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.

Investigating influences of incentives on implicit attitudes toward body size

A thesis

submitted partial fulfilment

of the requirements for the degree

of

Masters of Applied Psychology

at

The University of Waikato

by

Tokiko Taylor

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2017

Abstract

The Implicit Relational Assessment Procedure (IRAP) was designed to detect relational responding that cannot easily be accessed via traditional survey methods. The IRAP requires participants to meet speed and accuracy criteria during practice trials before proceeding to test trials, which has resulted in an attrition rate of approximately 20%, on average, in the existing research. Variables affecting the attrition rate have not been systematically investigated. I examined the influence of incentives (in this case a $20 voucher contingent on meeting performance criteria) on attrition rate and other IRAP performance measures. In addition, I examined whether the IRAP would reveal an implicit anti-fat bias in 82 university students. I found significant differences in the performance of the incentive group compared to the control group in their response accuracy and measurement of their implicit bias. The results indicated higher levels of bias compared with those from previous research studies, particularly in the incentive group. I did not find statistically significant differences in the attrition rate but found a low attrition rate in both groups. This study reveals the utility of incentives for improving performance on the IRAP, a procedure that demands accurate responses under time pressure for assessing spontaneous relational responding.

Acknowledgements

I would first like to thank my advisors, Dr Timothy Edwards and Dr Rebecca Sargisson, for their support and guidance. Their insightful and thought-provoking feedback were my guiding light throughout my research. I am thankful to my fellow students for their warm encouragement and support and to Judy for helping me with proofreading and feedback. Finally, a special thanks to my son, Louis, who went to school every day without being sick and went to bed early so his Mum could focus on her study. Without your support, I wouldn't have been able to finish writing up my thesis. Thank you all.

Table of Contents

Lists of Figures and Tables

Figures

Tables

Introduction

Attitudes and mentalism

Generally, 'attitudes' and 'values' are regarded as elusive concepts. Some of us consider them as mental entities that exist in our minds and govern our behaviour. This perspective is based on a simple linear causality of events, executing a 'you-say-you-do' programme, which is a commonly accepted notion that suggests if anyone wants to change their behaviour, they must change their attitudes first. I remember my elementary school teacher telling me, "Change your attitude!" and expecting me to behave properly and conform to the school's educational standards. According to Lloyd (1994), the "change-attitude-then-behaviour principle" (p.132) is widely accepted, even among psychologists, who have conducted various correlational studies focusing on relationships between attitudes and behavioural outcomes. Lloyd (1994) critically remarked that such "say-do studies" (p.136) were mostly based on heuristics that purported linear causality between verbal and non-verbal behaviour; he explained the high appeal of the "change-attitude-then-behaviour-principle" as resulting from the perception that it seems to be much easier and simpler to change attitudes (verbal behaviour) than to change non-verbal behaviour, which usually results in some resistance (Lloyd, 1994). The relationship between attitudes and behaviour is probably mediated by multiple internal and external stimuli that are not congruent with each other all the time.

Behavioural psychologists avoid "mentalisms" that attribute causality of events to a person's inner/mental elements. According to Baum (2005), what we call attitudes and values are all behaviour. Moore (2015) asserts that attitudes are nothing but "the probability of certain kinds of actions in certain kinds of

circumstances" (p.105). Attitude is not a mental entity that can cause certain behaviour. What we call 'attitudes' can be explained in behavioural terms as being generated by a person's experiences in the past and by current environmental variables (Baum, 2005; Moore, 2015).

According to Guerin (1994), attitudes and beliefs are verbal behaviour, just like non-verbal behaviour, which is shaped by operant behaviour and maintained by personal verbal communities that serve as discriminative stimuli for reporting one's own specific attitudes. Verbal behaviour is controlled by a non-verbal stimulus and maintained by social reinforcement (Skinner, 1957). Therefore, Guerin (1994) argues that an attitude is a mere tact of external and internal events. A tact is a category of verbal behaviour proposed by Skinner (1957), which refers to the act of labelling when a stimulus is presented or a description of a specific event as an observer. For example, uttering a statement like "Mum, I see a dead pukeko on the road!" is a tact; the verbal reporting itself is often reinforced by a listener (e.g., a mother) paying attention to the child for reporting an event. Then, the simple tact is eventually developed into a more complex form of tact by expressing one's own feeling or preference such as "Poor pukeko, I will rescue them if I see them on the road again," which results in receiving social reinforcement from the child's audience for expressing care for the animal. Our attitudes or evaluations of events (e.g., attributing values such as good and bad, like and dislike) are social behaviours shaped by our verbal communities. The mand is one category of verbal behaviour that refers to asking for something when one is controlled by deprivation or satiation (i.e., motivating operations). Guerin (1994) points out that an expression of attitude can function as a mand, which the author calls an "attitudinal mand" (p.157). For example, if a

boss says that he does not like anyone who does not recycle, his statement may change his workers' recycling behaviour at their workplace. Thus, attitudes are largely contingency-driven, meaning that behaviour is determined by its functionality (e.g., social attention or escape) and the motivating operations relevant to those functions. In the above example, if the statement were made by a co-worker, it would likely be less effective in eliciting recycling behaviour among other workers.

Explicit and implicit measures of attitudes

The study of attitudes has mainly been carried out by social psychologists. Around WWII, such studies were conducted to gauge people's attitudes toward the war effort (Moore, 2015). Attitude was defined based on three components including affect (i.e., mainly feelings about the evaluation of an event), cognition, and behaviour (Fazio & Olson, 2003). Measurements of attitudes have been commonly obtained from explicitly designed self-reports or questionnaires answered directly by study participants (Fazio & Olson, 2003; Ferguson & Bargh, 2007; Moore, 2015). The self-reporting method poses validity problems in terms of social contingencies involved during the process. The following quote from a website seems to depict the nature of attitudes very well. It shows that attitude is context-dependent and can be largely categorised into two separate aspects; one is 'explicit' and the other is 'implicit' attitude: "Attitude is like posing for a picture. We pose the way we want to be seen by others. But stolen shots are better, they capture the real you" (posted by Tayyeb S, 2014).

The first line in the quote is a good analogy for explaining direct measurement procedures such as self-reports, Likert scales, and questionnaires of attitudes, which are sometimes driven by social contingency; we tend to change

our attitudes depending on our audience and the situational variables affecting our private and public behaviour at the time. The second line suggests another aspect of attitudes that is free of social contingency; many researchers have attempted to capture the implicit attitudes inferred from physiological responses (such as skin contractions and blink rate) and/or latency responses in their participants (Vanman, Paul, Ito & Miller, 1997; Hart, Whalen, Shin, McInerney, Fischer& Rauch, 2000). Such behavioural outcomes are usually automatic and happen in the blink of an eye, not allowing enough time for participants to think of a socially appropriate response.

The quote above emphasises the authenticity of the captured implicit moment that may reveal *the real you*; however, an implicit measure is not necessarily a truer measure, nor does it capture the unconscious construct of a person (Fazio & Olson, 2003). Fazio and Olson (2003) argue that both explicit and implicit attitudes do "pre-exist" in one's memory (p.303). As seen in the quote from Fazio and Olson (2003), the notion "pre-exist" could be interpreted as somewhat mentalistic in the way that it explains where our attitudes "exist" in our cognition. It is also quite vague about defining the precise mechanisms for how attitudes are formed and maintained.

Functional contextualism is a behavioural analytic approach to seeking explanations for why-we-do-what-we-do by examining the context in which the behaviour is generated along with the purpose or function of the act (Biglan & Hayes, 1997; Hayes, Blakledge, & Barnes-Holmes, 2001). The analytic goal is to predict behaviour and its change by assessing and manipulating contexts (Barnes-Holmes, Barnes-Holmes, Power, Hayden, Milne, & Stewart, 2006; Hughes & Barnes-Holmes, 2011; Hughes, Barnes-Holmes, & De Houwer, 2011). The truth

criterion of functional contextualism is whether or not expected change has occurred after manipulating specific environmental variables. According to Biglan and Hayes (1996), the goals of prediction and influence are not separate criteria. There are two criteria to achieve prediction: "(a) it identified variables that permitted the prediction of the event in questions and (b) the identified variables would, if they could be manipulated, affect the probability of the event" (Biglan & Hayes, 1996, p.51). In short, through the conceptual framework, we look at how events (e.g., overt and covert verbal behaviour) function in specific contexts, identify variables to affect behaviour and test to see if it actually works or not by manipulating the identified contexts.

Correlational studies of attitudes are not considered to reside within the framework of functional contextualism because in correlational studies, any changes or influences on the behaviour under investigation would not be demonstrated; i.e., would not achieve the goal of prediction and influence (Hayes, 1993; Gifford & Hayes, 1999). However, according to Bigland and Hayes (1996), we should not discard correlational studies altogether from empirical studies based on functional contextualism, because such studies provide "clues as to the environmental events" (p. 51) or clues to identify variables that predict and influence behaviour. Within the framework of a functional approach, researchers have investigated attitudes with the analytic goal of identifying and manipulating specific mental constructs to predict and to influence behaviour (Hughes & Barnes-Holmes, 2011).

Consequently, researchers who study implicit attitude (or implicit evaluation) using the functional approach (i.e., using functional contextualism) tend to treat attitudinal behaviour as contingency-based (i.e., controlled by

contextual and functional variables) both in private and in public, which means that we can conduct language and cognition analysis by studying and manipulating contextual and functional variables. As indicated in the posing-for-a-picture quote above, depending on the contingency in which a person finds themselves, their overt attitudinal outcome can vary accordingly. The Implicit Relational Assessment Procedure (IRAP), which is based on behavioural theory (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008), can be used to study the influence of variables such as time, verbal communities (e.g., being with people or alone), and one's history of reinforcement for relevant behaviour. This approach is in line with Moore's (2015) claim that attitudes can be used as valid predictors of human behaviour, which corresponds to the analytic goal of functional contextualism: we can see that both explicit and implicit measurements of attitudes are significant indicators that signal the probability of a particular class of behaviour.

Explicit and implicit relations

Numerous researchers have investigated whether explicit and implicit attitude measurements correlate with one another or whether they are incongruent. Recent studies indicate that the relationships between explicit and implicit measures correlate more strongly when measuring socially non-controversial items (Fazio, Jackson, Dunton, & Williams, 1995; Fazio & Olson, 2003; Greenwald & Banaji, 1995) including phobia (Hughes et al., 2013). In implicit attitude studies using the Implicit Association Test (IAT), researchers generally found high correlations between explicit and implicit measures when used in a preference study on non-controversial stimuli such as flowers vs. insects (Greenwald et al., 1998). In contrast, if the two measured outcomes demonstrate

an incongruent result, the targets are often socially controversial issues such as racial profiling, the death penalty, global warming, or abortion. Explicit measurement is credited for its easy application and convenience in measuring people's beliefs and attitudes; however, there are issues relating to accuracy in self-reporting and, as a result, explicit measures for assessment of attitudes related to socially controversial issues are problematic. For example, people with paraphilia are secretive because of cultural taboos and legal sanctions; therefore, any information obtained in initial interviews during the assessment phase, along with self-reported data during the treatment, may be unreliable because self-preservation is a prime motivation for some offenders. Therefore, implicit measures may indicate attitudes less contaminated with socially appropriate answers or any other thoughts that lead to more discrepancy between private and public behaviour.

Implicit Relational Assessment Procedure (IRAP)

Various implicit measurement tools have been developed, including the word-fragment completion task (Warrington & Weiskrants, 1968; Bassili & Smith, 1986); the name-letter-preference task (Nuttin, 1985); Go/No-Go Association Tasks (Nosek & Banaji, 2001); and the IAT (Greenwald, McGhee, & Schwartz, 1998). These tests are mainly based on the assumption that the strength (or the degree of cognitive persistence) of a paired-stimulus association can be measured by participants' response latencies in time-pressured categorization tasks (Golijiani-Moghaddam, Hart, & Dawson, 2013). For example, the IAT (Greenwald et al., 1998) is designed to measure "automatically activated evaluation" (p.553) using a series of computerized categorization tasks. Participants are presented with task instructions (i.e., two opposing category

names such as a combination of “Fat” and “Thin” or “Pleasant” and “Unpleasant”), then provided with a sample stimulus (e.g., a series of pictures representing overweight and thin individuals or a list of adjectives conveying a sense of pleasant or unpleasant) at the centre of the monitor screen and asked to categorize the stimulus quickly and without errors by pressing the left or right key (see Figure 1.1). If the average response latency time is faster for the stimulus combinations of “fat – unpleasant” than for “fat – pleasant,” then an inference that the former association may reveal an implicit attitude is made.

