http://researchcommons.waikato.ac.nz/

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

A Replication and Extension of Hughes and Barnes-Holmes's (2011) Study of Induced Implicit Attitudes

A thesis
submitted in partial fulfilment
of the requirements for the degree
of
Master of Applied Psychology (Behaviour Analysis)
at
The University of Waikato
by
Douglas Miles Harvey

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2016

Abstract

In order to replicate Hughes and Barnes-Holmes's (2011) study, 12 undergraduate students were presented with two novel stimuli (the words "Cug" and "Vek") in one of three attitude-induction conditions. In the first condition, a training IRAP provided feedback that shaped the participants' brief-and-immediate relational responding over multiple discrete trials. In the second condition, a written narrative presented the two potential attitude objects (the words "Cug" and "Vek") as equivalent to a series of either positive or negative trait adjectives. In the third condition, participants were given both the training IRAP and the written narrative. Having completed the training phase, the participants then took a testing IRAP in which the words "Cug" and "Vek" were presented as being either "Similar" or "Opposite" to a range of verbal stimuli with either positive or negative connotations. As the results of Experiment 1 indicate, all three of the experimental manipulations generated differential relational responding, yet there was no repeat of the weakening in the 'IRAP effects' that was observed by Hughes and Barnes-Holmes. To find out if the engineered response biases persisted over a longer timescale than the 20 min it took to run the first testing IRAP, a second experiment was conducted in which the remaining participants (n = 11) retook the testing IRAP five days later (Experiment 2a). The results of Experiment 2a indicate that the response biases that were engineered in the first session persisted into the second session; furthermore, there was a weakening of the 'IRAP effects' across the three pairs of test blocks in all of the attitude-induction conditions. To find out if the engineered response biases could be strengthened by further training, following the second testing IRAP (Experiment 2a), the participants were presented with the same attitude-induction procedure they had been given during Experiment 1; they were then asked to complete a third and final

testing IRAP (Experiment 2b). The results of Experiment 2b indicate that response biases can be strengthened through further training. However, the ‘IRAP effects’ recorded in Experiment 2b were broadly weaker than the ‘IRAP effects’ recorded in the other two experiments.

Acknowledgements

Firstly, I would like to thank Dr Rebecca Sargisson and Dr Mary Foster for their help and guidance in producing this thesis: their support was invaluable and inspiring. Dr Sargisson, thank you so much for that week of intense proofreading: you helped me get this thesis into a readable state. I would also like to thank Joseph Graddy who assisted me in many elements of the research. Furthermore, the undergraduate students who took part in the two experiments also deserve thanks for donating their time so generously. And finally, I would like to thank the infinitely supportive Sean Hughes and Dermot Barnes-Holmes whose original study it was a pleasure to replicate and extend.

I take full responsibility for any ignorance or illogic within this thesis.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction and Literature Review

1.1 Defining the Attitude Construct

As a construct of interest to individuals working within a broad range of disciplines, 'attitude' is a psychological phenomenon that has been conceptualised and defined in a variety of ways. For example, one definition is that an attitude is a cognitive evaluation of an environmental event (Albarracín, Zanna, Johnson, & Kumkale, 2005; Eagly & Chaiken, 1993; Maio & Haddock, 2010) that may influence subsequent overt behaviour. Such evaluations occur whenever there is a stimulus (or attitude object; Eagly & Chaiken, 1993) with which to interact; and given how complex human existence is, these evaluations may provide a short-cut to navigating the world at large (Albarracín et al., 2005; Eagly & Chaiken, 1993). Furthermore, these evaluations may be positive, negative, or neutral with each attitude existing on a spectrum from extremely positive to extremely negative (Albarracín et al., 2005; Allport, 1935; Eagly & Chaiken, 1993; Maio & Haddock, 2010; Olson & Kendrick, 2008). The attitude object can be anything within the environment of an individual: an experimenter in a psychology lab, a person, something to be feared or desired, a complex sequence of events, or even a verbal stimulus presented within a written narrative.

As with any psychological construct, there are many alternate definitions to the one most approached by researchers (Eagly & Chaiken, 1993). Attitudes are variously described as being either physical or mental postures that prepare an organism for later action (Allport, 1935), a neurological change that alters how information is received in future (Warren, 1922, as cited in Allport, 1935), or a tendency to respond in a consistent way towards a particular stimulus (Eagly & Chaiken, 1993). This last definition assumes that such tendencies are

stored as dispositions acquired either during an individual's unique learning history or by way of genetic inheritance (Eagly & Chaiken, 1993; McGuire, 1985). Furthermore, this definition necessarily supposes that there is a hypothetical inner state which remains relatively stable over time despite the ongoing influence of novel learning situations (Hughes, Barnes-Holmes, & De Houwer, 2011). Yet, even though there has been some empirical support for the dispositional definition of the attitude construct, recent research has demonstrated that attitudes are more malleable than they were once thought to be (Cullen et al., 2009; Hughes & Barnes-Holmes, 2011; Ranganath & Nosek, 2006).

Perhaps one of the most utilised interpretations of the attitude construct is the Tripartite or Multicomponent Model (Cacioppo, Petty, & Green, 1989; Eagly & Chaiken, 1993; Maio & Haddock, 2010; Olson & Kendrick, 2008) which states that each attitude comprises elements of cognition, affect, and behaviour. Within the model, the cognitive component refers to what an individual thinks in regards to an attitude object (Eagly & Chaiken, 1993; Maio & Haddock, 2010; Olson & Kendrick, 2008): such thoughts include what they believe and what attributes are ascribed to the object to be evaluated. An example of attitudinal cognition would be an individual believing that a person of another race is inherently good or bad in accordance with a particular personality trait that the observed person may or may not actually possess in reality. The affective component refers to how an individual feels about an attitude object, including the emotions that the object elicits and any typical autonomic arousal that occurs in the presence of the attitude object (Olson & Kendrick, 2008). An example of attitudinal affect would be the arousal resulting from the presence of a feared object such as a spider or a snake and the summated feelings and moods that are experienced as a result of the encounter. The behavioural component includes any past response that has occurred in the presence of the attitude object and how that past response informs momentary 'online' behaviour (Olson & Kendrick, 2008); this last

component also includes memories of past actions and possibly elements of classical conditioning (Olson & Fazio, 2001), habit, and automatic responding. The Multicomponent Model has long been used within social psychology (and even Greek and Hindu philosophy) to describe the three different manifestations of the attitude construct (Eagly & Chaiken, 1993). However, the model becomes problematic when the evaluative responses of each component are indistinguishable and heavily convergent (Chaiken & Eagly, 1993; Cacioppo et al., 1989; Maio & Haddock, 2010). Indeed, in some circumstances the various distinctions become almost arbitrary (as occurs when what a person *thinks* is the same as how they *feel*), indicating that there may an inherent redundancy to the Multicomponent Model (Cacioppo et al., 1989; Maio & Haddock, 2010). Furthermore, the attitude construct can be readily defined in purely behavioural terms without the need for any hypothetical secondary structures and components (as occurs within the functional interpretation of attitudinal behaviour; Hughes et al., 2011; Hughes, Barnes-Holmes, & Vahey, 2012).

Beyond considering evaluative responses in affective and behavioural terms, there has been much recent work on the cognitive component of the attitude construct. Indeed, since the cognitive "revolution" (Goldstein, 2011, p. 12) of the 1950s and 1960s, there have been a great many cognitive theories of the attitude construct and several theoretical attitude models that use the computer-processing metaphor prevalent within the cognitive perspective (Crano & Prislin, 2006). Such mentalistic interpretations of evaluative responding almost always require the use of either a hypothetical structure within which attitudinal cognition occurs or some form of cognitive mechanism which influences and instructs overt attitudinal behaviour (Hughes et al., 2011). These cognitive interpretations of the "essential architecture" (Hughes et al., 2011, p. 465) of the attitude construct are most often developed to explain evaluative responding that has been observed and recorded under controlled experimental conditions. Yet, as Hughes et al. (2011) and Hughes and Barnes-Holmes (2013) assert, it is possible to

interpret such overt evaluative behaviour without resorting to the use of suppositional mental structures and processes. Furthermore, as all psychological research begins with the recording of observable behaviour anyway, the construction of secondary hypothetical constructs appears to complicate what is essentially just the observation and description of context-dependent evaluative behaviour. Indeed, it is the added complexity that is answered by the inherent parsimony of the behaviour-analytic interpretation of the attitude construct; within this interpretation, attitudinal responses are analysed to find out, not so much how they happen, but simply why they occur at a certain point in time and within a certain environment (Hughes et al., 2011; Hughes et al., 2012).

As is evident from what has just been discussed, how the construct is conceptualised influences entirely how one describes, explains, and researches attitudinal behaviour (Hughes et al., 2011). For example, if one conceptualises 'attitude' in solely functional terms (Hughes et al., 2011; Hughes et al., 2012) then there is no need to consider whether or not there are any cognitive elements involved in the evaluative process as they do not add value to the interpretation. Rather, the focus of inquiry would be on why the evaluative behaviour happens, what it achieves for the individual who is doing the evaluating, and under what contextual conditions the behaviour is performed (so that it can be predicted and possibly influenced; Biglan & Hayes, 1996).

1.2 The Two Types of Evaluation and Explicit Attitudes

Allport (1935) noted that an individual can hold more than one attitude towards a particular object or event and described such multiple evaluations as thinking one thing and saying something else. The evaluative duplicity Allport describes perhaps occurs because an individual readily knows how they should or should not respond to something despite their

'gut-feeling' telling them something different (Hughes et al., 2011; Olson & Fazio, 2001). Furthermore, a person's 'gut feeling' (Olson & Fazio, 2001) is not as readily accessible as the opinions or beliefs they express knowingly and overtly, a situation which requires contemporary attitude researchers to measure inaccessible attitudes through indirect means and the more accessible evaluations through direct questioning (Greenwald & Banaji, 1995; Nosek, 2007). In some circumstances, these two approaches to measuring attitudes produce contradictory results when used within the same experimental context (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010). As a result, there has been speculation as to how the attitude 'structure' promotes seemingly multiple evaluations under certain circumstances (McGuire, 1985). There has also been an attempt made by those working within behaviour analysis to explain in purely functional terms the way an individual can hold more than one attitude to an object at any one time (Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart, 2010; Hughes & Barnes-Holmes, 2013; Hughes et al., 2011; Hughes et al., 2012).

The contradiction between how people respond openly to an object and how they may respond to that same object when unaware of what is being asked of them may occur because in the process of evaluating something there are actually two types of attitude at play. Indeed, one of the principal distinctions made within mainstream attitude research is between evaluations that are explicit and those that are implicit (Nosek, 2007). The former type of attitude occurs deliberately and knowingly (Hughes & Barnes-Holmes, 2011; Payne & Gawronski, 2010), in a way that requires cognitive processing while the latter type occurs automatically with little cognitive effort (Greenwald & Banaji, 1995; Hughes & Barnes-Holmes, 2011; Payne & Gawronski, 2010); and because explicit attitudes operate only within consciousness and implicit attitudes occur solely without it, it is possible to hold two simultaneous attitudes to an object or event (Wilson, Lindsey, & Schooler, 2000). In Allport's (1935) assertion that each individual possesses two sets of attitudes, one public and

one private, there is the assumption that both types of evaluation are readily accessible. However, as a large body of research has shown, there are at least two types of evaluation and attitudes can be defined as being either explicit or implicit depending on whether or not they are accessed with conscious intent. Simply put, explicit attitudes are those that are readily accessible, self-reportable, and deliberate (Hughes et al., 2011; Nosek, 2007), and implicit attitudes are inaccessible, automatic, and unconscious (Greenwald & Banaji, 1995; Payne & Gawronski, 2010).

Explicit attitudes are cognitive evaluations that occur within specific contexts and in response to a demand situation in which controlled deliberation can occur. As Krosnick, Judd, and Wittenbrink (2005) explain, explicit attitudes are activated only when there is the required motivation to engage in controlled and deliberate evaluating and there are the cognitive resources available at that particular moment. In other words, an explicit evaluation takes place when an individual has the cognitive capacity to engage in the evaluative process and the innate need (or even 'desire') to do so (Krosnick et al., 2005). This dynamic may also be mirrored in the way explicit attitudes are formed, with such evaluations occurring in a context in which cognitive effort is able to be expended on the evaluative process and deliberate conscious judgements are engaged in (Bassili & Brown, 2005). Furthermore, explicit attitudes may be a by-product of 'socialisation' (Devos, 2008; Olson & Kendrick, 2008), which is the ever-present social interaction that occurs whenever an individual states an opinion that is received, processed, and reinforced by another individual who is then reinforced for their response to the original statement. An example of an explicit evaluation in such a socialisation process would be when a father tells his disinterested son which sports team he thinks will win an upcoming tournament. If the son believes that his father is a good predictor of future sporting outcomes, or if he has no better source of knowledge on the matter, or if he does not want to risk annoying his father, he may agree to the proffered

attitude in a process that is motivated, deliberate, controlled, and (as is the case with all social interactions) requiring of a certain amount of available cognitive resources. If the son is distracted or momentarily required to use his cognitive resources for another task, he may offer a more implicit attitude such as "I don't care about sports". Accordingly, it is evident that much social cognition involves the activation and expression of explicit attitudes. Indeed, many, if not all, explicit evaluative responses occur within some form of social interaction, be it the answering of a spoken request, the reading of an email, or the circling of a number on a Likert scale in response to a series of written instructions.

There are multiple theories as to how explicit attitudes are acquired. As Hughes et al. (2011) assert, explicit attitudes may be generated as a consequence of receiving new knowledge in the guise of information related to the immediate context. It is also possible that explicit attitudes are the behavioural products of cognitive reasoning that contains verbal rules (or propositions) activated within conscious awareness in an intentional way (Hughes et al., 2011; Olson & Kendrick, 2008). Furthermore, some theorists believe that explicit attitudes are actually evidence of an enduring predisposition to respond to a stimulus in a certain way and that the momentary context serves to only influence evaluative responding in a manner that disrupts the influence of the underlying disposition (Hughes et al., 2011). In the purely functional perspective that behaviour analysts work within, explicit attitudes are behavioural events which occur in response to stimuli operating within a specific environmental context (which can be the presence of a person, the spoken request for an opinion on a certain subject, or the reading of a written request for a response on a self-report measure). As such, explicit attitudes are not the manifestation of a stable disposition (Hughes et al., 2011) but an ever-changing 'act-in-context' (Biglan & Hayes, 1996; Hughes et al., 2012) with no intentional, conscious, and deliberate evaluation of an attitude object ever being the same as another evaluation.

1.2.1 'Direct' Measures of Explicit Attitudes

The measurement of attitudes occurs whenever there is an attempt to quantify an individual's evaluation of stimuli within their environment (Albarracín et al., 2005). Yet, being an unobservable psychological construct, an attitude is something that can only be inferred from overt behaviour, which in the 'direct' (De Houwer, 2006a, as cited in Hughes et al., 2011) measurement of explicit attitudes involves first asking an individual how they evaluate a stimulus and then interpreting just what that evaluation means in the context (Schwarz & Bohner, 2001). The 'directness' of the measurement is due to the absence of any non-social response-pressure on the behavioural system (such as the need to respond quickly or in contradiction to a learned rule) which would require automatic and uncontrolled responding (Hughes & Barnes-Holmes, 2013). In other words, within a 'direct' measurement approach, any request for an evaluation is directly stated and then answered directly. Considered *en masse*, 'direct' methods measure explicit attitudes because the individual is able to provide a deliberate and controlled response to the demand that is placed upon them. However, the 'direct' measurement of explicit attitudes presents some very serious threats to internal validity. For example, asking someone what their attitude is to something produces a behavioural outcome, be it a mark on a visual scale, the production of a series of synonyms in response to a verbal stimulus, or the circling of an individual digit within a series of numbers; yet, it is possible that these permanent products do not so much describe the true attitude at play as the socially mediated attitude that *should* be at play (Allport, 1935; Hughes et al., 2011).

The first self-report measures were designed during the initial flourishing of attitude research in the 1920s and 1930s (McGuire, 1985). These early 'direct' measures involved researchers presenting individuals with a large volume of questions that were selected to

access the attitude of interest (Krosnick et al., 2005). Indeed, as Allport (1935) reports, it was not unheard of for a questionnaire in the 1930s to include literally hundreds of questions. Such lengthy tests have been phased out over time in favour of single-question surveys (Krosnick et al, 2005) and tests involving the method first used by Likert (Allport, 1935). Single-question surveys involve visual measures such as meaningfulness scales, semantic-differential scales, and feeling thermometers (Hughes & Barnes-Holmes, 2011; Maio & Haddock, 2010), along with 'direct' questions in which an individual is asked to offer an explicit evaluation of a written stimuli. Likert scales present an individual with a series of opinion statements with the requirement being that they indicate how much they agree or disagree with the statement by selecting where on a 5-point scale their attitude lies (from strongly disagree through to strongly agree with a neutral value in third place on the scale; Allport, 1935; Maio & Haddock, 2010).

1.2.2 The Complexity of Measuring Explicit Attitudes

For a psychometric measure to be both valid and reliable it must consistently and accurately measure what it is developed and deployed to measure. Yet, as Hughes et al. (2011) report, a measure of explicit attitudes may not measure the attitude itself but the attitude plus some other inexplicable and unavoidable value. Furthermore, there is possibly an inherent inaccuracy in how an individual reports on their own introspectively accessed thoughts and feelings (Hughes et al., 2011), with the immediate context undermining what would otherwise be a fairly 'pure' private event. As Krosnick et al. (2005) state, there are many theorists who believe that attitudes are momentary constructions that only exist at the point in time in which the attitude is expressed; and this constructivist perspective is echoed by the "context dependency" (Schwarz & Bohner, 2001, p. 17) inherent in the interpretation

of evaluative behaviour offered by behaviour analysts and others working in accordance with the assumptions of functional contextualism (Hughes & Barnes-Holmes, 2013; Hughes et al., 2011; Hughes et al., 2012). So, rather than seek to address the extraneous variables inherent in explicit attitude testing (a task made near-impossible by the unobservable nature of such additional 'noise'), it may be best to merely accept them as a necessary part of the experimental transaction. In other words, the attitude is the evaluation plus the context (Krosnick et al., 2005; Schwarz & Bohner, 2001) and any other extraneous variables associated with the flaws that accompany the introspective process.

Perhaps one of the most commented on aspects of the measurement of explicit attitudes is whether or not an individual alters (either knowingly or unknowingly) their behavioural response to create a more favourable social outcome. Such biased responding has been labelled impression management, desirability bias, and "[strategic] self-presentation" (Hughes et al., 2011, p. 466). In addition to such desirability and demand biases (Hughes et al., 2011; Wittenbrink & Schwarz, 2007), there are other ways in which an individual's evaluations may be momentarily aberrant or out-of-character. In evaluating an attitude object, an individual may cognitively attend to only a certain aspect of the object in a manner that could influence their subsequent response (Schwarz, 2008; Schwarz & Bohner, 2001). For example, an individual could momentarily think of a solitary poor feature when asked to judge the design quality of a car that they otherwise like the look of. Alternatively, an individual may recall an episode of their past behaviour towards an attitude object that was anomalous or out-of-character (thus altering their current attitude), and they may attend to their past emotional and physiological responses to an attitude object in a manner that influences their momentary emotional and physiological state (and hence their 'online' evaluating of the object). Furthermore, and in a way that reflects the self-editing that occurs in impression management and demand-responding (Hughes et al., 2011), having decided on

what their explicit evaluation of an attitude object is, an individual may then alter their answer to best suit the context that the original question was asked in (Schwarz & Bohner, 2001).

It is common practice to label the methodologies intended to access explicit attitudes 'direct' measures (Hughes et al., 2011). But in determining what an individual's controlled and deliberate evaluation of an attitude object is, the presence of the explicit attitude can only be inferred from the behavioural outcome that occurs in response to the demand that is placed on the individual (Krosnick et al., 2005). When it comes to behaviours that occur in response to a multitude of contextual elements, the inference becomes a complicated and elaborate process. Therefore, it is reasonable to assert that, when it comes to attitude-testing, there is – in actuality – no such thing as a 'direct' measurement procedure (Krosnick et al., 2005) and that any form of self-report is contaminated by both the individual's inability to truly access how they evaluate an attitude object and the very context in which the question is asked and the answer is given. There is also a methodological complexity in using 'direct' measures to access an individual's explicit evaluation of an attitude object. As Schwarz (2008) and Schwarz and Bohner (2001) assert, having been presented with a self-report measure that uses a rating scale (such as a Likert system or a semantic-differential scale), an individual must first interpret the scale before stating what they believe their opinion is. They must then place that evaluation within the particular scaling system used (which is potentially a further point of pollution in the recording of the attitude). Furthermore, there may also be an added layer of complexity when written or spoken instructions are used with some 'direct' measures of explicit attitudes. For example, the use of semantic differential scales may be fairly self-explanatory to many experimental participants (they do after all contain a rating scale set between two clearly oppositional terms; Himmelfarb, 1993). However, being a (possibly anachronistic) metaphor, feeling thermometers require an additional level of thought to the

already increased cognitive processing required to answer the demand. Similarly, meaningfulness scales require secondary and tertiary evaluations regarding what it means to be 'meaningful' and whether or not the introspectively accessed attitude could be described as such.

So, in order to control for the "context dependency" (Schwarz & Bohner, 2001, P. 17) of the behavioural outcomes and to mitigate any impression management, social desirability, demand confusion, and methodological noise from the measurement of attitudes, researchers have recently turned to more 'indirect' measures (as was recommended by Allport in 1935).

1.3 Implicit Attitudes

Since the work of Freud in the early 20th century there has been an interest within psychological science with the 'unconscious' and what social psychologists often label as "implicit social cognition" (Payne & Gawronski, 2010, p. 2). However, it was during the flourishing of attitude research in the 1980s that the discourse began to focus on the contextual and cognitive conditions that produce implicit attitudes and those that produce more explicit evaluations (Payne & Gawronski, 2010). As Krosnick et al. (2005) summarise, implicit attitudes may exist when the occurrence of an environmental event (the "attitude object", p. 24), or a symbolic representation of that event, gives rise to a spontaneous and automatic judgement that takes place quickly and with little cognitive 'effort'. Because much of the research into attitudinal cognition is predicated on the existence of mental structures, it is commonly presupposed that implicit attitudes are stored in either long-term memory or within some other hypothetical mental compound (Amodio & Ratner, 2011; Krosnick et al., 2005). Furthermore, it has been postulated that such attitudes occur as a result of an associative process within that cognitive store (Fazio, 2007; Gawronski & Bodenhausen,

2011). In other words, when an individual is exposed to a particular stimulus, the presence of the stimulus (or attitude object) may activate the mental representation of itself and the mental representations of other stimuli with similar physical, verbal, or even arbitrary characteristics (as perceived by the individual) and/or spatial or temporal proximity to the initial stimuli within the environment (Hughes et al., 2011). It is in this association of one stimulus with another stimulus that evaluations are made (such as a picture of a spider activating the verbal stimulus *scary*). As Hughes et al. (2011) assert, within this cognitive reading of the attitude construct, implicit attitudes are different from explicit evaluations in the way in which they are retained: put simply, explicit attitudes exist as mental propositions that require intentional and deliberate processing whereas implicit attitudes exist as associations which are activated with little cognitive effort or conscious awareness (Hughes et al., 2011).

