FEAST) implicit scores; and, depending on the viewpoint taken, these can be seen as two separate attitudes and thus the lack of correlation might add support to the idea that the FEAST is indeed measuring something separate to an explicit score. The second viewpoint is that this lack of a strong correlation is not what would be expected since there should be a link between explicit and implicit attitudes.

Whilst it is important to remember that the experiment presented in this chapter never set out to discover or quantify the link between implicit and explicit attitudes, especially since this is a matter which cannot be agreed upon in the scientific community, it is a very important conclusion to be borne out of the correlations that

the FEAST method does appear to be sensitive to the three different dimensions tested. Despite the lack of significance with the correlations, there is evidence of trends, which with a larger participant pool may have easily yielded significance. There are some issues which became apparent both as a consequence of the debriefing process and of the analysis. Firstly, the stimuli chosen for presentation within the white frames to represent a spread of SAM scores across all three dimensions did not do a great job of representing the extremes of the three dimensions. This however was unavoidable due to two constraints: first, the ethical guidelines set by Bangor University did not allow us to show the extreme images from the IAPS; second, we wanted to use the same set of images across all three blocks of the experiment, and it was just not possible to get a set that represented the extremes of all three dimensions concurrently and equally. Despite this concern, by asking the participants to carry out the explicit ratings of the stimuli, it was possible to conclude that the stimuli were indeed strong enough to be rated on a rough sliding scale from high to low for each dimension. However, it is impossible to conclude whether a stronger set of stimuli may have provided more conclusive results without running a separate study with pictures rated at the extremes of the SAM scale. As was the case for each experiment presented as part of this thesis, the data for each individual participant was examined for each picture, and we found large individual differences in terms of their explicit ratings of the images across the different dimensions. This variance of interpretation undoubtedly added to the overall variance of the implicit measures. In future work, a greater degree of pilot (pre) testing could give a more robust "white pictures" set which would presumably lead to more robust implicit measurement results.
Secondly, with any experiment where the stimuli dimensions are a little ambiguous in meaning such as this (e.g. "dominance") there is always the question as to whether a

participant has responded based on their true ("gut") responses or instead only because they were simply agreeing to an experimenter-given 'correct' response. The data was checked for each picture to examine the variability and any unusual patterns of responding such as excessive use of one key press. For the white framed pictures it was expected that there would be a certain degree of learning involved, indeed this is encouraged through the practice sessions to increase the strength of the key mapping. It is also hoped that for the coloured framed pictures that no learning need take place as participants are responding based on colour, and any influence which the picture may have on response speed and accuracy is of an implicit nature.

Although the results were not as strong as hoped for, there was enough confidence in the method to proceed to a different type of question that was more applied in nature using the FEAST method. Specifically, as this experiment has demonstrated that the FEAST is sensitive enough to measure a set of images on numerous dimensions, we decided to apply this approach to measure several dimensions of implicit attitudes about photos of real-world products.

Chapter 5: Experiment 5

Experiment rationale

Experiment 5 applies the knowledge gained from chapter 4 about testing different dimensions of the same picture and tests 'real world' images to quantify multiple aspects of the country of origin (COO) effect. Experiment 5 introduces a new area not covered in the introduction, so before the next experiment is presented there is a brief overview of the COO effect.

Introduction

COO effects are related to how consumers perceive products coming from a particular country (Roth & Romeo, 1992). Specifically, COO effects can be said to take place when consumers draw conclusions about products or services from one particular country based primarily on the country of origin - often without factual information. Clearly, COO effects are related to stereotypes – they can be thought of as product stereotypes that are linked to stereotypes about a specific country. Common examples today might include ideas such as "products made in China are of an inferior quality" or "products made in Germany tend to be well engineered." Stereotypes are internalized and perpetuated through various methods, such as learning, experience, and even punishment. Through these processes most members of society naturally learn these stereotypes (Ehrlich, 1973). People naturally internalize any images, biases, or stereotypes toward other countries, and these stereotypes can influence their thoughts and behaviors.

COO effects have been shown to influence both consumer decision making as well as product evaluation (Schooler, 1965; Erickson, Johansson, & Chao, 1984; Lawrence, Marr, & Prendergast, 1992). Hong & Toner (1989) found that American consumers perceive Japanese products to be superior in quality. Over the past few decades, there has been a growing level of research on COO effects relating to consumers' attitudes, preferences, and behaviours (Peris, Newman, Bigne, & Chansarkar, 1993; Nooh & Powers, 1999).

Most early studies related to COO focused on examining consumers' attitudes toward imports. Almost all studies supported the fact that country stereotyping affects evaluations for products in general. For example, the findings from an early study by Reierson (1966) demonstrated American consumer stereotyping of foreign products, as well as a strong preference towards products from their own country. Schooler (1965) demonstrated how COO biases can influence specific product evaluations. Numerous similar studies have supported this work such as Reierson (1966), Not all COO effects (country stereotypes) are negative in nature. For example, numerous authors (e.g. Lawrence, Marr, and Prendergast (1992), Nooh & Powers (1999) and Nagashima (1970)), have reported the existence of a general positive bias towards domestic products.

One opinion is that stereotypes are often not factually based and tend to be relied on more when there is incomplete information concerning the product of interest (Lawrence, Marr, & Prendergast, 1992). Erickson, Johansson, and Chao (1984) considered COO as an image variable and analyzed its effect on formation of beliefs and attitudes, and on how this process influences product evaluation. They investigated the impact of a particular type of image variable, COO, on attitude and beliefs obtained through a survey of evaluations of automobile alternatives. The results indicated that COO affects belief formation rather than attitude. Erickson, Johansson, and Chao (1984) proposed that there are three different types of beliefs, as follows: descriptive, inferential, and informational. They suggested that although these beliefs are formed in different ways, all contribute to consumers' beliefs about a product's attributes. According to Fishbein and Ajzen (1975), descriptive beliefs derive from direct experience with a product. Namely, these beliefs connect physical characteristics with product perceptions. Likewise, information beliefs are based on outside sources of information such as commercials, family, friends, and so on. Erickson, Johansson and Chao attribute the remaining type of belief (inferential belief) as contributing to the formation of stereotypes. This belief is formed by making inferences (correctly or incorrectly) based on past information or experiences and trying to extrapolate how such past experiences relate to the current stimulus of interest (Fishbein and Ajzen, 1975). For instance, people whose experience suggests that Japanese electronic products are durable might infer this since Sony is a Japanese electronic product and, from past experience, Sony produces durable electronic products.

What is the difference between belief formation and attitudes?

At this point it is useful to briefly discuss the difference between an attitude and a belief, a fuller discussion about attitudes can be found in the general introduction. Attitudes are a highly discussed area of social psychological research, often described as being a hypothetical construct, one that represents an individual's degree of like or dislike for an item. Usually an attitude is on the scale of positive or negative. Attitudes can be considered using the ABC model (affect, behavior, and cognition), where the affective component is based upon an emotional reaction. The behavioral component intention refers to how an individual tends to behave as an indicator of their attitude, and finally the cognitive component is based upon a cognitive evaluation of the attitude forming object. Beliefs, however, are considered to be a simple mental representation, divided into core and dispositional beliefs.

Where core beliefs are those which are actively thought about and dispositional beliefs are not. Attitudes can be formed from beliefs. Several studies (e.g. Brunel, Greenwald, & Tietje, 2004; Maison, Greenwald, & Bruin, 2001) have shown that the IAT could be used as a method to assess consumer's attitudes and beliefs. The next experiment (Experiment 5) looks at the COO effect for three countries, the UK, the USA, and Japan. It uses the FEAST described in Experiments 3 and 4 to examine how the COO effect can alter FEAST scores using logos. Following on from Experiment 4, which examined whether the FEAST is sensitive enough to test various dimensions of the same set of stimuli, Experiment 5 tests three different dimensions for each logo using the FEAST method already established in the previous experiments. The FEAST design was viewed as an ideal measure of implicit attitudes towards pictures, in this case brands. Its flexibility in terms of the dimensions which can be tested in one sitting proves useful in the field of market research which largely relies on explicit testing. Brand logos were selected to represent the brands and for each brand we measured its implicit country of origin by using the FEAST. After the FEAST study measured implicit reactions, each participant gave explicit ratings of each brand through a rating test (similar to that used in all experiments reported). These explicit ratings – along with post-experiment debrief and feedback – gave valuable information about how each brand logo was explicitly perceived by our participants (Bangor University undergraduates). The particular dimensions chosern were selected to represent a cross section of issues which market research tends to focus upon when launching a new brand. Through consultation with the company sponsor of this work (Blah D Blah, Conwy) they confirmed that the two dimensions of cost and quality were key issues. The third dimension (country of origin) was of particular interest for theoretical reasons; particularly to see whether participants would view brands which they associated as being from their country in a more favourable light, i.e. lower cost yet of a higher quality. Whilst Experiment 5 is based on theory, the topic area is one of an applied nature. The goal is to create a method which will aid the pursuit of better understanding consumers and their motivations. In order to gain the most 'value' from a participant, a market researcher would hope to glean as much information about a product or brand as possible, this is one reason for using multiple dimensions. The other reason is that brand logos, which were used in this experiment described here, are multi faceted in terms of attitudes held about them. Any particular brand logo will represent past experiences of the brand (both good and bad), stories told about them (both true and untrue), associations with linked brands, to name but a few of the numerous sub-representations that make up an overall brand representation.. A consumer will hold a mental representation about the logo which is made up of all of these attitudes and beliefs. Thus, testing three dimensions would just scratch the surface, but it is likely that in terms of market research there would be a few key goal directed dimensions, so this study aims to mimic such a situation.

Hypothesis

Based on the fact that the previous experiment demonstrated the ability of FEAST to measure multiple dimensions, we again predicted that this time we should be able to measure multiple different dimensions on the same stimulus set. Furthermore, in terms of the particular questions of interest in this study (vis-à-vis COO effects), it was expected that the results would vary depending on the country of origin of the participant, (from the UK or USA). It was also predicted that the results would be mirrored depending on the participants country of origin: in other words if the mean result for UK participants is that UK related companies/products are high cost and USA low cost, it was anticipated that for USA participants the USA-related companies/products would be high cost and UK low cost.

Experiment 5

Method

Participants

This experiment was run using participants from both the UK and the USA. The USA participants were, for the most part, students who were now studying in Bangor, Wales (UK) but who had lived in the UK for a maximum of 18 months. This time limit was chosen to prevent a loss of familiarity/affection towards USA brands.

UK participants

Thirty psychology students (9 male, 21 female) participated as part of a course requirement. All participants had normal or corrected to normal vision. The mean age of the participants was 21.

USA participants

Twenty eight psychology students (11 male, 17 female) participated for payment. All participants had normal or corrected to normal vision. The mean age of the participants was 23.

Materials and measures

Images used in the coloured frames are company logos, either from the USA, the UK, or Japan. Some of the logos were chosen to be deliberately confusing as to their COO, for example, ASDA, a previously British company is now owned by Walmart. The logos were chosen to reflect a sliding scale of association with the UK or USA, the Japanese company Honda was included as a false mid-point, by this it is meant that it was hoped that Honda was neither associated with the UK or the USA and so would prove as a changing mid-point for each participant providing a separation for the companies associated as more UK or USA. Table 9 shows each logo image which was used in the coloured frames, with table 10 listing those seen within white frames.

Logo name	Logo picture used	Why selected
ASDA	ASDA	Now a USA company owned by Walmart, previously UK. It is worth noting that an old logo was used since the new ASDA logo includes Walmart
Jaguar	JAGUAR	Now owned by the USA company Ford, but previously a UK company and is seen by many as symbolizing 'Britishness'
Tesco	TESCO	A UK based company which begin in North East London, which now has global recognition
Virgin	Virgin	A UK based company with global recognition. The generic Virgin logo was used not one of its subsidiaries e.g. virgin money, holidays, airlines etc
Honda	HONDA	Japan – To provide a mid-point company with neither as association to the UK or USA
Colgate	Colgate	A company from the USA, with a global presence.
Ebay	ebay	A company from the USA, with a global presence.
Ford	Ford	A company from the USA, with a global presence.
McDonalds	McDonald's	A company from the USA, with a global presence.

degrees to the left; 4 males looking 15 degrees to the left; 4 females looking 15 degrees to the right and 4 males looking 15 degrees to the right. An example of one male and one female looking to both the left and right can be seen in Figure 43; see the appendix for all images used within the coloured frames.

Figure 43.

Female and male looking to the left

Female and male looking to the right

Procedure

The procedure is the same as for Experiment 6a.

Design and analysis

Analysis was carried out using the same procedures as described in Experiment 4 using the number and breakdown of trials as described in Experiment one.

Results

Overview

There was one latency greater than 10000ms, which was removed from the analysis. One participant had latencies faster than 300ms for more than 30% of the trials, so was also removed from the analysis. Consequently the analysis was performed on data from 27 participants.

Untransformed RT and accuracy data

The next two tables, 25 and 26, show the mean untransformed reaction times and error rates as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 26 also includes experiment half.

Stimulus Picture	**Extrinsic Response Dimension**	
	Untrustworthy	Trustworthy
Gazing Left		
Reaction Time (ms)	583 (86.8)	588 (87.0)
Percentage of Errors	11.3 (7.21)	14.5 (9.85)
Gazing Right		
Reaction Time (ms)	582 (81.4)	584 (92.4)
Percentage of Errors	11.6 (9.00)	12.5 (8.12)

Table 25.

The shaded area for RT data (shown in rows) suggests whether the stimulus was seen as trustworthy or untrustworthy. If the RT shaded is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded RT in the untrustworthy column. The shaded area in Accuracy data suggests whether the stimulus was seen as trustworthy or untrustworthy. If the shaded area is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded accuracy in the untrustworthy column.

There do not seem to be any large changes between the RT data, with the fastest responses to both the leftward and rightward images being with the response key mapped as untrustworthy. Participants also tended to respond more accurately when the extrinsic response dimension was untrustworthy for both gaze directions. This does not support the hypothesis as there is no opposing trend for each gaze direction.

Stimulus Picture	**Extrinsic Response Dimension**			
	Untrustworthy **Experiment Half**		Trustworthy **Experiment Half**	
	1	2	1	2
Gazing Left				
Reaction Time (ms)	595 (95.3)	570 (98.1)	594 (81.8)	581 (98.2)
Percentage of Errors	14.4 (13.4)	8.33 (7.35)	13.9 (11.0)	15.0 (11.9)
Gazing Right				
Reaction Time (ms)	589 (100)	577 (76.3)	581 (113)	583 (102)
Percentage of Errors	12.5 (10.1)	10.6 (11.2)	11.8 (10.6)	13.3 (11.7)

Table 26

Reaction times seemed to decrease between experiment halves, apart from trials which showed a person with eyes gazing to the right and when the extrinsic response dimension was trustworthy. The error rates also seem to change somewhat between experiment halves. The variation in this data was small and did not cause concern.

ANOVA on latencies

The 2 (Experiment Half) x 2 (Stimulus Picture) x 2(Extrinsic Dimension) repeated measure model was fit to the logarithm of the reaction time data (including error penalty with one significant term; the experiment-half and extrinsic dimension interaction: $F(1, 26) = 14.5$, $p < 0.001$, and a marginally significant main effect of experiment half: $F(1, 26) = 3.13$, $p = 0.08$.

D600 analysis

The mean D600 EAST score for pictures with people gazing leftward was D=-0.081. The mean D600 EAST score for picture with people gazing rightward was D=-0.03. One sample t-tests suggested that these D600 scores were not significantly different from zero. Furthermore, a two-sample paired t-test suggested the difference between the scores was not significantly different from zero.

Correlations

For trials with a stimulus picture which included a face with leftward-gazing eyes, the correlation between the D600 scores was $r=0.184$. For trials with a stimulus picture which included a face with eyes gazing rightward, the correlation between the scores was significant, $r=0.496$, $t(25)=2.85$, $p<0.001$.

Experiment rationale

It may appear to be strange to jump between testing the effects of expressions and facial gaze to testing for visual fields effects on the perceptions of trust; indeed it is acknowledged that they are three distinct areas of study. However, as mentioned in the introduction to this section, there have been numerous studies linking visual field presentations with emotional responses. Specifically, numerous previous studies have suggested a left visual field preference for negative emotions and either a right or non-differential response bias for positive emotions.

Hypothesis

It was predicted that photographs presented in the left visual field would be perceived as more negative and by association more untrustworthy. The expectation is that participants would respond more quickly and accurately when faces presented on the left are responded to with the negatively mapped key response and vice versa.

Experiment 6c

Method

Participants

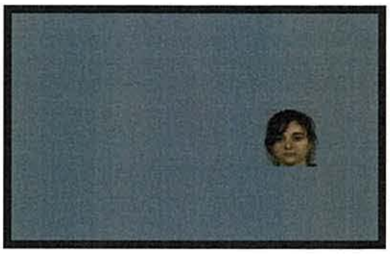
The same participants took part in versions 6a, b and c.

Materials and measures

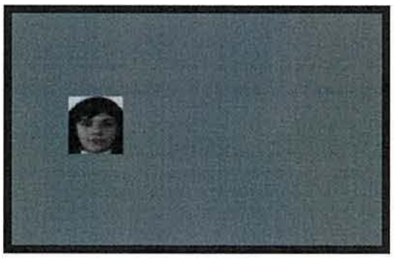
Face stimuli (which appear in blue/green frames) were taken from the same database as in Experiment 6a; the same 22 words (which appear in white frames) were also used in experiment 6c. Faces used were: 4 females in the left visual field; 4 males in the left visual field; 4 females in the right visual field and 4 males in the right visual field. An example of one right visual field and one left visual field image can be seen in Figure 44, the full catalogue of images used within the coloured frames can be seen in the appendix.

Figure 44.