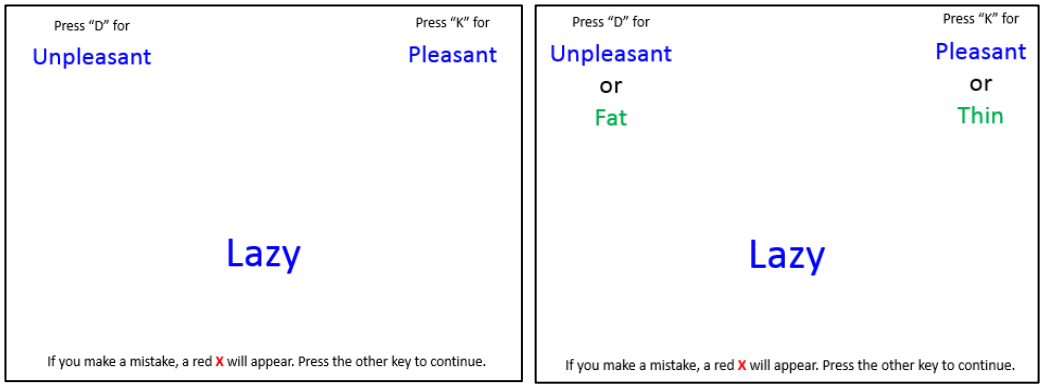


Figure 1.1. Hypothetical computer screen images of the IAT. Panel A demonstrates a simple categorization task. A participant is instructed to press either ‘D’ or ‘K’ keys on a keyboard to respond correctly and quickly to which category of either “Unpleasant” or “Pleasant” a presented sample stimulus (e.g., “Lazy”) belongs. The correct answer is to press the “D” key that belongs to the “Unpleasant” category. The sample stimulus presented in the centre of the monitor can be a picture image that can be categorized as either “Fat” or “Thin” in the same manner as explained previously. Panel B represents one of the IAT test blocks in which dual categories (e.g., “Unpleasant” and “Pleasant”, plus “Fat” and “Thin”) appear at the top left and right corners. A participant is instructed to press either the ‘D’ or ‘K’ keys on a keyboard to respond correctly and quickly in the same manner as described above.

The IAT is quite similar to IRAP; however, there are two major differences between them. First, the IAT can only provide measurements in two dimensions, which are “Fat-Unpleasant/Thin-Pleasant” and “Fat-Pleasant/Thin-Unpleasant” in the example above. Therefore, the IAT has been criticised because

it cannot reveal to what degree Stimuli A and B are liked or disliked (De Houwer, 2002; Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009). The IRAP is designed to measure four dimensions of such attitudes to reveal independent measures of "Fat-Unpleasant," "Fat-Pleasant," "Thin-Unpleasant," and "Thin-Pleasant." Thus, we can assess the directions of attitudes among four trial types.

Secondly, the IRAP is the only measure that is based on behavioural theory. It is built upon the Relational Frame Theory (RFT) proposed by Steven Hayes (Fletcher & Hayes, 2005), which is a behavioural analysis of language and cognition. In the IRAP measure, automatic evaluation is not established through simple association in memory, but rather is derived based on a few directly learned stimulus associations that have been formed earlier through exposure to multiple exemplar training. In contrast, the IAT measures associations without a conceptual framework that can help to explain the technical process of how the stimuli become associated and why the outcome occurs (De Houwer, 2002; Barnes-Holmes, Barnes-Holmes, Power, Hayden, Milne, & Stewart, 2006; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010).

Emergence of IRAP

Prior to the development of RFT, numerous empirical studies attempted to identify the mechanism of language acquisition, especially with regard to stimulus association. According to Burkholder and Pelaez (2000), Vygotsky (1962), who was a developmental psychologist, had a notion of the emergence of functional aspects of language through 'combining' one's language (e.g., cooing, crying) and needs (e.g., to be fed, or to get comfort) to manage social-environmental contingencies. This was similar to the verbal behaviour theory proposed by Skinner (1957). Some empirical studies were conducted to demonstrate different variables affecting the act of 'combining' symbols and needs (Barnes-Holmes,

Barnes-Holmes, Roche & Smeets, 2001; Barnes-Holmes et al., 2001; Brady & McLean, 2000; Hayes, & Hayes, 1993; Lipkens, Lowe, Horne, Harris, & Randle, 2002; Sidman, Willson-Morris, & Kirk, 1986;). Researchers such as Sidman (1986), Horne and Lowe (1996, 1997, 2002), Hayes (1994, 2001) and many other behaviour analysts who conducted empirical studies on stimulus equivalence attempted to identify the core factors responsible for the emergence of stimulus equivalence and theorized about the relationship between language development and stimulus equivalence.

Stimulus equivalence was extensively studied by Sidman (1982) and other researchers, including Steven Hayes and Dermot Barnes-Holmes. According to Sidman, there are three defining mathematical formulae to explain behavioural conditional discrimination: (1) A = A: Reflexivity, which indicates a relation in which a stimulus "A" is to be matched to itself; (2) If A = B, then B = A: Symmetry, indicating a reversal of a learned association; (3) If A = B and B = C, then A = C: Transitivity, referring to a member of an equivalence class (e.g., A=C) that emerges out of learned associations of the two different conditional discriminations (e.g., A=B and B=C). If an individual can derive all three relations, that individual is developing stimulus equivalence relations and forming an equivalence class.

Sidman (1982) mainly focused on describing the phenomenon of stimulus equivalence, whereas Hayes (1993) theorised a mechanism for the emergence of stimulus equivalence and other relations through his Relational Frame Theory (Gross & Fox, 2009). Hayes (1994) extended RFT to cover pragmatic aspects of language and cognition, suggesting that many types of private and public behaviour are based on derived relations instantaneously obtained from previously

learned associations. In addition, RFT led to the development of clinical benefits in RFT-based therapies (e.g., Acceptance & Commitment Therapy). The core elements of RFT are arbitrarily applicable relational responding that is defined by three core properties: mutual entailment, combinatorial entailment and the transformation of stimulus functions (Hayes et al., 2001). Mutual entailment refers to the bi-directional relation between two stimuli. For example, if one learns that stimulus A is the same as stimulus B, then one will derive that stimulus B is the same as A, which occurs without direct training. Combinatorial entailment arises from derived relations among two or more stimuli. For example, if A is bigger than B and C is smaller than B, then the derived relation can be that A is bigger than C. A subjective relational network linking various stimuli is formed. Many other relations (e.g., opposition, comparison, before/after) can be trained, and relevant relational responding associated with these relations can emerge instantaneously. Moreover, one significant aspect of RFT is a transformation of stimulus function that may take place among derived relations. For instance, if a person is scared of the sight or the sound of bees, the negative emotion attached to the insect can be transferred to another stimulus such as being in the woods or in someone's bushy backyard through its derived relations. The person feels scared being in the backyard without the presence of an actual bee. The transfer of function can occur without any direct training.

Sometimes, the emergent relational network becomes irrational (e.g., developing a fear of something that you have never been exposed to directly, or fearing that eating a slice of pizza will cause weight gain; see Figure 1.2). Fletcher and Hayes (2005) point out that such patterns of relational framing occur on their own and can easily become detached from the real physical or social environment.

Once such patterns get off track (creating psychopathology or irrational thoughts), people suffering from anxiety (or having anxiety-related disorders) can develop a tendency to adhere to an old paradigm (e.g., prediction of fear, or suppressing thoughts). Such adherence develops into a bias, which may lead people to be "psychologically inflexible." Barnes-Holmes (2006, 2010, 2013), Barnes-Holmes (2006, 2010), Power (2006), Hayden (2006), Milne (2006), Stewart (2006, 2010), Boles (2010), and Hughes (2013) regard psychological inflexibility observed among people as a bias or a "historically coherent relational network" (De Houwer, 2002; Barnes-Holmes et al., 2006; Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2013). The analytical goal of IRAP is to capture such inflexibility in people's perceptions.

Hughes and Barnes-Holmes (2013) define historically coherent relational networks as "brief and immediate relational responses (BIRRs)" (p.102) that are usually learned previously and have been through multiple instances of exemplar training of particular relations to certain stimuli (e.g., 'being obese is bad'), so they occur within a few seconds of time and precede "extended and elaborated relational responses (EERRs)" (p.102). The EERR represents the explicit attitude and it takes longer to occur because of the time spent searching for a socially appropriate response (Barnes-Holmes et al., 2006; Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2013). The researchers emphasise the use of the term "brief and immediate" instead of automatic or implicit because behavioural analysis usually excludes cognitive mediating processes and the functional definition of automatic and implicit is still somewhat mentalistic (Hughes & Barnes-Holmes, 2013; Moore 2012). However, Moore (2015) acknowledges the implicit measures of BIRRs (e.g., "A hairy worm is unpleasant") and EERRs

(e.g., “A hairy worm is pleasant” to impress one’s peer group at school) indicate a behavioural process that tracks a person’s actions (responses) as a result of contingencies. The measured response latency demonstrates that it simply takes longer to respond due to “covertly doing more” (p. 111), if a person finds a pair of stimuli to be inconsistent with their historically coherent relational network, which is a strongly held belief or bias (see Figures 1.2 and 1.3).

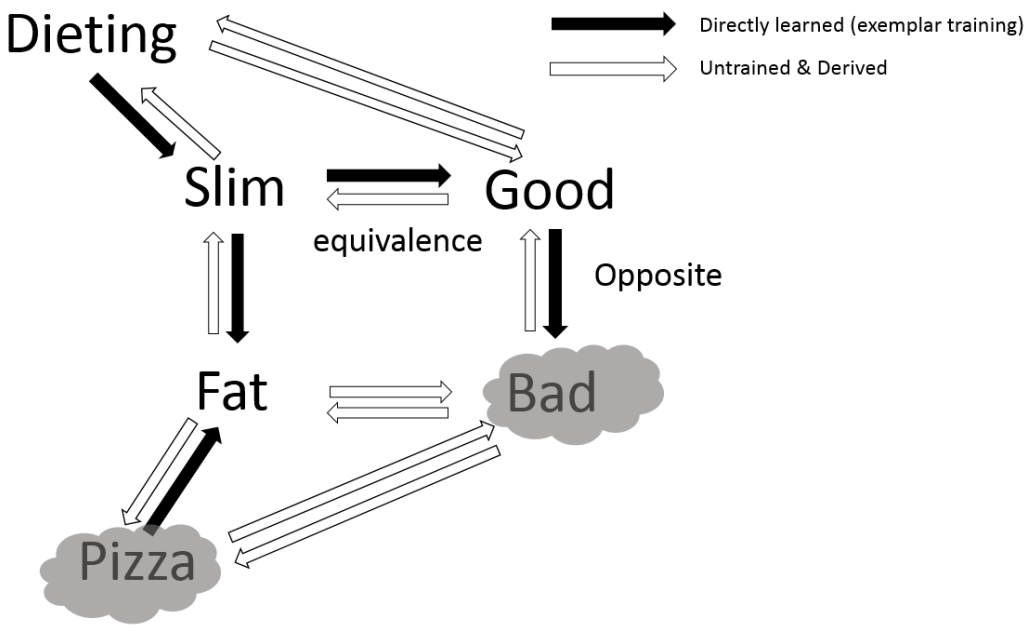


Figure 1.2. Relational frame theory (RFT) examples. For instance, we learn to associate the word “slim” with “good” through exemplar training, in which a stimulus equivalence is formed. At the same time, an association of “Good” with “Slim” is derived without training. According to RFT, this is called mutual entailment. In a similar manner, using the frames of opposition, we tend to learn the association between “Good” and “Bad” quite early in life (e.g., “Slim” and “Fat” are directly learned in the same frame). Then, instantly, “Fat” and “Bad” is derived because of the pre-established relational framing. Just a few instances of relational framing create networks of arbitrary relations that become the basis of our thoughts and feelings. The clouds represent negative valence (i.e., anxiety) attached to the word “bad,” which was learned directly in the past. Since there is already derived equivalence between “pizza” and “bad,” the negative cloud of feeling is automatically transferred to the word “pizza,” a process which is defined as transformation of stimulus functions in RFT, and can explain unlearned (not directly exposed) fear of some objects (e.g., beans, a snake, or bees). So, in this case, anyone who is dieting may feel afraid to eat a slice of pizza by believing that it would cause weight gain.