Four possible sources of implicit attitudes are traditionally postulated to be: the events of early childhood, past affective responses to specific events, the individual's culture, and the individual's need to evaluate events in a way that is consistent with what they otherwise think and believe (Olsen & Kendrick, 2004; Rudman, 2004). The first source of implicit attitudes refers to the evaluations that are gained during the observation of the environment that occurs before an individual is able to form, consolidate, and recall memories (Olson & Kendrick, 2008; Rudman, 2004). The second (affective) source refers automatic evaluations that are related to increased activation of the amygdalae (which are two groups of nuclei in the brain that are associated with affect and emotion; Olson & Kendrick, 2008). The third source references the myriad biased learning opportunities that occur within the ongoing socialisation process an individual experiences throughout the lifespan (Olson & Kendrick, 2008; Rudman, 2004). The final source focuses upon the hypothetical need an individual has to associate environmental events in a manner that is consistent with what has been learnt in

the past (Olson & Kendrick, 2008; Rudman, 2004).

According to a behaviour-analytic interpretation of implicit cognition, implicit attitudes are behavioural events that involve the cognitive-behavioural process of arbitrarily-applicable relational responding (Hayes, Barnes-Holmes, & Roche, 2001; Hughes & Barnes-Holmes, Hughes et al., 2011, Hughes & Barnes-Holmes, 2013). Within this interpretation, implicit evaluations are "brief and immediate relational responses" (BIRRs; Hughes & Barnes-Holmes, 2013, p. 102) in which two or more stimuli are placed within relational frames (such as frames of co-ordination, frames of opposition, or even just frames of difference; Hayes et al., 2001; Hughes & Barnes-Holmes, 2013). Furthermore, within the demand characteristics of the contexts that promote automatic and unconscious evaluative responding, there is little opportunity for an individual to elaborate on such simple relational frames, so there is meagre opportunity for the type of arbitrarily-applicable relational responding that is traditionally known as impression management or desirability bias (Hughes et al., 2011). Indeed, regardless of the specific procedures used, one of the defining characteristics of the implicit attitude measures that belong to the recent methodological "revolution" (Hughes & Barnes-Holmes, 2013, p. 100) is their inbuilt ability to nullify controlled and deliberate responding. Accordingly, the overwhelming majority of 'indirect' measures of implicit attitudes tend to create the conditions for brief-and-immediate responses to occur at the expense of more elaborated evaluative behaviour (Hughes & Barnes-Holmes, 2013).

1.3.1 'Indirect' Measures of Implicit Attitudes

As Barnes-Holmes et al. (2006) report, increasing focus on inaccessible, uncontrolled, and unconscious cognition has been accompanied by a vast amount of research to develop

measures that produce data from which the presence of 'concealed' evaluations can be inferred. Among some of these new methodologies are the Extrinsic Effective Simon Task (De Houwer, 2003), the Go/No-go Association Task (Nosek & Banaji, 2001), the Affect Misattribution Procedure (Payne & Lundberg, 2014), the Implicit Association Test (IAT; Greenwald, McGhee, & Schwarz, 1998), the various procedures that use evaluative and semantic priming (Krosnick et al., 2005; Wittenbrink, 2007; Wittenbrink & Schwarz, 2007), and a range of techniques that involve the measurement of physiological or neurological activity (such as galvanic skin responses and event-related potentials) following an individual's exposure to a stimulus (Barnes-Holmes et al., 2006; Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008; Maio & Haddock, 2010; Wittenbrink & Schwarz, 2007).

Despite the heterogeneous nature of implicit-attitude measures, there are several core similarities. For example, implicit-attitude measures tend to be delivered to participants using a computer-based testing programme; they often involve measuring response latency as the key dependent variable; they commonly involve a forced-choice scenario (such as having to decide whether to press one computer key instead of another; Wittenbrink & Schwarz, 2007); and they may contain an element of distraction or false demand in order to obscure the demand characteristics of the particular test (Wittenbrink & Schwarz, 2007). The most widely used procedures are the semantic and evaluative priming techniques and the IAT.

An evaluative priming task involves the on-screen presentation of a verbal stimulus to a participant who then presses either a key that signifies "good" or a key that signifies "bad". The 'strength' of an implicit attitude is then inferred from the time taken for the individual to respond to the presented stimulus (Fazio, 1990; Maio & Haddock, 2010; Wittenbrink, 2007). This inference is made on the assumption that there is a correspondence within the individual's memory between the attitude object (the stimulus presented on-screen) and some form of stored evaluation, and that this correspondence makes the evaluation easier to access

when the object is first presented as a priming stimulus (Wittenbrink, 2007). So, the faster the response time, the greater the inferred association between the evaluation and the attitude object and hence the stronger the individual's attitude towards that same object.

A common extension to evaluative priming involves the momentary on-screen presentation of a stimulus just before a verbal evaluation is shown (in the form of an adjective such as "good", "friendly", or "horrible"). The individual is then required to state (by pressing a set key on a computer keyboard) whether the verbal evaluation is positive or negative (Maio & Haddock, 2010; Wittenbrink, 2007). As with the original version of the test, the computer programme records the time it takes for the test-user to respond to each verbal evaluation. Furthermore, if the stimulus presented before the verbal evaluation is considered implicitly by the individual to be in agreement with the evaluation then their response time should be faster than if the opposite were true (Maio & Haddock, 2010; Wittenbrink, 2007). In this way, the stimulus presented before the evaluation acts as a prime for what follows it; accordingly, this variation on the original procedure is called a semantic priming task because the meaning and quality of one stimulus prepares the participant for responding to the meaning and quality of another stimulus (Maio & Haddock, 2010; Wittenbrink, 2007).

The IAT (Greenwald et al., 1998; Lane, Banaji, Nosek, & Greenwald, 2007) is also a computer-based assessment procedure in which response latency is the primary dependent variable. A typical IAT procedure will involve a series of trial blocks (Greenwald et al., 1998; Lane et al., 2007; Maio & Haddock, 2010) with a specific rule operating within each block. In each discrete trial, the participant is presented with a stimulus on a computer screen and is asked to press one of two keys on a keyboard using a single finger on either hand (such as pressing the "d" key with the left middle finger and the "k" key with the right middle finger). The computer programme records the time taken for the participant to respond, with the key

assumption being that faster response times indicate the presence of stronger mental associations and hence stronger implicit attitudes (Lane et al., 2007; Wittenbrink & Schwarz, 2007). So in essence, the IAT involves asking the test-user to emit a certain response for a particular class of stimuli and another response for a second class of stimuli with the primary outcome measure being the speed of the differential responding that occurs (Greenwald et al, 1998; Lane et al., 2007).

The IAT procedure was a direct precursor to the Implicit Relational Assessment Procedure (IRAP; Hughes & Barnes-Holmes, 2011), the 'indirect' measure used in Experiments 1 and 2 of my research project. The need for accurate and fast responding (Greenwald et al., 1998; Maio & Haddock, 2010), and the way in which differential responding is both promoted and recorded, are qualities shared by both procedures. But, as Barnes-Holmes et al. (2006), Chan, Barnes-Holmes, Barnes-Holmes, and Stewart (2009), and Hughes et al. (2011) assert, the IRAP measures attitudinal cognition in non-associative (and functional) terms whereas the IAT relies entirely upon the assumption that implicit attitudes are actually associative cognitive events.

As will be argued, behaviour-analytic implicit-attitude measures tend to be fairly similar in appearance to the mainstream techniques. For example, they often involve the presentation of verbal stimuli on a computer screen while participants make a forced choice between two or more options. However, the theoretical and philosophical assumptions that guide behaviour-analytic research methods are markedly different to those that direct the most common implicit-attitude measures. For example, behaviour analysts believe that all cognition is behaviour, which means that attitudes, whether explicit or implicit, need to be considered as types of behaviour in and of themselves and not just behavioural manifestations of a multi-component mentalistic construct (Hughes et al., 2011; Hughes et al., 2012). Furthermore, the effects recorded by the aforementioned associative or priming methods can

be explained in behaviour-analytic terms using solely behavioural principles (such as the arbitrarily-applicable relational responding of Relational Frame Theory; Hayes et al., 2001; Gross & Fox, 2009). Indeed, the inherent scope of the behaviour-analytic interpretation of attitudes means that it can readily describe and explain attitudinal behaviour performed under a broad range of experimental conditions, even when those conditions have been designed and implemented in non-behaviour-analytic terms.

1.3.2 The Complexity of Measuring Implicit Attitudes

'Indirect' measures of implicit attitudes have been offered as the solution to the issue of the context influencing explicit evaluations to the point that the 'true' attitude becomes hard to determine (Schwarz, 2008; Schwarz & Bohner, 2001; Wittenbrink & Schwarz, 2007). By accessing an unconscious and uncontrolled attitude in a way that does not allow for intentional deliberation and context perception, the vast majority of implicit attitude measures appear to reduce the extraneous 'noise' associated with the context-bound recording of attitudinal behaviour. As Hughes et al. (2011) and Wittenbrink and Schwarz (2007) assert, one such source of extraneous noise is the self-editing that an individual frequently (if not always) does when asked to express an attitude within a social interaction. 'Indirect' measures address this confounding variable by denying the individual a chance to self-edit through the use of either response-time limits or some form of procedural distraction. However, it must then be assumed that automatic responding is not also under the same control of the momentary context albeit in an inaccessible and unconscious way. As Schwarz (2008) states, there is some evidence that implicit attitudes are *not* free from contextual influence, in that environmental manipulations alter not only explicit attitudinal responding but also the expression of implicit attitudes.

In the same way that 'direct' measures of explicit attitudes produce ambiguous data that require interpretation, 'indirect' measures of implicit attitudes produce data that require a level of translation before any significant meaning can be ascribed to them. Even then, such an inference is merely that, a hypothetical reading of an attitude-behaviour relationship that is entirely unobservable. In other words, 'indirect' measures may evoke responding that is relatively free from social desirability bias and any other form of context-dependent impression management (Wittenbrink & Schwarz, 2007), but the responding evoked by 'indirect' measures cannot automatically be considered a proxy for whatever implicit attitudes an individual holds. So, in the same way that explicit, self-reported attitudes may not be what say they are, automatic and unconscious implicit attitudes (as inferred from quantitative data) may not be as readily apparent as they appear. An example of this ambiguity is the data produced by the IAT, in which it is assumed that shorter response times indicate stronger associations and hence stronger attitudes (Greenwald et al, 1998; Lane et al., 2007; Maio & Haddock, 2010). Yet does response latency truly represent the strength of stored associations? Furthermore, do such stored associations (if they exist) represent attitudes held towards objects? As there have been no answers to these questions from within the discourse, there are some researchers and theoreticians who believe that 'indirect' measures evoke evaluations and that these evaluations (or attitudes) are a form of behaviour in their own right (Hughes et al., 2011; Hughes et al., 2012), and it is this functional interpretation of the attitude construct that guides my research project.

A further issue regarding the use of 'indirect' procedures is whether or not they are suitably reliable and valid tools for measuring implicit attitudes and cognition. As Maio and Haddock (2010) report, a measure's reliability depends on how accurately it records what it is designed to record. Two key measures of reliability are test-retest reliability and internal consistency (Coolican, 2009; Golijani-Moghaddam, Hart, & Dawson, 2013; Maio &

Haddock, 2010). The former measures whether an assessment technique produces the same data when used to measure the same behaviour administered on repeated occasions (Coolican, 2009), the latter whether or not it consistently records what it is supposed to record within a single application (Coolican, 2009). As Golijani-Moghaddam et al. (2013) report, the reliability of 'indirect' measures of implicit attitudes tends to be lower than more 'direct' methodologies (with the IRAP – my research project's featured methodology – comparing favourably to some of the other implicit-attitude measures; Golijani-Moghaddam et al., 2013).

A measure's validity can be established by determining if its content is appropriate for what is being tested, whether it compares well with a pre-existing measure of the same construct, how well it predicts future behavioural outcomes, and whether or not it appears to measure what it has been designed to measure (Coolican, 2009; Golijani-Moghaddam et al., 2013). The IRAP has been found to be moderately valid in some of the aforementioned validity domains (Golijani-Moghaddam et al., 2013). Furthermore, it is only through the experimental reiteration of 'indirect' methodologies that their validity and reliability can be tested, which is why replication and extension experiments (of the kind featured in my thesis) are useful to any scientific discourse.

1.4 Theoretical Models of Attitudes

As was mentioned within the earlier discussion of attitude definitions, one of the most referenced theoretical models of attitudes is the Multicomponent or Tripartite Model (Eagly & Chaiken, 1993; Maio & Haddock, 2010; Olson & Kendrick, 2008; Zanna & Rempel, 1988). Within this model, attitudes are postulated to be summative evaluations comprised of cognitive, behavioural, and affective elements (Maio & Haddock, 2010; Olson & Kendrick,

2008). The cognitive component of an attitude involves any beliefs held about an attitude object; these beliefs contain learned attributions in which a positive or negative (or neutral) quality is ascribed to something that has been encountered in the environment. The behavioural component refers to how past behaviour influences present responding: for example, an individual is more likely to behave in a positive or negative way towards an attitude object if they have behaved in a positive or negative way to that object in the past (Eagly & Chaiken, 1993; Olson & Kendrick, 2008). The affective component of an attitude includes any emotional responses or autonomic arousal experienced in the actual or imagined presence of an attitude object and the feelings that those responses evoke. As Hughes et al. (2011) assert, these three elements may be summated into an evaluation that is positive or negative and in turn influences future behaviour towards the attitude object. Furthermore, the summated evaluation can be of any strength (Hughes et al., 2011).

The Multicomponent Model makes intuitive sense. Indeed, it balances the three core activities that define the lived experience: what one *thinks*, what one *feels*, and what one *does*. However, one possible limitation of the Multicomponent Model is the potential absence of any substantive differences between any of the three elements (Maio & Haddock, 2010). What a person *thinks* is arguably what they *feel* and thinking and feeling are both examples of how one *behaves* (at least within a radical behaviourist worldview).

One further limitation to the Tripartite or Multicomponent Model is that although it offers an accessible and useful description of what attitudes comprise, it does not account for the different behaviours that sometimes occur when an individual's evaluation of an attitude object is tested both 'directly' and 'indirectly'. Furthermore, the model does not address the way in which implicit and explicit evaluations each predict future behaviour in some contexts and not others (Devos, 2008). In some settings, an implicit-attitude measure accurately predicts future overt responding towards an attitude object whereas a 'direct' measure does

not (Hughes et al., 2011). Conversely, 'direct' measures sometimes better predict later behaviour than the currently used 'indirect' measures (Hughes et al., 2011).

To explain the way in which implicit and explicit attitudes sometimes disagree and how these two types of evaluation differently influence later responding, some theorists have proposed that implicit and explicit attitudes are two types of discretely stored and heavily stable mental disposition held simultaneously towards an attitude object (Hughes et al., 2011). Within this perspective, attitudes are fixed evaluations that are both persistent and resistant to changing circumstances. The dispositional perspective has been given empirical support by studies that have demonstrated the rigidity of implicit evaluations (Gregg, Seibt, & Banaji, 2006; Hughes et al., 2011). However, a good deal of recent research has demonstrated that rather than being stable dispositions, both implicit and explicit attitudes are malleable evaluations that are sensitive to changes in context (Gregg et al., 2006; Ranganath & Nosek, 2006), a notion that makes logical sense when one considers the ongoing need the human organism has to appraise the environment in a relevant and workable way (Albarracín et al., 2005). Accordingly, there is growing support for a more constructivist perspective in which attitudes are considered to be 'online' associative responses that occur within specific contexts (Gawronski & Bodenhausen, 2006; Hughes et al., 2011).

Regardless of whether attitudes are resistant to change or highly malleable evaluations, the aforementioned models do little to explain the cognitive processing that may or may not occur when an individual is required to evaluate an environmental event (Albarracín et al., 2005). Accordingly, single- and dual-process models have been offered to describe and explain in cognitive terms what may be happening during an implicit or explicit evaluation (Crano & Prislin, 2006; Hughes et al., 2011; Maio & Haddock, 2010). The single-attitude model is predicated on the assumption that all attitudes use a single piece of environmental information which can even be an element of the attitude object itself

(Kruglanski & Thompson, 1999). However, rather than being simple evaluations that occur in consistent ways, implicit and explicit attitudinal responses appear to require different levels of motivation and opportunity (Ajzen & Fishbein, 2005; Fazio, 1990a, as cited in Ajzen & Fishbein, 2005; Hughes et al., 2011; Krosnick et al., 2005). When a motivated individual is able to make a deliberate and controlled evaluation they will do so; but if the same individual is not motivated to evaluate something or is presented with the need to respond quickly, their evaluation will be automatic and unconsidered (Ajzen & Fishbein, 2005; Bassili & Brown, 2005). Dual-process models address this issue by assuming two discrete processes in implicit and explicit evaluative cognition (Crano & Prislin, 2006; Hughes et al., 2011; Maio & Haddock, 2010), with associative processing responsible for implicit cognition and propositional processing responsible for explicit cognition. Associative cognition involves the activation of a further stimulus (or network of stimuli) in response to an environmental event whereas propositional cognition involves the controlled and conscious deliberation of verbal rules (or propositions; Hughes et al., 2011).

A principle issue with the vast majority of cognitivist attitude models is that they assume a broad array of mental structures and processes that cannot be observed or measured, only inferred (Hughes et al., 2011; Uttal, 2000). Accordingly, there is an alternative interpretation that is gaining support within behaviour analysis in which a more functional (and parsimonious) approach to the attitude construct is offered. Within this interpretation, attitudes are cognitive-behavioural events called arbitrarily-applicable relational responses (Gross & Fox, 2009; Hayes et al., 2001; Hughes & Barnes-Holmes, 2011; Hughes et al., 2011; Hughes & Barnes-Holmes, 2011) in which a verbal representation of an attitude object is placed within a relational frame with one or more other stimuli. As I will discuss, arbitrarily-applicable relational responding can be used to explain both attitudinal behaviour and the discrepancies that occur when evaluations of the same attitude object are captured by

'direct' and 'indirect' methodologies (Hughes & Barnes-Holmes, 2013).

1.5 Attitudes as Behaviour

This discussion has predominantly presented attitudes as evaluative episodes of implicit or explicit cognition that may or may not feature affective and behavioural components (Eagly & Chaiken, 1993; Olson & Kendrick, 2008; Maio & Haddock, 2010). These evaluative events have been at times described and understood using hypothetical mental structures and processes (Hughes & Barnes-Holmes 2011; Hughes & Barnes-Holmes, 2013, Uttal, 2000). However, attitudes may also be considered as entirely behavioural events using a purely functional approach in accordance with the assumptions of functional contextualism (Hughes & Barnes-Holmes, 2011; Hughes & Barnes-Holmes, 2013; Hughes et al., 2012).The fundamental goal of the functional approach is to understand, predict, and influence behaviour with scope (explaining as many behaviours occurring within as many contexts as possible), precision (using a limited series of principles to explain a particular behaviour), and depth (applying a single interpretation across a range of disciplines and multiple levels of analysis from discrete behavioural events to the most extended patterns of responding). The application of this tripartite goal to the comparison of a 'typical' cognitive theory of attitudes and a more functional interpretation will often reveal a disparity between the two types of approach, with the cognitive interpretation being able to describe and explain (using hypothetical mental structures and processes) a specific aspect of the attitude construct with great precision whilst simultaneously being unable to adequately describe and explain other aspects of the construct or the same aspect within a different context (as is the case when a 'mentalistic' model is unable to account for the differential responding that occurs when both 'direct' and 'indirect' measures are used to capture the same evaluation);

conversely, the functional interpretation may describe and explain (using broad behavioural principles) what occurs when an individual is engaged in evaluative cognition and behaviour in a way that allows for the analysis to be generalised to a seemingly limitless range of contexts (Hayes et al., 2001; Hughes et al., 2011). Additionally, the cognitivist interpretation may not have the predictive power that the functional approach has, meaning that it is difficult to suppose the nature of future responding from a cognitive model or theory alone. Furthermore, the cognitivist interpretation requires the creation of a complex system of assumptions and inferences to describe and explain attitudinal behaviour (Hughes et al., 2011; Uttal, 2000) whereas the more pragmatic functional approach would simply analyse what it is that these evaluations achieve (Hughes et al., 2011). Indeed, the inherent pragmatism of the functional interpretation of attitudes directly counters the complexity of the hypothetical mental structures that are postulated to mediate both private evaluations and subsequent overt behaviour in the more 'mentalistic' interpretations (Hughes et al., 2012; Uttal, 2000).

Relational Frame Theory (RFT; Hayes et al., 2001) provides one way to define attitudes in behavioural terms (Barnes-Holmes et al., 2010; Hughes et al., 2011; Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). RFT explains how an individual may evaluate an attitude object in a positive or negative way. RFT also accounts for the way individuals form attitudes to objects, events, or concepts they have not directly encountered (Gross & Fox, 2009; Hayes et al., 2001) and explains why an individual sometimes responds to the demands of a 'direct' measure differently to how they respond to the demands of an 'indirect' measure (Hughes & Barnes-Holmes, 2013). According to RFT, arbitrarily-applicable relational responses are acquired through historical learning processes that inform the momentary responding that occurs within specific contexts; so there is no need for cognitive structures and suppositional mental processes (Hughes et al., 2012; Uttal, 2000). This absolute "context

dependency" (Schwarz, 2008; Schwarz & Bohner, 2001, p. 17) is in keeping with Skinner's assertion that all behaviour begins in the environment, not in the *psyche* (Skinner, 1965). Furthermore, the inherently pragmatic nature of the RFT analysis of attitudes means that it is not necessary to unpack *how* an attitudinal response occurs (especially using hypothetical mental components; Uttal, 2000) but it is important to establish *why* an individual evaluates an attitude object as they do.