Right visual field



Left Visual field



Procedure

The procedure is the same as for Experiment 6a. However, in this case, the faces appeared lateralized: thus, in half of the trials, a face or word appeared to the left (RVF) or right (LVF) of fixation. As before, all stimuli remained present until a correct response was made. It should be noted that although the stimuli are presented lateralized initially, eye movements were not monitored, and thus we cannot be sure whether the stimuli remained lateralized throughout the trials.

Design and analysis

Analysis was carried out using the same procedures as described in Experiment 4 using the number and breakdown of trials as described in Experiment one.

Results

Overview

There were no latencies greater than 10000ms. One participant had latencies faster than 300ms for more than 10% of the trials, so was removed from the analysis, all other participants data was included.

Untransformed RT and accuracy data

Tables 27 and 28 show the mean untransformed reaction times and error rates as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 28 also includes experiment half.

Stimulus Picture	**Extrinsic Response Dimension** Untrustworthy	Trustworthy
Left Visual Field		
Reaction Time (ms)	583 (99.7)	571 (88.3)
Percentage of Errors	11.6 (9.72)	11.0 (8.25)
Right Visual Field		
Reaction Time (ms)	584 (81.1)	580 (95.5)
Percentage of Errors	9.84 (7.93)	12.9 (11.4)

Table 27.

The shaded area for RT data (shown in rows) suggests whether the stimulus was seen as trustworthy or untrustworthy. If the RT shaded is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded RT in the untrustworthy column. The shaded area for Accuracy data suggests whether the stimulus was seen as trustworthy or untrustworthy. If the shaded area is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded accuracy in the untrustworthy column.

For trials with a stimulus picture of a person in the left visual field, participants tended to answer more quickly and slightly more accurately when the extrinsic response dimension was Trustworthy. For trials with a stimulus picture of a person in the right visual field, participants were more accurate when the extrinsic response dimension was Untrustworthy. This data does not support the hypothesis as the RT data does not produce opposing fast RT data depending on the key response, furthermore, the pattern of error rates is in an opposite direction to that expected.

Stimulus Picture	**Extrinsic Response Dimension**			
	Untrustworthy **Experiment Half**		Trustworthy **Experiment Half**	
	1	2	1	2
Gazing Left				
Reaction Time (ms)	595 (95.3)	570 (98.1)	594 (81.8)	581 (98.2)
Percentage of Errors	11.1 (10.0)	12.0 (12.8)	11.3 (11.2)	10.7 (9.47)
Gazing Right				
Reaction Time (ms)	589 (100)	577 (76.3)	581 (113)	583 (102)
Percentage of Errors	9.49 (9.71)	10.2 (9.03)	11.2 (11.5)	14.6 (14.5)

Table 28

Again, reaction times seemed to decrease between experiment halves, apart from trials which showed a person in the right visual field and when the extrinsic response dimension was trustworthy. The error rates also seem to change somewhat between experiment halves, although again these variations were small and did not cause alarm.

ANOVA on latencies

The 2 (Experiment Half) x 2 (Stimulus Picture) x 2(Extrinsic Dimension) repeated measure model was fit to the reaction times (including error penalty), and the logarithm of the reaction times. Neither model had any significant terms.

Discussion of Experiments 6a, b and c

There were three separate hypotheses for each of the subsections of Experiment 6 as each was using a very specific stimulus set to examine different but related issues. It seems that only one experiment confirmed the hypothesis, this being 6a and the effect of expression upon implicit reactions of trust. In agreement with the hypothesis, participants responded quicker to negative images (tongue sticking out) when the key response was mapped as negative, thus implying that photographs with no expression are seen as more trustworthy. However, despite the untransformed data displaying the expected trends, this was not confirmed statistically with no significant terms being found.

As for the other two experiments presented in this chapter, (6b and 6c), the hypothesis was not confirmed. A gaze bias was not seen in the untransformed data tables, with all data indicating fastest responses with the least errors using the same mapped key. It was not known which direction a gaze bias was likely to be, as there is opposing literature to support both a left and right bias, but the lack of any bias was not expected. Also there was no confirmation of the hypothesis for visual field where the results using untransformed data represented the opposite of what was expected to be found with faster reactions with the trustworthy key when images where shown in the left visual field.

The lack of any significant terms for both experiments 6a (Expression) and 6c (visual field), and a significant main effect of experiment half for 6b (gaze) all highlight the failure of this type of image and subject area being tested using the FEAST.

In terms of how these three studies inform the development of the FEAST method it seems to be case that the use of the FEAST is not terribly successful in this particular area.

The RT and error score differences tended to be very small, and this does create cause for concern. Close inspection of the RT and error scores for two of the experiments uncovered some unusual and unexpected results between the experiment halves. In Experiment 6a (which focuses on expression), the RT increases for trustworthy in the second half, with the opposite being the case for untrustworthy where RT decreases between the experiment halves. In Experiment 6b, which focused on gaze direction, the accuracy levels improve in the second half for untrustworthy responses, with the opposite being true for trustworthy responses. This may indicate a problem with the stimuli rather than being of theoretical interest.

The overall lack of significance and generally confusing results could be due to several different factors. First, there is a possibility that the stimuli were to blame: perhaps the words and/or faces chosen were not adequate. Ahead of testing there were no pilot studies to check on the optimum picture (size, visibility or how quickly and easily the expression was perceived). As was the case with all experiments presented a rating study was carried out on both the words and coloured picture images (as they appeared in the experiment) after administration of the three FEASTs. And, this served as an effective check that the stimuli were indeed perceived as was intended, particularly in relation to the white framed words since there were no expectations about the implicit values that would be obtained from the coloured framed pictures.

Second, it could be the case that using "trust" as the tested dimension was not a prudent choice to begin with. From the brief review of the area covering Experiments 6a, 6b and 6c it is easy to see that there are many conflicting arguments regarding visual field superiority effects. In hindsight, it might have been more prudent (based on the literature) to use happy and angry faces since this is where the bulk of the theoretical studies concentrate. Third, the poor results in the present study could have been due to lack or participants (i.e. general issues around statistical power). And, finally, if could be that using two different types of stimuli (words for white frames and pictures for coloured frames) could have been problematic. For example, the words might have generally engaged (or activated) areas of the left hemisphere and this pre-emptive hemispheric priming of the left hemisphere might have served to wash away any right-hemisphere dominance effects (as might be predicted with negative emotional responses).

Experiment introduction

Over the past decade, the rapid (nearly exponential) growth of online commerce has been shockingly fast. Some have gone so far as to claim that e-commerce is the future of *all* commerce (Berezowski, 2003; Tseng, Yue and Taylor, 2005). One of the few factors that has slowed the growth of e-commerce is a psychological one: namely, consumers' hesitancy to click "buy". Indeed, a lack of consumer trust in online shopping is frequently regarded as the most critical hindrance to the development of electronic-commerce (McKnight, Choudhury and Kacmar, 2002). Even with 200 million of the USA's 300 million population using the Internet regularly (CIA press release, 2004), e-commerce accounts for less than 2% of total retail nationwide (Greenspan, 2004).

Trust in trading across the Internet is a critical and current issue since usage is dramatically increasing. By 2011 the number of Internet users worldwide is expected to double (CIA press release, 2006). With enhanced prospect for growth in e-commerce, there is an increased urgency to understand, and perhaps even to overcome, issues of trust.

Consumer trust in e-commerce has been extensively researched in terms of explicit, evaluative ratings of rational elements, such as privacy policies and third party tags (e.g. 'TRUSTe' and 'VeriSign' logos appearing on websites). However, the 'implicit' (indirect and automatic) perception of a website has been greatly neglected, particularly in response to affective cues like trustworthiness.

The face is a very valuable source of social and emotional cues. Photographs of faces are indeed often used as selling points in advertisements. They are comprehensibly and almost immediately attended to (Riegelsberger, 2002), and have a higher likelihood of an affective response than text. Faces in print advertisements are also reported to be less open to critical thought than text used (Kroeber-Riel, 1996, cited in Riegelsberger, 2002). Thus, photographs of faces on websites may evoke an automatic, emotional and less critically evaluated reaction than text.

There is also considerable evidence that the presence of a face can enhance participants' trust in a transaction. In comparison to e-mail contact alone, face-to-face communication has been shown to increase cooperation among participants in the 'prisoner's dilemma' bargaining game (Zheng, Veinott, Bos, Olson and Olson, 2002). Photographs of the correspondents' faces presented alongside their e-mail exchanges proved to result in a more equally distributed outcome for each participant than when communicating solely by e-mail (Roth, 1995). Photographs of authors attached to their online articles have also been found to increase credibility (Fogg, 2002).

Some testing of photographs on e-commerce websites has been conducted, but with mixed outcomes. Steinbrück, Schaumburg, Duda and Krüger (2002) demonstrated that customers' perceived trustworthiness of an interactive banking website increased when an employee photograph was added. This was enhanced even further when the name of the customer service representative was placed next to the photograph. It seemed to act as a more tangible, personal, point of contact.

Other research has found adverse reactions to photographs of faces on e-commerce websites. Users have complained about photographs 'cluttering' the site interface and increasing the time it takes for the website to load (Riegelsberger and Sasse, 2002), without contributing any additional function. There have also been claims that photographs of faces distract users from their tasks, such as, browsing product lists. This may, however, be due to a problem of placement and implementation. With further research, more appropriate positioning could be advised. Riegelsberger, Sasse and Mc Carthy's (2003) research did not support the claim that photographs decrease task performance.

Riegelsberger and Sasse (2002) found that virtual re-embedding had a positive effect on user trust for those with little or intermediate levels of online shopping experience. The same was not found, however, for highly experienced Internet users or consumers who already distrusted online vendors. For heavy users particularly, it seemed that photographs attracted the most attention when first viewed, but were thereafter ignored (Riegelsberger, 2002). Those distrusting of e-commerce interactions interpreted the photographs as strategic manipulations.

It seems reasonable for some interface designers and HCI experts to object to the placement of photographs on e-commerce websites. Each user's reaction to photographs of faces on websites is dependent on multiple factors; therefore there is probably not a single best approach (Riegelsberger and Sasse, 2002). On the other hand, emphasising certain dimensions of photographs in websites with specific target audiences (based on age or gender, for instance) could serve to maximise on the trustworthiness factor of face-to-face communication.

The purpose of Experiments 7a and 7b is to explore how consumer trust in e-commerce websites is influenced by the addition of photographs of faces to checkout pages. In practice, consumers may browse through several websites in search of a particular product. If a good impression is not immediately created then there are plenty of other, highly competitive e-commerce sites that the consumer can move on to. No matter what each website actually has to offer, it seems critical that a positive first impression is vital.

Lindgaard et al. (2006) reasoned that implicit perceptions are lasting impressions. If a website is first seen positively, then its content will be seen in a positive light too (i.e. the 'halo effect'). Consumers will tend to revisit the website that first made a good impression to them in order to confirm their initial decision, thus attempting to prove themselves right (Lindgaard et al., 2006). A negative first impression can detrimentally bias a consumer's cognitive judgment of the website overall. It is likely to lead to future avoidance of the website (Lindgaard et al., 2006; Tractinsky, Cokhavi and Kirschenbaum, 2004).

The inverse can be predicted with low-involvement e-commerce websites dealing in inexpensive, routine products, for example, rubber gloves or books. We would expect people to judge whether or not to purchase from these sites based more on their first impression than on the explicit content of the website.

In purely electronic exchanges, consumers have to pay an invisible entity before receipt of the purchased item. This puts them at risk should anything go wrong with the transaction. There is not always post-purchase assurance that unwanted or faulty goods can be exchanged or their value refunded. The consumer has to trust that only the pre-determined amount of money will be debited, that the product they receive will be the one desired, that it will be received in terms mutually agreed upon, and that the privacy of their personal (contact and financial) details will be maintained. This trust is required without the consumer being able to personally inspect (e.g. touch) the product prior to purchase, thereby having to accept the vendor's word on condition and authenticity (without having even seen the vendor).

In an online environment, spatial and temporal gaps between the consumer and vendor, as well as the lack of face-to-face correspondence, may make it difficult for a consumer to assess a vendor's trustworthiness. Also, without a walk-in equivalent of the electronic shop, potential customers may have doubts over the vendor's history of prior exchanges (Bhattacherjee, 2002). Further still, varying regulations and legislations, dependent on the part of the world from which the business operates, may be a cause of confusion and uncertainty (Nielsen, 2000). Such obstacles serve to block consumer's trust in online shops and may make it more difficult for them to make initial purchases, repeat purchases, and maintain enduring online relationships (Lombard and Ditton, 1997).

Although essentially providing the same service, the way in which online and walk-in shops operate and are perceived by consumers is quite different. Walk-in shops have sales clerks in person to make the payment transaction for the consumer. If the consumer has any concerns or uncertainties the clerk can immediately offer advice and reassurance. In online shops, consumers must electronically dispatch their personal details and tick boxes to accept (detailed and lengthy) terms and conditions via an automated system. This difference in approach seems to have led consumers to trust walk-in shops more than their online equivalents. From a survey of online shoppers, 51% stated that they browse on the Internet, but purchase from walk-in shops (NPD Group, 2001).

As reported by Gordon (2000), 60-70% of online shoppers abandon their selected items at the point-of-sale on e-commerce sites. Of Americans who had never bought online before, 35% abandoned their shopping carts due to concerns of providing their personal information and 24% did so because of fears regarding credit card fraud (Kelley, Rhinelander and DeMoulin, 2001). Almost 95% of American Internet users had previously refused to provide their personal information when asked at the point of sale (Nielsen Media Research's, 1997). According to the same survey sample of 45 million, 40% lied when giving their demographic details (Nielsen Media Research's, 1997). Consequently, we posit that the vendor's trustworthiness needs to be enhanced at this point, on the checkout webpage.

An instinctive 'feel' of trust is acquired in a walk-in shop as it is in any environment. Immediate cues as to the establishment's professionalism, friendliness and competence are partially conveyed through the presence of its staff and clientele. Accordingly, efforts made to build trust on websites need to be based on elements appealing to emotional as well as rational decision making.

'Virtual re-embedding' (Riegelsberger and Sasse, 2002), that is, adding social cues to interactive media, is considered in this study as a plausible tactic for increasing consumers' trust in online-vendors. It is a mechanism that constructs a more personalised social interaction between people who are not in direct face-to-face contact (Hjarvard, 2002). The Internet is a prime example of a medium in which this is implemented. Examples include 'signatures' (i.e. the correspondent's electronic autograph, an icon, or quote that is automatically placed) at the end of e-mails. One of the most fundamental means of embedding social cues in websites is with the use of photographs of faces.

Following previous literature founded on explicit measures, participants explicitly claim to be more trusting of smiling faces than neutral faces (LaFrance and Hecht, 1995) and right-sided faces more than left-sided faces (Chen, German and Zaidel, 1997). According to Riegelsberger and Sasse (2002), using a sample likely to have high Internet use levels (such as the student sample in this study) may result in websites without faces being rated as more trustworthy than those with faces.

Substantial evidence suggests that men and women each have different reactions to facial expressions, dependent on the sex of the face presented. In fact, both men and women seem to have a predisposition towards primarily trusting members of their opposite sex. In online bargaining games, Scharlemann et al. (2001) found men to be more cooperative when a photograph of a female face represented their correspondent. Women were more cooperative when their counterpart was visually presented as a man. Social psychology proposes that this tendency is incited by a sexual interest (Scharlemann et al., 2001).

Furthermore, it seems that females cooperate at similar levels no matter what their correspondent's facial expression is. Males, on the other hand, cooperate more willingly with a smiling person (Scharlemann et al., 2001). This is thought to be a resultant factor of the male's evolutionary dominance over females. In modern society, where there is more sexual equality, those in subordinate roles smile more than those in powerful positions regardless of gender (Deutsch, 1990). It still remains, however, that males occupy more dominant positions, and so, may generally be more alert to smiles than females. Among women, smiling is commonly used, and consequently is perceived to have less credibility and honesty attached (Hinsz and Tomhave, 1991).

Besides the aforementioned sex discrepancies in receptiveness to facial expression, sex in itself may be predictive of people's tendency to cooperate. Women are stereotypically portrayed as the 'cooperative sex' (Sharlemann et al., 2001). Extensive data from male versus female bargaining games conversely shows an individual's decision to be dependent on several factors other than sex (Sell, 1997; Walters and Stuchlmacher, 1998). Women do not necessarily cooperate more often.

Experiment rationale

Although it was the case that implicit measures of trust were not obtained very successfully with simple more controlled stimuli (Experiments 6 a,b, and c) it was decided that the FEAST method may simply be better suited to designs where the stimuli were more complex. Given that the EAST method has been used in a variety of genres as described earlier, some of which are more successful than others, it seemed sensible to assume that the FEAST would be no different, performing better or worse depending on the stimulus type and dimensions tested.

Hypothesis

It was expected that the presentation of a photograph on a website would increase the feeling of trust towards that site, mainly due to feeling the social presence and accountability that a face provides a website with. Thus, when a website with a face was presented and required a key-response that was extrinsically mapped to "trustworthy" performance would be better (both in terms of reaction time and accuracy). And, when a website had no face present, correct responses would be better with the extrinsically "untrustworthy" key.

Experiment 7a

Method

Participants

Thirty five psychology students participated as part of a course requirement. All participants had normal or corrected to normal vision. Fifteen males and twenty females participated. Their mean age was 21.23 years, aged between 18 and 39 years (SD= 3.639).

Materials and measures

To make the stimuli appear real, the website stimuli were edited versions of already existing e-commerce checkout pages. We used edited versions to make it difficult for participants to be able to immediately identify them as coming from a specific site. Names, logos, website addresses, layout and so on were changed for this reason. Websites were displayed ten times: for two of these times, the website did not contain a face, while the other eight times the website displayed a photograph of a face. The faces that were utilized were created specifically for this experiment, as detailed in the next paragraph.