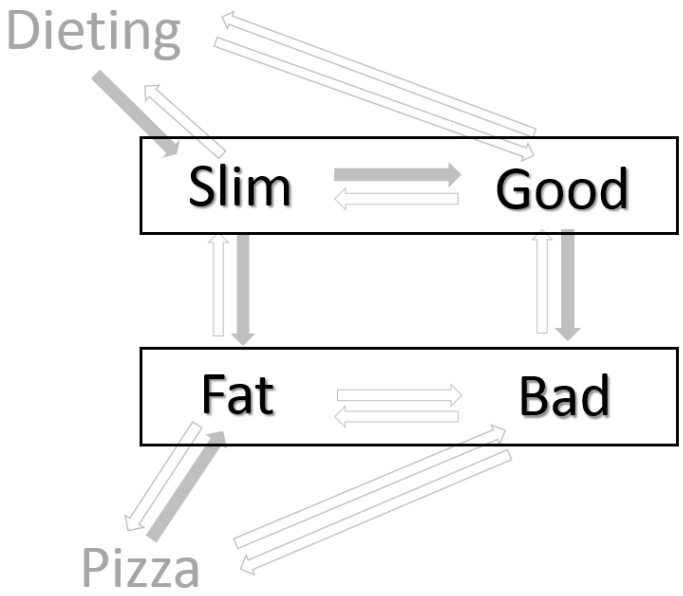


Figure 1.3. IRAP filters out the historically coherent relational network that is marked in black rectangles (see below for more explanation), which is regarded as 'bias' or 'automatic evaluation.' When a person responds to "Slim is Good," it is faster than their response to "Slim is Bad." We could infer that the person is cognitively doing more to respond to the latter association (Moore, 2015), indicated in the diagram as "two steps away" in order to reach the word "slim" related to "bad" compared to "slim" related to "good."

The psychological inflexibility observed in an individual's responding, which was explained as "historically coherent relational networks" by Barnes-Holmes and his fellow researchers (De Houwer, 2002; Barnes-Holmes et al., 2006; Barnes-Holmes et al., 2010; Hughes & Barnes-Holmes, 2013), was identified thorough laboratory-based empirical studies testing RFT. Barnes and Roche (1997) stated that "a person's prior history of social interaction may be revealed in the course of analysing derived relational responding" (Barnes & Roche, 1997, p.122). For example, in the RFT experiment conducted by Watt, Keenan, Barnes, and Cairns (1991), such derived relational responding was revealed. Their participants were English and Northern Irish people and were involved in two different equivalence training sessions explained below.

Equivalence Training 1: Catholic names with nonsense syllables

(matching-to-sample procedure)

Equivalence Training 2: Protestant symbols (e.g. Union Jack, Orange order etc.) with the same nonsense syllables used in Training 1.

The researchers tested for derivation of transitive responding - by presenting the Protestant symbol, the subjects were supposed to choose the corresponding Catholic name. Interestingly, the English subjects showed transitive responding by selecting the Catholic name, but Northern Irish subjects were not successful in transitive responding (12 out of 19 failed); most of them selected a Protestant symbol instead from three answer-options that consisted of two Catholic names and one novel Protestant symbol not used in the training. In other words, the Northern Irish subjects demonstrated their strong pre-established social learning that *Catholic was not equivalent to Protestant* (based on the historical context of their regional political/religious conflicts). The RFT equivalence training could not override the pre-established notions. Roche and Barnes-Holmes (1997) commented that this was just "one way of tapping into" pre-established verbal relations, and that there was a need to develop a "behavior-analytic methodology of measuring attitudes in a valid and reliable fashion" (p.125). Later, Barnes-Holmes and his fellow researchers created the IRAP program.

IRAP and research questions

IRAP is a computer-based task where participants are required to respond to a series of paired stimuli (i.e., words and images) by selecting true or false (e.g., same or opposite) as response options to whether a pair of words are consistent or inconsistent with the subject's historically coherent relational network. The participants are put under pressure to respond both quickly and accurately. The IRAP programme consists of practice blocks (usually three paired

blocks of consistent and inconsistent trials, each block containing about 24 trials) and test blocks (the same as the practice blocks). All response latencies are recorded in the IRAP program and used to calculate the D-IRAP score. The score algorithm is based on the improved IAT algorithm (Greenwald, Nosek & Banaji, 2003) introduced since the first introduction of IAT in 1998 (Greenwald, McGhee, & Schwartz, 1998). It consists of normalized indices of raw IRAP response latency (see Method section for the detailed translation process), which is similar to Cohen's effect size (Cohen's D). It indicates whether or not the differences in averaged response latency between consistent and inconsistent blocks are large enough to conclude that there is a real effect. The larger the D score value, the bigger the effect size (e.g., 0 is non-significant, and 0.6 to 0.8 indicates a large effect size).

Recently, using implicit measures, numerous studies have been conducted to measure anti-fat bias (Bessenoff & Sherman, 2000; Brochu & Morrison, 2007; O'Brien, Hunter, & Banks, 2007; Roddy, Stewart, & Barnes-Holmes, 2010; Teachman & Brownell, 2001; Expósito, López, & Valverde, 2015). There are growing concerns about how ideas of ideal body image (e.g., "thin ideal") are affecting people, especially teens (O'Brien et al., 2007), primarily in developed countries. The "thin ideal" may lead to eating disorders and low self-esteem and may limit the healthy growth of mind and body. Most of the anti-fat bias studies used IAT and a few have used IRAP. The studies typically found that people responded faster to 'slim' associated with positive words and 'fat' associated with negative words. Roddy et al. (2010) compared the IRAP and IAT in their study. Participants were presented with one of two relational response options (either the word "similar" or "opposite") and with one of a series of pictures of upper body

shots of either overweight or normal-weight men and women. The set of stimuli they used was taken from a previous implicit anti-fat bias study by Brochu and Morrison (2007). The paired-stimuli samples produced four trial types (slim-positive, fat-negative, fat-positive, and slim-negative) that were presented in both consistent and inconsistent blocks alternatively. The participants were asked to answer as if they were anti-fat and pro-slim in consistent trials and to answer with the opposite attitude in the inconsistent trials. The result indicated that participants responded faster in the consistent trials, producing the same result observed in their IAT test. However, the study did not show much difference in their responses between the fat-positive and fat-negative trials, which means that the participants were neither pro- nor anti-fat, but had a slightly more pro-fat attitude as evaluated by the picture stimuli used. A potential limitation of Roddy et al.'s (2010) study was the quality of the picture stimuli, which may have been somewhat obscure in the difference between overweight and normal weight individuals; also the people in the upper-body photos were all smiling which may have evoked different responses in the participants.

In an anti-fat study using IRAP, Expósito et al. (2015) attempted to investigate the attitudes of a young Spanish female population and used 12 target photographs of overweight and normal-weight individuals (the model's faces were blurred and non-recognizable). Results demonstrated that the participants responded faster for both the pro-slim and the anti-fat attitude, which is consistent with the results from Roddy et al. (2010); however, Expósito et al. (2015) also found a moderate pro-fat attitude which was significantly different from zero. According to the researchers, the moderate pro-fat attitude may be the result of

using blurred facial images to control participants' sensitivity toward attractiveness bias in picture stimuli.

Ferguson and Bargh (2007) specifically commented with regard to the implicit response that it "[does] not depend solely on the stimulus in isolation" (p.222). In reality, people are largely affected by the "rich context in which the stimulus is perceived" (p.222). The authors provided an example to explain the concept that pictures of a black person having a picnic and the same person standing in front of a graffiti-covered wall in the city elicit different responses from viewers. Therefore, I used silhouette images for my IRAP test stimuli to investigate the anti-fat bias, to eliminate 'noise' from compound stimuli (e.g. photo images) that may trigger divergent responses among participants, which may reduce the precision of the IRAP measurement.

Although there is an increasing trend toward using the IRAP for research purposes, only a small body of research has scrutinised the tool's validity and reliability. For example, unlike IAT, almost all studies indicate that the IRAP is not easily faked (McKenna, Barnes-Holmes, Barnes-Holmes, & Steward, 2007). Only one study countering McKenna et al. (2007) has been conducted so far (Drake, Seymour, & Habib, 2016). Drake et al. (2016) tested whether participants could fake the IRAP and established that the procedure could be faked if researchers provided the mechanism of the IRAP assessment strategy to their participants and asked them to fake by altering their response latency during the designated trial blocks. Another issue that could be a potential limitation to the use of the IRAP is attrition (Golijani-Moghaddam, Hart, & Dawson, 2013). The IRAP test requires participants to meet performance criteria during practice trials before proceeding to test trials. Most researchers specify that participants must

achieve a median response latency of 2,000 ms or less, with at least 80% response accuracy. To date, about 20% of participants in each study fail to pass the practice response criteria (see details in Hughes & Barnes-Holmes, 2013, p116-117). For example, Roddy et al.'s (2010) IRAP experiment on body-size bias resulted in 16 out of 80 participants failing to pass (i.e., 20% attrition rate) and 8 out of 40 participants (20%) failed in Expósito et al.'s (2015) study. So far, variables affecting the attrition rate have not been systematically investigated (Golijani-Moghaddam, et al. 2013; Hussey, Thompson, McEnteggart, Barnes-Holmes, & Barnes-Holmes, 2015).

Therefore, the independent variable in this research was the incentive for participants to improve their IRAP task performance. Money is the easiest and most commonly used form of incentive to promote research participation or improve work performance (Rudy, Estok, Kerr, & Menzel, 1994). A monetary reward is often used in research studies as an effective reinforcer to control behavioural outcomes. Such a mechanism to cause changes in behaviour is known as operant conditioning, which is a type of learning that occurs as a result of consequences of behaviour (Skinner, 1938). Monetary incentives may be effective positive reinforcers. For example, if a child receives $10 for washing a car, he or she is likely to repeat the task in the future (i.e., positive reinforcement). Gneezy and Rustichini (2000) conducted a simple experiment testing the effects of monetary incentives on the performance of 160 university students who were asked to correctly answer 50 questions taken from an IQ test. The participants were all paid 60 NIS (New Israeli Shekels), roughly equivalent to $20 NZD, for participating. Then, the researchers set up four groups to compare the participants' accuracy in providing correct answers: 1) the control group was asked to correctly

answer as many questions as possible; 2) a group was given additional payment of 10 cents per question answered correctly, 3) a group received an additional 50 cents (1 NIS) per correct answer; and 4) a group received an additional $1 (3 NIS) per correct answer. The study revealed a significantly higher accuracy among participants in Groups 3 and 4 in comparison with the control group and Group 2. Interestingly, the effect was consistent across high and low performances among their best and worst participants. Gneezy and Rustichini (2000) found that monetary incentives worked as effective reinforcers to change behavioural outcomes and to improve performance. In another study by Hayes, Brownstein, Zettle, Rosenfarb and Korn (1986), participants received points for pressing buttons and the points were exchanged for money at the end of their experiment. The maximum amount of money that the participants could earn from the experiment was $20. The researchers evaluated their participants' sensitivity to changes in contingency (i.e., offering different schedules of reinforcement) provided with four different verbal instructions. The participants were randomly assigned to one of the conditions, such as offering a minimal instruction to the participants, asking them to go slow or fast, and providing an accurate instruction on when to go slow or fast. They found that participants tended to adhere to the instructions, especially when they were verbally instructed to 'go slow.' On the other hand, without instructions (those who received minimal instruction), none of them earned enough points from the two programmed consequences (i.e., fixed-ratio and differential reinforcement of low rate). Thus, the verbal instructions overrode the programmed contingencies of pressing the button (i.e., different rate of scheduled reinforcement). Although the study by Hayes et al. (1986) demonstrated an influence of instructional control on human subjects who showed

insensitivity toward changes in contingency when provided with verbal instructions, the effects of monetary incentives on performance were not examined in their experiment.