1.5.1 Stimulus Equivalence, Derived Relational Responding, and Relational Frame Theory (RFT)

The phenomenon of stimulus equivalence was first observed when Murray Sidman and his colleagues were teaching two children with severe intellectual disability to read (Sidman & Cresson, 1973). The two boys were able to match printed words with spoken words having only been taught to match (a) spoken words with pictures, and, (b) the same pictures with printed words. So, the boys were able to match printed words with spoken words despite not having been taught how to do so (Sidman & Cresson, 1973). The ability of humans to correctly match stimuli in such a manner defies the principles of classical and operant conditioning, in that the derived stimulus relationships have not been previously acquired and there is no pre-existing matching behaviour (for the particular sample of stimuli) that has been reinforced in the past (Gross & Fox, 2009; Hayes et al., 2001; Törneke, 2010). In experimental settings, the untaught matching of stimuli often occurs during procedures in which the participant emits conditional discriminations, responses that are reinforced dependent upon the presence of another stimulus (Sidman, 2009). Since Sidman's work, there has been much research into how an organism learns such discriminations (Barnes-Holmes, Barnes-Holmes, Smeets, Cullinan, & Leader, 2004) and it is through

experimental procedures where participants perform conditional discriminations (in matching-to-sample tasks, for example) that stimulus equivalence is most often observed (Hayes et al., 2001).

In essence, stimulus equivalence is the learned relating of two or more stimuli through the mathematical principles of reflexivity, symmetry, and transitivity (Hayes et al., 2001; Törneke; 2010). Reflexivity refers to when one stimulus is correctly matched with an identical stimulus in a matching-to-sample task (Hayes et al., 2001). Symmetry refers to when one stimulus is matched with another stimulus in a way that reverses a previously reinforced relationship (Stimulus B = Stimulus A when only the matching of Stimulus A to Stimulus B has been reinforced previously; Hayes et al., 2001). Transitivity refers to when one stimulus is matched with a second stimulus which is then matched with a third stimulus and the first stimulus becomes matched with the third stimulus through a mutual relationship with the second stimulus (in other words, Stimulus A = Stimulus B, and, Stimulus B = Stimulus C, so, Stimulus A = Stimulus C, and, when combined with symmetry, Stimulus C = Stimulus A; Hayes et al., 2001).

Beyond this most basic definition of the phenomenon, stimulus equivalence has also been defined as the process that occurs when – through human language and cognition – verbal stimuli become associated in ways that have not been directly learnt and reinforced in the past (Barnes-Holmes, Barnes-Holmes, Smeets, Cullinan, & Leader, 2004; Hayes et al., 2001). This additional definition is of particular interest because it describes the core psychological process of RFT (Hayes et al., 2001), arbitrarily-applicable relational responding, or the relating of an object (and its verbal representation) with one or more objects (and their verbal representations) in a way that has not already been learnt by the individual doing the relating. Furthermore, because humans are able to emit verbal behaviour, they are also able to relate stimuli in arbitrary and non-formal ways (in other words, they are

able to relate stimuli that do not share similar physical properties; Hayes et al., 2001); and, as is the case with stimulus equivalence, such a cognitive-behavioural process goes against the principles of strict operant conditioning (Gross & Fox, 2009) because it is possible to relate objects in a way that has not been reinforced in the past.

RFT is a functional interpretation of human cognition and language which posits that both the private events that occur within active consciousness, and the overt behavioural events that occur when human language is used socially, involve the relating of stimuli in accordance with the behavioural principle of arbitrarily-applicable relational responding (Hayes et al., 2001; Hughes & Barnes-Holmes, 2011). As Hayes et al. (2000) and Törneke, 2010) state, there are a couple of postulated ways in which two or more objects can be related beyond the simple matching of identical stimuli. Firstly, *mutual entailment* occurs when a stimulus is related to another stimulus in a manner that reverses a previously acquired (and reinforced) relation. Secondly, when a stimulus is related to another stimulus through concurrent relations with a third stimulus then *combinatorial entailment* is said to be taking place (Hayes et al., 2001; Törneke, 2010). Mutual entailment and combinatorial entailment are similar to the processes of symmetry and transitivity in stimulus equivalence. However, there is an added complexity to the relations that occur when verbal stimuli are related than when non-verbal stimuli are matched on purely formal or associative grounds. For example, if a stimulus is related to another stimulus due to its greater size (i.e., one object is "larger" than the other object which is "smaller" than it), then in purely verbal terms neither the stimuli or their relationship are symmetrical (in other words, Stimulus A = Stimulus B, but within language and cognition, Stimulus B ≠ Stimulus A as "smaller" is not the same as "larger" and hence the relationship is not symmetrical). Similarly, the six concurrent associations that occur in transitivity work on the assumption that some sort of equivalence exists between the three stimuli that are matched (or equated) and so some form of

equivalence exists between the six stimulus associations (A = B, B = A, B = C, C = B, A = C, C = A); yet as was demonstrated in the first example, such an equivalence is not necessary within arbitrarily-applicable relational responding as it allows – through human verbal behaviour – for the arbitrary relating of stimuli without recourse to strict matching (Hayes et al., 2001).

When two or more stimuli become related through mutual or combinatorial entailment they are said to be in a relational frame, which is not so much a hypothetical mental structure as a term of reference for a particular cognitive-behavioural event that occurs during arbitrarily-applicable relational responding. Relational frames can be said to define the specific nature of the momentary relation between two or more stimuli with there being a myriad of possible relation types available (although the most prevalent frames are those of co-ordination or opposition; Hayes et al., 2001; Hughes & Barnes-Holmes, 2011). Furthermore, within such momentary behavioural events, a stimulus can assume the functions of another stimulus merely by being present within a relational frame with another stimulus (Hayes et al., 2001; Törneke, 2010). In other words, if a particular verbal stimulus becomes mutually or combinatorially entailed with another verbal stimulus, then the behavioural function of either stimulus may be transferred on to the other. This process can be witnessed in the 'real world' when an individual responds to a previously neutral object in a fearful way because that object has been placed in a relational frame with an object that the individual is afraid of (as is the case when an individual responds to an unattended suitcase in a public place as though it almost certainly is a bomb). The transference of behavioural function may also occur when novel attitudes are formed, something that can be recreated experimentally through the presentation of an attitude object with a verbal stimulus of either a positive or negative valence (such as trait adjective with positive or negative connotations).

RFT is a theoretical account of human language and cognition (Hayes et al., 2001)

that is deeply relevant to the behaviour-analytic (and functional) interpretation of the attitude construct. Indeed, within this perspective, arbitrarily-applicable relational responding *is* the implicit and explicit evaluative behaviour that occurs in the private and public expression of attitudes. And finally, arbitrarily-applicable relational responding adequately accounts for the divergences that often occur when an attitude is accessed using 'direct' and 'indirect' measures (Barnes-Holmes et al., 2010). As will be discussed in the next section, the Relational Elaboration and Coherence (REC; Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2013; Hughes et al., 2012) Model has been offered as a functional account of implicit and explicit attitudinal responding and an explanation of why the two forms of evaluation so often disagree.

1.5.2 The Relational Elaboration and Coherence (REC) Model

Despite being primarily a behaviour-analytic explanation as to why there are different types of relational responding, the Relational Elaboration and Coherence (REC) Model is offered here as a functional interpretation of implicit and explicit attitudes (Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). Like RFT, the REC Model is predicated on the assumption that all language and cognition involves relating combinations of stimuli in a behavioural process known as relational responding (or, to give the process its full title, arbitrarily-applicable relational responding; Hayes et al., 2001; Hughes & Barnes-Holmes, 2013). When an individual is given little time to respond to a stimulus, they may emit a "brief and immediate relational response" (BIRR; Hughes & Barnes-Holmes, 2013, p. 102), which is broadly synonymous to an automatically produced implicit evaluation. However, when an individual is given more time to respond to a stimuli, they may emit an initial BIRR which is then followed by a series of further relational

responses which may or may not feature elements of the original response (Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2013) along with relational responses to other salient stimuli in the proximal environment (a little like the controlled, deliberate, and context-dependent responding that occurs during explicit evaluations; Hughes & Barnes-Holmes, 2011; Payne & Gawronski, 2010).

To offer an example of this process, imagine a woman has returned to her hometown after an absence of several years in which she travelled the world and fell in love with her current partner who is also the father of her child. While driving down the main street, the woman sees an ex-partner whose infidelity resulted in her leaving the town several years earlier. When the woman first sees her ex-partner she may emit a BIRR in which he is evaluated negatively in verbal terms in accordance with the woman's learning history. (In other words, the ex-partner is in a learned or derived frame of co-ordination with any number of negative trait adjectives such as "*untrustworthy*", "*unfaithful*", or "*two-timing*".) However, as she continues to drive down the road, the woman's initial evaluation may be followed by further relational responses in which the ex-partner and his actions are given equivalency to such verbal responses as "*causing freedom*" and a range of positive memories associated with the woman's time overseas, her subsequent intimate relationship, and the birth of her child. This pattern of additional cognition is called an "extended and elaborated relational response" (or "EERR"; Hughes & Barnes-Holmes, 2013, p. 102) and if given enough time, EER-Responding should cohere into a network of relational responses that reflects both the woman's unique learning history and the momentary context she is in (Hughes et al., 2012). Furthermore, as the woman's most immediate relational response is added to by further relational responses, it is possible that some of the additional responses are oppositional, indicating that it is indeed possible to hold contradictory attitudes towards an attitude object within certain contexts.

As should be apparent from the above example, the two types of relational responding are not so much separate operations as different manifestations of the same behavioural process. The main distinction between the two forms is that for BIRRs to be automatic and unconsidered and EERRs to be deliberate and controlled, the former must occupy less time than the latter (Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). Simply put, because BIRRs are time-pressured in a way that EERRs are not, it should take longer to emit an extended and possibly complex pattern of relational responses than the short time it takes to emit a few automatic relational evaluations. However, even though BIRRs can be performed in less time than EERRs, it does not mean that the brief responses are always emitted (and completed) at an earlier point in time to more elaborate responses. Due to the nature of arbitrarily-applicable relational responding, EERRs can contain multiple BIRRs *and* be the antecedent stimuli to brief-and-immediate responses in and of themselves (Hughes & Barnes-Holmes, 2013). Accordingly, there is no limit to the complexity of implicit and explicit evaluations when such cognitive events are considered from a purely functional behaviour-analytic perspective without recourse to any mediating mental constructs (Hughes et al., 2011; Hughes et al., 2012).

The complex responding that is implied by the REC Model may account for incidences in which 'indirect' measures of implicit attitudes disagree with 'direct' measures of explicit attitudes (Barnes-Holmes et al., 2010; Hughes et al., 2012). For example, an individual may evaluate a person of the opposite sex negatively (due to their unique learning history) when presented with a procedure that sets the conditions for automatic and unconscious responding; but when presented with a request for the same attitude within a social setting, given time to respond, the individual may offer an answer that suggests they value people of the opposite sex in a more positive way. The REC Model explains such contradictory responding as follows: relational responding is either acquired through

mediation by the verbal community or performed as some form of derived (and unlearned response; Hughes et al., 2012). Accordingly, a stimulus is more likely to be equivalent to another stimulus when at least one other person reinforces the stated relational frame (Hayes et al., 2001; Hughes & Barnes-Holmes, 2013); however, such reinforcement is not a compulsory requirement of arbitrarily-applicable relational responding (Hayes et al., 2001). Furthermore, the social mediation of relational responses occurs on a broader and more elaborate scale when an individual's relational responding is mediated by the verbal community at large (as is the case when someone conforms to the 'norms' of their society). So, an individual's initial BIRR may involve relating two stimuli in a way that may or may not have been reinforced in the past by a single individual or a small subset of society, but, given both the time to respond elaborately and the presentation of the same two stimuli in a context containing many other latent relational networks, their EERR may result in a very different type of evaluation (and one that more readily coheres with the 'normative' response expected by the verbal community). So to return to the earlier example, the sexist individual may readily profess their negative opinion of the opposite sex when alone or with one other person, but they may become far more equalitarian when presented with a more complex social context and the time to respond elaborately (such as having to evaluate a person of the opposite sex in a television studio with a live audience, the majority of whom would not tolerate such sexist opinions).

On a purely pragmatic level, the REC Model is important because it allows for the two most-postulated types of evaluative cognitive event to be described, differentiated, and explained in functional terms without the need for either hypothesised mental structures or suppositional psychological processes (Hughes et al., 2012). However, as Hughes and Barnes-Holmes (2013) assert, the use of observable (and recordable) behaviour as a proxy for a hypothetical mental event is still prevalent within attitude research despite there being

adequate precision and scope in the explanations offered by those seeking to analyse the functional nature of attitudes and evaluative cognition. One possible reason for this reliance on 'mentalistic' constructs is the evident need of many researchers to explain in mechanistic terms what happens when an attitude is formed, retained, and activated. Yet functional interpretations of evaluative behaviour may readily be used to account for why people perform initial evaluations, why such evaluative behaviour persists over time, and why such behaviour responds so readily to changes in context (Hughes et al., 2011). Furthermore, the measures used by individuals working in accordance with the functional interpretation of attitudes access 'online' evaluations as the very behavioural events they are purported to be and not as proxies of unobservable cognition (Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). Accordingly, the following sections focus on a small sample of behaviour-analytic measures of implicit attitudes including the Relational Evaluation Procedure, which was a direct precursor to the IRAP.

1.6 Behaviour-Analytic Measures

For a research methodology to be valid, its conception, development, and deployment must be informed by the theoretical and philosophical assumptions that underpin the target construct and its associated field of inquiry. In other words, the scientific approach used to measure a psychological construct of interest (be it implicit attitudes, evaluative cognition, or arbitrarily-applicable relational responding) should occupy the same suppositional 'space' as the construct and operate within the same epistemological and ontological 'rules'. A fundamental reason for this philosophical and theoretical consistency is the scientific need for a research methodology to accurately access and record whatever it is designed and deployed to measure (Coolican, 2009; Golijani-Moghaddam et al., 2013). For example, in order to

successfully record episodes of the behaviour specifically associated with a particular psychological construct (in lieu of being able to directly access the construct itself), a methodology must account for how the construct is defined and how it influences overt and observable responding (if at all). However, in the design and application of psychological research methods, such an analysis can be taken too far into the theoretical and philosophical realms. Indeed, one significant problem with a good deal of 'direct' and 'indirect' measures of explicit and implicit attitudes is that they rely upon such a weight of supposition that they may, in fact, not be recording what they are intended to record (for example, it is possible that the IAT measures implicit associations rather than implicit attitudes; Hughes et al., 2011). This concern is arguably less common in the deployment of behaviour-analytic methodologies as the measures directly access what they are supposed to access: behaviour as it is performed 'online'. Accordingly, the following two sections outline a brief sample of behaviour-analytic procedures. As will become evident, the methods were not primarily designed to access implicit or explicit attitudes *per se*; however, the procedures all promote relational responding which is what evaluative cognition is in behaviour-analytic terms.

1.6.1 The Stimulus Equivalence Procedure

Unlike the two behaviour-analytic approaches that are discussed after this one, the stimulus-equivalence procedure is not a discrete, made-to-order assessment tool; rather it is a broad class of techniques that all involve matching-to-sample procedures and the use of conditional discrimination (the reinforcement of a correct response following the presentation of a sample stimulus alongside a series of comparison stimuli; Sidman, 2009). Accordingly, the stimulus-equivalence procedure used by Watt, Keenan, Barnes, and Cairns (1991) is offered here as a useful explanatory example of a prevalent form of behaviour-analytic

psychological testing. The procedure is also being discussed because it is a key precursor to the IRAP through its use of a computer-based programme to both train and test the participants in an experimental setting.

Watt et al.'s (1991) experiment featured a group of Northern-Irish Protestants, a group of Northern-Irish Catholics, and a group of English Protestants. The participants in all three groups were exposed to a double-barrelled training procedure in which the participants were continuously and then intermittently reinforced for correctly matching three Catholic names with three nonsense syllables and for correctly matching the same three nonsense syllables with three Protestant symbols (Watt et al., 1991). The training phase used a computer-based programme in which either a Catholic name or a nonsense syllable (the sample stimuli; Watt et al., 1991) was presented on-screen above a combination of nonsense syllables and Protestant symbols (the comparison stimuli; Watt et al., 1991). The participant was then asked to select one of the comparison stimuli by pressing the keyboard key that had been allocated to it. Feedback was then provided by the presentation of either the word "Correct" or the word "Wrong" and the sounding of either a high-pitched tone or a low-pitched tone depending on whether or not the preceding response was the right one. In the testing phase, the participants undertook the same matching-to-sample task; however, no feedback was presented.

One interesting outcome of the study was that, rather than matching Catholic names to Protestant symbols as per the contingency at play, some of the participants matched Protestant names to Protestant symbols with this type of incorrect response being performed by participants in all three experimental groups (Watt et al., 1991). This discrepancy offers some support for the Relational Elaboration and Coherence Model (REC; Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2011; Hughes et al., 2011; Hughes et al., 2012) in that some of the participants may have been responding to the stimuli as per their matching-to-

sample training but also in a way that cohered with the networks of stimulus relations that existed within the wider social context. In other words, in accordance with their individual learning histories, some of the participants related the Protestant names to Protestant symbols despite being trained not to. Accordingly, one of the limitations of Watt et al.'s (1991) study is that it is evidently not possible to accurately assess experimental learning when familiar stimuli are used. Furthermore, there is a possible ethical concern to reinforcing the matching of familiar stimuli in a non-normative way, especially when the stimuli have some level of association with social unrest or trauma (as is the case in Watt et al.'s experiment; 1991).

The aforementioned concerns with Watt et al.'s (1991) study aside, one of the general limitations of the stimulus-equivalence approach is that it only assesses whether or not an individual is able to cognitively associate two or more stimuli in the simplest sense and not whether or not they are able to cognitively relate the stimuli within either individual relational frames or more elaborate relational networks (which is purported to be more likely to happen within everyday language and cognition; Hayes et al., 2001; Hughes & Barnes-Holmes, 2013; Hughes et al., 2011; Hughes et al., 2012). This drawback was addressed in the conceptualisation and design of two measures of arbitrarily-applicable relational responding: the Relational Evaluation Procedure and the IRAP.

1.6.2 The Relational Evaluation Procedure as a Direct Precursor to the Implicit Relational Assessment Procedure (IRAP)

With the introduction of RFT (Hayes et al., 2001) came the premise that, through the behavioural process of arbitrarily-applicable relational responding, verbal and non-verbal stimuli are related within language and cognition to powerful effect; furthermore, such stimuli can be relationally responded to in a way that is either directly learned or entirely

derived from other learned relations (Törneke, 2010). As was mentioned earlier, the ability to relate stimuli in ways that have not been previously reinforced is similar to the behavioural phenomenon of stimulus equivalence (Barnes-Holmes et al., 2004). However, there is one important and irreconcilable difference between the principles at play within stimulus equivalence and those at play within RFT: unlike the strict reciprocity of reflexivity, symmetry, and transitivity, arbitrarily-applicable relational responding accounts for the bi-directional cognitive framing that occurs when two qualitatively different stimuli become related in a mutually entailed way (Hayes et al., 2001). For example, Stimulus B may be an object that is "taller" than Stimulus A (an object that is "shorter" than Stimulus B), so when a person learns to relate Stimulus B to Stimulus A (and Stimulus A to Stimulus B) it is not through some form of equivalence matching (they are after all both physically and *verbally* different) but through a process by which the bi-directional relating of stimuli becomes a learned behaviour (Hayes et al., 2001). This fundamental premise is extended in the process of combinatorial entailment (Hayes et al., 2001), whereby a third object (Stimulus C) becomes related to the first stimulus (A) in a way that is similar to how the second stimulus (B) relates to it regardless of the similarity or dissimilarity of the second and third stimuli (such as when Stimulus C is "taller" than Stimulus A but not quite as tall as Stimulus B).

The implications for psychometric testing (which includes the assessment of implicit and explicit attitudes) are that it is not possible to test for arbitrarily-applicable relational responding using only a matching-to-sample scenario (as occurs within the stimulus-equivalence procedure; Watt et al., 1991) or some form of associative paradigm (Hughes et al., 2011). Rather, it is more appropriate (and arguably more true-to-life; Hayes et al., 2001) to present combinations of stimuli to a participant who attends to the stimuli before selecting the appropriate relational term from an array of two (or possibly more) response options (such as the oppositional words "Same" and "Different"; Barnes-Holmes, Hayden, Barnes-

Holmes, & Stewart, 2008). This approach was first attempted by behaviour analysts with the Precursor to the Relational Evaluation Procedure (Cullinan, Barnes-Holmes, & Smeets, 2001); however, this seminal measure still used matching-to-sample testing and it was not until the Relational Evaluation Procedure (REP) that a methodology truly accessed arbitrarily-applicable relational responding in accordance with the principles of RFT (Barnes-Holmes et al., 2008; Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009).

The REP is a behaviour-analytic approach that is a direct precursor to the featured 'indirect' methodology of my research project; and like the IRAP, the REP is predicated on the assumption that both language and cognition are comprised of relational frames (Hayes et al., 2001). Within each trial of the REP, an individual is asked to report on the type of relationship that exists between two stimuli presented to them simultaneously (Barnes-Holmes et al., 2008; Vahey et al., 2009). For instance, two shapes may be shown concurrently with a pair of relational terms such as "Similar" and "Opposite" or "Same" and "Different". The test-user then selects the term that best describes the relation between the shapes (Barnes-Holmes et al., 2008; Stewart, Barnes-Holmes, & Roche, 2004; Vahey et al., 2009), a response that momentarily clears the screen in preparation for the presentation of the next stimuli and relational response options.

Unlike some of the 'indirect' measures of implicit attitudes being used at the time, the REP did not limit the time a test-user was given to respond. Accordingly, the REP is a pure measure of arbitrarily-applicable relational responding as defined by RFT (Barnes-Holmes et al., 2008). But, because an individual taking the REP can respond in a controlled and deliberate way within each trial, the procedure is unable to elicit the type of BIRRs that are described and explained by the REC Model (Barnes-Holmes et al., 2010; Hughes et al., 2012; Hughes & Barnes-Holmes, 2013; Hughes et al., 2011). Therefore, the procedure may also be unable to access any implicit attitudes that the participants may hold. Given that the REP was

not intended to produce automatic and uncontrolled responding (Hughes & Barnes-Holmes, 2013; Stewart et al., 2004), a further behaviour-analytic assessment tool was developed in which there is a need for participants to respond quickly and accurately to paired stimuli in relational terms (Barnes-Holmes et al., 2008). The resultant methodology is called the Implicit Relational Assessment Procedure (IRAP) and its deployment has allowed behaviour analysts to study implicit attitudes in functional (and hence relational) terms (Barnes-Holmes et al., 2008; Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2011; Hughes et al., 2011; Hughes et al., 2012).