A group of females were recruited from the student population to serve as models. Those photographed were instructed to present themselves neatly and professionally - as if their photographs were actually going to be placed on the Internet, with them acting as representatives of their company. Models were instructed to make four poses, two to be used to portray 'expression' (smiling and neutral), and the other two for the 'gaze' dimension. The latter involved the volunteers turning their heads so that either their left or right cheek was prominently in view, whilst they looked directly at the camera and maintained a neutral expression. To take these photographs the camera was positioned approximately 30cm away from the volunteers' faces. Photographs of four volunteers were taken, though we only used the two sets that portrayed the desired qualities most clearly. The selected models were both white and aged 24 and 22 years respectively. This served to limit race and age effects, considering that the majority of possible participants were also white and of a similar age. Everything surrounding the photographed faces was painted white so that it matched with the website background. An example of an expression and gaze website can be seen in Figure 45, the full catalogue of websites seen within coloured frames can be found in the appendix.

Figure 45.

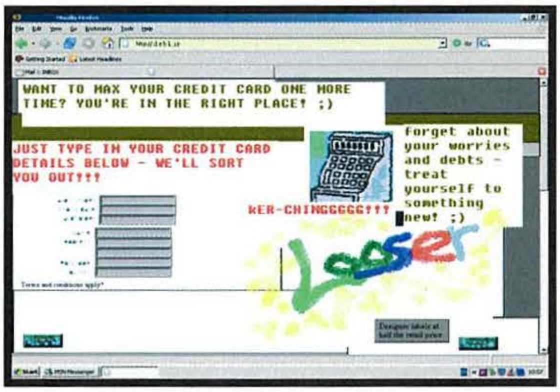
Expression version website - smiling

Gaze version website

Twenty-four white-framed primes were used to set the explicit keypress dimensions. Twelve of the 24 primes were made to appear ‘trustworthy’ and the remaining twelve were ‘untrustworthy’. ‘Trustworthy’ websites were tailored by adding positive, affirmative statements such as “Experts in Electronics. Here to Help You”, adding VeriSign and Trust-e logos or padlocks, and removing unnecessary clutter. ‘Untrustworthy’ websites created that impression by implying that the goods for sale were counterfeit, and/or by adding demanding, money orientated messages such as, ‘Show me the money!!’ They also were purposely unclear in a very obvious manner. Terms and conditions were written in a font too small to be legible. Text was scrambled, and/or misspelled, and/or coloured to be almost indistinguishable from the background. Examples of both a trustworthy and untrustworthy website can be seen in Figure 46.

Figure 46.

Untrustworthy website



Trustworthy website

Participants took part in both the gaze and expression versions of the experiment which were counterbalanced across participants to control for order effects. Websites were chosen to induce trust rather than using words, concepts, or pictures. This decision was primarily taken due to the PhD sponsors who work in website design. As was the case with all previous experiments reported, an explicit ratings study was carried out to ensure that there was confidence that the websites were viewed as trustworthy or not. The websites were created to represent trust or not in a very obvious manner using slogans and inappropriate pictures.

Procedure

As before, the coloured frame always appeared around the outermost edge of the display. The coloured frame was white, blue, or green. Participants were instructed that when the frame is white they must make a response to the website based on trust (i.e. they must classify it as either trustworthy or un-trustworthy). When the frame is either green or blue, participants were instructed to make a response based on the colour of the frame, that is, to classify whether the frame is green or blue, ignoring the website presented in the centre of the display. Both the frame and website remained on screen until a correct response was made. All other aspects remain consistent with procedures described in the previous experiments reported. No change for presentation time was implemented, although the framed websites are clearly more complex than simple word stimuli. The reason for this was that the websites presented within white frames were created to represent obvious extremes so it was thought that a participant would easily and immediately recognize whether the website was trustworthy or not.

Design and analysis

Analysis was carried out using the same procedures as described in Experiment 4. Websites were designed to represent trustworthy or not trustworthy by the use of criterion such as text scrambled, images, slogans, and difficult to read sites. This criterion was decided upon after discussion with the web design team at my PhD company sponsor. They advised about what was seen to make a site untrustworthy, and the sites used for the trustworthy dimension were established sites. As was the case with all previous experiments reported, the explicit rating study was carried out after the FEAST experiment and the results added weight to the assumption that the white framed websites were indeed clearly perceived as trustworthy or untrustworthy. The results are presented with all 'Expression' data first (neutral or happy), followed by 'Gaze' data (leftward or rightward facing). For each dimension, expression or gaze has separate analyses carried out using the D600 method. There are also two sets of analyses reported. The first looks at the effects of the dimension itself (either gaze to the right versus gaze to the left, or neutral versus happy expression), against 'no face' as a control. The second set of analyses looks at the effect of displaying a face versus no face, where the 'face' data includes all data where a face was present irrelevant of gaze or expression. Each participant completed 288 trials which were broken down by each coloured face (or no face website) being seen 12 times, and each white framed website also being seen a total of 12 times each. To describe this differently, there were 48 neutral and 48 smiling websites and 24 with no face presented. When the experiment was initially carried out and designed it was not intended that there should be a specific individual analysis carried out comparing websites with and without a face, hence the uneven numbers.

Results

Expression

Overview

There were no latencies greater than 10000ms. Participant 31 had latencies faster than 300ms for more than 30% of the trials, so was removed from the analysis. The analysis was therefore performed on data from the remaining 34 participants.

Untransformed RT and accuracy data

The next two tables, 29 and 30, show the mean untransformed reaction times and error rates as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 30 also includes experiment half. These tables show the latencies before any error penalty is introduced.

Stimulus Picture	**Extrinsic Response Dimension**	
	Untrustworthy	Trustworthy
Female Neutral Expression		
Reaction Time (ms)	581 (104)	576 (113)
Percentage of Errors	13.2 (12.7)	12.7 (11.0)
Female Smiling		
Reaction Time (ms)	573 (97.4)	594 (101)
Percentage of Errors	13.2 (13.8)	11.4 (13.3)
No Face		
Reaction Time (ms)	592 (120)	589 (102)
Percentage of Errors	22.4 (23.4)	11.0 (14.0)

Table 29

After examining the data presented in Table 29 it can be seen that for trials where the stimulus picture did *not* include a face, participants tended to respond more quickly and much more accurately when the extrinsic response valence was trustworthy, indicating a preference for no photograph. For trials when the stimulus picture included a female smiling, participants tended to respond more quickly when the extrinsic response valence was untrustworthy, however participants made more errors in this case. For trials when the stimulus picture included a female with a neutral expression, participants tended to respond more quickly and accurately when the extrinsic response dimension was trustworthy, however, these differences were relatively small, although once again they do suggest that when there is a face that a neutral expression is seen to be most trustworthy.

Stimulus Picture	**Extrinsic Response Dimension**			
	Untrustworthy		Trustworthy	
	Experiment Half		**Experiment Half**	
	1	2	1	2
Female Neutral Expression				
Reaction Time (ms)	570 (114)	587 (129)	567 (106)	587 (130)
Percentage of Errors	10.7 (14.5)	15.8 (16.1)	13.2 (12.7)	12.1 (15.5)
Female Smiling Expression				
Reaction Time (ms)	576 (127)	575 (114)	584 (93.1)	604 (125)
Percentage of Errors	12.5 (15.1)	14.0 (17.1)	9.56 (11.5)	13.2 (18.2)
No Face				
Reaction Time (ms)	607 (128)	572 (137)	577 (122)	601 (116)
Percentage of Errors	19.1 (21.4)	25.7 (29.2)	12.5 (18.7)	9.56 (16.3)

Table 30

Table 30 shows that the reaction times and error rates did change between experiment halves, however, the direction of the changes varied between trial types. The difference in reaction times and error rates for each extrinsic response dimension changed direction several times.

ANOVA on latencies

The 2 (Experiment Half) x 3 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures model was fit to the logarithms of the reaction times (including error penalties). The model had two significant terms: the main effect of stimulus picture: F

(2, 66) =3.20, p=0.05, and the interaction between stimulus picture and extrinsic response dimension: F (2, 66) =4.22, p=0.02.

D600 analysis

For trials with a stimulus picture of a female with a neutral expression, the mean D600 EAST score was D=0.022. For trials with a stimulus picture of a female smiling, the mean D600 EAST score was D=-0.023. One sample t-tests suggested there was no evidence that these scores were significantly different from zero. For trials with a stimulus picture which did not include a face, the mean D600 EAST score was D=0.24. Participants tended to respond more quickly and accurately when the extrinsic response dimension was trustworthy, t (33) =2.56, p=0.02.

Correlations

The correlation between the scores for trials which included a neutral face was significant, r=0.540, t (32) =3.63, p<0.001. The correlation between the scores for trials which included a smiling face was r=0.283. The correlation between the scores for trials which included no face was r=0.202.

Presence of a face

Untransformed RT and accuracy data

Tables 31 and 32 show the mean untransformed reaction times and error rates as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 32 also includes experiment half. These tables show the latencies before any error penalty is introduced.

Stimulus Picture	**Extrinsic Response Dimension** Untrustworthy	Trustworthy
Face		
Reaction Time (ms)	577 (97.3)	586 (101)
Percentage of Errors	13.2 (12.4)	12 (10.6)
No Face		
Reaction Time (ms)	592 (120)	589 (102)
Percentage of Errors	22.4 (23.4)	11.0 (14.0)

Table 31

For Table 31 the shaded area for RT data (shown in rows) suggests whether the stimulus was seen as trustworthy or untrustworthy. If the RT shaded is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded RT in the untrustworthy column. The shaded area for Accuracy data suggests whether the stimulus was seen as trustworthy or untrustworthy. If the shaded area is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded accuracy in the untrustworthy column. Participants did tend to respond more quickly when there was a face included in the website. Participants tended to answer much more accurately for websites which did not include a face when the extrinsic response dimension was trustworthy.

Stimulus Picture	**Extrinsic Response Dimension**			
	Untrustworthy		Trustworthy	
	Experiment Half		**Experiment Half**	
	1	2	1	2
Face				
Reaction Time (ms)	574 (110)	581 (116)	575 (90.3)	598 (120)
Percentage of Errors	11.6 (12.9)	14.9 (14.8)	11.4 (9.79)	12.7 (13.5)
No Face				
Reaction Time (ms)	607 (128)	572 (137)	577 (122)	601 (116)
Percentage of Errors	19.1 (21.4)	25.7 (29.2)	12.5 (18.7)	9.56 (16.3)

Table 32

Table 32 shows that, in most cases, participants tended to answer more quickly and accurately in the first half of the experiment. The direction of the difference in reaction times for extrinsic response dimension changed between experiment halves.

ANOVA on latencies

The 2 (Experiment Half) x 2 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures model was fit to logarithm of the reaction times (including error penalties). The model had two significant terms; the main effect of stimulus picture: $F(1, 33) = 4.35$, $p=0.04$ and the interaction between stimulus picture and extrinsic response dimension: $F(1, 33) = 5.59$, $p=0.02$. It also had a marginally significant term; the main effect of extrinsic response dimension: $F(1, 33) = 3.73$, $p=0.06$.

D600 analysis

D600 EAST scores were calculated for each participant for trials which used each stimulus picture. For trials with a stimulus picture which included a female face, the mean D600 EAST score was $D=-0.005$. For trials with a stimulus picture which did not include a face, the mean D600 EAST score was D=0.24. Participants tended to respond more quickly and accurately when the extrinsic response dimension was trustworthy, t (33) =2.56, p=0.02. A paired, two-sample, t-test suggested there was evidence to suggest a difference in between the two sets of scores: t (33) =2.12, p=0.04.

Correlation

The D600 EAST scores were calculated for each experiment half and the correlations between the two sets of scores for each stimulus picture group were also found. Only one correlation was significant, this was for the correlation between the trials which included a female face, r=0.51, t (32) =3.39, p<0.01.

Gaze

Overview

There were no latencies greater than 10000ms. Participant 31 had latencies faster than 300ms for 19% of the trials, so was removed from the analysis. The analysis was therefore performed on data from the remaining 34 participants.

Untransformed RT and accuracy data

Tables 33 and 34 show the mean untransformed reaction times and error rates as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 34 also includes experiment half. These tables show the latencies before any error penalty is introduced.

Stimulus Picture	Extrinsic Response Dimension Untrustworthy	Trustworthy
Female Gazing Left		
Reaction Time (ms)	618 (191)	590 (105)
Percentage of Errors	13.6 (14.7)	10.8 (9.14)
Female Gazing Right		
Reaction Time (ms)	581 (102)	585 (91.9)
Percentage of Errors	16.0 (18.3)	11.4 (10.9)
No Face		
Reaction Time (ms)	596 (115)	585 (93.7)
Percentage of Errors	17.3 (18.7)	11.4 (13.9)

Table 33

Table 33.

The shaded area for RT data (shown in rows) suggests whether the stimulus was seen as trustworthy or untrustworthy. If the RT shaded is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded RT in the untrustworthy column. The shaded area for Accuracy data suggests whether the stimulus was seen as trustworthy or untrustworthy. If the shaded area is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded accuracy in the untrustworthy column. Table 33 also shows that for trials where the stimulus picture included a female gazing to the left and in trials where the stimulus picture did not include a face, participants tended to respond more quickly and accurately when the extrinsic response dimension was trustworthy. For trials where the stimulus picture included a female gazing to the right, participant tended to answer more accurately when the extrinsic response dimension was trustworthy, although there were no large differences in reaction times.

Stimulus Picture	**Extrinsic Response Dimension**			
	Untrustworthy **Experiment Half**		Trustworthy **Experiment Half**	
	1	2	1	2
Female Gazing Left				
Reaction Time (ms)	582 (108)	657 (354)	580 (109)	603 (152)
Percentage of Errors	13.2 (17.1)	14.0 (17.9)	11.8 (13.4)	9.93 (11.0)
Female Gazing Right				
Reaction Time (ms)	586 (112)	576 (118)	563 (94.9)	606 (148)
Percentage of Errors	13.6 (15.8)	18.4 (23.3)	14.7 (14.6)	8.09 (11.9)
No Face				
Reaction Time (ms)	577 (101)	619 (168)	597 (123)	571 (103)
Percentage of Errors	14.7 (20.5)	19.9 (23.7)	12.5 (20.6)	10.3 (15.2)

Table 34

Table 34 shows that the reaction times and error rates did change between experiment halves, however, the direction of the changes varied between trial types. The direction of the difference in error rates between the between the extrinsic response dimension did change between experiment halves, often getting slower RT in the second half.

ANOVA on latencies

The 2 (Experiment Half) x 3 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures model was fit to the logarithms of the reaction times (including error penalties). The model had two marginally significant terms; the main effect of extrinsic response dimension: $F(1, 33) = 3.27$, $p = 0.08$ and the interaction between experiment half and extrinsic response dimension: $F(1, 33) = 3.54$, $p = 0.07$.

D600 analysis

For trials with a stimulus picture of a female gazing left, the mean D600 EAST score was D= 0.069. For trials with a stimulus picture of a female gazing right, the mean D600 EAST score was D=0.053. One sample t-tests suggested there was no evidence that these scores were significantly different from zero. For trials with a stimulus picture which did not include a face, the mean D600 EAST score was D=0.17. Participants tended to respond more quickly and accurately when the extrinsic response dimension was good, $t(33) = 1.80$, $p = 0.08$ (marginal significance).

Correlation

The correlation between the scores for trials which included a face gazing to the left was $r = -0.152$. The correlation between the scores for trials which included a face gazing to the right was $r = 0.159$. The correlation between the scores for trials which included no face was $r = 0.194$.

Presence of a face

Untransformed RT and accuracy data

The next two tables, 35 and 36, show the mean untransformed reaction times and error rates as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 36 also includes experiment half. These tables show the latencies before any error penalty is introduced.

Stimulus Picture	**Extrinsic Response Dimension**	
	Untrustworthy	Trustworthy
Face		
Reaction Time (ms)	599 (137)	588 (87.8)
Percentage of Errors	14.8 (15.8)	11.1 (8.54)
No Face		
Reaction Time (ms)	596 (115)	585 (93.7)
Percentage of Errors	17.3 (18.7)	11.4 (13.9)

Table 35

Table 35. The shaded area for RT data (shown in rows) suggests whether the stimulus was seen as trustworthy or untrustworthy. If the RT shaded is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded RT in the untrustworthy column. The shaded area for Accuracy data suggests whether the stimulus was seen as trustworthy or untrustworthy. If the shaded area is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded accuracy in the untrustworthy column. Table 35 also shows that participants tended to respond more quickly and accurately when the extrinsic response dimension was trustworthy whether there was a photograph present or not.

Stimulus Picture	Extrinsic Response Dimension			
	Untrustworthy		Trustworthy	
	Experiment Half		**Experiment Half**	
	1	2	1	2
Face				
Reaction Time (ms)	585 (97.5)	616 (210)	572 (94.0)	604 (117)
Percentage of Errors	13.4 (14.6)	16.2 (19.0)	13.2 (11.5)	9.01 (9.13)
No Face				
Reaction Time (ms)	577 (101)	619 (168)	597 (123)	571 (103)
Percentage of Errors	14.7 (20.5)	19.9 (23.7)	12.5 (20.6)	10.3 (15.2)

Table 36

Table 36 shows that, in most cases, participants tended to answer more quickly and accurately in the first half of the experiment.

ANOVA on latencies

The 2 (Experiment Half) x 2 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures model was fit to the logarithms of the reaction times (including error penalties). The model had one significant term; the interaction between experiment half and extrinsic response direction: $F(1, 33) = 5.53$, $p = 0.02$. It also had one marginally significant term; the main effect of extrinsic response dimension: $F(1, 33) = 3.72$, $p = 0.06$.

D600 analysis

For trials with a stimulus picture which included a female face, the mean D600 EAST score was D=0.060. There was no evidence to suggest that these scores were significantly different from zero. For trials with a stimulus picture which did not include a face, the mean D600 EAST score was D=0.17 (marginally significant $t(33) = 1.80$, $p = 0.08$). A paired two-sample t-test suggested there was no evidence of a difference between the two sets of scores.