In the present study, I focused on the influence of monetary incentives on participants' performance on the IRAP. I offered a $20 voucher as a performance-based monetary incentive to the participants in the experimental group in order to examine any changes in overall performance such as attrition rate, overall accuracy in response, and the latency difference between bias-consistent and bias-inconsistent trials (i.e., implicit bias) relative to measures obtained from the control group who were offered no incentive. I expected to see that attrition due to poor performance would be reduced by the $20 voucher in the group provided with the monetary incentive compared to the control group without the incentive. The attrition rate was also compared with the results from studies by Roddy et al. (2010) and Expósito et al. (2015). I also expected to see the latency differences (i.e., IRAP scores) of the incentive group diverge from those of the control group. Participants in the incentive group were expected to perform with higher accuracy and lower response latency, relative to the control group.

The present study also aimed to test the utility of IRAP in measuring implicit attitudes toward body size (slimness and fatness) among students at the University of Waikato. I expected to see some IRAP effect consistent with *anti-fat* and *pro-slim* attitudes in all four trial types, which would support the results of past research on prejudice toward overweight individuals (Bessenoff & Sherman, 2000; O'Brian, Hunter, & Banks, 2007; Teachman & Brownnell, 2001). In addition, I carried out a correlation analysis using the results from the implicit (IRAP) and explicit measures (Anti-Fat Attitudes (AFA: Appendix F), Behavioral

Intention Questionnaire (BIQ: Appendix G), and Body Mass Index (BMIs)). I expected to see a small divergence between the two measures (i.e., the results of the IRAP would indicate a strong *anti-fat* and *pro-slim* attitude and the AFA and the BIQ would indicate slightly less of an *anti-fat* and *pro-slim* attitude). Because anti-fat bias was considered to be a socially sensitive subject, responses from participants in this experiment would be affected by social contingency. The study of correlations among the participants' BMIs, IRAP, and other explicit measurements were largely exploratory at this stage.

Method

Participants

Eighty-two university students, ranging from 17-51 years of age (*M* = 23 years, 95% CI [21.72, 25.04]), participated in the study. The gender ratio was 77% female and 23% male and the majority were majoring in psychology in their undergraduate and postgraduate studies at the University of Waikato's Hamilton campus. Of the participants, 32% had previous exposure to implicit-measures procedures (i.e., IRAP). The participants were recruited through advertisements on bulletin boards and mail-group recruitment from the university's research participant database. No screening process took place in terms of the participants' cultural backgrounds, ethnicity, gender, or age. Anyone who consented to participate in the study was admitted. All participants were coded for anonymity and I did not make a name-matching list for reference because personal identification was unnecessary for this study. However, the code was marked on each consent form so that the researcher could track a participant's name if necessary. All participants were given the choice to receive a course credit (only for undergraduate psychology classes) or to enter a draw to win one of five $20 iTunes vouchers if they completed the experiment. The choice was offered at the end of the experiment.

Materials

Implicit Relational Assessment Procedure (IRAP).

In this study, I used the 2012 version of the IRAP programme (Barnes-Holmes & Hussey, 2012). I downloaded the free IRAP software from the following web site: https://contextualscience.org/implicit_relational_assessment_procedure_irap_web site.

I set it up for the experiment by following the 2012 IRAP manual (Barnes-Holmes & Hussey, 2012).

In this study, I used 12 silhouette stimuli as label stimuli (see Appendix C). The picture stimuli were six overweight male and female silhouette images and six slim male and female images. Each image was 235 x 300 ppi, which was the same size as the photo stimuli set used in by Roddy et al. (2010).

For consistency with previous research by Roddy et al. (2010), I used the same stimulus set that they did. The stimulus set contained 12 different adjectives with 'positive' or 'negative' connotations as the target stimuli in IRAP. The six positive-word stimuli were 'Desirable,' 'Active,' 'Disciplined,' 'Attractive,' 'Healthy,' and 'Good.' The six negative words were 'Undesirable,' 'Lazy,' Sloppy,' 'Ugly,' 'Ill,' and 'Bad.' In each IRAP trial, participants were provided with two relational response options; either "True" or "False" at the bottom of the screen (see Figure 2.1).

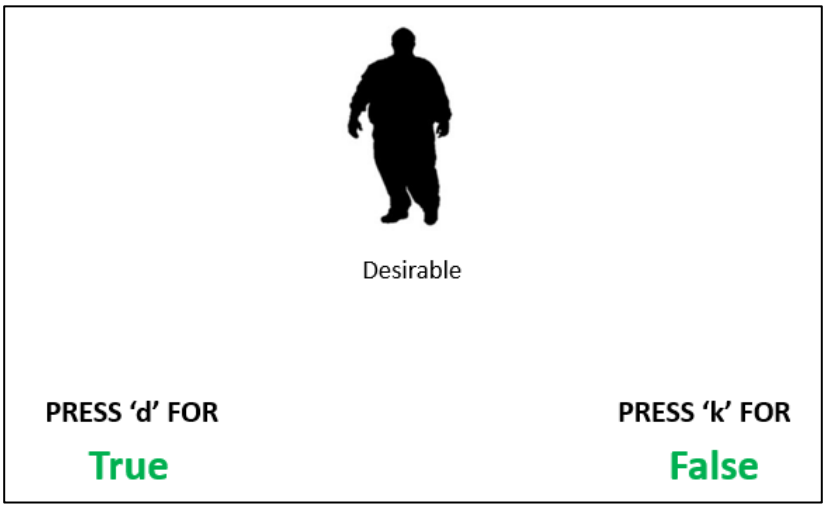


Figure 2.1. A screen shot image of a single trial of IRAP presented on a monitor.

To avoid position bias, I set the response option parameters as “moving positions” so that the randomization programme was activated to swap the left and right positions of the “True” and “False” options randomly. Also, I set the number of test blocks to provide 6 blocks (i.e., 3 paired blocks of consistent and inconsistent blocks). Each block contained 24 trials of stimulus presentations that were randomly arranged among the four different trial types (i.e., a “slim” silhouette image presented with a positive word, a “slim” image with a negative word, a “fat” image with a positive word and a “fat” image with a negative word) (see Figure 2.2). Participants had to meet specified criteria, which were set at a median latency of 2000 ms and 80% correct in each training block, in order to progress to the test blocks. Previous IRAP experiments generally used a 3 s median latency criterion; however, Barnes-Holmes and his colleagues found an improvement in internal reliability when they reduced the latency criterion from 3 to 2 s (Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart, 2010). I also set the programme to run a latency feedback feature that encouraged participants to maintain their optimum response speed during the test blocks. A red exclamation mark appeared on the screen for each test trial whenever the participant took longer to respond than the maximum latency criterion. The presentation of the feedback was set to start from the second pair of blocks onwards, not from the very beginning, because most of the participants tended to be slow in the first test blocks as they familiarized themselves with the trials.

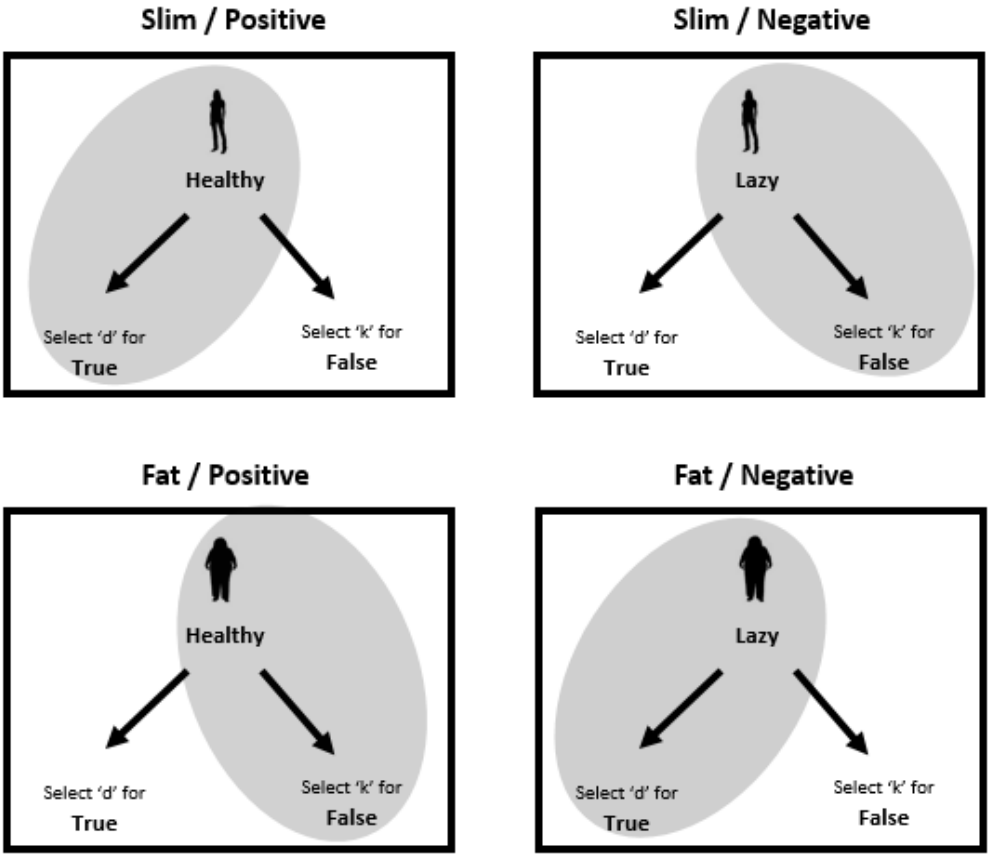


Figure 2.2. Four different IRAP trial types. The label stimulus for the silhouette image (either overweight or thin male or female image) appears at the top of the target stimulus for six positive and six negative adjectives randomly selected by the IRAP programme. The relational response options of "True" or "False" appears at the bottom of the screen. Participants follow two different rules for either consistent blocks, answering as if they are *anti-fat & pro-slim* or for inconsistent blocks, answering as if they are *pro-fat & anti-slim.* For example, when completing a consistent block, the correct answer you need to select is "True" for a combination of a thin silhouette image and a positive word (e.g., "Healthy") indicated in the grey circle. All other combinations with grey circles represent the correct answers required under the "consistent" rule. When completing an inconsistent block, the selection should be reversed.

Anti-Fat Attitudes (AFA).

The Anti-Fat Attitudes (AFA) scale was developed by Crandall (1994) to evaluate participants' explicit attitudes towards overweight individuals. The questionnaire has 13 items categorized in three sub-scales: "Dislike" reflects one's feeling or disposition toward overweight individuals; "Fear" reflects personal fear of being overweight; and "Willpower" reflects one's sense of control with regard to being overweight. The reliability estimates for each sub-scale are good: $\alpha = .84$ for the dislike subscale (7 items), $\alpha = .79$ for the fear subscale (3 items), and $\alpha = .66$ for the willpower subscale (3 items; Crandall, 1994). In other studies, the Cronbach's alphas for each subscale were good as well: $\alpha = .85$ for dislike, $\alpha = .84$ for fear, and $\alpha = .79$ for willpower (O'Brien, Hunter, & Banks, 2006) and $\alpha = .8, .74$, and $.74$ respectively (Roddy et al., 2010). The items are answered on a 9-point Likert scale ('very strongly disagree' to 'very strongly agree'), with higher total scores indicating a greater anti-fat bias. In this study, I used the AFA questionnaire in which the original term "fat" was replaced with "overweight" or "overweight or obese" (Setchell, Watson, Jones, Gard, & Briffa, 2014) to minimise reactions towards the word 'fat'.

Behavioural Intention Questionnaire (BIQ).