The IRAP is a computer-based measure of implicit (or brief-and-immediate) relational responding that requires participants to relate sample and target stimuli within a series of discrete trials. As Barnes-Holmes et al. (2008) and Vahey et al. (2009) assert, the IRAP combines the relational-response requirements of the REP with the alternating conditions and response-time requirements of the IAT (Greenwald et al., 1998). A participant using the IRAP is required to sometimes respond in a way that is consistent with their learning history and to sometimes respond in a way that is inconsistent with it (Barnes-Holmes et al., 2008). This is achieved by presenting blocks of trials in which the participants must agree with a verbal rule and blocks of trials in which they must agree with a second verbal rule that is oppositional to the first rule (Barnes-Holmes et al., 2008). Within each discrete trial, the participant is also required to either affirm or negate whether the presented stimuli conform to whichever rule is then in place (Hughes & Barnes-Holmes, 2011). Furthermore, throughout each testing block, there is a requirement for the participants to respond quickly and accurately with feedback provided for slow or incorrect responses.

Since its development in 2004 (Barnes-Holmes & Hussey, 2015), the IRAP has been used to evoke BIRRs to a broad range of verbal and pictorial stimuli in both experimental and applied settings. Among the attitude objects that have been targeted, the IRAP has been used

to access implicit attitudes to work and leisure (Chan et al., 2009), racial stereotyping (Barnes-Holmes et al., 2010), religion (Drake et al., 2010), children and adults as sexual beings (Dawson, Barnes-Holmes, Gresswell, Hart, & Gore, 2009), meat and vegetables (Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, 2010), gender (Drake et al., 2010), race (Drake et al., 2010; Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), spiders (Nicholson & Barnes-Holmes, 2012), the elderly (Cullen, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), young people (Cullen et al., 2009), obesity (Drake et al., 2010), self-esteem (Vahey et al., 2009), smoking (Vahey, Boles, & Barnes-Holmes, 2010), depression (Hussey & Barnes-Holmes, 2012), forgiveness (Bast & Barnes-Holmes, 2014), and verbal stimuli the participants would not have encountered prior to taking the assessment procedure (Hughes & Barnes-Holmes, 2011). Because the rest of this Master's thesis will feature a series of scientific reports on the experimental application of the IRAP, the minutiae of procedure's construction and application will not be discussed here. Suffice it say however, that there is a growing body of research that supports the use of the IRAP as a valid measure of implicit evaluative cognition.

1.7 Chapter Summary

This chapter was intended to provide a broad overview of scientific research into human evaluative behaviour. It began with a discussion on some of the more popular definitions and interpretations of the attitude construct before explicit and implicit attitudes were described alongside discussions on the measures used to access and record such evaluations and the particular complexities inherent in doing so. Following the presentation of a range of cognitivist models and theories, I proposed that attitudes can be considered as types of behaviour in and of themselves. To support this premise, the ensuing sub-sections

presented the related behavioural phenomena of stimulus equivalence and arbitrarily-applicable relational responding along with an overview of RFT (Hayes et al., 2001) and the REC Model (Barnes-Holmes et al., 2011; Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). In the previous two sections, there were discussions on two of the behaviour-analytic measures that influenced the featured methodology of this research project before a brief mention was made of the IRAP itself.

Chapter 2

Experiment 1: A Replication of Hughes and Barnes-Holmes's (2011) Study of Induced Implicit Attitudes

2.1 Introduction

The Implicit Relational Assessment Procedure (IRAP) is a computer-based measure of differential relational responding which records how long participants take to relate textual or pictorial stimuli presented on a screen or monitor. As Hughes and Barnes-Holmes (2013, p. 102) report, the IRAP evokes and records "brief and immediate relational responses" (BIRRs) on the assumption that each test user has a history of relational responding and that this history can be accessed by getting the test user to react both in accordance and in disagreement with it. The procedure contains trial blocks in which the test-user is asked to respond in a way that is consistent with what they have learned in the past and trial blocks in which they are asked to respond in way that is inconsistent with their previously acquired relational networks (Barnes-Holmes et al., 2008; Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). There is also the opportunity within each individual trial for the participant to either affirm or negate what is being presented to them in accordance with the particular rule at play (Hughes & Barnes-Holmes, 2011). Furthermore, to make sure that only BIRRS are accessed, the IRAP prompts the participants to react quickly and accurately and provides feedback when they either take too long or produce an incorrect response.

The requirement for each test-user to hastily produce the desired response within these alternating circumstances creates the conditions that result in the automatic cognition that is associated with brief-and-immediate relational responding (Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2013). In addition to generating implicit cognitive-

behaviour as a procedural outcome, the IRAP's response-demands produce multiple patterns of responding that differ in subtle ways. For example, it is suggested that test users react faster on trials in which they have to respond in a way that is consistent with what they have learnt than on trials in which they have to respond in way that is inconsistent with their learning history (Barnes-Holmes et al., 2008). Furthermore, it is also possible that test users react faster on trials in which they are asked to affirm a rule than on those trials in which they are required to negate it (Hughes & Barnes-Holmes, 2011). Accordingly, the principal outcome measure of the IRAP is the difference in mean response times between either consistent and inconsistent trial blocks (Golijani-Moghaddam et al., 2013) or the different trial types, with any bias in responding being considered an 'IRAP effect' (Barnes-Holmes et al., 2008; Hughes & Barnes-Holmes, 2011; Hussey, Thompson, McEnteggart, Barnes-Holmes, & Barnes-Holmes, 2014). Furthermore, it is suggested that the presence of an implicit attitude (or at least implicit cognitive-behaviour) can then be inferred from the presence of such an effect (Hughes & Barnes-Holmes, 2011; Hughes & Barnes-Holmes, 2013; Hughes et al., 2012).

The IRAP is a methodology that is based on both the core assumptions of Relational Frame Theory (RFT; Barnes-Holmes et al., 2006; Hayes et al., 2001; Hughes et al., 2011; Vahey et al., 2009) and the functional interpretation of the attitude construct (Golijani-Moghaddam et al., 2013; Hughes & Barnes-Holmes, 2013; Hughes et al., 2011; Hughes et al., 2012). Accordingly, the cognitive-behavioural responses evoked by the procedure are considered 'acts-in-context' (Biglan & Hayes, 1996) that can be described and explained in purely behaviour-analytic terms. These 'acts' are arbitrarily-applicable relational responses in which the test user reacts to the stimuli that are presented to them by selecting the response option that best describes the relation between the stimuli (Barnes-Holmes et al., 2006). The IRAP promotes (and evokes) arbitrarily-applicable relational responding due to the

participant needing to place the stimuli within relational frames in order to progress through the assessment procedure. These relational frames may take any form (for example, they may be frames of co-ordination, opposition, or difference; Hayes et al., 2001); however, the relational nature of the frames is ultimately decided upon by the experimenter and prompted by the procedure itself (Hughes et al., 2011). Furthermore, because the IRAP measures implicit cognitive events in functional and not mentalistic terms, it is not used to measure the strength of hypothetical mental associations and so the spatial and temporal features prevalent in the associative measures of implicit cognition (such as the various procedures that use evaluative and semantic priming; Hughes et al., 2011; Krosnick et al., 2005; Wittenbrink, 2007; Wittenbrink & Schwarz, 2007) are absent in the procedure. Rather, the IRAP presents the sample stimuli, target stimuli, and response options simultaneously in order to evoke a prompted relational response using verbal behaviour. The presence (or absence) of an implicit attitude is then inferred from the difference in mean response times between the consistent and inconsistent trial blocks (Barnes-Holmes et al., 2008).

The first use of the IRAP involved recording BIRRs towards a range of everyday verbal stimuli. Participants were presented with the words "Pleasant" and "Unpleasant" as sample stimuli, the words "Similar" and "Opposite" as response options, and – as target stimuli – six words that are normatively considered to be *pleasant* in meaning and six words considered to be *unpleasant* (Barnes-Holmes et al., 2006). Following two practice blocks of 24 trials, there were alternating consistent and inconsistent trial blocks (of 24 trials apiece) and an even distribution of trials in which the proposed stimulus relation needed to be either affirmed or negated in accordance with whichever one of two verbal rules was at play (with the two rules being: (1) that *pleasant* target stimuli are "Similar" to the sample stimulus "Pleasant" and *unpleasant* target stimuli are "Similar" to the sample stimulus "Unpleasant" on consistent trials, and, (2) that *pleasant* target stimuli are "Similar" to the sample stimulus

"Unpleasant" and *unpleasant* target stimuli are "Similar" to the sample stimulus "Pleasant" on inconsistent trials with the combinations of stimuli featuring the response option "Opposite" being the reverse of these stated relations for both rules). Multiple 'effects' were observed in the first IRAP study, as the participants took less time to respond in accordance with their learning histories than it took for them to respond in disagreement with them (Barnes-Holmes et al., 2006). In other words, when asked to assert that *pleasant* words were indeed "Pleasant" and unpleasant words were indeed "Unpleasant", the participants did so more quickly than when they were asked to assert that the opposite was true (Barnes-Holmes et al., 2006). Therefore, from the results of Barnes-Holmes et al.'s (2006) research, it is possible to infer that: (1) the *pleasant* and *unpleasant* target stimuli were more readily in frames of co-ordination with consistent sample stimuli than with inconsistent (and oppositional) stimuli, and that, (2) as a group, the participants held implicit attitudes towards the verbal stimuli that could be considered normative.

Not long after the first significant 'IRAP effects' were recorded in an experimental setting, research was undertaken to see whether differential relational responding still occurs when the participants are made aware of the demand characteristics of the methodology (McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart; 2007). Participants were given either explanations of why 'IRAP effects' happen or detailed strategies on how to fake the IRAP. Neither manipulation influenced their subsequent behaviour, however, and despite being told what was expected of them and how to respond differently, the participants still reacted faster on consistent trials than on inconsistent trials (McKenna et al., 2007). This shows that even being told the rule at play within each trial block fails to influence the participants' brief-and-immediate relational responding. Indeed, the results of McKenna et al.'s (2007) study suggest that the IRAP is a measure of implicit attitudes free from desirability bias or any form of impression management (Hughes et al., 2011), an assertion that has been built on by other

studies that have provided empirical support for the IRAP as a reliable and valid 'indirect' measure (Golijani-Moghaddam et al., 2013).

As well as recording well-established patterns of responding to combinations of familiar (and meaningful) stimuli, the IRAP can be used to train and test new relational responses and hence new implicit attitudes. One way this can be done is by altering a test-user's existing relational network by exposing them to novel relational frames (Hughes et al., 2012). For example, an individual who demonstrates biased responding against a particular attitude object may have that bias reduced through exposure to the disfavoured object in a frame of co-ordination with a favoured stimulus. This type of manipulation occurs in Cullen et al.'s (2009) study in which participants who were shown pictures of well-known and well-respected elderly people then demonstrated less ageist attitudes towards the sample stimulus "Old People" (p. 597). However, the alteration of an individual's prior history of learning is not an easily controlled process as it is difficult to account for every single relational frame within their network of stimulus relations (Hughes et al., 2012). Accordingly, the formation of implicit attitudes (and new relational responses) is best demonstrated when participants are exposed to things that they have had no opportunity to encounter in the past, something that requires researchers to create novel pictorial or textual stimuli with no obvious similarity to any pre-existing attitude objects.

There has been little work undertaken to access newly engineered implicit attitudes in an experimental setting; yet this is exactly what Hughes and Barnes-Holmes (2011) did in their study of the "formation and persistence of implicit attitudes" (p. 391). Their experiment involved the use of the IRAP and written materials to engineer the sort of differential relational responding from which the presence of implicit attitudes can be inferred (Hughes & Barnes-Holmes, 2011). In order to acquire these inferred implicit attitudes, participants were exposed to two novel stimuli (the words "Cug" and "Vek") in one of three attitude-induction

conditions. In the first condition, a training IRAP provided feedback that shaped the participants' brief-and-immediate relational responding over multiple discrete trials. In the second condition, a written narrative presented the two potential attitude objects (the words "Cug" and "Vek") as equivalent to a series of either positive or negative trait adjectives. In the third condition, participants were presented with both the training IRAP and the written narrative. All three of these experimental manipulations generated implicit attitudes of varying strengths; yet with continued exposure to the testing IRAP, the newly engineered attitudes weakened over time (Hughes & Barnes-Holmes, 2011). Accordingly, the aim of my first experiment was to see whether such a weakening occurred again during a strict replication of the original study.

2.2 Method

In order to replicate Hughes and Barnes-Holmes's (2011) experiment as strictly as possible, I contacted the original study's authors to seek advice on administering the IRAP and to confirm any experimental details not overtly discussed in their published paper. As in the original study, the sample stimuli presented to the participants in Experiment 1 were the nonsense words "Cug" and "Vek" and the response options were the words "Opposite" and "Similar". Furthermore, the IRAP was administered in Experiment 1 using similar response-option parameters to those used in the earlier experiment and exactly the same mastery, response-latency, and accuracy criteria as those reported by Hughes and Barnes-Holmes. However, because Barnes-Holmes could not locate the experimental materials (personal communication, May 11, 2015), my participants were exposed to a different written narrative, a different set of target stimuli, and a different set of written instructions to the corresponding items that were presented to the participants in the original study.

2.2.1 Participants

The participants (n = 12) were all undergraduate students enrolled in the School of Psychology on the University of Waikato's Hamilton campus. They were recruited via a posting on the School of Psychology's online news forum and through direct contact via email (for those students who had previously expressed an interest in being involved in post-graduate psychological research). Because the *D*-IRAP transformation (which is the first part of the statistical process for determining the presence of an 'IRAP effect'; Drake et al., 2010; Hughes & Barnes-Holmes, 2011; Hussey, Barnes-Holmes, & Barnes-Holmes, 2015) accounts for individual differences in age, cognitive ability, and motor skills (Hughes & Barnes-Holmes, 2011) no demographic details were taken. However, English was not a first language for two of the participants. All of the participants undertook Experiment 1 in the same room in the School of Psychology. The participants were asked on the consent form whether they wished to be contacted about the results of the research and those that did were presented with a summary of the experimental findings via email. Ethical approval was granted by the School of Psychology Research and Ethics Committee (#15:64).

2.2.2 Materials and Equipment

2.2.2.1 'Indirect' Training and Testing Procedures

The IRAP was used to both engineer implicit attitudes in the contingency-learning and combined-learning conditions and to access and record brief-and-immediate relational responding by all participants in all three conditions. In each trial of the IRAP the participant was presented with a sample stimulus (which was either the word "Cug" or the word "Vek"),

a target stimulus (which was any one of 12 positive or negative trait adjectives or their equivalent nouns), and a series of response options (the words “Similar” and “Opposite”; see Figure 2.1 for a 4-picture representation of the above description; the various verbal stimuli are presented in Figure 2.2). For each trial, the participant was required to attend to the sample and target stimuli and then select one of the two response options by pressing either the “d” key on a keyboard with the index or middle finger of the left hand or the “k” key with the index or middle finger of the right hand. If the participant’s response was incorrect they were shown a red “X”, which was presented in the middle of the screen to prompt the participant to make a different response. If the correct response was given, the screen was cleared for 400 ms before the next combination of stimuli was presented. In order for the programme to test implicit relational responding that was brief and immediate (Hughes & Barnes-Holmes, 2013), there was a time restriction for each response (with a maximum median latency of 3,000 ms across each trial block as per the original experiment; Hughes & Barnes-Holmes, 2011). Accordingly, if a participant responded too slowly after the first pair of practice blocks in the testing procedure or after the first pair of trial blocks in the training procedure, they were met with a custom message which read “Too slow!” The need to respond quickly was also stated in all the participants’ written instructions.

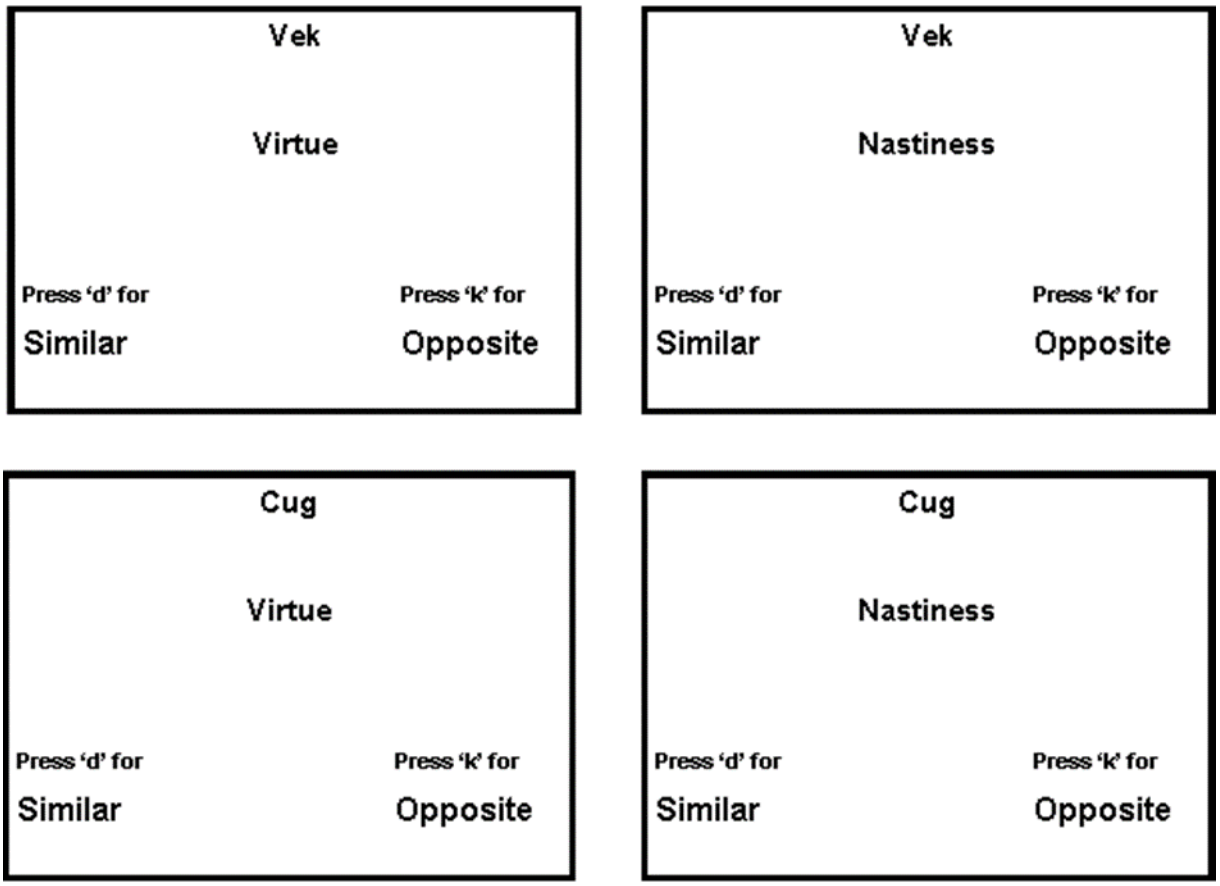

Figure 2.1. Example screen presentations of the four different trial types that occur within the IRAP's testing programme. The two sample stimuli are the words "Vek" and "Cug"; the words "Virtue" and "Nastiness" are two of the twelve positively or negatively valenced target words that were used in the testing programme; and the words "Similar" and "Opposite" represent the two available response options (with the nearby prompting phrases used to instruct the participant which key to press).

To determine whether a participant had acquired the desired BIRRs, a required accuracy level of 95% for the training blocks and 80% for the testing blocks was implemented for each application of the IRAP. Furthermore, participants in all three conditions were given between one and four pairs of practice blocks to help them master relational responding in accordance with whichever verbal rule was in place. These practice blocks were presented as part of the testing IRAP (regardless of condition) and were particularly important for the participants who underwent verbal instruction as they had received no other exposure to the IRAP programme and no previous exposure to inconsistent trial blocks (in which they were expected to respond in a way that contradicted what they

knew; Barnes-Holmes et al., 2006). Within all three experimental conditions, an incorrect response in the testing IRAP resulted in the presentation of a red "X" and the opportunity to respond differently; however, in the two applications of the training programme, the correct relational term was also shown alongside the red "X" so that the participant would be prompted to respond differently *and* be reminded of the verbal rule at play. Failure to meet the accuracy and response-time criteria in any one of the practice and training blocks produced an inter-block screen which displayed the participant's accuracy score and median-response time for the previous trial block as well as a reminder of what the various response criteria were. Failure to meet the response criteria within any given practice or training block also meant that the participant had to take a further pair of consistent and inconsistent trial blocks (in the practice phase of the testing programme) or a single consistent block (in the training programme) until a total of eight blocks had been taken; and if the participant failed to meet the response criteria by the last training block or the last pair of practice blocks they were thanked for their involvement and their data were removed from the later analysis. Of the 12 participants in Experiment 1, only one failed to meet the various response demands of the practice phase of the testing programme.

Table 2.1
The Verbal Stimuli Presented On-screen during the Training and Testing IRAPs

	Sample Stimuli	Target Stimuli	Response Options
Relational Training Procedure (IRAP)	Cug, Vek	**Positively valenced trait adjectives:** virtuous, wise, good, peaceful, caring, generous **Negatively valenced trait adjectives:** nasty, ignorant, evil, violent, cruel, selfish	Similar, Opposite
Relational Testing Procedure (IRAP)	Cug, Vek	**Positively valenced nouns:** virtue, wisdom, goodness, peace, kindness, generosity **Negatively valenced nouns:** nastiness, ignorance, evil, violence, cruelty, selfishness	Similar, Opposite

In both Experiment 1 and Experiment 2, the same two words were used as sample stimuli. The target words were altered between training and testing procedures (the use of adjectives in the training IRAP ensured the target stimuli in the computer-based assessment were the same as the comparison stimuli in the written narrative of the verbal-instruction condition). The same two relational response options appeared in all applications of the IRAP, although these relational terms swapped positions at random throughout each procedure to avoid any form of fixed responding. In order for the participants to acquire the rule that certain target words were related to a sample word in a particular way and that other target words were related to another sample word in a particular way, the IRAP programme presented multiple combinations of the three types of stimuli before providing – contingent on the participant's response – either progress to the next trial or negative feedback. As Hughes and Barnes-Holmes (2011) assert, it takes 48 trials for each of the 12 target stimuli to be presented with each of the two sample stimulus and response options at least once (accordingly, 48 trials equal a single trial block). There are four different ways to combine the different types of stimuli. These four ways can be stated using capital letters to denote the sample stimuli, *equals* signs and *not-equals* signs to denote the two relational response terms,

and lowercase letters to denote the various target stimuli: trial-type 1: *A* = {abcdef} / trial-type 2: *A* ≠ {ghijkl} / trial-type 3: *B* = {ghijkl} / trial-type 4: *B* ≠ {abcdef}. As in Hughes and Barnes-Holmes's study, the sample stimuli were "Cug" and "Vek", the response options were the words "Similar" and "Opposite", and the target stimuli were either positive or negative trait adjectives or their equivalent nouns (depending on the type of IRAP procedure being run). So, using only two of the 12 target stimuli (the oppositional terms "Virtue" and "Nastiness") by way of an example, the four trial types are shown in Figure 2.1. Each of the trials places one of the sample stimuli (e.g. "Vek") within a frame of co-ordination or opposition with a target stimulus (e.g. "Virtue") in accordance with the rule at play.