Correlation

The correlation between the scores for trials which included a face was r=0.116. The correlation between the scores for trials which included no face was r=0.194.

Experiment 7b

Experiment rationale

As a result of discussions with the PhD sponsor it was clear that there were different expectations from their clients about how they wanted the design of their website to both look and to 'feel'. Interestingly it seemed that photographs were often not used as there was a fear about creating the wrong impression of a company, often this choice was over the use of a male or female face and what age of face to include. From informal discussions it seemed that there was a view that a male photograph could be seen as too domineering and less trustworthy, whilst a female face was perceived as more trustworthy but less professional. It is important to note that these were just informal observations made and not tested; this was the basis for testing using male faces.

Method

Participants

Thirty-eight students were recruited at random from the University of Wales, Bangor experimental pool, who took part in order to gain course credits. Ten males and twenty eight females participated. Their mean age was 20.55 years, aged between 18 and 30 years (SD= 5.29).

Materials and measures

In Experiment 7a all faces were female, in Experiment 7b all faces are male. The male photographs were taken from the GTAV database used in Experiment 6. Images were taken from the GTAV database instead of using photographs taken from the university population due to difficulty finding willing male participants. The photographs in both 7a and b are in the same style so this was not expected to alter the results. All other manipulations of the faces regarding both gaze and expression remain the same. Also, all websites used remained the same: in essence, this experiment simply replaced the female faces with male faces.

Procedure

All details remained identical to Experiment 7a.

Design and analysis

Analysis was carried out using the same procedures as described in Experiment 4 using the number and breakdown of trials as described in Experiment one.

Results

Expression

Overview

There were no latencies greater than 10000ms. Participant 28 had latencies faster than 300ms for 25% of the trials, and so was removed from the analysis. Participant 6 had a great deal of missing data, so was also removed from the analysis. The analysis was thus performed on data from the remaining 36 participants.

Untransformed RT and accuracy data

In Tables 37 and 38 the mean untransformed reaction times and error rates are shown as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 38 also includes experiment half. These tables show the latencies before any error penalty is introduced.

Stimulus Picture	Extrinsic Response Dimension: Untrustworthy	Extrinsic Response Dimension: Trustworthy
Male Neutral Expression		
Reaction Time (ms)	619 (159)	604 (141)
Percentage of Errors	14.6 (16.0)	17.0 (12.3)
Male Smiling		
Reaction Time (ms)	641 (171)	606 (131)
Percentage of Errors	12.5 (13.4)	13.2 (11.0)
No Face		
Reaction Time (ms)	607 (160)	608 (133)
Percentage of Errors	18.1 (16.8)	10.1 (13.0)

Table 37.

The shaded area for RT data (shown in rows) suggests whether the stimulus was seen as trustworthy or untrustworthy. If the RT shaded is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded RT in the untrustworthy column. The shaded area for Accuracy data suggests whether the stimulus was seen as trustworthy or untrustworthy. If the shaded area is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded accuracy in the untrustworthy column. Table 37 shows that the error rates and reaction times tend to disagree (i.e. that there is a speed-accuracy tradeoff which is often seen throughout the experiments presented as part of this thesis). In trials where the stimulus picture included a male with either a neutral expression or a male smiling, participants tended to answer more quickly when the extrinsic response dimension was trustworthy, however, participant tended to make more errors in these cases. For trials which did not include a face, participant made fewer errors when the extrinsic response dimension was trustworthy.

Stimulus Picture	Extrinsic Response Dimension			
	Untrustworthy Experiment Half		Trustworthy Experiment Half	
	1	2	1	2
Male Neutral Expression				
Reaction Time (ms)	631 (221)	605 (138)	618 (155)	594 (150)
Percentage of Errors	13.9 (18.8)	15.3 (18.4)	19.4 (15.4)	14.6 (12.9)
Male Smiling Expression				
Reaction Time (ms)	651 (192)	628 (187)	612 (152)	600 (131)
Percentage of Errors	11.5 (15.3)	13.5 (15.6)	12.80 (13.2)	13.5 (12.8)
No Face				
Reaction Time (ms)	601 (201)	604 (148)	608 (155)	607 (139)
Percentage of Errors	18.1 (23.6)	18.1 (22.0)	9.03 (14.8)	11.1 (16.3)

Table 38

Table 38 shows that the reaction times and error rates did change between experiment halves, however, the direction of the changes varied between trial types. In most cases, participants answered more quickly in the second half of the experiment.

D600 analysis

For trials with a stimulus picture which included a male with a neutral expression, the mean D600 EAST score was D=-0.040. For trials with a stimulus picture which included a male smiling, the mean D600 EAST score was D=0.048. For trials with a stimulus picture which did not include a face, the mean D600 EAST score was D=0.14. One sample t-tests suggested there was no evidence to suggest these scores were significantly different from zero.

Correlations

The correlation between the scores for trials which included a neutral face was r=0.123. The correlation between the scores for trials which included a smiling face was r=0.277. The correlation between the scores for trials which included no face was r=0.211.

Presence of a face

Untransformed RT and accuracy data

Tables 39 and 40 show the mean untransformed reaction times and error rates as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 40 also includes experiment half. These tables show the latencies before any error penalty is introduced.

Stimulus Picture	Extrinsic Response Dimension	
	Untrustworthy	Trustworthy
Face		
Reaction Time (ms)	631 (157)	605 (134)
Percentage of Errors	13.5 (14.1)	15.1 (10.2)
No Face		
Reaction Time (ms)	607 (160)	608 (133)
Percentage of Errors	18.1 (16.8)	10.1 (13.0)

Table 39.

The shaded area for RT data (shown in rows) suggests whether the stimulus was seen as trustworthy or untrustworthy. If the RT shaded is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded RT in the untrustworthy column. The shaded area for Accuracy data suggests whether the stimulus was seen as trustworthy or untrustworthy. If the shaded area is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded accuracy in the untrustworthy column. Table 39 shows that for trials when the stimulus picture included a face, participants tended to respond more quickly when the extrinsic response dimension was trustworthy; however they made more errors in this case. For trials when the stimulus picture did not include a face, participants answered much more accurately when the extrinsic response direction was trustworthy.

Stimulus Picture	Extrinsic Response Dimension			
	Untrustworthy		Trustworthy	
	Experiment Half		**Experiment Half**	
	1	2	1	2
Face				
Reaction Time (ms)	641 (182)	616 (158)	615 (146)	597 (133)
Percentage of Errors	12.7 (15.5)	14.4 (15.4)	16.1 (11.9)	14.1 (11.0)
No Face				
Reaction Time (ms)	601 (201)	604 (148)	608 (155)	607 (139)
Percentage of Errors	18.1 (23.6)	18.1 (22.0)	9.03 (14.8)	11.1 (16.3)

Table 40

Table 40 shows that, in most cases, participants tended to answer more quickly in the second half of the experiment. There was a small difference in error rates between experiment halves.

ANOVA on latencies

The 2 (Experiment Half) x 2 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures model was fit to the logarithm of the reaction times (including error penalties). There were no significant terms.

Gaze

Overview

There was one latency greater than 10000ms; this was removed from the analysis. Participants 23 and 28 had latencies faster than 300ms for more than 18% of the trials, so were also removed from the analysis. Participant 17 had lots of missing data, so was also removed. The analysis was performed on data from 35 participants.

Untransformed RT and accuracy data

Tables 41 and 42 show the mean untransformed reaction times and error rates as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 42 also includes experiment half. These tables show the latencies before any error penalty is introduced.

Stimulus Picture	**Extrinsic Response Dimension**	
	Untrustworthy	Trustworthy
Male Gazing Left		
Reaction Time (ms)	636 (150)	610 (113)
Percentage of Errors	17.1 (19.7)	13.2 (16.2)
Male Gazing Right		
Reaction Time (ms)	623 (138)	608 (123)
Percentage of Errors	17.0 (16.3)	14.5 (10.7)
No Face		
Reaction Time (ms)	639 (137)	630 (143)
Percentage of Errors	17.5 (19.5)	14.6 (15.9)

Table 41.

The shaded area for RT data (shown in rows) suggests whether the stimulus was seen as trustworthy or untrustworthy. If the RT shaded is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded RT in the untrustworthy column. The shaded area for Accuracy data suggests whether the stimulus was seen as trustworthy or untrustworthy. If the shaded area is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded accuracy in the untrustworthy column. Table 41 shows that for all trials, participants seemed to respond more quickly and accurately when the extrinsic response dimension was trustworthy.

Stimulus Picture	Extrinsic Response Dimension			
	Untrustworthy		Trustworthy	
	Experiment Half		**Experiment Half**	
	1	2	1	2
Male Gazing Left				
Reaction Time (ms)	636 (140)	628 (202)	607 (113)	615 (127)
Percentage of Errors	16.1 (18.8)	18.2 (23.7)	14.6 (17.8)	11.8 (16.9)
Male Gazing Right				
Reaction Time (ms)	624 (172)	621 (135)	611 (132)	606 (134)
Percentage of Errors	18.9 (19.3)	15.0 (18.9)	14.6 (14.7)	14.3 (14.3)
No Face				
Reaction Time (ms)	620 (142)	657 (179)	671 (253)	596 (106)
Percentage of Errors	19.3 (23.6)	15.7 (21.9)	15.7 (18.3)	13.6 (18.5)

Table 42

Table 42 shows that the reaction times and error rates did change between experiment halves, however, the direction of the changes varied between trial types. The direction of the difference in error rates between the between the extrinsic response dimension did not often change between experiment halves.

ANOVA on latencies

The 2 (Experiment Half) x 3 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures model was fit to the logarithm of the reaction times (including error penalties. There were no significant terms.

Presence of a face

Untransformed RT and accuracy data

Tables 43 and 44 show the mean untransformed reaction times and error rates as a function of the stimulus picture group and extrinsic dimension (Trustworthy or Untrustworthy). Table 44 also includes experiment half. These tables show the latencies before any error penalty is introduced.

Stimulus Picture	**Extrinsic Response Dimension**	
	Untrustworthy	Trustworthy
Face		
Reaction Time (ms)	629 (131)	609 (115)
Percentage of Errors	17.1 (16.9)	13.8 (12.3)
No Face		
Reaction Time (ms)	639 (137)	630 (143)
Percentage of Errors	17.5 (19.5)	14.6 (15.9)

Table 43.

The shaded area for RT data (shown in rows) suggests whether the stimulus was seen as trustworthy or untrustworthy. If the RT shaded is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded RT in the untrustworthy column. The shaded area for Accuracy data suggests whether the stimulus was seen as trustworthy or untrustworthy. If the shaded area is in the trustworthy column this infers a trustworthy picture, the opposite being the case for shaded accuracy in the untrustworthy column. Table 43 shows that participants tended to respond more quickly and accurately when the extrinsic response dimension was trustworthy. The reaction times tended to be longer when the stimulus picture did not include a face: suggesting that faces somehow sped up overall processing of the webpages (or, at least, speed up the overall decision-making and responding process).

Stimulus Picture	Extrinsic Response Dimension			
	Untrustworthy		Trustworthy	
	Experiment Half		**Experiment Half**	
	1	2	1	2
Face				
Reaction Time (ms)	629 (142)	632 (153)	610 (115)	611 (125)
Percentage of Errors	17.5 (16.9)	16.6 (19.6)	14.6 (14.2)	13.0 (13.1)
No Face				
Reaction Time (ms)	620 (142)	657 (179)	671 (253)	596 (106)
Percentage of Errors	19.3 (23.6)	15.7 (21.9)	15.7 (18.3)	13.6 (18.5)

Table 44

Table 44 shows that the reaction times and error rates for trials which included a face did not change dramatically between experiment halves. For trials which did not include a face, the reaction times changed considerably between experiment halves.

ANOVA on latencies

The 2 (Experiment Half) x 2 (Stimulus Picture) x 2 (Extrinsic Dimension) repeated measures model was fit to the logarithm of the reaction times (including error penalties. However, there were no significant terms.

Chapter discussion

Experiment 7a produced some interesting results. Some of these seen in the existence of significant results, and in other cases the lack of significant results proved interesting (in that they did not demonstrate the expected behaviour). It was anticipated that the presence of a face on a website would increase a viewer's feeling of trust towards that site as it should have provided a personal touch-- as if they were dealing with a person and not a machine. However, in the comparison of the untransformed data for trials with a face present compared to no face at all, it is the occasions where there is no face that produces the fastest RTs when using the trustworthy mapped key response. This is in accordance with Riegelsberger and Sasse (2002), who argued that using a sample likely to have high Internet use levels may result in websites without faces being rated as more trustworthy than those with faces. In the comparison of which facial expression produces the most trustworthy website it appears that having a neutral expression (as opposed to a smile) evoked stronger feelings of trust as RTs using the trustworthy mapped key response were fastest when viewing the neutral compared to the smiling face.

This result may be a product of the type of website being used as the stimulus. Had the website reflected a homepage for example, rather than a checkout where a smile may not be greeted in the same way as an opening 'Hello', the results may have reflected a different trend.

Regarding the effect of gaze, the untransformed data indicated again that having no face made the site appear more trustworthy than having a face present. In terms of whether it is better to present a leftward or a rightward gazing face, the data indicated a trend for leftward gazing faces to be responded to with faster RTs when the response key used is mapped as trustworthy.

However, these same trends were not replicated when using male faces (7b) where there was a small indication that webpages with male faces present are viewed as more trustworthy than when there is no face presented. Regarding gaze: with male faces all stimulus types, (leftward gazing, rightward gazing, and no face at all) were responded to fastest when the response key was also mapped with the trustworthy key. It is curious that there should be such different responses between Experiments 7a and 7b given that they are identical in every way other than the gender of the photographs displayed.

Where participants claim to be more trusting of smiling faces than neutral faces (LaFrance and Hecht, 1995) and right-sided faces more than left-sided faces (Chen, German and Zaidel, 1997) these were not within a web checkout setting which may be the cause of this non-compliance with previous findings. The context of these photographs cannot be underestimated as the images are not being singled out by the participant, no mention was made of the differences in faces as part of the instructions, indeed the participants are told to ignore the webpage as it is not important to the task of responding to the coloured frame, although this of course does not prevent basic processing of the webpage being carried out. Given that there is evidence to suggest that both men and women seem to have a predisposition towards primarily trusting members of their opposite sex the uneven ratios of males-to-females between Experiments 7a and 7b may have had a bearing on the results, with 7a having a more even ratio of males to females (15:20) compared to 7b (10:28).

PART 4. GENERAL DISCUSSION

Chapter 7.

Overall Goals of this Thesis Research

When beginning the research reported in this thesis, the primary goals were to firstly modify an existing implicit measurement technique for applied use, and secondly to test its effectiveness using a variety of stimulus types and theoretical questions.

In the design of our new technique (the FEAST) we assumed that an implicit measure remains implicit until the participant considers their specific answer in response to a stimulus. Or, put another way, a measure will remain implicit until a participant attends to that aspect of the stimulus that is being measured. Once attention is wholly focused on the stimulus of interest, cognitive processes will come into play and any response will be a mixture of both implicit and explicit processing. If the task indicator and stimulus are too closely linked, then the experimenter runs the risk of having attention overly focused onto that stimulus, thus creating a certain amount of considered thought about it. By making the stimulus irrelevant to the task, the stimulus receives scant thought and so the effects that are measured could be argued to be of a "more" implicit nature.

FEAST as an Implicit Measurement Technique

In Experiment 1, we replicated DeHouwer's EAST, with one important methodological change: we used a higher ratio of white words to coloured words to strengthen the extrinsic key mapping. In Experiment 2, the addition of a frame as the task indicator changed EAST to FEAST (Framed Extrinsic Affective Simon Task). This modification results in diffused attention, where the participant would focus their attention on the frame presented in the periphery of the screen, and the stimulus in the center of the screen would be of secondary interest.

It is worth noting that whilst the intention was that attention was removed from the participants 'conscious' attention, after debriefing several of these participants doubt was cast as to whether this was actually the case. As is common in psychology studies, many of the participants spend much of the experiment trying to react in a way that they believe is expected of them. Rarely in the case of the studies reported in this thesis were the participants correct in their guesses about the role of the study (although they were sometimes close). However, this does not alter the fact that these participants were consciously, or to put it another way, explicitly attending to the stimulus. The results from participants who reported awareness of the study-goals did not yield any results that were different enough to cause alarm, but it is certainly worthy of note.

As stated earlier, when beginning the research reported in this thesis, the first goal was to modify an existing implicit measurement technique for applied use. We can see that this primary goal was achieved in Experiments 1-3.

Experiment 1 was required to set the groundwork for all of the other experiments presented. The results match what has been previously reported in various EAST studies, thus giving us the confidence required to move on to the modified version, using frames. The first experiment also allowed the exploration of an updated version of the relevant statistical analyses (the D600 method) rather than simply following the methods employed by DeHouwer (2003).

Experiment 2, which used the same stimuli as Experiment 1, utilized the framed method. All other details remained the same, even down to using the same participants so that there can be confidence that the only change in results is due to a switch from the non-framed to the framed version. When direct comparisons are made between Experiments 1 and 2 it is easy to see that there are a large number of parallels. In both experiments, both accuracy and RT data show a speeded response/greater accuracy when the stimulus valence and extrinsic response valence are congruent with each other. There are very similar results in terms of how many words shows the expected valence direction (either positive or negative), this is both in terms of accuracy and RT EAST scores. Interactions between stimulus valence and extrinsic response valence are significant at the 0.001 level for both experiments. Finally, the D600 scores for each word showed similar results: with Experiment 1 showing 16 out of 18 words in the expected valence direction, and Experiment 2 showing 17 out of 18 words in the expected valence direction. After looking at these results, there was satisfactory confidence to conclude that the framed version was producing results that were sufficiently similar to the original version to continue with changes to this framed version. The next stage was the addition of a pictorial stimulus instead of words.