Based on previous studies by Brochu and Morrison (2007) and Roddy et al. (2010), participants were presented with four silhouette images selected from the 12 images used for the IRAP in this study, which were silhouette images of overweight and thin people. The aforementioned studies both used two photographs of an overweight and an average weight female image. For each of the silhouette images, the participants were asked to answer five questions evaluating their degree of willingness to interact with the person depicted in the

image, using a 7-point rating scale (1 = very unlikely to 7 = very likely). In the present study, randomly, I named the four silhouette images Joy (a thin female figure), Lisa (an overweight female), Daniel (a thin male) and Seth (an overweight male) to represent everyday life situations that were depicted in the questions that were phrased exactly the same as in the Brochu and Morrison (2007) and Roddy et al. (2010) studies. The questions were:

1) How likely is it that you would want to get to know "Joy" better? (i.e., compared to the same-sex counterpart, in this case "Lisa").
2) How likely is it that you would ask Joy if you could copy her notes from a class you missed?
3) How likely is it that you would want to work on a class project with Joy?
4) How likely is it that you would invite Joy to join a study group for an exam?
5) How likely is it that you want to become friends with Joy?

Scores could range from 5 to 35 and higher scores indicated willingness to socially interact with the person depicted in the silhouette image. The Cronbach's alpha for the scale was assessed at .86 by Brochu and Morrison (2007).

General information questionnaire.

This questionnaire (see Appendix E) contains five general questions about gender identity, age, history of dieting, pre-exposure to IRAP tests and frequencies of exposure to the IRAP.

Materials for determination of body mass index.

A tape measure was used to measure the heights of the participants who agreed to have their BMI determined. It was attached to the wall in the experimental room. An electronic scale was used to measure the weights of those participants who were not sure of their weight and agreed to have it measured.

Lastly, a box where participants deposited their anonymized weight (on a piece of paper with their participant code) was used.

Procedure

This research project was approved by the School of Psychology Research and Ethics Committee of the Faculty of Arts and Social Sciences, University of Waikato (Protocol #16:12). Participants went through the experimental procedure individually in a small room equipped with a desktop computer (DELL Optiplex 780 running Windows Vista) and a monitor (1920 x 1080 resolution), along with two desks (one for the participant and the other for myself; I was present in the room during the entire experiment), and two chairs. The room was not acoustically insulated, so to control noise I placed three warning posters outside in the hallway to encourage passers-by to be quiet. In the course of the experiment, we did not encounter any issues with noise.

Participants were asked to read an information sheet that explained the features of the study before they signed a statement of informed consent. Before obtaining consent to participate in this experiment, all potential participants were informed about being exposed to silhouettes of slim and overweight people, given the opportunity to ask questions about those images, and asked to report their height and weight, which would be handled anonymously. In addition, the researcher obtained verbal consent from the potential participants to ensure that they still wanted to participate. All participants filled in a general questionnaire, then performed the IRAP tasks, followed by the completion of two explicit questionnaires (AFA and BIQ) and self-reported their weight and height on a piece of paper. Two different groups were set up: the control group (n = 41) and the incentive group (n = 41), who would receive a NZ$20 voucher from a local

department store. Both groups received a participation reward of either course credits or entry in a draw to win a $20 iTunes voucher.

Before commencing the IRAP task, participants were presented with a set of general instructions for IRAP tasks (see Appendix D, which is a verbatim record of the instructions used during the IRAP experiment). The IRAP programme contained practice and test blocks. The practice blocks provided a maximum of four sets of consistent/inconsistent blocks. Each block contained 24 trials. The test blocks gave three sets of consistent/inconsistent blocks, each with 24 trials. Two types of IRAP sequences were used to control for order bias, and participants were randomly assigned to one of two sub-groups: one group started with *anti-fat/pro-slim* trials and the other group started with *pro-fat/anti-slim* trials (see Table 2.1). These two sequences were randomly assigned to an equal number of participants in each independent group (i.e., the control or the incentive group).

Table 2.1

Number of Participants in Each Group and Sub-group

	Control Group (*N*)	Incentive Group (*N*)
Anti-fat / pro-slim trials first	20	20
Pro-fat / anti-slim trials first	21	21
Total	41	41

All participants were told that they needed to meet the criteria (median latency of 2000 ms and 80% accuracy) during the practice trials, but if they failed, then they went through the test blocks anyway. For the incentive group, all the sequences and procedures were the same as for the control group, except that I mentioned the possibility of obtaining a $20 voucher, in addition to the participation rewards, if participants passed the speed and accuracy criteria for the

IRAP task for the whole IRAP experiment. For the detailed verbatim experimenter script, please see Appendix D**.**

After the whole IRAP task was completed, the participants were asked to fill out AFA and BIQ questionnaires as this was the same test sequence used in previous IRAP studies (Expósito et al., 2015; Roddy et al., 2010). I provided a pen. When participants had completed the explicit questionnaires, I asked their height. If the participant was not sure, I measured their height using a tape measure attached to the wall of the room. I then asked participants to report their own weight on a piece of paper, which was marked with the participant's code (not their name), and put it in a box with the other participants' weight records. If they were not sure about their weight, they were offered the opportunity to measure their own weight using an electric scale. Before weighing they were advised to take off their shoes and jacket if applicable. The researcher subtracted 0.8 kg for women and 1.2 kg for men for the relevant clothing weight adjustment (Whigham, Schoeller, Johnson, & Atkinson, 2013). To maintain participant privacy, weight data were handled anonymously. While participants were weighing themselves, I looked away. The participants recorded their weight on a piece of paper as described above.

Before participants left, I gave a friendly reminder that the participant should not share or discuss the details of the experiment with anyone else. Then, I thanked and rewarded participants (see details above).

Data Analysis

The IRAP captured response latency, which was recorded as elapsed time (in ms) in trials with correct responses; latency was recorded from the onset of a stimulus presentation (e.g., a combination of a silhouette image and a target label

of either a positive or negative word) until the participant pressed the correct key (i.e., either the “d” or “k” key for true or false relational response stimuli) in each trial. The raw scores from the response latency were transformed into differential scores (D-IRAP scores) based on the D-score algorithm proposed by Greenwald, Nosek, and Banaji (2003), who developed and adjusted the algorithm to control for individual variations related to extraneous factors such as age, outliers, and speed-accuracy trade-offs (Greenwald et al., 2003). The calculation is similar to Cohen’s effect size (Cohen’s D) which indicates whether or not the difference found between the consistent (i.e., “anti-fat and pro-slim” attitude) and inconsistent (i.e., “pro-fat and anti-slim”) blocks was large enough to be considered a significant effect. Simply put, the D-IRAP score indicates which trial blocks produce faster responses: the larger the D-score, the bigger the effect size, which indicates a historically coherent relational network associated with one or more trial type. If the D-score is zero, it means there is no significant difference between the consistent and inconsistent blocks, and if it is between 0.6 and 0.8, it indicates a large effect size, implying a historically coherent bias.

The following steps were taken to transform the raw latency scores to D-IRAP scores, as developed by Barnes-Holmes, Waldron, Barnes-Holmes, and Stewart (2009) based on the D-algorithm (Greenwald et al., 2003). The process was introduced by Barnes-Holmes et al. (2009) as the standard calculation method, using Microsoft Excel (Microsoft Excel 2016, Version 16.0.7329.1045, 32-bit) file.

1) Only latency data from the test blocks were used.
2) Latencies over 10,000 ms were removed from the dataset.

3) If the participant responses had latencies shorter than 300 ms in more than 10% of the test-block trials, their datasets were removed.
4) 12 *SD*s for the four trial types were computed: four for the response latencies from Test Blocks 1 and 2, four for latencies from Test Blocks 3 and 4, and another four from Test Blocks 5 and 6.
5) 24 mean latencies were calculated for the four trial types in each test block.
6) Difference scores were calculated for each of the four trial types, for each pair of test blocks, by subtracting the mean latency of the *pro-slim* & *anti-fat* block from the mean latency of the corresponding *pro-fat* & *anti-slim* block.
7) Each difference score was divided by its corresponding *SD* from Step 4, yielding 12 D-IRAP scores; one score for each trial type for each pair of test blocks.
8) Four overall trial-type D-IRAP scores were calculated by averaging three scores for each trial type across the three pairs of test blocks (for participants who had failed to maintain the accuracy and latency criteria in any block, only the data from valid block pairs were averaged).
9) An overall relative D-IRAP score was calculated by averaging all 4 D-IRAP scores from Step 9.

A positive D-IRAP score indicated an anti-fat & pro-slim attitude and a negative one indicated a pro-fat & anti-slim attitude.

Furthermore, based on the body mass index scale used by the World Health Organization (2006), I categorized the participants into two groups to see

if overweight and normal-weight groups performed differently: an overweight group (BMI ≥ 25) and a normal weight group (BMI < 25). The calculation method for BMI for each individual was weight (kg) divided by the square of their height (m^2).

Results

Implicit anti-fat attitudes

D-IRAP Scores.

I conducted a preliminary analysis of the block-sequence effect on a group taking the *anti-fat & pro-slim* trial first and on another group taking the *pro-fat & anti-slim* trial first. The analysis showed no significant effect on performance, $t(318) = -.54$, $p = .37$. Thus, I assumed there was no order bias. Internal consistency for all D-IRAP scores was calculated using the split-half method in which the D-scores for odd- and even-numbered trials were compared and analysed. Spearman-Brown reliability for the D-IRAP scores showed moderate correlation for internal consistency ($r = .46$, $p < .001$).

Figure 3.1 shows the overall D-IRAP scores in the present study for each of the four trial types. The mean total D-IRAP score was .19, 95% CI [.14, .24], which was significantly different from zero, $t(304) = 7.74$, $p < .001$. A strong pro-slim IRAP effect was observed for the *Slim & +ve words* trial type (participants were faster to press 'true' in consistent trials than to press 'false' in inconsistent trials when presented with a series of slim silhouette images and positive words) and for the *Slim & -ve words* (participants were faster to press 'false' in inconsistent trials than 'true' in consistent trials when presented with slim silhouette images with negative words). The D-IRAP score for the *Slim & +ve words* was .49, 95% CI [.39, .58] and for *Slim & -ve words* .24, 95% CI [.15, .32]. A one-sample *t* test indicated that the two D-IRAP trial scores were significantly different from zero (*Slim & +ve words*, $t(76) = 9.99$, $p < .001$; *Slim & -ve words*, $t(76) = 5.58$, $p < .001$).

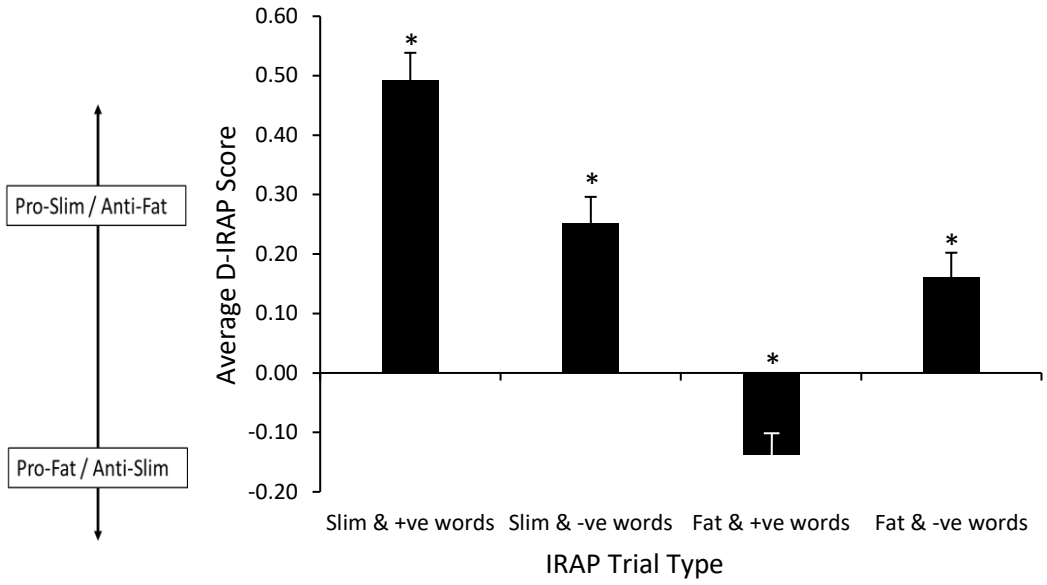


Figure 3.1. Mean D-IRAP scores across the four IRAP trial-types. Positive scores indicate a pro-slim/anti-fat attitude and negative scores indicate anti-slim/pro-fat. * $p < .05$. The error bars represent the standard error of the mean.