In order to induce biased relational responding in the contingency and combined-learning conditions, I used an IRAP training procedure. The training IRAP contained only consistent trial blocks on the assumption that exposure to inconsistent trial blocks during the training phase inhibits differential responding (Hughes & Barnes-Holmes, 2011). To further ensure that the desired BIRRs are engineered, the training IRAP can be adjusted so that near-perfect relational responding is required across multiple trials (as was the case in Hughes & Barnes-Holmes's, 2011, study). The training programme also provides within-trial and between-block prompts for accurate and fast responding and negative feedback for incorrect relating and the failure to attain the various criteria for an entire trial block. I set the training programme's maximum median latency at 3,000 ms for each trial block and the response-option parameter at a maximum hold of three trials. The last variable refers to how frequently the relational response terms switch sides across each trial block, which is an important specification that has been shown to influence the size of any 'IRAP effect' with static placements of the relational terms producing weaker differential responding (Campbell, Barnes-Holmes, Barnes-Holmes, & Stewart, 2011).

Finally, I used IRAP software which was conceived, designed, built, extended, and

updated by Dermot Barnes-Holmes between 2004 and 2012 (Barnes-Holmes & Hussey, 2015). The particular programme was a 2012 version available for download from an IRAP research-website that no longer exists (retrieved from http://IRAPresearch.org in January, 2015).

2.2.2.2 'Direct' Verbal-Instruction Procedure

The participants in the verbal-instruction and combined-learning conditions were exposed to a written narrative which offered a fictional account of the use of the words "Cug" and "Vek" by the speakers of an ancient language which is no longer used. An individual with a set of six particular attributes was described as a "Cug" by speakers of the ancient language whereas an individual with an oppositional set of six attributes was called a "Vek". Using a counterbalanced structure, graphic imagery (Hughes & Barnes-Holmes, 2011), and repetition, the narrative presented the reader with contrasting descriptions of the two different types of people; furthermore, the words "Cug" and "Vek" were presented frequently throughout the narrative, meaning that the reader was given multiple opportunities to learn to relate the two novel words with a certain series of familiar attributes. To provide counterbalancing across conditions, half of the participants in the verbal-instruction and combined-learning conditions were provided with a written narrative in which the word "Cug" was used to describe an individual with positive attributes and the word "Vek" was used to describe an individual with negative attributes; the other half were provided with a narrative in which the attributes were reversed.

The written narrative is as follows (with the word "Cug" used to describe a person with positive attributes by way of example only):

> There is an ancient language which has long since fallen from use in the region where it was once spoken. In this ancient language there are two

> words used to describe two very different sorts of people. The word "Cug" was used to describe a person who was virtuous, wise, and good; a "Cug" was a peaceful person who would never beat his slaves or his family members; a "Cug" was caring and generous, even to his enemies. Indeed, a "Cug" was a good person to have around, not least because they were caring, generous, good, wise, virtuous, and peaceful! However, in those distant times not everyone was a "Cug": the ancient language also has a word which was used to reference a very different sort of person. The word "Vek" was used to describe an individual who was nasty and ignorant; unlike a "Cug" a "Vek" *would* be violent towards his slaves and family; a "Vek" would behave in ways considered evil, even in the savage times in which the ancient language was spoken; a "Vek" was known for being cruel and selfish (even to a "Cug"), and if anyone was ever in need, a "Vek" would not do a single thing to help them. Indeed, a "Vek" was a bad person to have around, not least because they were cruel, violent, evil, ignorant, nasty, and selfish!

Having read the written narrative, the participants were asked to complete two memory checks (Hughes & Barnes-Holmes, 2011) to confirm that they had learned to relate the words "Cug" and "Vek" with the correct trait adjectives. In the first check, participants indicated whether they believed the words "Cug" and "Vek" to be either "positive" or "negative". In the second check, participants wrote three words that could be used in place of "Cug" and three words that could be used in place of "Vek". Having completed these tasks, the participant summoned the experimenter. If the participant had failure to respond to the tasks in a way that reflected the two verbal rules they were exposed to, the experimenter asked the participant to re-read the narrative and complete the memory checks again (however, no participants were required to do so).

2.2.2.3 Post-experiment Self-Report Measures

Having taken the testing IRAP, the participants in all three conditions were then asked to complete three pen-and-paper self-report measures, which were a set of feeling thermometers, two semantic-differential scales, and a meaningfulness scale. Each of the two feeling thermometers had an 11-point scale on its right-hand side numbered from 0˚ to 100˚

in 10˚ intervals. The word “Cug” was printed above one thermometer and the word “Vek” above the other. Underneath the thermometers were printed instructions asking the participants to draw a line across each one at a point roughly in accordance with how they felt about the word printed above it. Unlike the thermometers used by Hughes and Barnes-Holmes (2011), there was no reference to 0˚ being ‘cold’ and 100˚ being ‘hot’. Accordingly, the demands of this measure proved to be unclear to some of the participants (who asked for further instruction from the experimenter). The semantic-differential scales were 7-point Likert scales that were numbered from 1 (located just to the right of the word “negative”) on the left-hand side of the scale through to 7 (which was just to the left of the word “positive”) on the right-hand side. Each scale had either the word “Cug” or the word “Vek” printed above it and instructions for the participants to circle the number that best represented how they felt about each word. The final self-report measure was a meaningfulness scale, a 7-point Likert scale that was numbered from 1 (which was just to the right of the phrase “Not at all”) on the left-hand side of the scale through to 7 (located just to the left of the word “Extremely”) on the right-hand side. Printed instructions underneath the scale requested that the participant circle the number that best reflected how meaningful they believed their attitudes towards the words “Cug” and “Vek” were. As was the case in Hughes and Barnes-Holmes’s (2011) experiment, the feeling thermometers and semantic-differential scales were used to determine the participants’ explicit attitudes towards the previously novel stimuli and the meaningfulness scale was used to establish how ‘real’ the participants’ acquired attitudes ‘felt’ to them.

2.2.3 Experimental Procedures

Each participant was asked to attend an individual experimental session in a room in

the School of Psychology at the University of Waikato. Upon arrival, the participant was greeted and presented with a consent form and a copy of the information sheet (which had also been sent to the participants prior to the experiment via email; see Appendix A). When the participant had read and signed the consent form they were then presented with the written instructions for their condition and advised to ask questions about anything they did not understand. The specific procedures are described in the following sections.

2.2.3.1 Contingency-Learning Condition

In the contingency-learning condition, the participants (n = 4) were immediately presented with a training IRAP in which there was only one rule to be followed (which was that the words "Cug" and "Vek" were either "Opposite" or "Similar" to any one of six positive or negative trait adjectives). The instructions on how to successfully complete the training programme were contained in the participants' written instructions and read as follows:

> You are about to take a computer-based learning task which involves reading combinations of two words that are presented simultaneously on the screen. Your task is to state whether the words are similar in meaning or opposite.
>
> For each trial you will need to look at the word at the top of the screen, the word presented below it, and then the words "Similar" and "Opposite" which will be presented horizontally at the bottom of the screen. Using either the "d" key or the "k" key you must then choose one of the two words (or response options) that are presented at the bottom of the screen.
>
> When making your response, use either the index or middle finger of your left hand to press the "d" key and either the index or middle finger of your right hand to press the "k" key.
>
> The response options "Similar" and "Opposite" will not remain in the same place on the screen for each trial: these words will swap places in a way that cannot be predicted.

> In order for you to relate the various words presented simultaneously above the two response-words the computer programme will provide feedback by telling you if you have made an incorrect response. An incorrect response is indicated by the presentation of a red cross.
>
> **It is important that you make few errors and respond as quickly as you can. To do this, try to focus on accurate responding first and foremost; when you are giving a correct response on the majority of the trials, try to speed up your responding until you taking as little time to relate the words as possible.**
>
> Go well!
>
> Please press the space bar to continue with the learning task and summon the experimenter when prompted to do so.

When the participant had read the instructions and pressed the space bar they were presented with an on-screen statement which advised them that an incorrect response would result in a red cross being shown and that only a correct response would allow them to proceed to the next trial. Having once again pressed the space bar, the participant was then presented with the first combination of stimuli, which featured a sample stimulus (which was either the word "Cug" or the word "Vek"), one of the trait adjectives as a target stimulus, and the two response options including an instruction as to which key to press to select them (see Figure 2.1 for how these stimuli were positioned on-screen). The participant then had to press one of the two response keys without knowing which one produced the correct response. However, as the participant proceeded through the blocks of training trials, feedback was given in accordance with the particular rule at play. There were between one and eight training blocks available to the participants. If a participant did not respond in a way that met the required response criteria by the end of the eighth trial block they were thanked for their involvement and presented with a course-credit slip. Participants who successfully completed the training IRAP were asked to read the next set of written instructions while the test programme was loaded onto the laptop. The instructions for completing the test IRAP read as follows:

You are about to take a computer-based test that is similar in appearance to the computer-based learning task that you took earlier.

As before, the computer programme will ask you to follow a rule which will become apparent through the feedback that is given following each incorrect response.

At some points in the following test the feedback provided will make sense to you; at other points, the feedback will contradict what you have previously learnt. This is part of the task and you will be required to respond in different ways.

It is important that you make few errors and respond as quickly as you can. To do this, try to focus on accurate responding first and foremost; when you are giving a correct response in the majority of the trials, try to speed up your responding until you taking as little time to relate the words as possible.

Go well!

Please press the space bar to continue with the test and summon the experimenter when prompted to do so.

To expose the participants to the trial blocks in which they were asked to respond in way that was in disagreement with what they had earlier learnt, the testing IRAP contained a series of practice blocks in which feedback was provided for both consistent and inconsistent responding (across alternating blocks of 48 trials). This was necessary because without such an exposure the participants in the contingency-learning condition would have had no prior experience of responding in a way that contradicted what they had learnt during the training procedure (and so would have responded poorly on the inconsistent trial blocks during the testing phase). If a participant had failed to meet the response criteria during the practice phase, they would have been thanked and their involvement in the experiment would have been finished; however, no participants in this condition failed to successfully complete the practice blocks within the allotted amount. When the participants finished the testing IRAP they were asked to complete the three self-report measures before being thanked for their participation and reminded of their next session time.

2.2.3.2 Verbal-Instruction Condition

Participants (n = 4) in the verbal-instruction condition were given a printed copy of the written narrative. Having read the narrative, the participants were asked to complete a series of memory checks which the experimenter inspected to make sure the answers were congruent with the experimental manipulation. If a participant had responded to the questions in a way that suggested they had not acquired the two verbal rules present in the narrative, they would have been asked to re-read it and complete the memory check a second time (however, none of the participants were required to re-read the narrative). The final part of the experiment involved the participants taking a testing IRAP that was identical to the one that was presented to the participants in the contingency-learning condition. The same response criteria applied through the practice phase of the testing programme with one participant failing to meet the required standards for accuracy and speed (they were thanked for their involvement in the experiment and provided with a course-credit slip; the participant's data was also removed from the later analysis). When they had finished the testing IRAP, the remaining participants in the verbal-instruction condition were then asked to complete the three hard-copy self-report measures before being thanked for their participation and reminded of their next session time.

2.2.3.3 Combined-Learning Condition

Participants (n = 4) in the combined-learning condition were presented with the attitude-induction procedures that were used in the contingency-learning and verbal-instruction conditions. They were first presented with the written narrative and its associated memory checks before being asked to take the training IRAP that was used in the

contingency-learning condition and then the same testing IRAP that was used in the other two conditions. Having successfully completed the testing programme and the self-report measures, the participants were thanked and reminded of their next session time.

2.3 Results

Transformation of raw data. The IRAP programme records the time it takes for a test user to relate two or more stimuli in multiple discrete trials. As Hughes and Barnes-Holmes (2011) state, to account for individual differences, the raw latency scores can be transformed into *D*-IRAP scores (something which the IRAP software does automatically; Barnes-Holmes & Hussey, 2015). This transformation allows for test data to be analysed without the need to factor in a broad range of extraneous variables (such as age or cognitive ability; Dawson et al., 2009; Hughes & Barnes-Holmes, 2011). Furthermore, the *D*-IRAP scores can be used to produce group mean scores for each attitude-induction condition and each pair of test blocks.

Accuracy Criterion. In order to confirm that the participants were responding in accordance with the rule at play, the accuracy criterion was set at 80% for the testing IRAPs used in all three conditions. However, three of the participants (one in the Verbal-Instruction Group and two in the Combined-Learning Group) attained accuracy scores just below 80% in some of their trial blocks (overall, there were five trial blocks in which the minimum-accuracy criterion was not met: three consistent trial blocks and two inconsistent trial blocks). Accordingly, the minimum-accuracy level was lowered to 79% after the event with any scores below that being discarded from the analysis (the data from one pair of trial blocks in the verbal-instruction condition were removed on this basis).

Formation and persistence of differential relational responding. As Figure 2.2

indicates, the three attitude-induction conditions produced differential relational responding with the participants reacting faster on consistent trials than on inconsistent trials. A 3 (contingency-learning, verbal-instruction, combined-learning) x 3 (1st, 2nd, and 3rd pair of test blocks) repeated-measures ANOVA indicated that there was no significant effect for attitude-induction condition, $F(2, 2) = .04$, $p = .96$, and no significant effect for the pair of test blocks, $F(2,2) = 14.78$, $p = .06$; furthermore, there was no significant interaction between the pair of test blocks and the attitude-induction condition, $F(4, 4) = .18$, $p = .9$. The absence of any significant effect may be due to the small number of participants and the substantial amount of variance within the group data (particularly in the 2nd pair of trial blocks in the verbal-instruction condition). In order to account for these limitations and any missing values (for those trial blocks in which the accuracy criterion was not met), the data for the 1st and 2nd trial blocks were combined in order to create two levels: (1) trial blocks 1 to 4 ('early' trial blocks), and, (2) trial blocks 5 & 6 ('late' trial blocks). Accordingly, a 3 (contingency-learning, verbal-instruction, combined-learning) x 2 ('early' trial blocks, 'late' trial blocks) mixed repeated-measures ANOVA was conducted which indicated that there was no significant effect for attitude-induction condition, $F(2, 8) = .38$, $p = .70$, $r = .29$, no significant effect for time/trial block, $F(1, 8) = .003$, $p = .96$, $r = 0$, and no significant interaction between attitude-induction condition and time/trial block, $F(2, 8) = 1.23$, $p = .34$, $r = .48$. The observed power was also very low for the attitude-induction (.092) and time/trial block (.05) effects, meaning that if there *were* real effects due to these manipulations they would only be detected 9% and 5% of the time, respectively.

One of the aims of Experiment 1 was to see whether there was any weakening of the 'IRAP effects' across the three pairs of test blocks. A visual inspection of Figure 2.2 will indicate that there was no substantial decrease in the differential responding performed by the participants in the three attitude-induction conditions. Indeed, the Verbal-Instruction Group

produced stronger 'IRAP effects' as the testing IRAP progressed. Furthermore, the 3 x 2 mixed repeated-measures ANOVA produced a series of estimated marginal means that show a decrease in the *D*-IRAP scores over time for the contingency-learning and combined-learning conditions but an increase in the *D*-IRAP scores for the verbal-instruction condition.

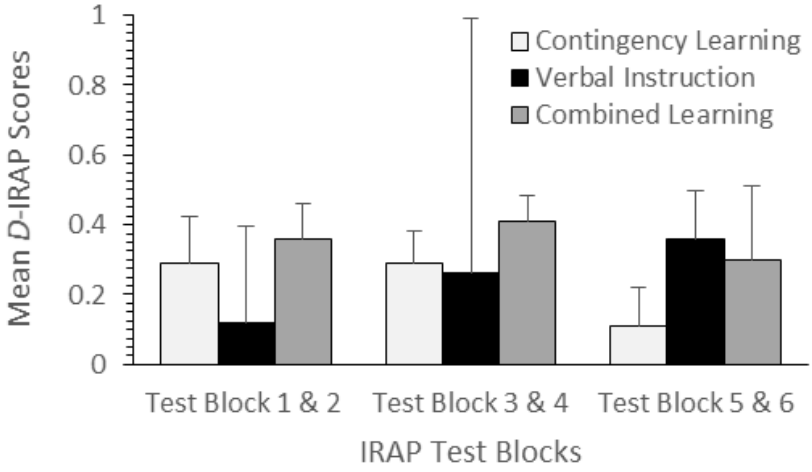

Figure 2.2. The mean *D*-IRAP scores for the three attitude-induction conditions in Experiment 1. If the mean *D*-IRAP score is above or below zero, then there was a bias in participant responding that may represent the presence of an implicit attitude. Standard error bars reflect the standard error of the mean.

Affirmation and negation trial types. As part of their analysis, Hughes and Barnes-Holmes (2011) compared the mean *D*-IRAP scores of the trials in which the participants were asked to affirm the proposed stimulus relation within a particular trial and the mean *D*-IRAP scores of the trials in which the participants were asked to negate the proposed stimulus relation. As Figure 2.3 portrays, there was little difference in the mean *D*-IRAP scores in Experiment 1; indeed, a one-way repeated-measures ANOVA indicated that there was no significant effect for trial type, $F(1, 10) = .64$, $p = .44$.

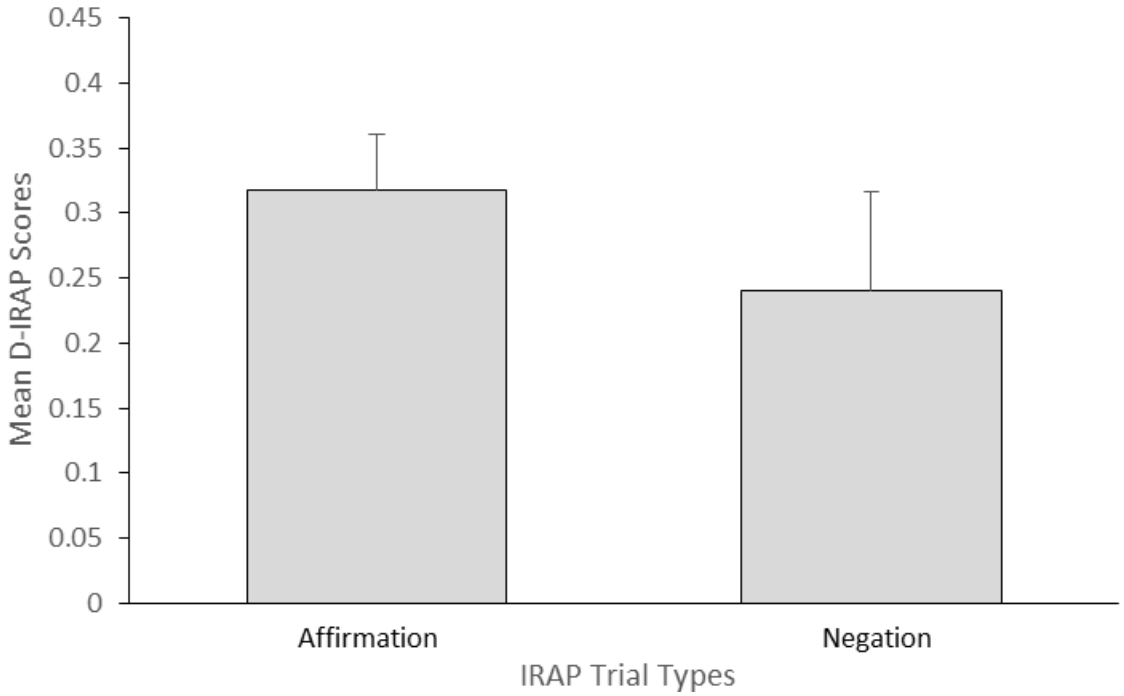

Figure 2.3. The mean *D*-IRAP scores for the trial types in which the participants were required to affirm the rule at play and the mean *D*-IRAP scores for the trials in which they were required to negate it. Standard error bars reflect the standard error of the mean.

Explicit Attitudes. Two of the participants in Experiment 1 failed to respond on the post-test feeling thermometers in a way that was consistent with the attitude-induction process they had undergone. These participants were both in the contingency-learning condition. In their study, Hughes and Barnes-Holmes (2011, p. 402) discarded the data for participants who "failed to produce self-reported attitudes that were consistent with the induction procedures". However, given the limited number of participants in my experiment, the two participants' IRAP data was retained in order for analyses of variance to be conducted. The data was also retained for a single participant in the combined-learning condition who rated the sample stimuli "Cug" and "Vek" equally on both the feeling thermometers and the semantic-differential scales. Furthermore, one participant in the verbal-instruction condition took their written instructions home so their self-report data was lost. Using the data produced by the feeling thermometers, a 2 (word valence) x 3 (contingency-learning, verbal-instruction, combined-learning) mixed repeated-measures ANOVA indicated

that there was no significant effect for valence, $F(1, 2) = 3.15$, $p = .22$, and induction condition, $F(2, 4) = .42$, $p = .68$, and no significant interaction between induction condition and word valence, $F(2, 4) = 2.41$, $p = .28$. A one-way repeated-measures ANOVA with the data from the induction-conditions collapsed indicated that there was a significant effect for word valence, $F(1, 10) = 7.14$, $p = .023$, which suggests that, despite there being no significant effect for the method used, exposure to any one of the attitude-induction procedures tended to produce an explicit evaluation.

In the ten pairs of semantic differential scales that were retained by the experimenter, 80% of the participants in Experiment 1 assigned a score at the midpoint or higher for the sample stimulus that had been related to positive trait adjectives or nouns within their attitude-induction procedure. Conversely, 60% of the participants assigned a score at the midpoint or lower for the sample stimulus that had been related to negative trait adjectives or nouns within their attitude-induction procedure.