Experiment 3 used IAPS pictures within the same frame layout as that used in Experiment 2. The results were encouraging with most of the implicit classifications being as expected, but with others more difficult to explain. In terms of performance accuracy, the implicit classifications were all as expected: participants made fewer errors when the stimulus valence and extrinsic response valence were congruent. For RT, however, this was not the case: although any differences were very small (e.g. 6.9 ms). Another unexpected result was that responses were slower in the second half, with more errors when the stimulus valence was positive. For the FEAST scores for accuracy, all picture valences were in the expected direction, and for the RT data 11 out of 18 were in the expected direction. As was the case with Experiments 1 and

2, there was a significant interaction between stimulus valence and extrinsic response valence at the $p<0.001$ level. Finally, in looking at the D600 scores for each picture, 17 out of the 18 were in the expected valence direction.

In terms of the overall logic (separating the task indicator from the stimulus), there is no evidence to suggest that this diffused attention has resulted in a 'better' method. However, this is not a failure of the pictorial modified frame version, far from it, the results show that it is competent at an equal level to the original version with words, and the framed version with words. In summary, it would be difficult to argue that this diffused attention has caused any enhancement in the results as was initially hoped for; however, they were promising enough to begin to test the sensitivity of this pictorial version with diffused attention by looking at the different emotional dimensions within one stimulus picture.

Experiment 4 examined three different affective dimensions of the same set of stimulus pictures: these were arousal, dominance, and valence. The main aim of this experiment was to push the FEAST a little further by testing its sensitivity, would the results for the three affective dimensions turn out differently and so indicating sensitivity of the FEAST method or would they display similar trends? Despite the large number of analysis carried out on the different data groups, the area that was of greatest importance in terms of testing the method, was the comparison between the three dimensions for each picture. After studying Figure 16, it is clear to see that the FEAST has created an individual path for each dimension. There are pictures that produced similar results such as 'Sky' and 'Mountain 2', but other such as 'Aimed gun' 'Babies' and 'Attack' display stark differences which was very pleasing to see in terms of how *sensitive* the FEAST is.

It would be foolish not to acknowledge that there are some surprising results which do not fit into the expected pattern, making them difficult to explain. However, if the overall aim of this experiment was to assess whether the scores produced would be different from each other, and could be used in a predictive capacity, then the aim has been achieved. Moving onto Experiment 5, this experiment aimed to capitalize on the predictive capabilities of the FEAST for individual pictures on several dimension, but this time in an applied setting.

FEAST: Additional Observations

In these experiments (4) we demonstrated that this new "framed" version of the EAST is effective at measuring implicit attitudes in a variety of contexts and that it was sensitive enough to measure reactions to the same stimuli with different "dimensions" of interest (e.g. valence, arousal, dominance).

There are several important additional observations that should be made about the FEAST technique. The first is related to the theoretical importance of extending the paradigm from the "affective" domain to a variety of other domains (consumer/not related). The original EAST paradigm was envisioned as an implicit "affective" test: a test that would quantify an affective response to a stimulus. Here we have demonstrated that this quantification need not be only on the "affective" domain. But, instead, it could be on domains as diverse as "arousing", "dominating", "expensive", "British", and "trustworthy". This takes the EAST – and perhaps the purported underlying mechanisms out of the domain of affective responses and shows both the EAST and the FEAST to be more related to techniques such as priming. In other words, it is not the immediate approach/withdraw type of response that is foundational to the success of these techniques. Instead, it is likely something akin to response priming that facilitates responding to congruent stimuli.

Another important observation about the FEAST has to do with its sensitivity to context and dimensions of attention. Recall that we have demonstrated FEAST to be sensitive enough to measure various aspects (dimensions) of the same set of stimuli (e.g. as shown in the valence/arousal/dominance study). This suggests that it is not simply the image-properties that drive responding, but instead the mind-set of the participant (set via context and attention) is crucial in determining the response profile.

Related to the context and attention issue is the use of (and importance of) the "white" items. The specific choice of white items is crucial in setting the dimensions of interest. This set of stimuli will set the endpoints of any dimension being measured, and thus the right choices are crucial. Although it was not explicitly tested in these experiments, it should be noted that the set of white words can very easily either compress or expand differences of interest. For example, it is possible that including such ridiculously un-trustworthy websites in the various faces studies (Experiments 6 and 7) might have served to compress any possible differences that could be found in the "coloured" webpages. In other words, the differences between a smiling face and a neutral face might be tiny in comparison to the difference between a "show me the money!" website and a "eTrust" site. This is just one example of the importance (and possible impact associated with) choosing the proper set of "white stimuli". Future work should include more extensive pre-testing of the "white words" and perhaps even the coloured items (e.g. for visibility, explicit ratings, etc etc).

FEAST in Applied Settings

After establishing that the FEAST paradigm worked with well-controlled laboratory stimuli (words and IAPS images), we then moved on to use the paradigm in several more applied settings (in Experiments 4-7). Within these experiments, the main question was: how successful is this technique and how might it be used within both theoretical and applied settings?

Several studies such as Brunel, Greenwald, & Tietje, 2004; Maison, Greenwald, & Bruin, 2001 have shown that another implicit measure (the IAT) could be used to assess consumer's attitudes and beliefs. The results of the experiments presented in this thesis demonstrate that the FEAST could also be a useful method to evaluate beliefs and attitudes of consumers.

In Experiment 5 we used the FEAST to evaluate the beliefs and attitudes of consumers towards different companies using well-recognised brand logos. Furthermore, we included participants from two different cultural groups (students from the UK and from the USA) to see if the FEAST was sensitive enough to measure different attitudes in the two groups. While the FEAST was able to identify differences between the two groups of participants, we again encountered many of the same problems as in Experiment 4 when trying to draw 'hard' conclusions from the data. For example, there were again many cases where the FEAST-scores based on reaction time data told a different story to those based on accuracy data.

The main issue with this particular experiment is whether any confident conclusions can be made about which logos are seen as more costly, greater quality, or associated with a particular country. Based on the data we obtained, it would be complicated and probably incorrect to draw any grand conclusions about how a particular brand is perceived. However, it is possible to be quite confident about several of the more robust findings. For example, if one were interested in whether a particular logo (ASDA for example) is seen as costly in the UK and USA, it is clear from looking at Figures 21 and 22 that ASDA is seen as more costly by UK consumers than those from the USA. The difference is small, but of importance.

It is apparent that the FEAST is producing interesting data, some of which reaches significance at a high level. However, the use of the FEAST is questionable in its current form, unless some clarity involving the differences between the RT and accuracy scores can be drawn. This is one of the reasons that some authors have decided to use the D600 measure instead of these – as the D600 incorporates both RT and accuracy data. However, in our tests, it was not always clear whether the D600 was "the best" metric to use. In some cases, the implicit measures as quantified through the lens of the D600 had better agreement with explicit measures. However, in other cases it was the implicit scores calculated based on RT or accuracy data that showed the best agreement with explicit scores. Thus, based on the numerous studies conducted in this thesis, we still believe that there is no clear "winner" in terms of the best way to quantify the data from the FEAST. Further testing will be required before we can find the best way to quantify FEAST data so that they compliment the explicit ratings data.

Of course, the previous discussion has an implicit assumption: that there *should be* a correlation between implicit and explicit attitudes. As has been discussed several times in the thesis, it is likely that implicit and explicit attitudes will sometimes be highly correlated and, in other situations, they may diverge wildly. Thus, it would be fair at this point to question why explicit tests are not enough to measure consumer attitudes. Differences between explicit and implicit measurement results can be argued as reflecting different aspects of an attitude or indeed different attitudes being measured (Bohner & Wanke, 2002). Strack and Deutsch (in press) propose that both explicit and implicit attitudes jointly influence behaviour. However, the type of behaviour, or the instances in which this combined effort may occur, has not yet been uncovered. It seems sensible to assume that bigger purchase decisions (with more risk involved) will generally be considered and deliberated over more than routine, low-risk purchase decisions. Such detailed and in-depth cognitive deliberation may have an effect of negating differences between implicit and explicit attitudes. For example, if the decision process involves large sums of money and requires careful cost-benefit analysis, then it is unlikely that the final decision will be greatly altered by implicit attitudes. That is, their explicit views may more closely match the implicit views when making a high-risk decision. That said, it should be noted that even in these cases there might be a possibility of measuring implicit attitudes that differ vastly from the explicit attitudes but simply ignored or "pushed out" of the decision processes.

If high-risk decisions are not deeply driven by implicit attitudes, what about all of the low risk decisions that we make every day? What about our choice of chocolate bar at lunchtime or the items in our baskets when we do the weekly shopping? What about the websites we so quickly click away from when we glimpse them in a preview? What role might implicit attitudes have in these contexts? Or, even more broadly, what drives us to buy? Of course there are the obvious pointers, such as eye catching displays and price offers, but if all similar items were a similar price and displayed on the same shelf, what drives you to pick one over the other. This is why we need implicit techniques to test what drives us beyond what is 'controlling' us consciously. In the pursuit of continuing to develop the FEAST method for both theoretical and applied usage, as well as with the aim of tightening the results, we conducted Experiment 6. This experiment employed a further modification: using a combination of both words (to create the extrinsic response dimension) and pictures within frames to look at a research topic that is central to both real and virtual commerce: the concept of "trust." Specifically, in this experiment we looked at how visual field, gaze, and expressions in photographs can affect our perception of trustworthiness.

The effect of a neutral face in comparison to a smiling face was examined in Experiment 6a. In this experiment words were used as the white-framed stimuli (serving to create the underlying dimensions of interest), and photographs of faces were presented within the coloured frames. It was unknown whether this would cause extra confusion amongst the results or serve to strengthen them. Unfortunately, it seems that it was the former, with D600 analyses producing poor results. The same participants also took part in Experiment 6b which examined the role of facial gaze on trustworthiness. Previous research has suggested that facial gaze has an impact on explicit ratings of trustworthiness. Thus, we wanted to see if the same kinds of effects could be found using our implicit (FEAST) measure. The effect of a face looking to the left in comparison to a face looking to the right was examined in Experiment 6b. In this experiment the results are very closely matched for both gaze right and gaze left, and the lack of statistical significance did not allow us to draw any strong conclusions.

In Experiment 6c we examined how the left and right visual fields impact the perception of trustworthiness. Again, based on previous research in the field, there was a prediction that more emotionality might be found in left visual field presentations and we hoped to quantify this using the FEAST. While we did show a little more variability between the visual fields with the D600 method, significance levels were again very poor.

While the ultimate "problem" with Experiments 6a, b, and c remains unknown, there are several possible candidates. First, it could be that the effects we were trying to measure are just too small to be captured in this paradigm. Second, we may not have had enough power. Third, the effects might have gone away after repeated exposures to the stimuli. Fourth, we did not measure eye-movements and thus stimuli were not well controlled for lateralization. Fifth, the stimuli we used might have been sub-optimal (in terms of issues such as visibility, expression, etc). And, finally, it might well be that mixing words and pictures was a poor choice. As mentioned in the discussion to these experiments (6a,b,c), the use of words could have primed or excessively activated the left hemisphere (the "more verbal" hemisphere) and this might have served to negate the right hemisphere bias we had been hoping to quantify. Consequently the final set of experiments presented (7a and b) move towards using photographs within websites, still focusing upon the area of trust, but in an applied sense.

The purpose of Experiments 7a and 7b was to explore how consumer trust in e-commerce websites can be influenced by the addition of photographs of faces to checkout pages. In practice, consumers may browse through several websites in search of a particular product. If a good impression is not immediately created, then there are plenty of other e-commerce sites to which the consumer can go. No matter what each website actually has to offer, it is critical that they create a positive first impression.

The results from Experiments 7a and b produced much more encouraging results with many more statistically significant findings across the different manipulations, these results are important in and of themselves, but they are also important as a proof-of-concept. They demonstrate that it is possible to use the FEAST to perform fairly detailed implicit attitude testing on webpages. This kind of detailed "attitude" testing is rarely carried out in web design work, which instead often favours methods such as explicit questionnaires, focus groups, and more recently eye tracking studies. The data from studies such as 7a and b can assist web designers in making design decisions which are otherwise based only on guesswork (and of course past experience).

The limitation of participant intention

No implicit attitude measurement technique is without its limitations, and the FEAST is no exception. The FEAST method provides an effective measure for use with a wide range of both words and pictorial stimuli. However, it shares many of the limitations that apply to the EAST and has some others of its own. We will briefly review one of the main limitations next, that of participant intention.
What this means is that the FEAST encourages speedy responses, and participants have a 50/50 chance of hitting the correct key. However, they may sometimes respond correctly by mistake. For example, a participant sees the word "death" with a blue frame. Instead of hitting "q" for "blue," she chooses "q" because of its prior/white association with "bad." How can we know her intention? (If "death" had had a green frame, her "q" would have been incorrect.) In this example, the "accidental" correct blue would be grouped with intentional correct answers and analyzed as such.

It was hoped that the separation of the task indicator and stimulus would enhance the results found; however, this has not been shown to be the case. Support for the hypothesis that task and stimulus separation would enhance results was provided by Voss and Klauer (2007) who researched the cognitive processes behind the EAST, in terms of a task switching account. They used a mix of both pictorial (both simple and complex line drawings) and word stimuli. They found support for the theory that the EAST does not simply work due to confusion between the colour and evaluation classification tasks. It was argued that the EAST uses the automatic effects of task switching, and is driven by interference between two conflicting tasks.

Extra analyses were carried out to examine whether effects were stronger or weaker depending on how many white frames were seen before a coloured frame. In other words, when the task switched from evaluative to colour, the number of white frames before a coloured frame was presented varied between one and three. No significant differences were found in RT's or accuracy rates dependant on the number of white frames preceding a coloured frame.

An issue of analysis

Throughout this thesis the accuracy and RT data have produced very different results. The main reason for their inclusion is that this analysis was carried out by DeHouwer (2003), and since the FEAST method is an extension of the work from this paper, it seemed reasonable to follow the same flow of analysis. However, that being said more recently several authors (including DeHouwer himself) have called this original form of analysis into question. The authors Greenwald, Nosek, and Banaji (2003) have recommended the D600 technique as a better way of analyzing EAST/FEAST data. Indeed it is often the case that research reports will report error latencies but that they are not included in any analysis. Greenwald, Nosek, and Banaji (2003) questioned this practice especially when the error and RT scores have very different effects. They actually found that IAT measures showed higher implicit/explicit correlation when the error data was included in the analysis.

A further suggestion which was part of the same research paper reporting this new scoring algorithm, the D600, is that pilot data should be included in the analysis. For the Experiments reported as part of this thesis this procedure was not followed. The main reason for omitting this recommendation is that often it was the case that participants asked for a repeat of the practice trial since this was an offered option as part of the instruction process. Also, there were occasions when the practice trials were undertaken with a degree of assistance, this was especially the case when participants were not first language English (it is worth noting that those not first language British only took part in non-word versions of the FEAST).

So, where do we go from here...the future of implicit testing?

The EAST method has already proved successful in many areas, to name but a few, food preference (De Houwer and Bruycker, in press), attitudes toward "green" consumer behavior (Vantomme et al, 2006), and with clinical populations, such as heavy alcohol users (De Houwer et al, 2004). It is without doubt a useful method of measuring implicit attitudes, avoiding some of the problems seen with the more widely used IAT. There have also been developments in terms of other implicit measures which are cited as alternatives to the EAST method such as the Affective Misattribution Procedure (AMP) reported by Payne, Cheng, Govorun and Steward, (2006) which is a variant of the affective priming paradigm. This method was reported as being flexible in terms of adaptability to both unipoloar and multi-target concepts, but is reported as being highly susceptible to 'faking' so still in need of further development.

The IAT itself is also under continuous development to address its short comings. Recently there have been two new variants of the IAT, the first is a MultiFactor Trait IAT (MFT-IAT) published by Greenwald (2005) allows for multi construct testing such as with the Big Five. The second development is the use of Breif IATS (Sriram and Greenwald, 2007) these as the name suggests abbreviates the number of trials, they concentrate on the procedure of the relevant association and are reported to reach similar reliability and validity with fewer trials.

Despite these developments the FEAST method presented in this thesis approaches the use of implicit testing from a slightly different perspective and end goal. That is to open up the use of testing to bridge the gap between business and psychology, rather than using the implicit test in the pursuit of theoretical advance. Whilst theoretical knowledge is of course a necessary part of its development, and indeed it is only through understanding the theoretical motivations behind why the FEAST modifications work that the FEAST will continue to develop, it has never been the primary driver behind the experimental progress.

The FEAST has been shown throughout this thesis to opens up a plethora of new possibilities. However, it also creates several questions which need to be addressed before confidence in the results can truly be achieved.

1. The first being the issue mentioned above about participant intent. It is always difficult to address this issue in any study which is open to the possibility of participants misleading either intentionally or unintentionally. Quite frequently participants report that they had 'worked out what you were looking for in the study so responded accordingly' rather than simply

responding as you had instructed. A methodological change needs to be applied to address this problem. Differences in response keys could be altered to a 'key press down' 'key press up' rather than a 'Q' and 'P' method, or using the possibility of pressure pads to monitor force of responses.

2. The second is that there is a large issue around the differences seen in both accuracy and RT data, often providing conflicting results. The results throughout this thesis use both the standard De Houwer (2003) analysis as well as a suggestion of analysis by Greenwald et all (2003) (D600 method). In Experiments 1 to 3 the D600 method worked to amplify the results, however, the results for all other experiments presented are not so clear cut, in some cases the D600 analysis often results in the statistics becoming weaker.
3. Larger sample sizes are required due to data and participant loss.
4. Further understanding is required concerning the underlying processes for why the EAST works at all. Is it a task switching issue as advocated by Voss and Klauer (2007), or is there another underlying driver such as response interference (Gawronski et al, 2008)? Possibly going back to basics, using the framed version but with simple stimuli, not to answer a question, but to explore the differences between results when simple changes are made. For example, a stroop version of the test using blocks of colours within the frames.
5. It is widely accepted that mood state and personality affects how emotional stimuli is attended to, an example of which is with anxious participants who are much more likely to show an attentional bias towards threatening stimuli Mineka and Sutton, 1992). No consideration was given to this in any of the

studies presented; the use of a simple anxiety questionnaire may have helped to account for any bias in the data due to anxiety levels.