A mild pro-fat IRAP effect was observed for the *Fat & +ve words* trial-type (participants faster to press 'true' rather than 'false' when presented with a series of overweight silhouette images and positive words). The D-IRAP score for the *Fat & +ve words* was -.12, 95% CI [-.2, -.04] (see Figure 3.1), and was significantly different from zero, $t(76) = 3.17$, $p < .001$. For the *Fat & -ve words* trial type, a mild anti-fat IRAP effect was observed (participants faster to press 'true' rather than 'false' when presented with a series of overweight silhouette images and negative words). The D-IRAP score for the *Fat & -ve words* was .16, 95% CI [.07, .23] (see Figure 3.1), and was significantly different from zero, $t(76) = 5.58$, $p < .001$. Thus, according to the IRAP scores, there was a relatively strong pro-slim implicit attitude towards slim silhouette images and a moderate anti-fat implicit attitude towards overweight silhouette images presented with negative words. On the other hand, the participants showed moderate pro-fat attitudes

towards overweight silhouette images presented with positive words. In short, the participants were pro-slim and anti-fat when they were quick to associate the fat silhouettes with negative words, but at the same time, they seemed to be pro-fat as they associated the fat silhouettes with positive words quickly.

Explicit anti-fat attitudes

The Anti-Fat Attitude (AFA) scale.

Average scores on the whole AFA scale (n = 82) were 3.00, 95% CI [2.77, 3.22]. The scale has three sub-categories. The average scores for each category were 1.79, 95% CI [1.57, 2.02] for dislike, 4.48, 95% CI [4.01, 4.95] for fear of overweight, and 4.42, 95% CI [4.07, 4.77] for willpower. The possible range of scores for the three subscales were from 0 to 9. Thus, participants demonstrated that they had relatively weak or moderate levels of explicit anti-fat attitudes.

The Behavioural Intention Questionnaire (BIQ) scale.

Scores for each of the five question items were added to produce a total scale score. The range of scores was 5 to 35 and higher scores indicated willingness to socially interact with people of each body type (i.e., Fat male image [FM-Seth], Fat female image [FF-Lisa], slim male image [SM-Daniel] and slim female image [SF-Joy]). The averaged willingness indexes were as follows: FM-Seth (M = 22.37, 95% CI [21.11, 23.62]); FF-Lisa (M = 23.59, 95% CI [22.3, 24.87]); SM-Daniel (M = 24.09, 95% CI [22.8, 25.37]); and SF-Joy (M = 27.78, 95% CI [23.45, 26.11]). A two-way analysis of variance was conducted on the difference in effect of the two variables (body type, gender) on the obtained scores of BIQ scale. Body type included two levels, which were fat and slim, and the gender included two levels of male and female. The main effect for body type indicated a significant main effect of body type, $F(1, 324) = 4.91$, $p = .027$, Fat:

M = 22.98, 95% CI [22.08, 23.88], Slim: *M* = 24.43, 95% CI [23.51, 23.35]. There was no main effect of gender, $F(1, 324) = 2.12$, $p > .05$, Male: M = 23.23, 95% CI [22.33, 24.13]. Female: M = 24.28, 95% CI [23.36, 25.2]. There was no significant interaction effect, $F(1, 324) = 0.16$, $p > .05$. Under the hypothetical scenario, participants indicated more openness to socially interact with slim characters than overweight characters.

The Body Mass Index and other variables.

In my study, 27 out of 82 participants (33%) were overweight according to the WHO criteria, *M* = 32.06, 95% CI [29.89, 34.23] and the rest of the participants were in the normal weight criteria, *M* = 21.63, 95% CI [21.06, 25.04]. Information on other variables collected from the general information questionnaires is summarized in Table 3.1.

Table 3.1

Descriptive Statistics for Body Mass Index (BMI) and Age, Number of Male and Female Participants, Number of Participants who had Pre-Exposure to Other IRAP experiments, and Number of Participants who had Previously Tried to Lose Weight

	N	*M*	*95% CI*
BMI ≥ 25 Overweight individuals	27	32.06	29.89, 34.23
BMI < 25 Normal weight individuals	55	21.63	21.06, 22.20
Age	82	23.38	21.72, 25.04
Male	19	-	-
Female	63	-	-

	N	*Yes*	*No*
Participated in other IRAP experiments (previous exposure to IRAP)	82	26	56
Experience of dieting	82	61	21

Relationship between IRAP scores and other variables.

A multiple linear regression was calculated to predict AFA from the D-IRAP scores of the four different trial types, age, and BMI. The regression model was insignificant, $F(6, 75) = 1.71$, $p = .13$, $R^2 = .12$. I conducted another regression analysis using the same predictor variables and BIQ score as the outcome variable; however, it was also insignificant, $F(6, 75) = 1.78$, $p = .11$, $R^2 = .13$. Thus, the D-IRAP scores, age, and BMI were not able to significantly predict scores from either explicit measure, AFA or BIQ.

The reason for conducting two separate regression analyses was that the outcome variables both violated one of the assumptions of multiple regression. The scores of the BIQ and AFA tests were significantly correlated, $r = .26$, $p = .02$ (See Table 3.2). In addition, based on the results of the correlation analysis, the two explicit measurements of AFA and BIQ were both moderately related to the D-IRAP scores obtained in the "Fat" and positive word trial type; the correlations of AFA and BIQ were $r = .24$, $p = .03$ and $r = -.22$, $p = .04$, respectively.

Table 3.2

Pearson's Correlations between D-IRAP scores and Explicit measures of AFA and BIQ, Age, BMI, and History of Dieting

Variable	AFA	BIQ	Age	BMI	Diet-ing	IRAP Slim+	IRAP Slim-	IRAP Fat+	IRAP Fat-
AFA	-								
BIQ	**-.26***	-							
Age	-.16	.18	-						
BMI	-.06	-.05	**.51****	-					
Dieting	.18	.01	.09	**.35****	-				
IRAP Slim+	-.01	.05	-.02	**-.25***	-.02	-			
IRAP Slim-	-.03	-.17	.08	.05	.21	.20	-		
IRAP Fat+	**.24***	**-.22***	-.10	-.03	-.06	.02	.18	-	
IRAP Fat-	.18	.06	.19	-.00	.02	.02	-.06	.17	-

*$p < .05$. **$p < .01$

The positive correlation obtained for the AFA score with the D-IRAP score for the *Fat & +ve words* trial-type indicated more inclination toward the anti-fat implicit attitude as the scores on the AFA increased. The negative correlation obtained for the BIQ with the D-IRAP score with the aforementioned trial-type also indicated the anti-fat implicit attitude was stronger as the BIQ score decreased (i.e., less favourable feelings toward an overweight image). Based on the results of the correlations with the other variables, the BMI was also moderately related to the D-IRAP scores obtained from the *Slim & -ve words* trial type, $r = -.25$, $p = .03$. This meant that an overweight individual was more likely to have a pro-fat implicit attitude. As expected, the BMI was strongly correlated with age, $r = .51$, $p < .001$ and the history of dieting, $r = .35$, $p < .001$.

IRAP summary for each individual

In Figure 3.2, each participant's D-IRAP score for the four trial types is presented. From this graph, it can be seen that large D-IRAP scores (or strong effects) were obtained for some individual participants and others demonstrated individual differences in the attitudinal directions across the four trial types. Scores above or below the two dashed lines at either .5 or -.5 indicate a strong implicit attitude suggested by the IRAP test for that participant.

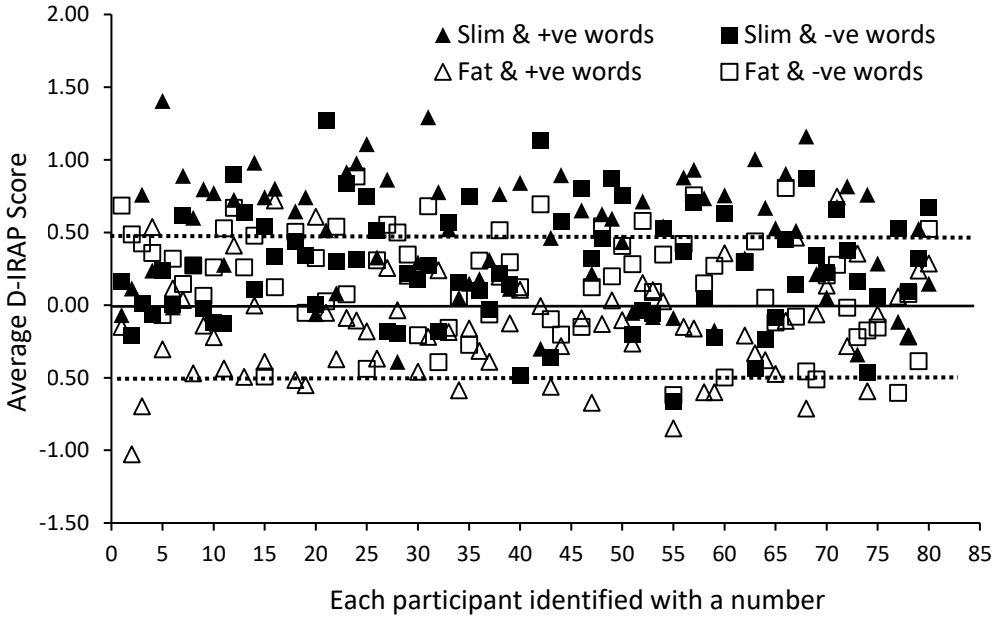


Figure 3.2. Mean D-IRAP scores across participants (n = 82). Each participant completed four IRAP trial types (Slim & +ve words, Slim & -ve words, Fat & +ve words, and Fat & -ve words) and produced four D-IRAP scores each. Scores above or below the two dashed lines at either .5 or -.5 indicate a strong implicit attitude. Positive scores indicate a pro-slim/anti-fat attitude and negative scores indicate anti-slim/pro-fat attitudes.

In Table 3.3, I list individuals who had an extremely strong D-IRAP score, above .8 or below -.8, along with the scores from the two explicit measurements, AFA and BIQ, for a comparison. I separated the participants into control and extreme groups; however, there were no obvious differences in numbers of individuals or patterns in the array. A number of participants (n = 21) produced strong D-IRAP scores in the trial of slim and positive words, which indicated a very strong pro-slim implicit attitude. On the other hand, only three indicated a strong anti-fat implicit attitude (#24, #66, and #81) and one (#61) had a strong pro-fat attitude. In addition, among all the participants, only #12 and #81 indicated congruent results with the explicit measurements (i.e., showing strong pro-slim/anti-fat IRAP effect and the corresponding attitudes being captured in the AFA and BIQ scores). The rest of the participants with the extremely strong D-

IRAP scores showed incongruent relationships between implicit and explicit measures on attitudes toward body size.