Only 36% of the participants rated the meaningfulness of their attitudes towards the words “Cug” and “Vek” at or above the midpoint of the meaningfulness scale. However, in Hughes and Barnes-Holmes’s (2011, p. 402) experiment, 71% of the participants rated the “perceived realism” of their attitudes at or above the midpoint (indicating that there may have been an important difference in how the two self-report measures and their accompanying instructions were worded). Despite there being a limited number of participants within each group, there was an interesting trend in the data from the meaningfulness scales: in the contingency-learning condition, 25% of the participants rated the meaningfulness of their attitudes at or above the midpoint, with 37% of participants in the verbal-instruction condition and 50% of the participants in the combined-learning condition responding in a similar manner.

2.4 Discussion

The aim of Experiment 1 was to see if there was a weakening in the participants' differential relational responding that was similar to the attenuation observed in Hughes and Barnes-Holmes's (2011) study. In order to replicate the conditions in which such a weakening could occur, it was first necessary to evoke biased relational responding using a series of attitude-induction procedures. As Figure 2.2 shows, the contingency-learning, verbal-instruction, and combined-learning procedures produced 'IRAP effects' across all three pairs of trial blocks. These 'IRAP effects' occur whenever a mean *D*-IRAP score is above or below zero, with a positive value indicating faster responding during consistent trial blocks and a negative value indicating faster responding during inconsistent trial blocks. With the mean *D*-IRAP scores ranging from 0.11 to 0.41, the 'IRAP effects' in my experiment were broadly as strong as those recorded by Hughes and Barnes-Holmes. However, there was greater variance in my data which brings into question the significance of my results. Furthermore, in my experiment there was no evident weakening in the 'IRAP effects' across the three pairs of trial blocks in all three attitude-induction conditions, an outcome that contradicts the findings of Hughes and Barnes-Holmes.

The absence of any weakening in the 'IRAP effects' indicates that there was a robustness to the implicit attitudes that were induced in my experiment (or, more accurately, there was a robustness to the differential relational responding from which the presence of such attitudes are inferred). Indeed, within my application of the testing IRAP there was an equal number of opportunities for the participants to acquire the opposite relational rule as there were opportunities for a restatement of the original rule to be deemed 'correct'. Yet the engineered response bias persisted throughout the six trial blocks of the testing IRAP despite the intermittent reinforcement of contradictory relational behaviour. A possible cognitivist

explanation for the persistence of the differential relational responding would be that once formed, each implicit attitude remained as a stable mental entity and the response bias was evidence of that entity (Eagly & Chaiken, 1993; McGuire, 1985). When considered in associative-conditioning terms, the re-occurring 'IRAP effects' may be evidence of a series of strong mental associations that supported fast relational responding even when joined by a series of secondary associations that were formed through exposure to the competing contingency (Hughes & Barnes-Holmes, 2011).

In purely functional terms, the engineered response bias may have persisted simply because the 'correct' relational behaviour (which was fast, accurate responding in accordance with the rule at play) received more reinforcement than an 'incorrect' alternative response over time (making the 'correct' behaviour more likely to occur). Furthermore, when the participants were exposed to the contingency in which the alternative relational response became part of the 'correct' behaviour, they could still produce the alternative response (and received reinforcement for doing so), but not as readily as if they had been asked to emit the original 'correct' response that was so heavily reinforced within the training phase of the experiment (in at least two of the attitude-induction conditions). However, if the 'IRAP effects' were solely a product of differential reinforcement over time, there should have been no strong response bias performed by the participants in the verbal-instruction condition (which, as Figure 2.2 indicates, was not the case), as they were exposed to identical amounts of the two contingencies and were not presented with a training IRAP. So it is possible that, in accordance with the REC Model (Hughes & Barnes-Holmes, 2013; Hughes et al., 2012), the 'IRAP effects' actually resulted from the performance of two different types of evaluative cognitive-behaviour: (1) the brief-and-immediate relational responding that was performed during the consistent trial blocks, and, (2) the elaborate and more time-consuming relational responding required in order to affirm or negate a stated relation in accordance with a verbal

rule that is a reversal of another rule.

The main limitation of my experiment was the small number of participants in each attitude-induction condition (n = 4 in each). Furthermore, there were substantially fewer participants in my experiment overall (n = 12) than in Hughes and Barnes-Holmes's (2011) study (n = 64). As well as the limited number of participants in each group, the need to remove data for the trial blocks in which the minimum-accuracy criterion was not met meant that the results could not be analysed using some tests of statistical significance. For example, it was not possible to directly compare group variance when the data for a particular group contained only a single value (as was the case in the verbal-instruction condition of my experiment). Accordingly, the results of the various ANOVAs should be considered with caution.

As a strict replication, my experiment may also be limited by the way in which the participants were exposed to a different written narrative, a different set of target stimuli, and a different set of written instructions to those used in Hughes and Barnes-Holmes's (2011) study. When it comes to replicating research on arbitrarily-applicable relational responding, similarities in experimental materials and stimuli matter because there may be significant differences between the relational responding that occurs in the presence of one word and the relational responding that occurs in the presence of another. For example, there may be stronger differential responding when sample stimuli are paired with emotionally-laden adjectives such as "delightful" than when they are paired with emotionally-neutral adjectives such as "nice". Similarly, a commonly used word such as "nice" may produce faster responding through familiarity and meaningfulness than a relatively unknown synonym such as "congenial". In order to address this limitation, a future study could look at how synonyms influence differential relational responding.

In summary, the participants in my experiment were able to relationally respond

within the competing contingencies in a way that suggested the presence of implicit attitudes. Furthermore, this experiment supports the use of the IRAP as a way of inducing biased relational responding (and hence implicit attitudes) using novel stimuli. However, within my replication of Hughes and Barnes-Holmes's (2011) experiment there was no weakening of the 'IRAP effects', which indicates a certain level of persistence in the engineered response biases. Accordingly, a second experiment was conducted in order to find out just how persistent the induced response biases were.

Chapter 3

Experiment 2: An Extension of Hughes and Barnes-Holmes's (2011) Study of Induced Implicit Attitudes

3.1 Introduction

The results of Experiment 1 indicate that differential relational responding can be engineered and then accessed using the IRAP. Furthermore, given certain key assumptions, it is also possible to infer the presence of an engineered implicit attitude from such biased responding. However, my first experiment did not see a repeat of the weakening in the 'IRAP effects' observed by Hughes and Barnes-Holmes (2011). Indeed, the lack of any attenuation in the differential relational responding suggests that the engineered biases were both persistent across time and resistant to the influence of the IRAP's alternating contingencies. Yet the testing IRAP can be administered in under 20 min, which is possibly too short a timescale to confirm with certainty the longevity and robustness of recently acquired cognitive-behaviour. Accordingly, within my thesis work there was a need for the participants' relational responding to be assessed again following an extended period in which they had not been exposed to the experimental stimuli, instructions, and materials and the testing procedure itself (including the contingencies that are present within it).

If implicit attitudes are assumed to be valid psychological constructs, their persistence raises questions about how such evaluations are stored and how resistant they are to change. Indeed, there are some who believe that implicit attitudes remain stable over time while others assert that implicit attitudes are ever-changing acts of automatic cognition (Gregg et al., 2006; Hughes et al., 2011; Schwarz, 2008). Although there is some support for the stability of implicit cognition (Hughes & Barnes-Holmes, 2011; Ranganath & Nosek, 2006),

the malleability of implicit attitudes makes logical sense when one considers the ongoing need a human has to appraise the environment in a way that is adaptable (Albarracín et al., 2005). Accordingly, an implicit attitude may not persist as some form of enduring mental quality but as the heightened probability of a particular evaluative behaviour occurring in response to a particular stimulus (Hughes et al., 2012). In other words, an implicit attitude may reoccur (i.e., persist) simply because it is a behaviour that has been performed and reinforced under broadly similar conditions in the past, which is something that may happen when implicit attitudes are trained and then tested using the IRAP.

One further explanation for the persistence of implicit attitudes is offered by the REC Model (Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). Within this account of human evaluative behaviour and cognition, implicit attitudes are defined as being behavioural acts in which stimuli are related within individual relational frames and more expansive relational networks (Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). Accordingly, the persistence of an implicit attitude may be due to the ongoing presence of the attitude object within 'online' relational framing. To illustrate this point with a fictional example, following my first experiment, one of the participants in my research may have experienced a cognitive-behavioural event in which they automatically placed the word "virtuous" in a frame of co-ordination with the word "Cug" having heard the word "virtuous" spoken at a dinner party. The same participant may then have proceeded to relate the word "Cug" with other verbal stimuli, such as "good", "similar", and "experiment". For the individual who engages in such relational responding, the implicit attitude is more likely to continue over time (and into the second experimental session). Indeed, the 'strength' of the individual's evaluation of the word "Cug" may even increase as the word enters into progressively more elaborate networks of stimulus relations and a greater number of relational frames (Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). Conversely, a

participant who, having taken part in the first experiment, did not then perform any relational responses that included the word "Cug" may have lost their implicit attitude to the sample stimulus as a result. So, in accordance with the REC interpretation of the attitude construct, the persistence of implicit attitudes may actually be the repetition of relational behaviour over time.

Having replicated Hughes and Barnes-Holmes's (2011) study in my first experiment, I undertook a second experiment after a waiting period of five days. The aims of my second experiment were (1) to find out whether the response biases that were engineered in Experiment 1 persisted beyond the time it took to conduct the first testing IRAP (in Experiment 2a), and, (2) to find out whether the persistent response biases were strengthened by further training (Experiment 2b). Accordingly, the participants that were involved in Experiment 1 were asked to attend a second session in which their implicit relational responding was assessed using a testing IRAP that was identical to the one used in the earlier experiment. The participants were then exposed to the same training procedure that they encountered during the first experiment before being presented with the same testing IRAP.

3.2 Method

In order to determine how persistent the engineered response biases were after five days of non-exposure to either the written materials or the IRAP programme, Experiment 2 began with an application of the testing IRAP to all the participants regardless of which experimental condition they were in. Having had their (experimentally unrevised) relational responding accessed and recorded in Experiment 2a, the participants were then exposed to identical written materials and identical training and testing procedures to the ones they had encountered five days earlier (Experiment 2b). In order to keep the experimental session

times to under an hour, there was no re-administration of the feeling thermometers, semantic-differential scales, and meaningfulness scales that were deployed at the end of the first experiment. However, this planned shortening of the various experimental procedures was not actually required as the participants all completed the two testing IRAPs and the intervening training in under 30 min.

3.2.1 Participants

The participants (n = 11) for Experiment 2 were the same as in Experiment 1 (less the one participant who did not meet the accuracy criterion during the practice phase of the testing IRAP in the first experiment). Upon arrival each participant was shown a copy of the consent form they had signed at the start of Experiment 1 and a copy of the information sheet before being given the written instructions for the particular condition they were in. To allow for the direct comparison of experimental data, the participants remained in the same condition they had been randomly assigned to in Experiment 1; this between-experiment consistency extended to the ordering of consistent and inconsistent trial blocks in the first administration of the IRAP, with those exposed to inconsistent trials blocks first in Experiment 1 having to respond 'inconsistently' first in Experiment 2a and those given consistent trials blocks first in Experiment 1 having to first respond in accordance with what they had earlier learnt. Furthermore, this fixed ordering of the trial blocks continued during the final administration of the IRAP (which took place after the retraining phase of the experiment).

3.2.2 Materials and Equipment

Aside from the non-administration of the three post-experimental self-report measures, in all three attitude-induction conditions, the training and testing procedures were identical to those used in Experiment 1.

3.2.3 Experimental Procedures

The participants in all three attitude-induction conditions began Experiment 2 by re-taking the testing IRAP that they had completed in Experiment 1. Having successfully completed the test (Experiment 2a), the participants were exposed to the same attitude-induction procedures that they had undergone at the first session. All three training procedures were identical to the ones used during Experiment 1.

When they had taken the attitude-induction training for the second time in 5 days, the participants in all three experimental conditions were then asked to complete a testing IRAP (Experiment 2b). This final assessment of the participants' differential relational responding was identical to both the testing programme they had taken at the start of the session and the testing programme that they had taken 5 days earlier. When they had completed the last IRAP, the participants were thanked for their involvement in the experiment and were either given a course-credit slip or a lottery ticket for a prize draw.

3.3 Results

Accuracy Criterion. In order to confirm that the participants were responding in accordance with the rule at play, the accuracy criterion was originally set at 80% for the

testing IRAPs used in all three conditions. However, four of the participants (three in the Verbal-Instruction Group and one in the Combined-Learning Group) attained accuracy scores below 80% in some of their trial blocks for Experiment 2a; furthermore, one of the participants in the Combined-Learning Group did not meet the accuracy criterion in one of the trial blocks of Experiment 2b. So in Experiment 2a and Experiment 2b, there were five trial blocks in which the minimum-accuracy criterion was not met (with all five blocks being inconsistent trial blocks). Accordingly, the minimum-accuracy level was lowered to 79% after the event, with any scores below that being discarded from the analysis (the data from one pair of trial blocks in the combined-learning condition and three pairs of trial blocks in the verbal-instruction condition were removed on this basis from Experiment 2a, with Experiment 2b losing the data from one pair of trial blocks in the verbal-instruction condition).

Formation and persistence of differential relational responding.

Experiment 2a. As Figure 3.1 indicates, the participants in the three attitude-induction conditions produced differential responding with the participants reacting faster on consistent trials than on inconsistent trials in all but two of the pairs of test blocks. I could not run a 3 (contingency-learning, verbal-instruction, combined-learning) x 3 (1^{st}, 2^{nd}, and 3^{rd} pair of test blocks) repeated-measures ANOVA for Experiment 2a as one of the pairs of test blocks for the verbal-instruction condition contained only one datum (the other data had been removed from the analysis). So in order to account for the missing values (for those trial blocks in which the accuracy criterion was not met), the data for the 1^{st} and 2^{nd} trial blocks were combined in order to create two levels: (1) trial blocks 1 to 4 (‘early’ trial blocks), and, (2) trial blocks 5 & 6 (‘late’ trial blocks). Accordingly, a 3 (contingency-learning, verbal-instruction, combined-learning) x 2 (‘early’ trial blocks, ‘late’ trial blocks) mixed repeated-measures ANOVA was conducted which indicated that there was no significant effect for

attitude-induction condition, $F(2, 8) = .91$, $p = .44$, $r = .43$, and no significant interaction between attitude-induction condition and time/trial block, $F(2, 8) = .69$, $p = .53$, $r = .38$. However, there was a significant effect for time/trial block, $F(1, 8) = 5.82$, $p = .04$, $r = .65$. The observed power was very low for the attitude-induction effect (.16) and the interaction between attitude-induction and time/trial block effect (.13), meaning that if there *were* real effects they would only be detected 16% and 13% of the time, respectively. The observed power was still fairly low for the significant effect for time/trial block (.56) but the effect size was large ($r = .65$).

A visual inspection of Figure 3.1 will indicate that there was a substantial decrease in the biased relational responding performed by the participants in the contingency-learning and verbal-instruction conditions. Indeed, the 5th and 6th test blocks saw a reversal of the 'IRAP effects' that were recorded in the first two pairs of blocks for these conditions. Furthermore, the 3-x-2 mixed repeated-measures ANOVA produced a series of estimated marginal means that indicated a decrease in the *D*-IRAP scores over time for all three conditions. This means that despite the mid-test increase in differential responding by the participants in the combined-learning condition, there was still a decrease in the response bias shown by these participants over time.

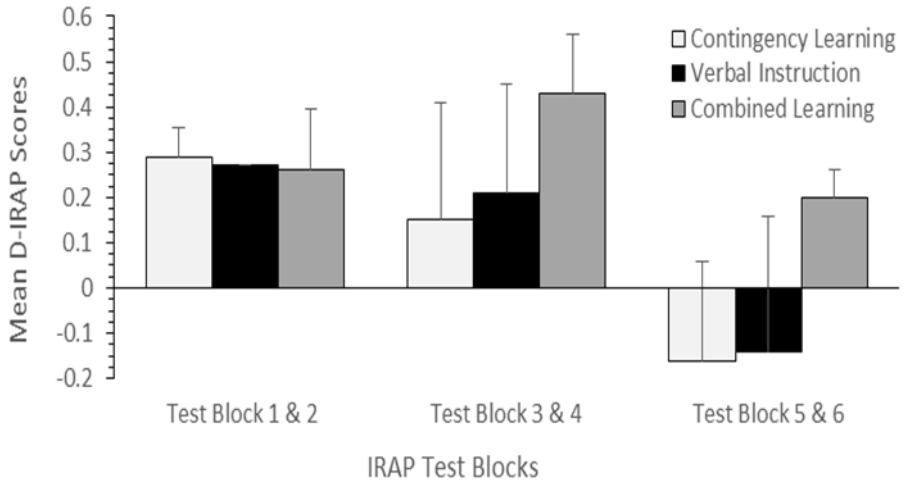

Figure 3.1. The mean *D*-IRAP scores for the three attitude-induction conditions in the first testing phase of Experiment 2. If the mean *D*-IRAP score is above or below zero, then there was a bias in participant responding that may represent the presence of an implicit attitude. Standard error bars reflect the standard error of the mean.

Experiment 2b. As Figure 3.2 portrays, the participants in the three attitude-induction conditions produced differential relational responding throughout the test with the participants reacting faster on consistent trials than on inconsistent trials in all but one of the pairs of test blocks. In order to account for the values missing from the trial blocks in which the accuracy criterion was not met, the data from Experiment 2b underwent the same 'chunking' as occurred with the data from Experiment 2a. Accordingly, a 3 (contingency-learning, verbal-instruction, combined-learning) x 2 ('early' trial blocks, 'late' trial blocks) mixed repeated-measures ANOVA was conducted which indicated that there was no significant effect for attitude-induction condition, $F(2, 8) = 1.66$, $p = .25$, $r = .54$, no significant effect for time/trial block, $F(1, 8) = .31$, $p = .59$, $r = .19$, and no significant interaction between attitude-induction condition and time/trial block, $F(2, 8) = .14$, $p = .87$, $r = .18$. The observed power was also very low for the attitude-induction (.25), time/trial block (.08), and the interaction between attitude-induction condition and time/trial block (.07) effects, meaning that if there *were* real effects they would only be detected 25%, 8%, and 7% of the time, respectively.

A visual inspection of Figure 3.2 will indicate that there was no decrease in the biased relational responding performed by the participants in the contingency-learning condition. Indeed, the 'IRAP effects' increased across the pairs of test blocks in the contingency-learning condition. Furthermore, the 'IRAP effects' were generally weaker in Experiment 2b than those recorded in Experiment 2a.

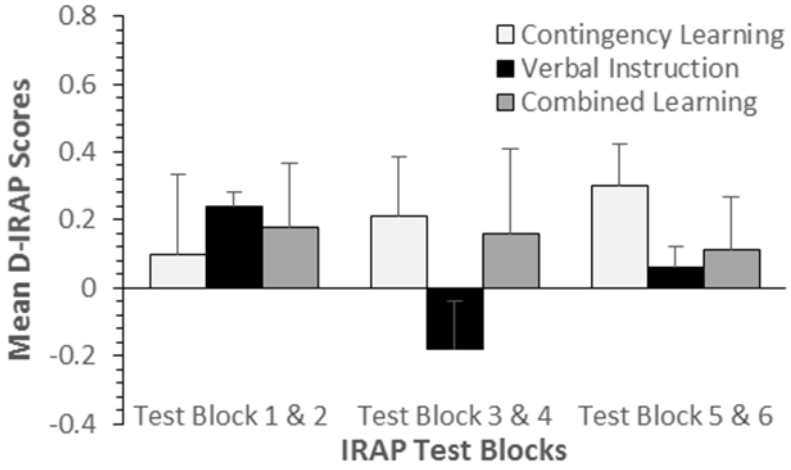

Figure 3.2. The mean *D*-IRAP scores for the three attitude-induction conditions in the second testing phase of Experiment 2. If the mean *D*-IRAP score is above or below zero, then there was a bias in participant responding that may represent the presence of an implicit attitude. Standard error bars reflect the standard error of the mean.

Affirmation and negation trial types. Figure 3.3 (Experiment 2a) and Figure 3.4 (Experiment 2b) portray the mean *D*-IRAP scores of the trials in which the participants were asked to affirm the proposed stimulus relation and the mean *D*-IRAP scores of the trials in which the participants were asked to negate the proposed stimulus relation. A visual inspection of these two figures will reveal a substantial decrease in differential responding on negation trials across the two experiments. Indeed, a one-way repeated-measures ANOVA indicated that there was no significant effect for trial type in Experiment 2a, $F(1, 10) = .48$, $p = .5$; yet a second one-way repeated-measures ANOVA indicated that there was a significant effect for trial type in Experiment 2b, $F(1, 10) = 16.82$, $p = .002$.

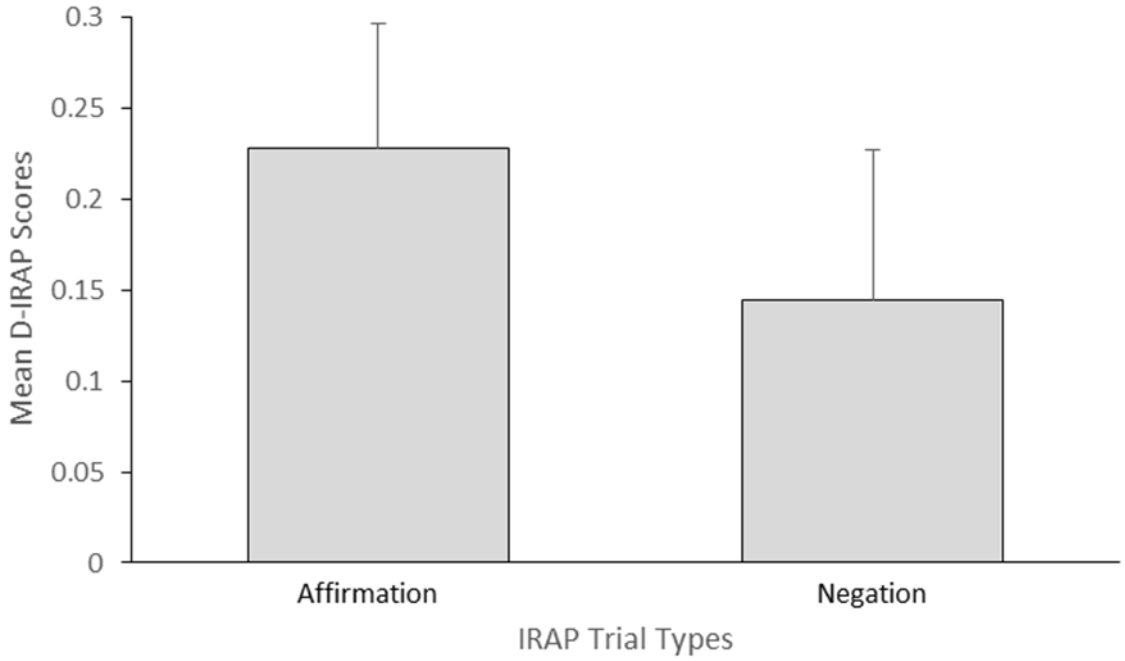

Figure 3.3. The mean *D*-IRAP scores for the Experiment 2a trial types in which the participants were required to affirm the rule at play and the mean *D*-IRAP scores for the trials in which they were required to negate it. Standard error bars reflect the standard error of the mean.