Since many researchers do not yet know or agree about what factors affect implicit attitudes, or even if they are a separate attitude at all, it seems reasonable that the fewer manipulation of a stimulus, the better. The use of a frame, as in the FEAST method, keeps manipulation of the stimulus to a minimum. There are undoubtedly problems with this technique as it stands, and areas of psychology where it does not seem to prove useful (Experiment 6 a to c). However, the experiments presented in this thesis show how this modified technique can be used to test in a new way, aiding the business world in its difficult choices, looking at the trends produced from the FEAST, as well as in the pursuit of theoretical advance.

References

Arieli, A. Sterkin, A., Grinvald, A., and Aertsen, A (1996) Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science*, 273(5283), 1868-1871

Allport, G. (1935) Attitudes. In C. Murchison (Ed), *Handbook of social psychology* (pp. 798-844). Worcester, MA: Clark University Press.

Anderson, N. (1974) Cognitive algebra. In L. Berkowitz (Ed), *Advances in experimental social psychology*, 7, 1.101. New York: Academic press.

Askegaard, S., & Ger, G. (1998). Product-country image: Toward a contextualized approach. *European advances in consumer research*, 3, 50 – 58

Asthana, H.S., & Mandal, M.K. (2001). Visual-field bias in the judgment of facial expression of emotion. *Journal of General Psychology, 128*, 21-29.

Bagozzi, R. P. (1982). A field investigation of causal relations among cognitions, affect, intentions, and behaviour. *Journal of Marketing Research, 19*, 562-584.

Bargh, J. A., Chen, M, & Burrows, L. (1996). Automaticity of social behaviour: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology*, 71, 230 – 244.

Batra, R., Ramaswamy, Alden, D.L., Steenkamp, J.-B.E.M., & Ramachander, S. (1999). Effects of brand local/ non-local origin on consumer attitudes in developing countries. *Journal of consumer psychology*, 9, 83 – 95.

Berezowski, M. (2003). An introductory discussion of e-commerce. Retrieved on 18th June 2006 from www.cs.valberta.ca/nwfb/300/projects/berezowski/

Bhattacherjee, A. (2002). Individual trust in online firms: Scale development and initial test. *Journal of Management Information Systems, 19,* 1, 213-243.

Bilkey, W. J., & Nes, E. (1982), COO effects on product evaluation, *Journal of international business studies, 13(spring)*. 89 – 99.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6,* 242-261.

Blascovich, J. (2000) Using physiological indexes of psychological processes in social psychological research. In Reis, Harry T.; Judd, Charles M. (2000). *Handbook of research methods in social and personality psychology*, 117-137. New York, NY, US: Cambridge University Press.

Bohner, G., & Wanke, M. (2002), *Attitudes and Attitude Change*, New York: Psychology Press.

Borod, J. C., Kent, J., Koff, E., Martin, C. and Alpert, M. (1988). Facial Asymmetry while posing positive and negative emotions: Support for the right hemisphere hypothesis. *Neuropsychologia, 26(5),* 759-764.

Borod, J. C. (1993) Cerebral mechanisms underlying facial, prosodic and lexical emotional expression: a review of neuropsychological studies and methodological issues. *Neuropsychology,* 7, 445-465.

Borod J.C., Cicero B.A. and Obler L.K. (1999) Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology*, 12, 446–58.

Bradburn, N., Sudman, S., and Wansink, B. (2004) *Asking questions* (2nd Ed) San Francisco: Jossey-Bass.

Bradley, M.M., & Lang, P.J. (1999). Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report C-1, Gainesville, FL. The Center for Research in Psychophysiology, University of Florida

Bradely, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P.J. (2001) Emotion and motivation I: defensive and appetititive reactions in picture processing. *Emotion,* 1(3), 276-298.

Brunel, F. F., Tietje, B. C., and Greenwald, A. G. (2004) Is the Implicit Association Test a Valid and Valuable Measure of Implicit Consumer Social Cognition? *Journal of Consumer Psychology*, (4), 385-404.

Cacioppo, J. T. Petty, R. E., Losch, M. E., and Kim, H. S. (1986) Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, 50(2), 260-268.

Cacioppo, J. T. Bush, L. K., Tassinary, L. G. (1992) Microexpressive facial actions as a function of affective stimuli: Replication and extension. *Personality and Social Psychology Bulletin*, 18(5), 515-526.

Cacioppo, J. T. and Petty, R. E. (1979) Attitudes and cognitive response: An electrophysiological approach. *Journal of Personality and Social Psychology*, 37(12), 2181-2199.

Cacioppo, J. T. and Petty, R. E. (1981) Electromyographic specificity during covert information processing. *Psychophysiology,* 18(5), 518-523.

Cacioppo, J. T. Crites, S. L., Berntson, G. G., and Coles, M. G. (1993) If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science,* 4(2), 108-112.

Cacioppo, J.T., Berntson, G. G., Klein, D. J., and Poehlmann, K.M (1997) The psychophysiology of emotion across the lifespan. *Annual review of gerontology and geriatrics,* 17, 27-74.

Cacioppo, J. T. , and Tassinary, L. G. (1990) Inferring psychological significance from physiological signals. *American Psychologist,* 45(1), 16-28.

Chen, A. C. German C. and Zaidel, D. W. (1997). Brain asymmetry and facial attractiveness: Facial beauty is not simply in the eye of the beholder. *Neuropsychologia, 35,* 471–476.

Cheung, W. L., & Denton, L. (1995), National stereotypes and product evaluations about Japan and the United States: a Hong Kong Perspective, *Journal of International Consumer Marketing, 7(4)*, 59 – 79.

CIA Press Release (2004, September). Worldwide Internet users will top 1 billion in 2005. Retrieved on January 5th 2006 from http://www.c-i-a.com/pr0904.htm

CIA Press Release (2006, January). Worldwide Internet users top 1 billion in 2005. Retrieved on August 8th 2006 from http://www.c-i-a.com/pr0106.htm

Clark, V. P., and Hillyard, S. A. (1996) Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of cognitive neuroscience*, 8, 387-402.

Constantine, R. McNally, R. J. and Hornig, C. D. (2001) Snake fear and the pictorial emotional Stroop paradigm. *Cognitive Therapy and Research.* Vol 25(6), 757-764

Cowell, A. J. and Stanney, K. M. (2003). On manipulating nonverbal interaction style to increase anthropomorphic computer character credibility. Retrieved on 29 December, 2005 from http://www.vhml.org/workshops/aamas2003/papers/cowell/cowell.pdf.gz

Crayenest, M., Crombez, G., De Houwer, J., Deforche, B., Tanghe, A., & De Bourdeaudhuij, I. (2005). Explicit and implicit attitudes toward food and physical activity in childhood obesity. *Behaviour Research and Therapy, 43*, 1111-1120.

Crosby, F., Bromley, S., and Saxe, L. (1980) Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological bulletin*, 87, 546-563.

Cunningham, W. A., Raye, C. L., Gatenby, J.C., Gore, J.C., and Banaji, M. R. (2004) Separable neural components in the processing of black and white faces. *Psychological science*, 5, 806-813.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for white Americans : Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36*, 316 – 328.

Davidson, R.J., Ekman, P., Saron, C.D., Senulis, J.A. and Friesen, W.V. (1990). Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology I. *Journal of Personality and Social Psychology, 58,* 330–341.

Davidson, R. J. (1993a). Cerebral asymmetry and emotion: Conceptual and methodological conundrums. *Cognition and Emotion*, 7, 115–138.

Degner, J., Wentura, D., Gniewosz, B. and Noack, P (2007) Hostility-related prejudice against Turks in adolescents: Masked affective priming allows for a differentiation of automatic prejudice. *Basic and Applied Social Psychology.* Vol 29(3), pp. 245-256

De Houwer, J. (2001). A structural and process analysis of the implicit association test. *Journal of Experimental Social Psychology*, 37, 443 - 451

De Houwer, J. (2003). The Extrinsic Affective Simon Task. *Experimental Psychology, 50 (2)*, 77 – 85.

De Houwer, J., & De Bruycker, E. (2007). Implicit attitudes toward meat and vegetables in vegetarians and non-vegetarians. *International Journal of Psychology, 42*, 158-165.

De Houwer, J., & De Bruycker, E. (2007). The Implicit Association Test outperforms the Extrinsic Affective Simon Task as a measure of interindividual differences in attitudes. *British Journal of Social Psychology, 46*, 401-421.

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion, 12*, 45-61.

De Houwer, J., Hermans, D (1994) Differences in the affective processing of words and pictures. *Cognition & Emotion.* 8(1), 1-20.

De Houwer, J., Crombez, G., Koster, E. H. W., & De Beul, N. (2004). Implicit alcohol-related cognitions in a clinical sample of heavy drinkers. *Journal of Behavior Therapy and Experimental Psychiatry, 35*, 275-286.

De Houwer, J. and Moors, A. (2006) Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin,* 132(2), 297-326.

De Houwer, J. and Moors, A. (2007) How to Define and Examine the Implicitness of Implicit Measures. In Wittenbrink, Bernd; Schwarz, Norbert (2007). Implicit measures of attitudes. (pp. 179-194). New York, NY, US: Guilford Press. x, 294 pp.

Deutsch, F. M. (1990). Status, sex, and smiling: The effect of role on smiling in men and women. *Personality and Social Psychology Bulletin, 16,* 531–540.

Dovidio, J, F. and Fazio, R.H. (1992) New technologies for the direct and indirect assessment of attitudes. In J.M. Tanur (Ed), *Questions about questions* (pp. 204- 236). New York: Russell Sage.

Dovidio, J, F., Kawakami, K., Johnson, C., Johnson, B., and Howard, A. (1997) On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, 33(5), 510-540.

Dovidio, J, F., Kawakami, K., and Gaertner, S.L. (2002) Implicit and explicit prejudice and interracial interaction. *Journal of personality and social psychology*, 82(1), 62-68.

Duda, P.D. and Brown, J. (1984) Lateral asymmetry of positive and negative emotions. *Cortex*, 20, 253–261.

Eckel, C. C. and Wilson, R. K. (2003). The human face of game theory. In: E. Ostrom and J.M. Walker, Eds, *Trust and reciprocity: Interdisciplinary lessons from experimental research.* New York: Russell Sage Foundation.

Ehrlich, H. J. (1973). *The social psychology of prejudice*. New York: Wiley.

Erickson, G, M., Johansonn, J. K., & Chao, P. (1984), Image variables in multi-attribute product evaluations: COO Effect, *Journal of consumer research (11)*, 694 – 699.

Fabiani, M., Gratton, G., and Coles, M. G. H. (2000) Event related brain potentials. In J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Eds). *Handbook of psychophysiology* (2nd ed., pp. 53-84). Cambridge, UK: Cambridge University Press.

Farah, M. J. (1996). Is face recognition 'special'? Evidence from neuropsychology. *Behavioural Brain Research, 76,* 181-189

Fere. C. (1888) Note sur les modifications de la resistance electrique sous l'influence des excitations. *CR Soc. Biol*, Paris.

Fishbein, M., & Ajzenn, I.(1975), *Belief, Attitude, Intention, and Behavior: An introduction to theory and research*, Addison-Wesley, Reading, MA.

Fazio, R.H. (1990) Multiple processes by which attitudes guide behaviour: The MODE model as an interactive framework. In M.P. Zanna (Ed), *Advances in* *experimental social psychology* (Vol.23, pp. 75-109). San Diego, CA: Academic press.

Fazio, R.H. (2001). On the automatic activation of associated evaluations: An overview, *Cognition and Emotion*, 15(2), 115-141.

Fazio, R. H, & Dunton, B. C. (1997). Categorization by race : The impact of automatic and controlled components of racial prejudice. *Journal of Experimental social Psychology, 33*, 451 – 470.

Fazio, R. H. and Olson, M. A. (2003) Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*. Vol 54, 2003, pp. 297-327.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229-238.

Frings, C. and Wentura, D. (2003) Who is watching Big Brother? TV consumption predicted by masked affective priming. *European Journal of Social Psychology*. Vol 33(6), Nov-Dec 2003, pp. 779-791

Fogg, B. J. (2002). *Persuasive Technology: Using computers to change what we think and do.* San Francisco: Morgan Kaufman.

Gawronski, B., Deutsch, R., LeBel, E. P., and Peters, K. R. (2008) Response interference as a mechanism underlying implicit measures. *European journal of psychological assessment*, 24, 218-225.

Gordon, R. (2000). Does the New Economy Measure Up to the Great Inventions of the Past. *Journal of Economic Perspectives, 14,* 4, 49-74.

Greenspan, R. (2004). Q1 '04 U.S. E-Com Sales = $15.5B. Retrieved on December 15, 2005 from www.clickz.com/stats/sectors/retailing/article.php/3361411.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition : Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4 – 27.

Greenwald, A. G., & Farnham, S. D. (2002). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79, 1022 - 1038

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464 – 1480.

Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003) Understanding and using the Implicit Association Test: I. An improved scoring algorithm": Correction to Greenwald et al. *Journal of Personality and Social Psychology*, 85(3), 481.

Hjarvard, S. (2002). Mediated encounters. An essay on the role of communication media in the creation of trust in the "global metropolis". In: G. Stald and T. Tufte, Eds, *Global encounters: media and cultural transformation*. Luton: University of Luton Press, 69–84.

Han, C. M. (1989). Country Image : Halo or Summary Construct?. *Journal of Marketing Research*, 26, 222 – 229.

Hansen, C. H. Hansen, R. D. (1988) Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, 54(6), 917-924.

Hinsz, V.B. and Tomhave, J.A. (1991). Smile and (half) the world smiles with you, frown and you frown alone. *Personality and Social Psychology Bulletin, 17,* 586–592.

HI Europe – Global Market Research Institute. (2003). *Sony Tops the List in Inaugural Best Brands Survey In Great Britain* (Issue volume No. 1). Brentford, UK.

Hong, S. T., & Toner, J. F. (1989). Are there gender differences in the use of COO in the evaluation of product?. In Thomas, K. S. (Ed.). *Advances in consumer research, 16*, Provo, Utah : Association for consumer research.

Hong, S. T., & Wyer, Jr., R. S. (1989). Effects of COO and product-attribute information on product evaluation: An information processing perspective. *Journal of consumer research, 16*, 175 – 187.

Hugenberg, K. and Bodenhausen, G. V. (2003) Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science,* 14(6), 640-643.

Huijding, J., & de Jong, P. J. (2005a). A pictorial version of the Extrinsic Affective Simon Task: Sensitivity to generally affective and phobia-relevant stimuli in high and low spider fearful individuals. *Experimental Psychology*, 52, 289–295.

Huijding, J., & de Jong, P. J. (2005b). A modified Extrinsic Affective Simon Task (EAST) to assess the affective value of pictorial stimuli: No influence of age or educational level. *Psychologica Belgica*, 45, 241–255.

Ivry, R. & Robertson, L. (1997). The two sides of perception. Cambridge, MA: MIT Press.

Janda, S., Rao, C. P. (1997). The effect of COO related stereotypes and personal beliefs on product evaluation. *Psychology & Marketing, 14(7)*, 689 – 702.

Janiszewski, C. (1988) Preconscious processing effects: The independence of attitude formation and conscious thought. *Journal of consumer research*, 15, 199-209.

Kaufmann, G., Drevland, G. C. B., Wessel, E., Overskeid, G. and Magnussen, S. (2003). The importance of being earnest: displayed emotions and witness credibility. *Applied Cognitive Psychology, 17(1),* 21-34.

Keating, C. F. (1985). Gender and the physiognomy of dominance and attractiveness. *Social Psychology Quarterly, 48,* 61–70.

Kelley, C. M., Rhinelander, T. and DeMoulin. G. (2001). Driving sales with improved carts: Consumer technographics North America. Retrieved on 13 December, 2005 from www.forrester.com/ER/Research/Report/0,1338,11348,FF.html.

Kindt, M. and Brosschot , J. F (1997) Phobia-related cognitive bias for pictorial and linguistic stimuli, *Journal of Abnormal Psychology*

Kornblum, S. (1992) Dimensional overlap and dimensional relevance in stimulus-response and stimulus-stimulus compatibility. Stelmach, George E.; Requin, Jean (1992). Tutorials in motor behavior, 2. *Advances in psychology*, Vol. 87. (pp. 743-777). Oxford, England: North-Holland. xviii, 960 pp

Korthalls, M. (2001). Taking consumers seriously: Two concepts of consumer sovereignty. *Journal of Agricultural and Environmental Ethics, 14,* 201–215.

LaFrance, M. and Hecht, M. (1995). Why smiles generate leniency. *Personality and SocialPsychology Bulletin, 21,* 207-214.

Lavy, E., Van Den Hout, M., and Arntz, A. (1993) Attentional bias and spider phobia :conceptual and clinical issues, *Behaviour research and therapy*, 31, 17-24.

Lau, S. (1982). The effect of smiling on person perception. *Journal of Social Psychology, 117,* 63–67.

Lawrence, C., Marr, N. E., & Prendergast, G. P. (1992), COO stereotyping: A case study in the New Zealand motor vehicle industry, *European Journal of Marketing, 26(3).*

Li, W. K., & Wyer, Jr., R. S. (1994). The role of COO in product evaluations : Informational and standard-of-comparison effects. *Journal of consumer psychology, 3*, 187 – 212.

Liefeld, J. P. (1993), *Experiments on COO effects : Review and Meta-analysis of effect size (p. 117 – 156)*, In paradopoulos, N & Heslop, L. A. (Eds.), Product-Country images : impact and role in international marketing, New York : International Business Press.