Table 3.3

A List of Participants with Extreme D-IRAP Scores, Above .8 and -.8, and Their Corresponding AFA and BIQ Scores

Participant	Slim +	Slim -	Fat +	Fat -	AFA	BIQ Fat	BIQ Slim
Control Group							
#9	**0.80**	-0.02	-0.14	0.06	3.62	**19**	**20**
#23	**0.91**	**0.83**	-0.09	0.07	2.69	26	26
#25	**1.11**	0.74	-0.18	-0.44	2.31	**19.5**	34
#27	**0.86**	-0.18	0.25	0.55	2.69	28	21.5
#31	**1.29**	0.27	-0.22	0.68	3.15	33	35
#61	**0.86**	0.05	**-0.81**	0.27	2.38	21	22
#63	**1.00**	-0.44	-0.33	0.44	0.92	32	32
#81	**0.91**	-0.04	**0.86**	-0.14	**5.23**	**19**	23
#21	0.51	**1.27**	-0.05	0.03	3.46	20.5	22.5
#49	0.60	**0.87**	0.03	0.20	4.46	**17.5**	25.5
Incentive Group							
#12	**0.82**	0.59	0.45	0.49	**5.46**	**17.5**	33
#14	**0.98**	0.11	0.00	0.48	2.85	**18**	**20**
#24	**0.97**	0.31	-0.11	**0.89**	1.46	26.5	**10.5**
#40	**0.84**	-0.49	0.10	0.12	3.23	26	21
#44	**0.89**	0.57	-0.28	-0.20	2.23	**16**	**16.5**
#56	**0.88**	0.37	-0.15	0.42	3.85	25	30.5
#66	**0.90**	0.45	-0.11	**0.81**	3.15	25	22
#68	**1.16**	0.78	-0.18	-0.02	2.62	23	24.5
#72	**0.82**	0.38	-0.28	-0.02	4.08	**11**	25
#82	**0.90**	-0.04	0.15	0.08	4.08	**16**	26
#46	0.65	**0.8**	-0.09	-0.15	2.85	30.5	27.5

Note. The possible range of scores for the *AFA* were from 0 to 9, with higher scores indicating a greater inclination toward the *anti-fat* attitude. The *BIQ Fat* indicates an average score obtained by participants (between 5 and 35) against silhouette images of overweight male and female. Any score less than 20 suggests less willingness to socially interact with the body type image. Similarly, the *BIQ Slim* indicates the average scores obtained from the silhouette images of slim male and females.

Between-groups comparison of IRAP performance analysis (Control vs. Incentive groups)

I conducted a simple between-group comparison between the control (n = 41) and incentive (n = 41) groups to see whether there were any differences in participants' performances. The incentive group members were told that they

could obtain a $20 voucher if they could pass the performance criteria in practice blocks and maintain a good performance throughout the test IRAP task, whereas the control group members were given no such incentive. The details of the performance differences are given below.

Attrition.

In my IRAP study, I set the passing criteria as a median latency of 2000 ms and 80% correct in each block in order to progress to the test blocks. In the incentive group, only one person failed to pass the IRAP performance criteria during the practice phase (97.5% pass). However, five participants in the control group failed to pass (87.8% pass). Statistically, the difference between the groups was not significant, $\chi^2(1) = 2.87$, $p = .09$, Cramer's V = .004.

Accuracy.

A Chi-Square test was conducted to determine whether there were any differences between the numbers of correct responses in the two groups. The incentive group made significantly fewer mistakes; the numbers of error responses were 482 in the incentive group, compared to 556 for the control group, $\chi^2(1) = 5.78$, $p = .02$, Cramer's V = .22, which indicated midway between a small and medium effect size according to Cohen's conventions (Field, 2013).

Number of latency responses longer than 2000ms.

The total number of latency responses that exceeded 2000ms was 957 for the control group and 883 for the incentive group; the numbers that exceeded 3000ms were 196 and 177, respectively. A Chi Square test indicated no significant differences in the numbers of errors above 2000ms between the control and incentive groups, $\chi^2(1) = .036$, $p = .85$, Cramer's V = .004.

D-IRAP score.

As indicated in Table 3.4 and Figure 3.3, the mean differences in D-IRAP scores between the control and the incentive groups diverged in "fat silhouette & positive words" and "fat silhouette & negative words" trial types.

Table 3.4

Standard Deviation and Confidence Interval of the Mean of D-IRAP Scores across Four Trial Types in Control (n = 41) and Incentive Groups (n = 41)

	Control		Incentive	
Trial Type	M (*SD*)	95% CI	M (*SD*)	95% CI
Slim & +ve	.49 (.43)	[.36, .62]	.49 (.42)	[.36, .62]
Slim & -ve	.23 (.40)	[.11, .36]	.24 (.33)	[.14, .34]
Fat & +ve	-.18 (.38)	[-.29, -.06]	-.09 (.34)	[-.19, .01]
Fat & -ve	.05 (.33)	[-.05, .15]	.25 (.31)	[.16. .35]

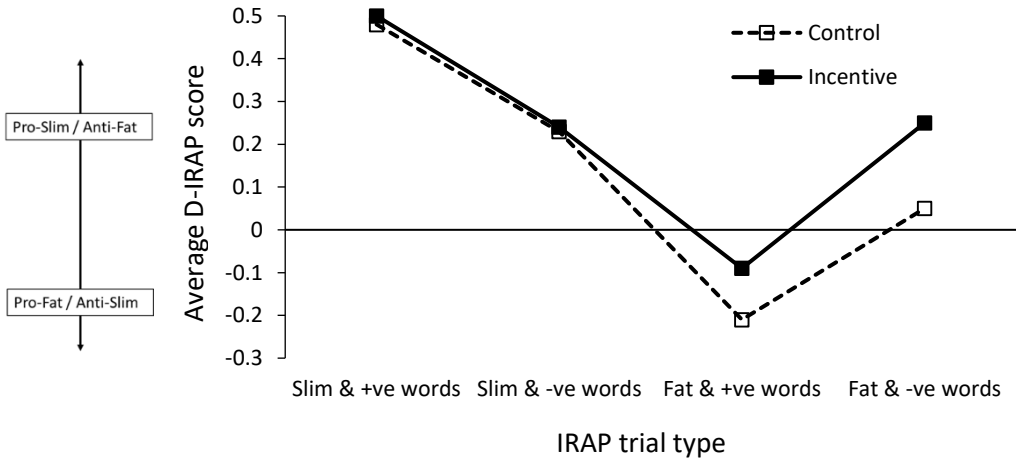


Figure 3.3. Comparisons of mean D-IRAP scores between control and incentive groups.

A 2-x-2 ANOVA with two IRAP trial types (fat & +ve, and fat & -ve) and the two independent groups (control and incentive) revealed a statistically significant main effect of group membership on D-IRAP scores from fat silhouette trials, $F(1, 160) = 7.54$, $p = .007$, $R^2 = .45$. The incentive group's mean D-IRAP scores for fat & +ve words, $M = -.09$, and fat & -ve words, $M = .25$, were both towards the positive side (i.e., indicating greater implicit *anti-fat* attitude) in

Figure. 3.3, compared to the control group, *M* =. -.18 and .05, respectively. The differences in the mean D-IRAP scores were observed only for the trials in which the "fat" silhouette was presented, not for the trials that involved the "slim" silhouette image. The main effect of differences in the two trial types was found to be statistically significant, $F(1, 160) = 29.22$, $p < .001$, $R^2 = .15$, which indicates that, overall, D-IRAP scores associated with the Fat & -ve trial type are more anti-fat than those associated with the Fat & +ve trial types. The interaction effect between the groups and the two trial-types was found to be non-significant, $F(1, 160) = 1.14$, $p = .28$, $R^2 = .007$.

The D-IRAP scores were designed to indicate the mean difference between the total response latency in the two different (consistent and inconsistent) blocks. Thus, the current result indicated that the incentive group did respond comparatively faster than the control group in the consistent blocks (i.e., responding 'false' to the combination of "fat" silhouette with positive words) under the *Fat & +ve words* trials, and they were also faster in the consistent block (responding 'true' to the combination of fat image with negative words) under the *Fat & -ve words* trials.

Discussion

I examined several hypotheses. Firstly, I expected to see that a group of participants whose performance was incentivized with a $20 voucher would demonstrate a lower attrition rate than a control group (who received no incentive) in that more of the participants in the incentive group would meet the IRAP performance criteria of maintaining 80% or greater accuracy and a median latency of 2000 ms or less. Secondly, I expected that the group that received the $20 voucher incentives would show better overall IRAP performance compared to the control group, especially in accuracy and latency of responses. Thirdly, I expected to see some divergence in the D-IRAP scores between the experimental and control groups because lower latency would minimize the influence of "extended and elaborated relational responses (EERRs)" (Hughes and Barnes-Holmes, 2013, p.102) on IRAP responding. Fourthly, I examined whether or not the implicit attitudes toward body size in the New Zealand context and the expected data outcome would reveal the *anti-fat* and *pro-slim* attitudes that were observed in previous implicit studies on body-size in other countries. Finally, I questioned whether or not explicit measurements such as AFA and BIQ and other variables such as BMI, age, and gender were congruent or incongruent with implicit measurements, D-IRAP scores.

Monetary Incentive Effects on IRAP performance

In my IRAP study, the overall attrition rate was surprisingly low. I observed a 97.5% pass rate among participants who received the incentive and an 87.8% pass rate among participants in the control group. Statistically, no significant difference was found between the two groups; however, considering only one person failed to pass in the incentive group, compared to five in the

control group, there is some evidence to support the hypothesis that a monetary incentive enhances data integrity by retaining higher numbers of participants. In addition, there may be a ceiling effect that prevented the detection of an effect on attrition rate. The likelihood of any individual failing the IRAP is usually high, thus it would limit its utility for the assessment of individuals. The high attrition rates seen in previous IRAP studies have been one of the limiting factors in reaching the full potential of this method (Golijani-Moghaddam, Hart, & Dawson, 2013). To date, about 20% of participants in each IRAP study had failed to pass the practice response criteria (Hughes & Barnes-Holmes, 2013). Strict pass criteria need to be applied for IRAP research because they are the main way of eliminating social contingency (i.e., EERR) from the measured responses. Without these strict criteria, IRAP results may be similar to those obtained from explicit measures (e.g., a self-reporting questionnaire), in which participants have sufficient time to engage in socially appropriate responses.

However, the strict criteria may cause a dilemma due to the possibility of systematically rejecting some participants from IRAP studies. This potential sampling bias, in which conclusions are based on the population after a high rate of systematic attrition, could produce a biased picture of the population characteristics to some extent. The systematic elimination of individuals from the study may exclude individuals with a strong implicit bias (i.e., BIRRs) who may perform poorly in trials that are inconsistent with such bias (Golijani-Moghaddam, Hart, & Dawson, 2013). For example, if one has a strong anti-fat bias against overweight individuals, he or she may find it very difficult to meet criteria in IRAP's inconsistent trials because such individuals may feel resistant to answering as if they were someone with a "*pro-fat* and *anti-slim*" attitude during

the inconsistent trials (e.g., selecting "true" for a fat silhouette image presented with positive words and "false" for the fat silhouette image presented with negative words).

Thus, finding a pragmatic approach to lowering the attrition rate in the IRAP test is very important. In previous studies using the IRAP to assess attitudes related to body size, both Roddy et al. (2010) and Expósito et al. (2015) found an 80% pass rate (features of the two studies are summarized in Table 4.1 below). The criteria used by Roddy et al. (2010) were a median latency of 3000 ms and minimum accuracy of 80% correct. Expósito et al. (2015) set a median latency of 2500 ms and minimum accuracy of 80% correct. Therefore, my study yielded a lower attrition rate compared to other studies and achieved this under stricter criteria, a median latency of 2000 ms and minimum accuracy of 80% correct.

For the effect of monetary incentives on promoting accurate responses, my study revealed a statistically significant improvement in response accuracy for participants who received the incentive. The participants in the incentive group followed the specific IRAP rules administered in both consistent and inconsistent trials more accurately than the control group.

In addition, a statistically significant difference in the D-IRAP scores was observed. Among several variables investigated for the effect of the monetary incentive, the mean D-IRAP score was different in *fat & positive words* and *fat & negative words* trials only, not in the trials involving stimuli with a *slim* silhouette. The difference in scores observed only in trials involving "fat" stimuli may be because the "fat" trials could be especially difficult or challenging due to the fact that the participants were asked to respond to socially sensitive stimuli. In some cultures (e.g., western culture), it is generally more socially acceptable to

associate positive traits with slim people than it is to associate negative traits with overweight people. For example, in a consistent trial (i.e., the rule was to answer as if the participant were "*anti-fat and pro-slim*"), participants were required to select "false" to the stimulus set presented with a *fat* silhouette and a positive word (e.g., healthy, attractive) and to select "true" to the combination of the *fat* silhouette with a negative word (e.g., lazy, ill).

Compared to the control group, the IRAP scores of the incentive group indicated that they revealed a greater *anti-fat* implicit attitude in the trials that included the presentation of stimuli that could provoke participants' social sensitivity or associated feelings (i.e., those associated with *anti-fat* bias). This result suggests that the monetary incentive may have functioned as an establishing operation, in that it altered the functions of the stimuli associated with meeting or failing to meet the strict criteria (i.e., 80% accuracy and 2000 ms median latency) during the IRAP task. For example, indication of an exclamation mark on a computer screen during the IRAP trials may function as a more effective aversive stimulus, thus resulting in participants' speeding up their subsequent responses.