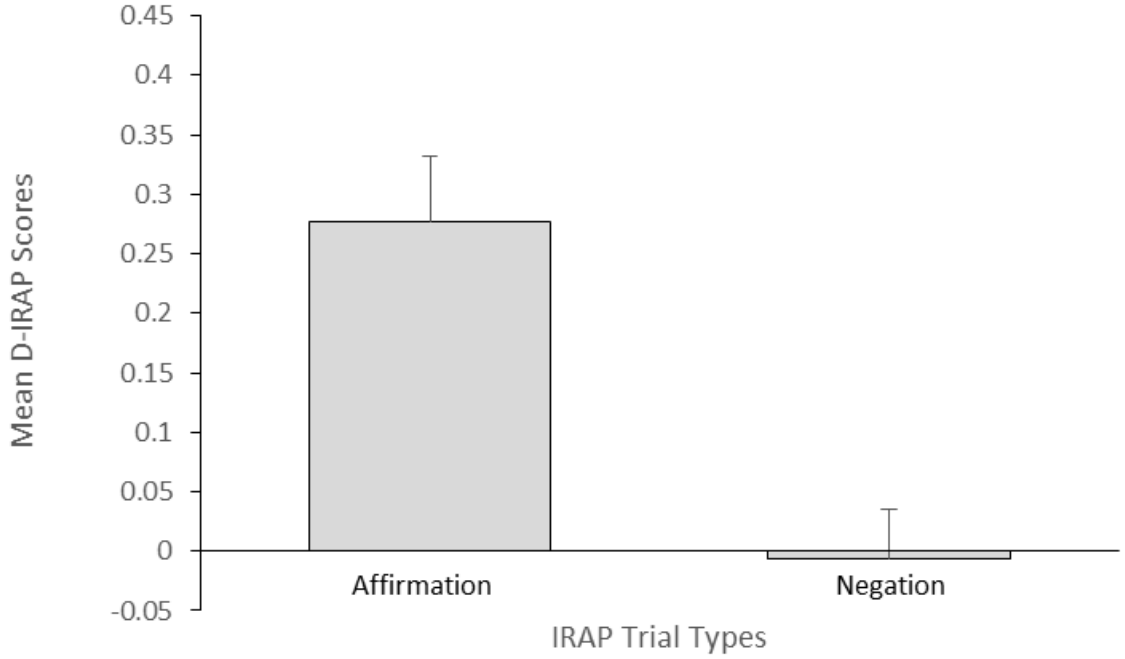

Figure 3.4. The mean *D*-IRAP scores for the Experiment 2b trial types in which the participants were required to affirm the rule at play and the mean *D*-IRAP scores for the trials in which they were required to negate it. Standard error bars reflect the standard error of the mean.

3.4 Discussion

Experiment 2a was conducted in order to see if the response biases that were engineered in Experiment 1 persisted after five days. Having presented the participants with the first testing IRAP (Experiment 2a), the participants were then given the same attitude-induction procedure that they had been asked to take in Experiment 1. The retraining phase was then followed immediately by Experiment 2b, which was a second application of the testing IRAP conducted in order to find out if the retraining procedures had altered the participants' differential relational responding. As Figure 3.1 portrays, the participants in Experiment 2a continued to perform differential relational responding with 'IRAP effects' occurring in all three pairs of test blocks for all three conditions. As was the case in Hughes and Barnes-Holmes's (2011) experiment, there was no significant effect for attitude-induction condition in Experiment 2a; there was, however, a significant effect for time/trial blocks, with the level of differential responding decreasing from early to late trial blocks. Accordingly, Experiment 2a successfully recreates the weakening of the 'IRAP effects' that was observed by Hughes and Barnes-Holmes. Furthermore, the weakening continued until the 'IRAP effects' were reversed in the final pair of test blocks for the contingency-learning and verbal-instruction conditions. This reversal means that the participants within these conditions were responding faster on inconsistent trials than on consistent trials.

There were no significant effects for attitude-induction method or time/trial blocks in Experiment 2b. The recorded 'IRAP effects' were also smaller than the 'effects' recorded in the two other tests. Furthermore, there was a reversal of the 'IRAP effect' in the middle block of the verbal-instruction condition with the participants responding faster on inconsistent trials than on consistent trials. The relatively weak 'IRAP effects' and the single reversal in differential responding is of particular interest because both response phenomena happened

after the retraining phase of Experiment 2. Indeed, it should be expected that after the re-application of the attitude-induction procedures, the 'IRAP effects' would be strengthened or at least returned to a positive value (for those 'effects' that had been reversed in Experiment 2a). Yet rather than echo the clear and persistent differential responding of Experiment 1, the results of Experiment 2b indicate the presence of more complex patterns of behaviour.

The results of Experiment 2a indicate that the response biases engineered in Experiment 1 persisted over five days. This conclusion is made because there would have been little opportunity for the participants to rehearse the multiple stimulus relations away from the experimental setting (other than through their own cognitive-behaviour) and certainly no opportunity to receive reinforcement or negative feedback for doing so. Furthermore, the participants received no exposure to either the sample or target stimuli in the second session until the first testing IRAP of Experiment 2a. However, as is the case with any application of the testing IRAP, the participants were asked to complete a series of practice blocks before Experiment 2a began. Indeed, the pre-test exposure to the various stimuli across multiple trial blocks may have reactivated the response biases (and any implicit attitudes) before they were accessed and recorded by the testing IRAP. Accordingly, it may be worth recreating my two experiments without the practice blocks in Experiment 2a. It may also be inappropriate to confirm the persistence of the engineered implicit attitudes without this additional research being done.

A limitation to Experiment 2 is that one of the participants had the written instructions in their possession for the five days between the two sessions. While it is possible that the instructions were not looked at, it is also possible that the participant re-read the written narrative along with their answers to the memory checks and self-report measures. If this additional exposure did happen, then the participant's revision of the various stimulus relations may have produced an appreciable outcome in an experiment with such a small number of

participants. Accordingly, although no additional analysis was made of the participant's data, it is possible that exposure to the written instructions produced a significant effect in the participant's differential relational responding when compared to the rest of their group.

One further limitation with Experiment 2 concerns the lack of any measurement of the participants' explicit-attitudes. The self-report measures that were used in Experiment 1 were not included in the second experiment to save time for the participants, who were charged with completing two testing IRAPs and one training procedure within an hour. However, all the participants completed Experiments 2a and 2b within 30 min, meaning that there was a missed opportunity to further determine the participants' explicit attitudes towards the sample stimuli. Indeed, it would have been valuable to know whether, having taken two testing IRAPs and a training procedure, the participants still held explicit attitudes that were congruent with their implicit evaluations. Furthermore, given that there were two participants in Experiment 1 who did not respond in accordance with the training procedure, it is not certain that the participants would have expressed explicit attitudes that agreed with what they had been taught.

In summary, the participants in Experiment 2a continued to show biased relational responding to the sample stimuli even though five days had passed since they had last been exposed to them. However, the participants' biased relational responding diminished through Experiment 2a's three pairs of testing blocks until those in the contingency-learning and verbal-instruction conditions were responding faster on inconsistent trials than on consistent trials. Experiment 2b produced 'IRAP effects' across all test blocks in all attitude-induction conditions; however, the effects were weaker than those observed in Experiment 2a and there was a reversal in the 'IRAP effect' within the verbal-instruction condition. The results indicate that there is a level of persistence to the response biases (and hence implicit attitudes) that are engineered using the IRAP and a written narrative, but the engineered

biases are not necessarily strengthened by re-exposure to any particular attitude-induction procedure.

Chapter 4

General Discussion

The aims of this thesis were: (a) to find out if the weakening of the 'IRAP effects' in Hughes and Barnes-Holmes's (2011) study occurred again during a strict replication (Experiment 1), (b) to find out whether the response biases that were engineered in Experiment 1 persisted beyond the time it took to conduct the first testing IRAP (Experiment 2a), and, (c) to find out whether any persistent response biases were strengthened by further training (Experiment 2b).

The results of Experiment 1 show that biased relational responding can be engineered using the IRAP. There was, however, no weakening in the 'IRAP effects' of Experiment 1 despite the participants being exposed to competing contingencies in the testing IRAP. The results of Experiment 2a indicate that the response biases that were engineered in Experiment 1 persisted into the testing IRAP that was conducted five days after the first session. Furthermore, the response-bias attenuation that was observed in Hughes and Barnes-Holmes's (2011) study occurred within my second experiment, with the last block of test trials containing two reverse 'IRAP effects'. Retraining the participants using several attitude-induction procedures produced definite differential responding in Experiment 2b, although there was a reverse bias in one of the trial blocks for one of the conditions and the recorded 'IRAP effects' were generally weaker than those observed in Experiments 1 and 2a.

The results of my two experiments support the use of the IRAP as a measure of differential relational responding that can also be used to induce implicit and explicit cognitive behaviour (in the guise of either brief-and-immediate or extended-and-elaborated relational responses; Hughes & Barnes-Holmes, 2013; Hughes et al., 2012). Indeed, despite the lack of significance in the majority of the observed response biases, there were positive

'IRAP effects' in 24 of the 27 available pairs of test blocks. Furthermore, the presence of a negative response bias does not signify the failure to produce an 'IRAP effect'; rather it indicates a shift in the predominant pattern of relational responding. Given the alternating contingencies at play (Hughes & Barnes-Holmes, 2011), there is always the potential for participants to respond faster on inconsistent trial blocks than on consistent trial blocks (which produces a negative 'IRAP effect'). But whether or not this means that the participants' implicit attitudes change during such reversals in response bias depends upon how the biased responding is interpreted. If relational responses are proxies for an underlying attitude, then clearly the attitude is altered with any change in response bias. However, shifts in differential responding can be explained in purely functional terms without any need to reference the attitude construct (Hughes & Barnes-Holmes, 2011; Hughes et al., 2011; Hughes et al., 2012). For example, each trial within an inconsistent trial block requires the participant to relate the sample and target stimuli in a way that is oppositional to what they know (e.g., having to state that "Cug" is "Similar" to "Virtuous" when they have been trained to relate "Vek" with "Virtuous"). Yet this elaborated relational response is a behavioural act which is likely to become easier to perform with practice. Accordingly, it stands to reason that as more inconsistent trial blocks are presented, the participants are able to speed up their extended-and-elaborated relational responding to the point that it may even negate an earlier response bias (Hughes & Barnes-Holmes, 2013; Hughes et al., 2012).

Although the IRAP is capable of assessing new and pre-existing relational response biases, exposure to the procedure may alter the very evaluations that are being accessed. Indeed, by presenting multiple reversals of the relational rule at play, the IRAP is able to influence the arbitrarily-applicable relational responding from which the presence or absence of an implicit attitude is inferred (Hayes et al., 2011; Hughes & Barnes-Holmes, 2011). For example, a participant may understand the word "Vek" to be "Similar" to the word

"Virtuous" on an implicit level, yet if the participant were to have this learned verbal behaviour tested using the IRAP, for 50% of the testing trials they would be given negative feedback (i.e. a red "X") if they agreed with the learned rule (that the word "Vek" is "Similar" to word "Virtuous") and for the other 50% of the trials they would be given positive feedback (i.e., progress) if they responded in a way that disagreed with that same rule. So it stands to reason that any pre-existing bias in relational responding would be lessened in the face of such competing (and oppositional) contingencies, which could explain why the 'IRAP effects' appeared to weaken in both my Experiment 2a and Hughes and Barnes-Holmes's (2011) study. Accordingly, the IRAP may be an effective measure of relational response biases during a limited number of trial blocks, but with extended use it may overly influence the behaviour it is intended to access and measure.

The IRAP's alternating contingencies may also limit the persistence of the same differential responding the procedure was used to engineer. This assertion has been made because the IRAP demands relational responses that contradict the induced response bias on at least as many discrete trials as it took to engineer it. Indeed, in my Experiments 1 and 2b it was possible for a participant to be exposed to a greater number of inconsistent trials in the testing IRAP than the number of trials they took to acquire the initial rule in the training procedure. This discrepancy occurred whenever a participant only required one or two training blocks to meet the response criteria in the contingency-learning and combined-learning conditions. The results of Experiment 2a suggest that the response biases that had persisted over five days weakened through exposure to the testing IRAP; furthermore, even though the participants had just been presented with a training procedure, Experiment 2b saw at least one reversal in the expected response bias and a low level of differential responding overall. Accordingly, there may be a need to regulate (or even increase) the amount of training each participant gets before they are presented with the competing contingencies of

the testing IRAP (Hughes & Barnes-Holmes, 2011).

Some mention should be made of the differential responding that occurred during trials in which the participants had to affirm the proposed stimulus relation and the trials in which the participants had to negate the stimulus relation that was presented to them. Hughes and Barnes (2011) analysed the results of their study in these terms and found that affirmation trials produced a much larger 'IRAP effect' than negation trials. There was no significant difference between the amount of differential responding by trial type in my Experiments 1 and 2a. However, the results of Experiment 2b indicate there was a significant effect for trial type, with a relatively large 'IRAP effect' in the affirmation trials and a slight reverse 'effect' for negation trials. This outcome suggests that, in the final testing IRAP, the participants were quicker to affirm a stated relation (i.e., confirm that the sample and target stimuli were indeed "Similar") on consistent trial blocks than on inconsistent trial blocks and slightly quicker to negate a stated relation (i.e., confirm that the sample and target stimuli were indeed "Opposite") on inconsistent trials blocks than on consistent trial blocks. As Hughes and Barnes-Holmes (2011) assert, a possible explanation for this secondary response bias is that frames of co-ordination occur more often within everyday language and cognition than frames of opposition. Accordingly, a participant using the IRAP is able to state that something is "Similar" to something else quicker than the time it takes them to state that something is "Opposite" to something else. Yet whether this particular response bias is mitigated by practice is unknown at this stage.

A possible limitation of this research concerns the intermittent use of a written script during the testing and training components of the experiment. As became apparent within the first experimental session, the IRAP is a procedure that places a high demand on participants to respond in an entirely prescribed way, a dynamic which creates a certain urgency for participants to emit correct responses over many discrete trials. Accordingly, and in order to

support some of the participants to meet the response criteria, it was necessary to offer further feedback beyond that offered by the IRAP (such as recommending to a participant that they respond slower in order to improve their accuracy). Whether this was a limitation to the written instructions or the instructions that were presented on-screen is not known as no feedback was sought from the participants. However, given the results of McKenna et al.'s (2007) study, it is unlikely that any minor deviations from the experimental script significantly influenced the participants' differential responding in my three experiments.

One further limitation regards the written narrative that was used in the verbal-instruction and combined-learning conditions of Experiments 1 and 2. An oversight in the development of the narrative meant that it contained one sample stimulus six times and the other sample stimulus seven times. Accordingly, in the training components of the two conditions featuring the written narrative there was one more opportunity for the participants to learn the verbal rule associated with the first sample stimulus than for them to learn the verbal rule associated with the second; furthermore, this disparity occurred a second time within the training component of Experiment 2. Although no analysis was conducted, it is possible that this variance in sample presentation affected the participants' responding.

Finally, a mention should be made of the limitations of the meaningfulness scale used in Experiment 1. This self-report measure produced predominantly low scores with only 36% of the participants rating the meaningfulness of their attitudes towards the words "Cug" and "Vek" at or above the midpoint of the scale. There may be several reasons for why the participants offered such low scores. For example, some of the participants may have been uncertain about what it meant for something to be 'meaningful' and so offered a low score in place of a true appraisal. Furthermore, the instructions on how to complete the scale may have been misleading or even inadequate. These issues aside, it is possible that a participant's perceived meaningfulness of the words "Cug" and "Vek" was influenced by whether or not

they had come into contact with the sample stimuli in an 'enriched' context (as occurred within the written narrative, for example). Furthermore, the presence of the words "Cug" and "Vek" in the written narrative appeared to instil additional 'meaning' to the sample stimuli when the training IRAP was also used, an effect that can be explained by RFT and the REC model (Hayes et al., 2001; Hughes et al., 2012; Hughes & Barnes-Holmes, 2013). Indeed, the highest meaningfulness scores were given by the participants who had read the written narrative *and* completed the training IRAP. Accordingly, it is possible that, in order to make engineered response biases strong, persistent, and 'meaningful', an 'enriched' context should be used alongside the IRAP whenever it is possible to do so.

References

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behaviour. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of Attitudes* (pp. 173-222). Mahwah, NJ: Lawrence Erlbaum Associates.

Albarracín, D., Zanna, M. P., Johnson, B. T., & Kumkale, G. T. (2005). Attitudes: Introduction and Scope. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of Attitudes* (pp. 3-20). Mahwah, NJ: Lawrence Erlbaum Associates.

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology: Volume two*. New York: Russell & Russell.

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*, *20*, 143-148.

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist*, *32*(7), 169-177.

Barnes-Holmes, D., Barnes-Holmes, Y., Smeets, P. M., Cullinan, V., & Leader, G. (2004). Relational Frame Theory and stimulus equivalence: Conceptual and procedural issues. *International Journal of Psychology and Psychological Therapy, 4*(2), 181-214.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) Model. *The Psychological Record, 60*, 527-542.

Barnes-Holmes, D., Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The Implicit Relational Assessment Procedure (IRAP) as a response-time and event-related-potentials methodology for testing natural verbal relations: A preliminary study. *The*

Psychological Record, 58(4), 497-516.

Barnes-Holmes, D., & Hussey, I. (2015). *IRAP user manual v2.4*. Retrieved from http://IRAPresearch.org

Barnes-Homes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure: Exploring the impact of private versus public contexts and response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record, 60*, 57-66.

Barnes-Holmes, D., Murtagh, L., Barnes-Holmes, Y., & Stewart, I. (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes toward meat and vegetables in vegetarians and meat-eaters. *The Psychological Record, 60*, 287-306.

Bassili, J. N., & Brown, R. D. (2005). Implicit and explicit attitudes: Research, challenges, and theory. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of Attitudes* (pp. 543-574). Mahwah, NJ: Lawrence Erlbaum Associates.

Bast, D. F., & Barnes-Holmes, D. (2014). A first test of the Implicit Relational Assessment Procedure as a measure of forgiveness of self and others. *The Psychological Record, 64*, 253-260.

Biglan, A., & Hayes, S. C. (1996). Should the behavioral sciences become more pragmatic? The case for functional contextualism in research on human behavior. *Applied and Preventive Psychology, 5*, 47-57.

Cacioppo, J. T., Petty, R. E., & Geen, T. R. (1989). Attitude structure and function: From the tripartite to the homeostasis model of attitudes. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 275-310). Hillsdale, NJ: Lawrence Erlbaum Associates.

Campbell, C., Barnes-Holmes, Y., Barnes-Holmes, D., & Stewart, I. (2011). Exploring screen

presentations in the Implicit Relational Assessment Procedure (IRAP). *International Journal of Psychology and Psychological Therapy*, *11*(3), 377-388.

Chan, G., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). Implicit attitudes to work and leisure among North American and Irish individuals: A preliminary study. *International Journal of Psychology and Psychological Therapy*, *9*(3), 317-334.

Coolican, H. (2009). *Research methods and statistics in psychology*. Abingdon, OX: Routledge.

Crano, W. D., & Prislin, R. (2006). Attitudes and persuasion. *Annual Review of Psychology*, *57*, 345-374.

Cullen, C., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) and the malleability of ageist attitudes. *The Psychological Record*, 59, 591-620.

Cullinan, V. A., Barnes-Holmes, D., & Smeets, P. M. (2001). A precursor to the relational evaluation procedure: Searching for the contextual cues that control equivalences responding. *Journal of the Experimental Analysis of Behaviour*, *76*(3), 339-349.

Dawson, D. L., Barnes-Holmes, D., Gresswell, D. M., Hart, A. J., & Gore, N. J. (2009). Assessing the implicit beliefs of sexual offenders using the Implicit Relational Assessment Procedure: A first study. *Sexual Abuse: A Journal of Research and Treatment*, *21*(1), 57-75.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology*, *50*, 77-85.

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation*, *37*, 176-187.

Devos, T. (2008). Implicit attitudes 101: Theoretical and empirical insights. In W. D. Crano, and R. Prislin (Eds.), *Attitudes and attitude change*. London: Psychology Press.

Dixon, M. R., Dymond, S., Rehfeldt, R. A., Roche, B., & Zlomke, K. R., (2003). Terrorism and

Relational Frame Theory. *Behavior and Social Issues*, *12*, 129-147.

Dixon, M. R., Rehfeldt, R. A., Zlomke, K. R., & Robinson, A. (2006). Exploring the development and dismantling of equivalence classes involving terrorist stimuli. *The Psychological Record*, *56*, 83-103.

Drake, C. E., Kellum, K. K., Wilson, K. G., Luoma, J. B., Weinstein, J. H., & Adams, C. H. (2010). Examining the implicit relational assessment procedure: Four preliminary studies. *The Psychological Record*, *60*, 81-10

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*, 603-637.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology*, *44*. 59-127.

Goldstein, E. B. (2011). *Cognitive Psychology: Connecting the mind, research, and everyday experiences* (3rd ed.). Belmont, CA: Wadworth, Cengage Learning.

Golijani-Moghaddam, N., Hart, A., & Dawson, D. L. (2012). The Implicit Relational Assessment Procedure: Emerging reliability and validity data. *Journal of Contextual Behavioral Science*, *2*, 105-119.

Greenwald, A. G. (1989). Why are attitudes important? In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 1-10). Hillsdale, NJ: Lawrence Erlbaum Associates.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4-27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences

in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90*(1), 1-20.

Gross, A. C., & Fox, E. J. (2009). Relational Frame Theory: An overview of the controversy. *The Analysis of Verbal Behavior, 25*, 87-98.