Lindgaard, G., Fernandes, G., Dudek, C. and Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour and Information Technology, 25(2),* 115-126.

Logan, G. D.(1989) Automaticity and cognitive control.In Uleman, James S.; Bargh, John A. (1989). Unintended thought. (pp. 52-74). New York, NY, US: Guilford Press. xxvi, 481 pp.

Lombard, M. and Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-mediated Communication 3(2).*

Maison, D, & Bruin, R. (1999) The Implicit Association test (IAT) – A method for the investigation of implicit attitudes. *Studia Psychologiczne* 37, 61 – 80.

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin, 2*, 61 – 79.

Mandal, M. K., Tandon, S. C. and Asthana, H. S. (1991) Right brain damage impairs recognition of negative emotions. *Cortex*, 27(2), 247-53.

McGregor, I., and Marigold, D. C. (2003) Defensive Zeal and the Uncertain Self: What Makes You So Sure? *Journal of Personality and Social Psychology*, 5(5), 838-852.

McKnight, D. H., Choudhury, V., and Kacmar, C. (2002). The impact of initial consumer trust on Intentions to transact with a web site: a trust building model. *Journal of Strategic Information Systems, 11,* 297-323.

Mehrabian, A. (1967). Orientation behaviours and nonverbal attitude communication. *Journal of Communication, 4,* 324-332.

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift fur Experimentelle Psychologie*, 48, 107 – 122.

Mierke, J., Nooh, S. M., & Powers, T. L. (1999), The impact of COO on product choice: A developing country perspective, *Journal of practical grobal business, 1.*

Morris, J. S., Ohman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature, 393,* 467-470.

Nagashima, A., (1970) A comparison of Japanese and US attitudes towards foreign products. *Journal of marketing, 34(1)*, pp 68 – 74.

Nicholls, M. E. R., Clode, D., Wood, S. J. and Wood, A. G. (1999). *Laterality of expression in portraiture: Putting your best cheek forward.* Proceedings of the Royal Society of London, 226, 1517–1522.

Nielsen, J. (2000). *Designing Web Usability.* Indianapolis: New Riders Publishing.

Nielsen Media Research. (1997). Nielsen Media Research/CommerceNet Internet

Demographics Study. Retrieved on 5th January 2006 from

http://www.nielsenmedia.com/com-mercenet/

Nooh, S. M., & Powers, T. L. (1999). The Impact of COO on Product Choice: A Developing Country Perspective. *Journal of Practical Global Business*, 1, 18 – 40.

Nosek, Brian A.1; Banaji, Mahzarin R.1 (2001) The Go/No-go Association Task. *Social Cognition*. Vol 19(6), Dec 2001, pp. 625-666

NPD Group (2001, July 17). NPD E-visory® report shows offline sales benefit from online browsing. [Press Release]. Retrieved on January 2nd 2006 from http://www.npd.com/dynamic/releases/press_010717.htm.

Obermiller, C., & Spangenberg, E. (1989). Exploring the effects of COO labels: An information processing framework. *Advances in consumer research, 16*, 454 – 459.

Olivers, N. L. and Nieuwenhuis, S. (2005) The Beneficial Effect of Concurrent Task-Irrelevant Mental Activity on Temporal Attention. *Psychological Science*. Vol 16(4), 265-269

Otta, E., Pereira, B., Delavati, N., Pimentel, O. and Pires, C. (1993). The effects of smiling and of head tilting on person perception. *The Journal of Psychology, 128,* 323-331.

Patzer, G. L. (1983). Source Credibility as a Function of Communicator Physical Attractiveness. *Journal of Business Research, 11,* 2, 229-241.

Paulhus, D. L. (1984) Two component models of socially desirable responding. *Journal of personality and social psychology*, 46 (3), 598-609.

Paulhus, D. L. and Levitt, K. (1987) Desirable responding triggered by affect: Automatic egotism? *Journal of Personality and Social Psychology*. Vol 52(2), 245-259.

Perugini, M. (in press). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology.*

Peris, S. M., Newman, K., Bigne, E., & Chansarkar, B. (1993), Aspects of Anglo-Spanish Perceptions arising from COO image, *International Journal of Advertising, 12(12)*, 131 – 142.

Perugini, M. (in press). Predictive models of implicit and explicit attitudes, *British Journal of Social Psychology.*

Petty, R. E. and Cacioppo, J. T. (1986). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag.

Poehlman, T. A., Uhlmann, E. L., Banaji, M. R. Greenwald, A. G. (2009) Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17-41.

Polanyi, M, (1967) The Tacit Dimension, New York, Doubleday.

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

Pratto, F., John, O. P. (1991) Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, 61(3), 380-391.

Reierson, C. (1966), "Are Foreign products seen as national stereotypes?". *Journal of Retailing, 42*. 33 – 40.

Riegelsberger, J. (2002). *The effect of facial cues on trust in e-commerce systems.* Proceedings Volume 2 of the 16th British HCI Conference London.

Riegelsberger, J. and Sasse, M. A. (2002). *Face it: Photographs Don't Make Websites Trustworthy.* Proceedings of the Conference on Human Factors in Computing Systems, 742-743.

Riegelsberger, J., Sasse, A.M., and McCarthy, J. D. (2003). *Shiny happy people building trust? Photos on e-commerce websites and consumer trust.* Proceedings of the SIGCHI conference on Human factors in computing systems, 121-128.

Rousseau, D. M., Sitkin, B. B., Burt, R. S. and Camerer, C. (1998). Not so different after all: A Cross-disciplinary view of trust. *The Academy of Management Review, 23(3),* 393–404.

Roskos-Ewoldsen, D. R., Fazio, R. H. (1992). On the orienting value of attitudes: Attitude accessibility as a determinant of an object's attraction of visual attention. *Journal of Personality & Social Psychology*. Vol 63(2), 198-211.

Ross, E. D., Homan, R. W., & Buck, R. (1994). Differential hemispheric lateralization of primary social emotions: Implications for developing a comprehensive neurology for emotions, repression, and the subconscious. *Neuropsychiatry, Neuropsychology, and Behavioural Neurology*, 7, 1–19.

Roth, M. S., & Romeo, J. B. (1992). Matching product category and country image perceptions : A framework for managing COO effects. *Journal of International Business Studies, 23(3)*, 477 – 497.

Roth, A. (1995). Bargaining Experiments. In J. H. Kagel and A. E. Roth (Eds.), *The handbook of experimental economics,* 295-304. Princeton, NJ: Princeton University Press.

Rothermund, K., Wentura, D. (2004) Underlying Processes in the Implicit Association Test: Dissociating Salience From Associations. *Journal of Experimental Psychology: General*, 133(2), 139-165.

Sato, W., & Aoki, S. (2006). Right hemispheric dominance in processing of unconscious negative emotion. *Brain and Cognition, 62,* 261-266.

Schiff, B.B., & Lamon, M. (1994). Inducing emotion by unilateral contraction of hand muscles. *Cortex, 30*, 247-254

Schirillo, J.A. (2000). Hemispheric asymmetries and gender influence Rembrandt's portrait orientations. *Neuropsychologia, 38(12),* 1593-1606.

Schmukle, S. C. Egloff, B. (2006) Assessing anxiety with extrinsic Simon tasks. *Experimental Psychology*, 53(2), 149-160.

Schooler, R. D. (1965), Product bias in the central American common market, *Journal of Marketing Research, 2*, 394 – 397.

Sell, J. (1997). Gender, strategies, and contributions to public goods. *Social Psychology Quarterly, 60(3),* 252–265.

Sharlemann, J. P.W., Eckel, C.C., Kacelnik, A. and Wilson, R.K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22(5),* 617-640.

Sherif, H. C. (1947) *The psychology of ego-involvements, social attitudes & identifications*, M J. Wiley & Sons, Inc.

Shields, S. A., Harriman, R. E. (1984) Fear of male homosexuality: Cardiac responses of low and high homonegative males. *Journal of Homosexuality. Special Issue: Homophobia: An overview*, 10(1-2), 53-67.

Sherif, Muzafer; Cantril, Hadley (1947). *The psychology of ego-involvements: Social attitudes and identifications*. (pp. 60-91). Hoboken, NJ, US: John Wiley & Sons Inc. viii, 525 pp.

Shiffrin, R.M., and Schieder, W. (1977) Controlled and automatic human information processing: Perceptual learning, automatic attending, and general theory. *Psychological review*, 84, 127-190.

Shimp, T. A., & Sharma, S. (1987). Consumer ethnocentrism: Construction and validation of the CETSCALE. *Journal of Marketing Research, 24*, 280 – 289.

Signal, H., and Page, R. (1971) Current stereotypes: A little fading, a little faking. *Journal of personality and social psychology*, 18, 247-255.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 81(1), 174-176

Skoda's image gets clearer. (2005, March 29). *The independent*, from http://motoring.indepent.co.uk

Smilek, D, Enns, James T, Eastwood, J, and Merikle, P. M.(2006) Relax! Cognitive strategy influences visual search. Relax! Cognitive strategy influences visual search. *Visual Cognition.* Special Issue: Visual search and attention. Vol 14(4-8), 543-564

Smith, S.D., Dixon, M.J., Tays, W.J, and Bulman-Fleming, M.B. (2004). Anomaly detection in the right hemisphere: The influence of visuospatial factors. *Brain and Cognition, 55,* 458-462

Solomon, H., Zener-Solomon, L., Arnone, M., Maur, B., Reda, R. and Roth, E. (1981). Anonymity and helping. *The Journal of Social Psychology, 113,* 37-43.

Steenkamp, J. (1989). *Product quality.* Van Gorcum, Assen (NL).

Steinbrück, U., Schaumburg, H., Duda, S. and Krüger, T. A. (2002). A picture says more than a thousand words- Photographs as trust builders in e-commerce web sites. Proceedings of CHI, USA, 20, 748-749.

Strack, F., and Deutsch, R. (in press). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review.*

Strutton, D., & Pelton, L. (1993). Southeast Asian consumer perceptions of American and Japanese imports: The influence of COO effects, *Journal of international consumer marketing, 6(1)*. 67 – 86.

Tarrés, F. and Rama, A. "GTAV Face Database" available at http://gpstsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm

Thomas,N. A., Burkitt, J. A. And Saucier, D.M. (2006) When advertising turns cheeky, *Laterality*, 11(3), 277- 286.

Thompson, M., Zanna, M. P., and Griffin, D. W. (1995) Let's not be indifferent about (attitudinal) ambivalence. In Petty, Richard E.; Krosnick, Jon A. *Attitude strength: Antecedents and consequences*. Ohio State University series on attitudes and persuasion, 4, 361-386. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Tractinsky, N., Cokhavi, A. and Kirschenbaum, M. (2004). *Using Ratings and Response Latencies to Evaluate the Consistency of Immediate Aesthetic Perceptions of Web Pages.* Proceedings of Third Annual Workshop on HCI, Washington D.C., USA, 40-44.

Tseng, Y., Yue, W. L., Taylor, M. A. P. (2005). The role of transportation in logistics chain. Proceedings of the FEASTern Asia Society for transportation studies, 5, 1657-1672.

Uleman, James S.; Bargh, John A. (1989). Unintended thought. (pp. 3-51). New York, NY, US: Guilford Press. xxvi, 481 pp.

Walters, A., Stuchlmacher, A. and Meyer, L. (1998). Gender and negotiator competitiveness: A meta-analysis. *Organizational Behavior and Human Decision Processes, 76,* 1–29.

Wilson, T., D., Lindsey, S., and Schooler, T. Y. (2000) A model of dual attitudes. *Psychological review*, 107(1), 101-126.

Wilson, T. D., Hodges, S. D., LaFleur, S. J. (1995) Effects of introspecting about reasons: Inferring attitudes from accessible thoughts. *Journal of Personality and Social Psychology*, 69(1), 16-28.

Winkielman, P. and Cacioppo, J. T. (2001) Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology*, 81(6), 989-1000.

Wolff, W. (1933). The experimental study of forms of expression. *Character and Person, 2,* 168–173.

Vantomme D. Geuens M. De Houwer J. De Pelsmacker P. 2006. Implicit attitudes toward green consumer behavior. *Psychologica Belgica*. 45 (4) : 217 -239.

van Strien, J. W., & van Beek, S. (2000). Ratings of emotion in laterally presented faces: Sex and handedness effects. *Brain and Cognition*, 44, 645–652.

Verlegh, P. W. J, & Steenkamp, J. E. M., (1999), A review and meta-analysis of COO research, *Journal of Economic Psychology, 20*, 521 – 546.

Voss, A and Klauer, K, C. (2007) Cognitive processes inj the extrinsic affective simony ask: A task set switching account of EAST effects, *Experimental psychology*, 54, 71-82.

Zaidel, D. W., Bava, S., and Reis, V. A. (2003). Relationship between facial asymmetry and judging trustworthiness in faces. *Laterality, 8,* 225-232

Zajonc, R. B. (1980). Feeling and thinking. Preferences need no inferences. *AmericanPsychologist, 35,* 151-175.

Zheng, J., Veinott, E., Bos, N., Olson, J. S., and Olson, G. M. (2002). *Trust without Touch: Jumpstarting long-distance trust with initial social activities.* Proceedings of the Conference on Human Factors in Computing Systems, NY, USA, 293-294.

Zikmund, W. G. (2003), *Essentials of Marketing Research (2nd ed)*. New York: THOMSON, Inc.

Appendix A – Stimulus pictures

Experiment 3

Coloured framed pictures


Treat


Sports cart


Sky


Playing children


Monkies


Flowers


Attack


Fireworks


Bride


Boat


Wires


Snake


Prison


Missiles


Knives


Gun


Flood


Car Crash

White framed pictures


Sunset


Seal


Puppy


Polar bear


Happy Child


Dolphins


Cub


Country


Bunny


Beach


Baby

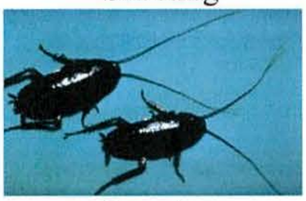

Tornado


Skulls


Shooting


Scream


Roach


Insects


Flies on pie


Fight


Cocaine


Angry face


Toxic

Experiment 4

Coloured framed pictures

Sky

Skier

Rabbit

Aimed Gun

Nature

Mountain1

Mountain2

Lamp

Kids

Earth

Beach

Basket

Babies

Attack

White framed pictures

Arousal, Valence and Dominance


Sinking ship


Plane crash


Police


Attack


Knife


Aimed gun


Shark


Snake


Parrots


Windows


Spoon


Nature


Flowers


Cow


Gannet


Men

Experiment 5

Coloured framed pictures

Experiment 6a – expression

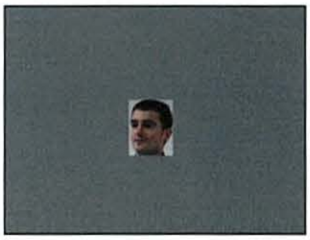
Coloured framed pictures

Experiment 6b – gaze

Coloured framed pictures











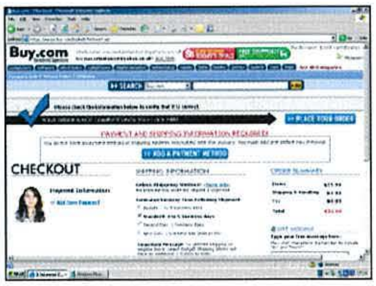


Experiment 6c – visual field

Coloured framed pictures

Experiment 7a – gaze

Coloured framed pictures

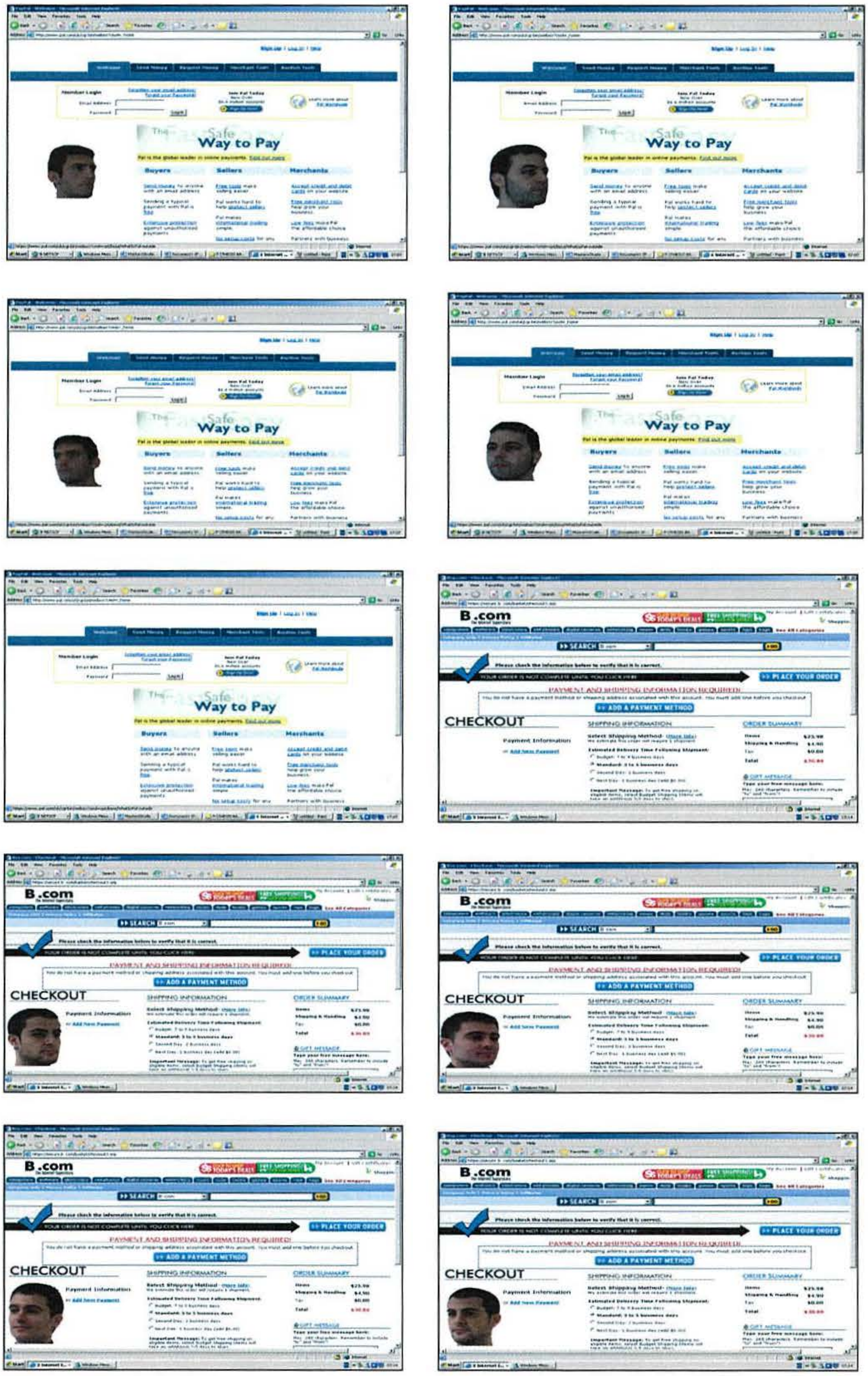


Experiment 7a - expression

Coloured Framed pictures

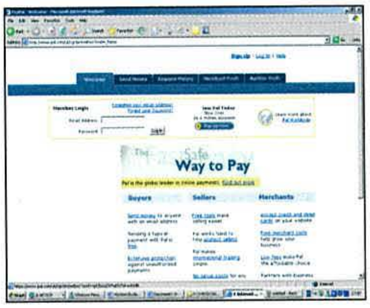
Experiment 7b - gaze

Coloured framed pictures



Experiment 7b – expression

Coloured framed pictures



Appendix B – Data tables for Experiment 4 showing untransformed data, and as a function of experiment half

Arousal

Picture	Measure	Extrinsic Response Dimension	
		High	Low
Aimed Gun	Reaction Time (ms)	551 (115)	566 (147)
	Percentage of Errors	8.62 (16.7)	26.7 (28.3)
Attack	Reaction Time (ms)	559 (134)	564 (141)
	Percentage of Errors	7.76 (13.5)	23.3 (26.7)
Babies	Reaction Time (ms)	573 (133)	606 (158)
	Percentage of Errors	22.0 (23.5)	10.8 (15.6)
Basket	Reaction Time (ms)	555 (141)	549 (99.6)
	Percentage of Errors	10.8 (15.6)	19.0 (18.5)
Beach	Reaction Time (ms)	577 (131)	560 (128)
	Percentage of Errors	9.77 (15.9)	14.2 (18.8)
Earth	Reaction Time (ms)	541 (110)	570 (138)
	Percentage of Errors	15.5 (22.6)	17.2 (19.0)
Kids	Reaction Time (ms)	546 (116)	558 (109)
	Percentage of Errors	13.4 (18.3)	16.8 (17.1)
Lamp	Reaction Time (ms)	551 (121)	564 (117)
	Percentage of Errors	*12.5 (19.5)*	*7.33 (11.3)*
Mountain1	Reaction Time (ms)	567 (121)	554 (140)
	Percentage of Errors	14.7 (20.6)	13.8 (19.6)
Mountain2	Reaction Time (ms)	536 (109)	559 (121)
	Percentage of Errors	11.2 (15.8)	21.6 (21.9)
Nature	Reaction Time (ms)	584 (164)	553 (116)
	Percentage of Errors	11.2 (17.1)	15.5 (19.4)
Rabbit	Reaction Time (ms)	605 (141)	561 (121)
	Percentage of Errors	17.7 (22.0)	18.1 (23.0)
Skier	Reaction Time (ms)	556 (133)	575 (146)
	Percentage of Errors	12.9 (15.8)	13.2 (17.5)
Sky	Reaction Time (ms)	597 (232)	545 (121)
	Percentage of Errors	15.5 (20.5)	17.2 (19.0)

Table 1

Stimulus Picture	Extrinsic Response Dimension			
	High		Low	
	Experiment Half		Experiment Half	
	1	2	1	2
Aimed Gun				
Reaction Time (ms)	516 (99.9)	571 (156)	570 (175)	567 (175)
Percentage of Errors	10.3 (24.6)	6.90 (17.5)	28.4 (36.4)	25.0 (31.3)
Attack				
Reaction Time (ms)	565 (145)	552 (158)	552 (106)	570 (170)
Percentage of Errors	8.62 (23.4)	6.90 (17.5)	19.8 (30.9)	26.7 (34.0)
Babies				
Reaction Time (ms)	610 (155)	543 (145)	600 (254)	610 (162)
Percentage of Errors	30.2 (33.7)	13.8 (26.4)	9.48 (19.4)	12.1 (21.8)
Basket				
Reaction Time (ms)	577 (195)	535 (153)	573 (144)	531 (107)
Percentage of Errors	12.1 (21.8)	9.48 (19.4)	19.0 (24.7)	19.0 (24.7)
Beach				
Reaction Time (ms)	606 (169)	559 (151)	564 (144)	560 (142)
Percentage of Errors	13.8 (22.7)	5.17 (15.5)	18.1 (27.5)	10.3 (24.6)
Earth				
Reaction Time (ms)	536 (138)	543 (123)	563 (155)	574 (145)
Percentage of Errors	12.9 (25.5)	18.1 (30.5)	19.0 (28.1)	15.5 (27.1)
Kids				
Reaction Time (ms)	558 (132)	532 (127)	559 (122)	554 (138)
Percentage of Errors	12.9 (21.8)	13.8 (26.4)	14.7 (26.3)	19.0 (28.1)
Lamp				
Reaction Time (ms)	594 (152)	507 (93.8)	589 (173)	546 (129)
Percentage of Errors	9.48 (19.4)	15.5 (27.1)	6.90 (17.5)	7.76 (17.8)
Mountain1				
Reaction Time (ms)	546 (126)	586 (160)	543 (170)	568 (156)
Percentage of Errors	12.9 (25.5)	16.4 (27.0)	8.62 (19.2)	19.0 (31.1)
Mountain2				
Reaction Time (ms)	560 (127)	511 (124)	587 (254)	537 (107)
Percentage of Errors	10.3 (20.6)	12.1 (21.8)	21.6 (33.9)	21.6 (31.1)
Nature				
Reaction Time (ms)	589 (203)	558 (125)	541 (110)	565 (189)
Percentage of Errors	12.90 (21.8)	9.48 (23.5)	10.3 (20.6)	20.7 (31.4)
Rabbit				
Reaction Time (ms)	605 (183)	587 (153)	555 (134)	557 (130)
Percentage of Errors	18.1 (24.0)	17.2 (33.5)	19.0 (28.1)	17.2 (30.7)
Skier				
Reaction Time (ms)	554 (117)	550 (170)	596 (197)	560 (161)
Percentage of Errors	13.8 (26.4)	12.1 (25.5)	10.3 (20.6)	15.5 (23.5)
Sky				
Reaction Time (ms)	609 (321)	545 (152)	544 (201)	545 (165)
Percentage of Errors	8.62 (19.2)	22.4 (31.6)	18.1 (27.5)	16.4 (28.6)

Table 2

Dominance

Picture	Measure	Extrinsic Response Dimension	
		High	Low
Aimed Gun	Reaction Time (ms)	604 (302)	642 (315)
	Percentage of Errors	2.92 (7.83)	30.8 (25.2)
Attack	Reaction Time (ms)	597 (181)	596 (176)
	Percentage of Errors	10.8 (18.2)	28.3 (30.6)
Babies	Reaction Time (ms)	598 (159)	586 (98.7)
	Percentage of Errors	14.0 (20.0)	10.4 (15.4)
Basket	Reaction Time (ms)	579 (151)	582 (122)
	Percentage of Errors	14.2 (19.3)	10.0 (15.5)
Beach	Reaction Time (ms)	576 (152)	589 (161)
	Percentage of Errors	12.9 (24.7)	18.3 (22.7)
Earth	Reaction Time (ms)	599 (149)	566 (147)
	Percentage of Errors	9.17 (12.3)	12.9 (18.1)
Kids	Reaction Time (ms)	609 (184)	594 (171)
	Percentage of Errors	21.7 (21.5)	12.5 (19.4)
Lamp	Reaction Time (ms)	592 (157)	576 (121)
	Percentage of Errors	12.9 (17.5)	16.7 (22.1)
Mountain1	Reaction Time (ms)	619 (357)	588 (160)
	Percentage of Errors	11.2 (16.2)	9.44 (17.1)
Mountain2	Reaction Time (ms)	549 (113)	576 (128)
	Percentage of Errors	7.92 (13.3)	20.4 (19.8)
Nature	Reaction Time (ms)	579 (151)	610 (224)
	Percentage of Errors	10.4 (12.3)	16.1 (16.9)
Rabbit	Reaction Time (ms)	569 (131)	613 (119)
	Percentage of Errors	9.58 (15.3)	10.0 (12.5)
Skier	Reaction Time (ms)	605 (160)	579 (122)
	Percentage of Errors	13.3 (19.4)	10.0 (15.5)
Sky	Reaction Time (ms)	561 (115)	557 (211)
	Percentage of Errors	10.8 (15.7)	14.6 (21.3)

Table 3

Stimulus Picture	Extrinsic Response Dimension: High, Experiment Half 1	High, Experiment Half 2	Low, Experiment Half 1	Low, Experiment Half 2
Aimed Gun				
Reaction Time (ms)	619 (371)	588 (250)	611 (324)	616 (255)
Percentage of Errors	1.67 (9.13)	4.17 (13.3)	30.0 (33.7)	31.7 (33.4)
Attack				
Reaction Time (ms)	605 (199)	583 (190)	568 (213)	615 (202)
Percentage of Errors	12.5 (25.2)	9.17 (19.1)	31.7 (38.2)	25.0 (34.1)
Babies				
Reaction Time (ms)	611 (233)	598 (164)	582 (127)	596 (154)
Percentage of Errors	12.5 (25.2)	15.0 (26.7)	6.67 (17.3)	14.2 (22.4)
Basket				
Reaction Time (ms)	571 (195)	578 (147)	585 (169)	583 (151)
Percentage of Errors	20.0 (31.1)	8.33 (19.0)	11.7 (21.5)	8.33 (19.0)
Beach				
Reaction Time (ms)	569 (145)	569 (172)	598 (170)	588 (268)
Percentage of Errors	13.3 (29.2)	12.5 (28.4)	25.0 (28.6)	11.7 (28.4)
Earth				
Reaction Time (ms)	599 (145)	589 (203)	558 (137)	577 (251)
Percentage of Errors	9.17 (19.1)	9.17 (19.1)	9.17 (19.1)	16.7 (27.3)
Kids				
Reaction Time (ms)	607 (225)	605 (270)	588 (183)	588 (202)
Percentage of Errors	22.5 (28.1)	20.8 (24.6)	10.8 (20.4)	14.2 (26.0)
Lamp				
Reaction Time (ms)	599 (204)	571 (121)	577 (173)	583 (177)
Percentage of Errors	10.0 (24.2)	15.8 (25.0)	18.3 (27.8)	15.0 (26.7)
Mountain1				
Reaction Time (ms)	575 (144)	660 (635)	597 (238)	582 (162)
Percentage of Errors	15.0 (23.3)	7.50 (17.6)	10.0 (20.3)	8.33 (19.0)
Mountain2				
Reaction Time (ms)	559 (128)	540 (135)	595 (154)	552 (142)
Percentage of Errors	8.33 (23.1)	7.50 (17.6)	19.2 (27.6)	21.7 (28.4)
Nature				
Reaction Time (ms)	569 (161)	580 (192)	588 (180)	652 (451)
Percentage of Errors	9.17 (19.1)	11.7 (21.5)	17.5 (27.2)	15.8 (26.7)
Rabbit				
Reaction Time (ms)	558 (141)	584 (173)	604 (128)	616 (157)
Percentage of Errors	7.50 (21.9)	11.7 (25.2)	10.0 (20.3)	10.0 (20.3)
Skier				
Reaction Time (ms)	596 (149)	630 (219)	588 (189)	592 (177)
Percentage of Errors	5.00 (15.3)	21.7 (31.3)	8.33 (19.0)	11.7 (25.2)
Sky				
Reaction Time (ms)	570 (139)	544 (140)	551 (220)	529 (116)
Percentage of Errors	6.67 (17.3)	15.0 (26.7)	15.0 (23.3)	14.2 (26.0)

Table 4

Valence

Picture	Measure	Extrinsic Response Dimension	
		High	Low
Aimed Gun	Reaction Time (ms)	642 (179)	601 (152)
	Percentage of Errors	19.4 (24.4)	11.1 (16.0)
Attack	Reaction Time (ms)	635 (229)	581 (183)
	Percentage of Errors	19.4 (23.3)	8.33 (15.5)
Babies	Reaction Time (ms)	597 (136)	668 (235)
	Percentage of Errors	9.72 (12.2)	17.1 (20.6)
Basket	Reaction Time (ms)	581 (116)	592 (129)
	Percentage of Errors	10.6 (15.8)	24.1 (24.5)
Beach	Reaction Time (ms)	624 (247)	559 (140)
	Percentage of Errors	10.2 (15.9)	10.6 (16.5)
Earth	Reaction Time (ms)	575 (116)	606 (185)
	Percentage of Errors	7.41 (13.5)	14.8 (19.9)
Kids	Reaction Time (ms)	576 (142)	568 (137)
	Percentage of Errors	17.1 (19.3)	9.72 (14.8)
Lamp	Reaction Time (ms)	602 (186)	607 (176)
	Percentage of Errors	10.6 (15.8)	16.2 (18.3)
Mountain1	Reaction Time (ms)	572 (123)	596 (174)
	Percentage of Errors	8.33 (17.0)	13.0 (17.5)
Mountain2	Reaction Time (ms)	569 (149)	581 (132)
	Percentage of Errors	5.56 (12.7)	13.9 (16.0)
Nature	Reaction Time (ms)	579 (120)	599 (124)
	Percentage of Errors	5.56 (12.7)	15.7 (21.0)
Rabbit	Reaction Time (ms)	588 (126)	616 (145)
	Percentage of Errors	8.80 (13.8)	23.1 (22.9)
Skier	Reaction Time (ms)	604 (153)	639 (196)
	Percentage of Errors	8.64 (12.5)	7.41 (11.6)
Sky	Reaction Time (ms)	574 (122)	552 (120)
	Percentage of Errors	5.56 (12.7)	7.41 (15.2)

Table 5

Stimulus Picture	Extrinsic Response Dimension			
	High		Low	
	Experiment Half		Experiment Half	
	1	2	1	2
Aimed Gun				
Reaction Time (ms)	653 (159)	626 (367)	595 (144)	599 (184)
Percentage of Errors	16.7 (27.7)	22.2 (32.0)	10.2 (24.3)	12.0 (25.4)
Attack				
Reaction Time (ms)	644 (295)	627 (210)	585 (169)	577 (256)
Percentage of Errors	11.1 (21.2)	27.8 (34.9)	8.33 (18.3)	8.33 (23.0)
Babies				
Reaction Time (ms)	596 (136)	596 (192)	640 (234)	723 (469)
Percentage of Errors	10.2 (19.9)	9.26 (19.8)	13.9 (26.3)	20.4 (28.6)
Basket				
Reaction Time (ms)	583 (147)	573 (134)	588 (162)	593 (179)
Percentage of Errors	7.41 (22.8)	13.9 (26.3)	25.9 (32.1)	22.2 (32.0)
Beach				
Reaction Time (ms)	580 (156)	678 (379)	562 (156)	555 (172)
Percentage of Errors	9.26 (24.2)	11.1 (21.2)	13.9 (22.3)	7.41 (22.8)
Earth				
Reaction Time (ms)	588 (169)	568 (124)	660 (330)	589 (182)
Percentage of Errors	6.48 (16.4)	8.33 (18.3)	16.7 (27.7)	13.0 (26.3)
Kids				
Reaction Time (ms)	589 (224)	568 (175)	546 (133)	583 (189)
Percentage of Errors	21.3 (24.7)	13.0 (26.3)	8.33 (18.3)	11.1 (21.2)
Lamp				
Reaction Time (ms)	652 (348)	565 (114)	634 (242)	598 (239)
Percentage of Errors	6.48 (16.4)	14.8 (27.1)	18.5 (28.2)	13.9 (22.3)
Mountain1				
Reaction Time (ms)	566 (117)	577 (167)	602 (206)	589 (181)
Percentage of Errors	8.33 (23.0)	8.33 (18.3)	16.7 (24.0)	9.26 (19.8)
Mountain2				
Reaction Time (ms)	556 (197)	585 (163)	575 (159)	581 (138)
Percentage of Errors	7.41 (18.1)	3.70 (13.3)	13.9 (29.7)	13.9 (22.3)
Nature				
Reaction Time (ms)	555 (123)	603 (157)	613 (141)	602 (181)
Percentage of Errors	2.78 (10.6)	8.33 (18.3)	16.7 (27.7)	14.8 (23.3)
Rabbit				
Reaction Time (ms)	573 (131)	608 (173)	610 (151)	618 (170)
Percentage of Errors	8.33 (18.3)	9.26 (19.8)	20.4 (31.8)	25.9 (29.0)
Skier				
Reaction Time (ms)	618 (230)	603 (155)	645 (216)	628 (231)
Percentage of Errors	9.26 (24.2)	9.26 (19.8)	7.41 (18.1)	7.41 (18.1)
Sky				
Reaction Time (ms)	589 (190)	559 (121)	560 (144)	539 (160)
Percentage of Errors	5.56 (16.0)	5.56 (16.0)	6.48 (16.4)	8.33 (20.8)

Table 6