Hayes, S. C., Barnes-Holmes, D., & Roche, B., (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition*. New York: Kluver Academic/Plenum Publishers.

Himmelfarb, S. (1993). The measurement of attitudes. In A. H. Eagly, & S. Chaiken (Eds.), *The psychology of attitudes* (pp. 23-88). Fort Worth, TX: Harcourt Brace Jovanovich.

Hughes, S., & Barnes-Holmes, D. (2011). On the formation and persistence of implicit attitudes: New evidence from the Implicit Relational Assessment Procedure (IRAP). *The Psychological Record, 61*, 391-410.

Hughes, S., & Barnes-Holmes, D. (2013). A functional approach to the study of implicit cognition: The Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) Model. In B. Roche & S. Dymond (Eds.), *Advances in Relational Frame Theory: Research and Application* (pp. 97-125). Oakland, CA: New Harbinger.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record, 61*, 465-496.

Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research.

Journal of Contextual Behavioral Science, 17-38.

Hussey, I., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure as a measure of implicit depression and the role of psychological flexibility. *Cognitive and Behavioral Practice*, *19*, 573-582.

Hussey, I., Barnes-Holmes, D., Barnes-Holmes, Y. (2015). From Relational Frame Theory to implicit attitudes and back again: Clarifying the link between RFT and IRAP research. *Current Opinion in Psychology*, *2*, 11-15.

Hussey, I., Thompson, M., McEnteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2014). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science*, *4*, 157-162.

Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of Attitudes* (pp. 21-76. Mahwah, NJ: Lawrence Erlbaum Associates.

Kruglanski, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodal. *Psychological Inquiry*, *10*, 83-109.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 59-102). New York, NY: The Guilford Press.

Maio, G. R., & Haddock, G. (2010). *The psychology of attitudes and attitude change*. Thousand Oaks, CA: Sage.

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology: Volume II; Special fields and applications* (pp. 233-346). New York: Random House.

McKenna, I. M., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2007). Testing the fake-

ability of the Implicit Relational Assessment Procedure (IRAP): The first study. *International Journal of Psychology and Psychological Therapy, 7*(2), 253-268.

Nicholson, E., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record, 62*, 263-278.

Nosek, B. A. (2007). Implicit-explicit relations. Current Directions in Psychological Science, 16, 65-69.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*, 625-666.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*(5), 413-417.

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

Payne, K., & Lundberg, K. (2014). The Affect Misattribution Procedure: Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass, 8*(12), 672-686.

Power, P., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) as a measure of implicit relative preferences: A first study. *The Psychological Record, 59*, 621-640.

Ranganath, K. A., & Nosek, B. A. (2006). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Sciences, 19*(3), 249-254.

Rudman, L. A. (2004). Sources of implicit attitude. *Current Directions in Psychological Science, 13*, 80-83.

Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser & N. Schwarz

(Eds.), *Blackwell Handbook of Social Psychology: Interpersonal processes* (pp. 436-457). Oxford, UK: Blackwell.

Sidman, M. (2000). Equivalence relations and reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, *74*(1), 127-146.

Sidman, M. (2009). Equivalence relations and behavior: an introductory tutorial. *The Analysis of Verbal Behavior*, *25*, 5-17.

Sidman, M., & Cresson, O., Jr. (1973). Reading and crossmodal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency*, 77, 515-523.

Skinner, B. F. (1965). *Science and human behavior*. New York: The Free Press.

Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the Relational Evaluation Procedure. *The Psychological Record, 54*, 531-552.

Uttal, W. R. (2000). *The war between mentalism and behaviorism: On the accessibility of mental processes*. Mahwah, NJ: Lawrence Erlbaum Associates.

Vahey, N. A., Barnes-Holmes, D. Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record, 59*(3), 371-388.

Vahey, N., Boles, S., & Barnes-Holmes, D. (2010). Measuring adolescents' smoking-related social identity preferences with the Implicit Relational Assessment Procedure (IRAP) for the first time: A starting point that explains later IRAP evolutions. International *Journal of Psychology and Psychological Therapy*, *10*(3), 453-474.

Wittenbrink, B. (2007). Measuring attitudes through priming. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 17-58). New York, NY: The Guilford Press.

Wittenbrink, B., & Schwarz, N. (2007). Introduction. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 1-16). New York, NY: The Guilford Press.

Appendix A

The information sheet that was emailed to individuals who expressed an interest in the research project and presented to each participant (along with the consent form) before they began Experiment 1.

A Replication and Extension of Hughes and Barnes-Holmes's (2011) Study of Induced Implicit Attitudes

Information Sheet

My name is Douglas Harvey and I am currently enrolled on the Masters of Applied Psychology (Applied Behaviour Analysis) course at the University of Waikato. It is my pleasure to invite you to participate in my replication and extension of an earlier experiment which tested engineered implicit attitudes using a novel, computer-based assessment tool.

The Implicit Relational Assessment Procedure (IRAP) is a computer-based assessment tool that has been used extensively to measure pre-existing implicit attitudes (understood in this context as being implicit relational responding) which have been acquired through historical learning processes. However, until Hughes and Barnes-Holmes's (2011, p. 391) experiment on the "formation and persistence of implicit attitudes", there had been no research in which the IRAP was used to measure implicit attitudes that had been engineered in a laboratory. This study intends to recreate the earlier experiment while also extending it to see if the engineered responses persist over several days. As a further extension of the original study, this research will see whether implicit attitudes that have been weakened by the passing time can be strengthened through further training.

As an undergraduate, your participation may allow you to receive up to 2% of course credit (if available to you); those participants who attend both sessions and who do not wish to receive course credit can opt to go into the draw to win one of four Warehouse vouchers instead.

As a participant you will be asked to:

- Attend two experimental sessions spaced five days apart.
- In the first experimental session you will learn a series of verbal stimulus relations through either direct contingency learning using the IRAP, through the use of a written narrative, or through the combined use of both methods. Depending on which experimental condition you are in you may then be asked to complete two short written 'memory checks'. Finally, all participants will undergo a testing IRAP and be asked to complete three brief self-report measures.
- The second experimental session will require you to take an identical testing IRAP before being exposed to the same training procedure you underwent in the first session. You will then take a final testing IRAP thus completing your participation in the study.

As a participant your data will remain anonymous:

- After the final experimental session any of the contact details used to schedule the time and date of your participation will be safely discarded along with your name and any other identifying information.
- Upon arrival at the computer lab you will be given a participant identifier – this identifier will be attached to your data report with no other form of identification being used.
- Consent forms and data sheets will be kept in a locked cabinet at the researcher's office for a period of three years – they will then be securely destroyed.

As a participant your involvement is totally voluntary:

- You may withdraw from this experiment at any time.
- There is no academic requirement (or otherwise) for you to participate in this study.

If you have any questions regarding your involvement in this research please contact either the researcher, Douglas Harvey, or his supervisor, Dr Rebecca Sargisson.

Researcher: Douglas Harvey
School of Psychology, University of Waikato
dmh45@students.waikato.ac.nz
021 273 3309
Supervisor: Dr Rebecca Sargisson
School of Psychology, University of Waikato
rebeccas@waikato.ac.nz

This research project has been approved by the School of Psychology Research and Ethics Committee of the Faculty of Arts and Social Sciences, University of Waikato. Any questions about the ethical conduct of this research may be sent to the convenor of the Research and Ethics Committee (currently Dr James McEwan, phone 07 838 4466 ext. 8295, email: jmcewan@waikato.ac.nz)

Reference:
Hughes, S., & Barnes-Holmes, D. (2011). On the formation and persistence of implicit attitudes: New evidence from the Implicit Relational Assessment Procedure (IRAP). *The Psychological Record*, *61*, 291-410.

Micro-Glossary

Implicit Attitude: A way of responding to something in a positive or negative way that may be automatic – you may even be unaware that you hold an implicit attitude towards something.

Implicit Relational Responding: A behaviour-analytic term for responding to something in a positive or negative way (in other words, having an implicit attitude). In the world of behaviour analysis, an implicit attitude *is a type of behaviour in and of itself*.

Implicit Relational Assessment Procedure: The Implicit Relational Assessment Procedure (IRAP) is a computer-based measure of implicit attitudes – or in behaviour-analytic terms, implicit relational responding – which records how long participants take to respond to meaningful textual stimuli presented on-screen. The IRAP programme can be used to test both pre-existing attitudes and to train new implicit responses.

Appendix B

An extract of the experimenter's script for both Experiment 1 and Experiment 2 with the combined-learning condition offered as an example.

Experiment 1

For all participants (regardless of condition), when the participant arrives, greet them, introduce yourself, and then ask them what their name is. Check the timetable to determine the participant's experimental condition and to confirm their details, give the participant their identity card, pick up a consent form and the written instructions for the participant's condition, and then ask the participant to sit in front of the laptop. Advise the participant that they are about to receive a consent form and an information sheet and that they have no time limit in which to read and sign the consent form. If the participant has signed the form place the written instructions for their condition in front of them and take the consent form. If they are not willing to sign the form, thank them for their attendance and remove all materials from the desk. For those participants continuing with the experiment, load the particular IRAP programme they will use on the laptop and present any written materials to them with the verbal instruction that they should summon the experimenter if they need any further assistance.

The written instructions for each condition are as follows (less the condition titles, the experimenter's dialogue, and the experimenter's instructions):

Combined-Learning Condition

YOU ARE ABOUT TO READ A WRITTEN NARRATIVE. PLEASE ATTEND TO WHAT IS PRESENTED IN THE NARRATIVE AS YOU WILL BE TESTED ON THIS AT A LATER POINT IN THE EXPERIMENT.

The participant should then read the narrative. They will then be prompted to turn the page and complete several tasks.

THANK YOU FOR READING THE NARRATIVE AND COMPLETING THE TWO ACCOMPANYING TASKS. PLEASE SUMMON THE EXPERIMENTER.

Check the participant's responses to the memory checks. If their answers are not consistent with what they have been exposed to, provide the participant with another memory-check question sheet and ask them to re-read the narrative before answering the questions for a second time. If the participant is still not answering in a consistent way then they should be thanked for their involvement in the experiment and told that their participation has ended.

Any remaining participants should then turn the page of their written instructions.

YOU ARE ABOUT TO TAKE A COMPUTER-BASED LEARNING TASK WHICH INVOLVES READING

COMBINATIONS OF TWO WORDS THAT ARE PRESENTED SIMULTANEOUSLY ON THE SCREEN. YOUR TASK IS TO STATE WHETHER THE WORDS ARE SIMILAR IN MEANING OR OPPOSITE.

FOR EACH TRIAL YOU WILL NEED TO LOOK AT THE WORD AT THE TOP OF THE SCREEN, THE WORD PRESENTED BELOW IT, AND THEN THE WORDS "SIMILAR" AND "OPPOSITE" WHICH WILL BE PRESENTED HORIZONTALLY AT THE BOTTOM OF THE SCREEN. USING EITHER THE "D" KEY OR THE "K" KEY YOU MUST THEN CHOOSE ONE OF THE TWO WORDS (OR RESPONSE OPTIONS) THAT ARE PRESENTED AT THE BOTTOM OF THE SCREEN.

WHEN MAKING YOUR RESPONSE, USE EITHER THE INDEX OR MIDDLE FINGER OF YOUR LEFT HAND TO PRESS THE "D" KEY AND EITHER THE INDEX OR MIDDLE FINGER OF YOUR RIGHT HAND TO PRESS THE "K" KEY.

THE RESPONSE OPTIONS "SIMILAR" AND "OPPOSITE" WILL NOT REMAIN IN THE SAME PLACE OF THE SCREEN FOR EACH TRIAL: THESE WORDS WILL SWAP PLACES IN A WAY THAT CANNOT BE PREDICTED.

IN ORDER FOR YOU TO RELATE THE VARIOUS WORDS PRESENTED SIMULTANEOUSLY ABOVE THE TWO RESPONSE-WORDS THE COMPUTER PROGRAMME WILL PROVIDE FEEDBACK BY TELLING YOU IF YOU HAVE MADE A CORRECT OR INCORRECT RESPONSE. AN INCORRECT RESPONSE IS INDICATED BY THE PRESENTATION OF A RED CROSS.

IT IS IMPORTANT THAT YOU MAKE FEW ERRORS AND RESPOND AS QUICKLY AS YOU CAN. TO DO THIS, TRY TO FOCUS ON ACCURATE RESPONDING FIRST AND FOREMOST; WHEN YOU ARE GIVING A CORRECT RESPONSE IN THE MAJORITY OF THE TRIALS, TRY TO SPEED UP YOUR RESPONDING UNTIL YOU TAKING AS LITTLE TIME TO RELATE THE WORDS AS POSSIBLE.

GO WELL!

PLEASE PRESS THE SPACE BAR TO CONTINUE WITH THE LEARNING TASK AND SUMMON THE EXPERIMENTER WHEN PROMPTED TO DO SO.

When the participant summons you (having finished the training IRAP), load the testing IRAP and advise the participant to turn to the next page of the written instructions:

"Thank you for completing the training task. Please read the next page of instructions."

YOU ARE ABOUT TO TAKE A COMPUTER-BASED TEST THAT IS SIMILAR IN APPEARANCE TO THE COMPUTER-BASED LEARNING TASK THAT YOU TOOK EARLIER.

AS BEFORE, THE COMPUTER PROGRAMME WILL ASK YOU TO FOLLOW A SERIES OF RULES WHICH WILL BECOME APPARENT THROUGH THE FEEDBACK THAT IS GIVEN FOLLOWING EACH INCORRECT RESPONSE.

AT SOME POINTS IN THE FOLLOWING TEST THE FEEDBACK PROVIDED WILL MAKE SENSE TO YOU; AT OTHER POINTS, THE FEEDBACK WILL CONTRADICT WHAT YOU HAVE PREVIOUSLY LEARNT. THIS IS PART OF THE TASK AND YOU WILL BE REQUIRED TO RESPOND IN DIFFERENT WAYS.

IT IS IMPORTANT THAT YOU MAKE FEW ERRORS AND RESPOND AS QUICKLY AS YOU CAN. TO DO THIS, TRY TO FOCUS ON ACCURATE RESPONDING FIRST AND FOREMOST; WHEN YOU ARE GIVING A CORRECT RESPONSE ON THE MAJORITY OF THE TRIALS, TRY TO SPEED UP YOUR RESPONDING UNTIL YOU TAKING AS LITTLE TIME TO RELATE THE WORDS AS POSSIBLE.

GO WELL!

PLEASE PRESS THE SPACE BAR TO CONTINUE WITH THE TEST AND SUMMON THE EXPERIMENTER WHEN PROMPTED TO DO SO.

When the participant summons you, say the following:

"Thank you for completing the test. Please turn the page and complete the three self-report measures and let me know when you are finished."

When the participant has finished the self-report measures, take the instruction booklet, thank the participant, and confirm the time of their next attendance.

Extension Experiment

For all participants (regardless of condition), when the participant arrives, greet them, re-introduce yourself, and then ask them for their identity card (or if they do not have it with them ask them what their name is). Check the timetable to determine the participant's experimental condition and to re-confirm their details, pick up the written instructions for the participant's condition, and then ask the participant to sit in front of the laptop. Load the IRAP programme the participant will use on the laptop and present any written materials to them with the verbal instruction that they should summon the experimenter if they need any further assistance.

The written instructions for each condition are as follows (less the condition titles, the experiment's dialogue, and the experimenter's instructions):

Combined-Learning Condition

YOU ARE ABOUT TO TAKE A COMPUTER-BASED TEST THAT IS SIMILAR IN APPEARANCE TO THE COMPUTER-BASED TEST THAT YOU TOOK AT THE EARLIER EXPERIMENTAL SESSION.

AS BEFORE, THE COMPUTER PROGRAMME WILL ASK YOU TO FOLLOW A SERIES OF RULES WHICH WILL BECOME APPARENT THROUGH THE FEEDBACK THAT IS GIVEN FOLLOWING EACH INCORRECT RESPONSE.

AT SOME POINTS IN THE FOLLOWING TEST THE FEEDBACK PROVIDED WILL MAKE SENSE TO YOU; AT OTHER POINTS, THE FEEDBACK WILL CONTRADICT WHAT YOU HAVE PREVIOUSLY LEARNT. THIS IS PART OF THE TASK AND YOU WILL BE REQUIRED TO RESPOND IN DIFFERENT WAYS.

IT IS IMPORTANT THAT YOU MAKE FEW ERRORS AND RESPOND AS QUICKLY AS YOU CAN. TO DO THIS, TRY TO FOCUS ON ACCURATE RESPONDING FIRST AND FOREMOST; WHEN YOU ARE GIVING A CORRECT RESPONSE IN THE MAJORITY OF THE TRIALS, TRY TO SPEED UP YOUR RESPONDING UNTIL YOU TAKING AS LITTLE TIME TO RELATE THE WORDS AS POSSIBLE.

GO WELL!

PLEASE PRESS THE SPACE BAR TO CONTINUE WITH THE TEST AND SUMMON THE EXPERIMENTER

WHEN PROMPTED TO DO SO.

When the participant summons you, say the following:

"Thank you for completing the test. Please turn the page of the written instructions and read them while I load the next programme."

While the participant is reading the written instructions, load whichever testing programme the participant is to take.

YOU ARE ABOUT TO READ A WRITTEN NARRATIVE. PLEASE ATTEND TO WHAT IS PRESENTED IN THE NARRATIVE AS YOU WILL BE TESTED ON THIS AT A LATER POINT IN THE EXPERIMENT.

The participant should then read the narrative. They will then be prompted to turn the page and complete several tasks.

THANK YOU FOR READING THE NARRATIVE AND COMPLETING THE TWO ACCOMPANYING TASKS. PLEASE SUMMON THE EXPERIMENTER.

Check the participant's responses to the memory checks. If their answers are not consistent with what they have been exposed to, provide the participant with another memory-check question sheet and ask them to re-read the narrative before answering the questions for a second time. If the participant is still not answering in a consistent way then they should be thanked for their involvement in the experiment and told that their participation has ended.

Any remaining participants should then turn the page of their written instructions.

YOU ARE ABOUT TO TAKE A COMPUTER-BASED LEARNING TASK WHICH INVOLVES READING COMBINATIONS OF TWO WORDS THAT ARE PRESENTED SIMULTANEOUSLY ON THE SCREEN. YOUR TASK IS TO STATE WHETHER THE WORDS ARE SIMILAR IN MEANING OR OPPOSITE.

FOR EACH TRIAL YOU WILL NEED TO LOOK AT THE WORD AT THE TOP OF THE SCREEN, THE WORD PRESENTED BELOW IT, AND THEN THE WORDS "SIMILAR" AND "OPPOSITE" WHICH WILL BE PRESENTED HORIZONTALLY AT THE BOTTOM OF THE SCREEN. USING EITHER THE "D" KEY OR THE "K" KEY YOU MUST THEN CHOOSE ONE OF THE TWO WORDS (OR RESPONSE OPTIONS) THAT ARE PRESENTED AT THE BOTTOM OF THE SCREEN.

WHEN MAKING YOUR RESPONSE, USE EITHER THE INDEX OR MIDDLE FINGER OF YOUR LEFT HAND TO PRESS THE "D" KEY AND EITHER THE INDEX OR MIDDLE FINGER OF YOUR RIGHT HAND TO PRESS THE "K" KEY.

THE RESPONSE OPTIONS "SIMILAR" AND "OPPOSITE" WILL NOT REMAIN IN THE SAME PLACE OF THE SCREEN FOR EACH TRIAL: THESE WORDS WILL SWAP PLACES IN A WAY THAT CANNOT BE PREDICTED.

IN ORDER FOR YOU TO RELATE THE VARIOUS WORDS PRESENTED SIMULTANEOUSLY ABOVE THE TWO RESPONSE-WORDS THE COMPUTER PROGRAMME WILL PROVIDE FEEDBACK BY TELLING YOU IF YOU HAVE MADE A CORRECT OR INCORRECT RESPONSE. AN INCORRECT RESPONSE IS INDICATED BY THE PRESENTATION OF A RED CROSS.

IT IS IMPORTANT THAT YOU MAKE FEW ERRORS AND RESPOND AS QUICKLY AS YOU CAN. TO DO THIS, TRY TO FOCUS ON ACCURATE RESPONDING FIRST AND FOREMOST; WHEN YOU ARE GIVING A CORRECT RESPONSE IN THE MAJORITY OF THE TRIALS, TRY TO SPEED UP YOUR

RESPONDING UNTIL YOU TAKING AS LITTLE TIME TO RELATE THE WORDS AS POSSIBLE.

GO WELL!

PLEASE PRESS THE SPACE BAR TO CONTINUE WITH THE LEARNING TASK AND SUMMON THE EXPERIMENTER WHEN PROMPTED TO DO SO.

When the participant summons you (having finished the training IRAP), load the testing IRAP and advise the participant to turn to the next page of the written instructions:

"Thank you for completing the training task. Please read the next page of instructions."

YOU ARE ABOUT TO TAKE A COMPUTER-BASED TEST THAT IS SIMILAR IN APPEARANCE TO THE COMPUTER-BASED LEARNING TASK THAT YOU TOOK EARLIER.

AS BEFORE, THE COMPUTER PROGRAMME WILL ASK YOU TO FOLLOW A SERIES OF RULES WHICH WILL BECOME APPARENT THROUGH THE FEEDBACK THAT IS GIVEN FOLLOWING EACH INCORRECT RESPONSE.

AT SOME POINTS IN THE FOLLOWING TEST THE FEEDBACK PROVIDED WILL MAKE SENSE TO YOU; AT OTHER POINTS, THE FEEDBACK WILL CONTRADICT WHAT YOU HAVE PREVIOUSLY LEARNT. THIS IS PART OF THE TASK AND YOU WILL BE REQUIRED TO RESPOND IN DIFFERENT WAYS.

IT IS IMPORTANT THAT YOU MAKE FEW ERRORS AND RESPOND AS QUICKLY AS YOU CAN. TO DO THIS, TRY TO FOCUS ON ACCURATE RESPONDING FIRST AND FOREMOST; WHEN YOU ARE GIVING A CORRECT RESPONSE IN THE MAJORITY OF THE TRIALS, TRY TO SPEED UP YOUR RESPONDING UNTIL YOU TAKING AS LITTLE TIME TO RELATE THE WORDS AS POSSIBLE.

GO WELL!

PLEASE PRESS THE SPACE BAR TO CONTINUE WITH THE TEST AND SUMMON THE EXPERIMENTER WHEN PROMPTED TO DO SO.

When the participant summons you, thank the participant for their involvement and provide them with a lottery ticket if they have opted to receive one.