The D-IRAP scores of the incentive group were based on their "better" performance in accuracy and latency (i.e., higher accuracy and shorter latency of response), compared to the control group. Thus, the *anti-fat* attitude observed in the incentive group may be closer to their true *implicit* attitude toward body-size. In other words, a monetary incentive may have helped the participants to access "brief and immediate relational responses (BIRRs)" (Hughes and Barnes-Holmes, 2013, p.102) and minimized the influence of "extended and elaborated relational responses (EERRs)" (p.102). In my study, I found that the monetary incentive had the potential to be effective in lowering attrition rates, promoting participants to

respond accurately, and to do better in overall performance; it therefore may help to enhance data integrity in IRAP measurements.

D-IRAP score comparisons with previous IRAP studies on the anti-fat bias

In my study, a clear *pro-slim* implicit attitude was observed, which was consistent with previous IRAP studies (Roddy et al., 2010; Expósito et al., 2015). The D-IRAP scores for each trial type (*Slim & +ve words, Slim & -ve words, Fat & +ve words* and *Fat & -ve words*), which was the combined results of the control and experiment groups, were all statistically significant, indicating that the IRAP effect was found in all trial types. The results from each study are presented in Figure. 4.1.

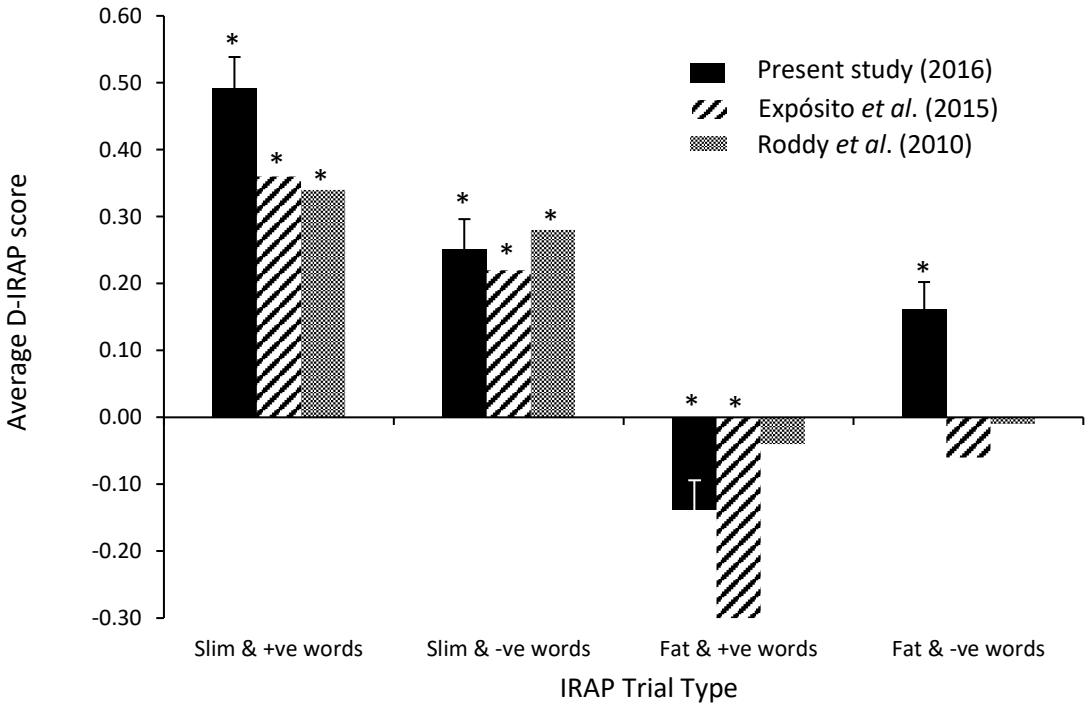


Figure 4.1. Comparisons with previous IRAP studies on anti-fat bias. The *y* axis indicates the average D-IRAP score of each IRAP trial type. Positive scores indicate a pro-slim and/or anti-fat attitude; negatives scores indicate a pro-fat and/or anti-slim attitude. Asterisks indicate that a score is significantly different from zero ($p < .001$ for my study, $p < .05$ for Roddy et al. (2009), and Expósito et al. (2015)).

All studies demonstrated moderate *pro-fat* attitudes among participants in the *Fat & +ve words* trial type. For the same trial type, the D-IRAP score in the

investigation by Roddy et al. (2010) was not significant, meaning that they found a neutral attitude; there was no significant difference found between a consistent implicit attitude (i.e., an *anti-fat* attitude indicated by faster "false" than "true" responses when pictures of overweight models were presented together with positive words) and an inconsistent implicit attitude (i.e., a *pro-fat* attitude indicated by faster "true" than "false" responses when presented with the same stimulus sets). However, my results and Expósito et al.'s (2015) were both statistically significant. A slightly weak *pro-fat* attitude was observed in my study compared to the moderate *pro-fat* attitude found by Expósito et al. (2015). A difference was observed in trials using the "fat" silhouette presented with a negative word. A moderately anti-fat attitude was found only in my study and a neutral implicit attitude was observed in the other two studies (Roddy et al, 2010; Expósito et al., 2015).

In regards to the *pro-fat* attitude (i.e., faster affirmation following the presentation of a fat silhouette and a positive word than negation), IRAP may have captured the implicit attitude that was somewhat obscured in the IAT measurement, which provides measurements only in two dimensions ("Fat-Unpleasant/Thin-Pleasant" and "Fat-Pleasant/Thin-Unpleasant"). Thus, IAT has a limiting factor in that it cannot reveal to what degree Stimuli A and B are liked or disliked (De Houwer, 2002; Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009). The IRAP study by Expósito et al., (2015), along with my study, revealed that people have a pro-fat implicit attitude, with positive framing toward the "fat" image. It may capture our 'ingrained' association of stimuli through a series of exemplar trainings that we may frequently associate a *fat* image with a positive connotation. For example, a baby and Santa Claus are perhaps both regarded as

strongly positive images by most people in western culture (e.g., a big jolly Santa), and in recent movies, characters representing a plump body shape such as Bay-Max in Big Hero 6 (2014), Melissa McCarthy in Ghost Busters (2016), Jack Black and many others all present positive images of being funny, warm, caring, and strong. It would be interesting to investigate further the effect of such "*fat*" characters on our attitudinal behaviour.

In the present study, the *anti-fat* implicit attitude was observed in the IRAP trial in which participants were presented with a *fat* silhouette with a negative word. In the previous IRAP studies, the researchers did not obtain any significant IRAP effect with this trial type. My results may have been affected by controlling various discriminative elements in the compound stimuli used in the previous IRAP studies. However, I designed the present IRAP experiment based on the specifics used in the two aforementioned studies: Table 4.1 outlines the key features of these studies.

Table 4.1

Comparisons of Specific Features Used in the Studies on Anti-Fat Attitudes. Summary of Different IRAP Stimuli Used in Each Study and Differences in Attrition Rates.

	Picture stimuli	Labelling stimuli	Contextual cues	Participants (*n*)
Current Study (2016)	16 silhouette images of slim and fat male and female figures	6 positive and 6 negative words	"*True*" or "*False*"	*n* = 82 (university students) Attrition = 7%
Roddy et al. (2009)	12 upper-body photos of either slim or fat male and female models	A pair of positive and negative words (e.g., "*good*" and "*bad*")	"*Same*" or "*Opposite*"	*n* = 80 (university students) Attrition = 20%
Expósito et al. (2015)	12 photos of either under/overweight young female models whose faces were blurred	A pair of positive and negative words (e.g., "*pleasant*" and "*unpleasant*")	"*True*" or "*False*"	*n* = 40 (university students – female only and Spanish) Attrition = 20%

I used silhouette images for the IRAP stimuli to attempt to eliminate 'noise' factors from compound stimuli (e.g. photo images) that could evoke wider ranges of responses among participants. People usually associate a smiling face with positive connotations, which may produce more pro-fat attitudes among participants. Expósito et al. (2015) were concerned about the same point and partially controlled their picture stimuli of young females by blurring their faces; however, their IRAP scores indicated a neutral attitude toward the "fat" image with negative words. Expósito et al. (2015) mentioned that further investigation would be required to determine the factors causing the absence of anti-fat bias found in their study. My study may indicate that selection of the stimuli is a key to eliminating EERRs and thus may help to narrow the range of responses.

In regard to the explicit measures (AFA and BIQ), no significant prediction based on my regression model was obtained between the explicit and implicit measures, or with other external variables such as age and history of dieting with the IRAP. It was observed that BMI was moderately related to the D-IRAP scores obtained in the trial type presenting a "slim" image with a negative word, meaning that an overweight individual tended to have an anti-slim implicit attitude. The results were consistent with those of the previous IRAP studies by Roddy et al. (2010) and Expósito et al. (2015). The AFA and BIQ explicit measures are designed to specifically evaluate an *anti-fat* attitude, and do not focus on the *pro-slim* aspect. Thus, only a weak correlation was observed among the scores on the whole AFA, the BIQ on the fat silhouette image and D-IRAP scores in the fat-&-positive-words trial type, and there were no correlations to the D-IRAP scores in the trials including slim images. Also, the incongruences between explicit and implicit measures observed in the current study, particularly

those associated with overweight silhouettes, are in alignment with other findings showing low correlations between these measures with socially controversial issues, but if such approaches are used in a preference study on non-controversial stimuli, they produce congruent results. The incongruence was also indicated in individual records that showed extremely high D-IRAP scores (over .8 or -.8) on each IRAP trial type. There were only two individuals who demonstrated congruence with the external measures (individuals #12 and #81); the remaining participants with extremely high D-IRAP scores had results that were incongruent with their explicit measurements. However, the lack of correlation does not serve to validate the findings of the IRAP effects. In order to test its validity, it would be critical to use a separate behavioural measure associated with attitudes related to body weight, such as tendency of participants to discriminate against people based on their weight.

A possible limitation in this study was that the course credit offered to all participants was already a highly valued incentive for the undergraduate student participants at the time of the study. The University of Waikato had changed its grading criteria, becoming slightly stricter from semester A 2016, so that for many students, receiving one point towards their course grade was very valuable because it might push up C+ to B- or B+ to A- and so on. Thus, the lower attrition observed in my IRAP study in both incentive and control groups could have been caused in part by the course credit that was offered to all participants for participation. This incentive to participate might have resulted in participants in both groups performing well, which may have led to the aforementioned ceiling effect. This may have influenced my ability to detect the influence of the

independent variable by pushing all participants' performances closer to the ceiling.

In conclusion, I investigated the effectiveness of a monetary incentive on IRAP performance because I was particularly concerned by the high attrition and variability in past IRAP research. In addition, generally low to moderate levels of internal consistency have been found in implicit research (Golijani-Moghaddam, Hart, & Dawson, 2013). Identifying some reliable control factors that may improve attrition and participants' overall performance in IRAP and also the applicability of the IRAP to individual assessments were my main interests in the current research. Thus, I selected a form of monetary incentive to alter the effectiveness of the existing reinforcer in the hope of improving the accuracy of IRAP performance. Through my IRAP research, I found that administration of a monetary incentive for meeting accuracy and latency criteria could be considered as a practical and convenient strategy to improve participants' performance, hence it may also improve the data quality, and lower the attrition rate. Furthermore, careful selection of the IRAP stimulus sets (e.g., pictures and words) appears to be critical, as the material chosen might change the outcome drastically, potentially reducing the validity of the IRAP effect. IRAP requires further testing to improve the reliability of the technique via replication of existing studies; for this reason, I conducted my IRAP study on anti-fat bias, based on two previous studies using different population samples. A strong *pro-slim* implicit attitude was obtained in all three IRAP studies; Roddy et al., (2010), Expósito et al., (2015) and my research. It would be useful to further investigate the *anti-fat* implicit attitude in a different cultural context, especially in a culture where people do not share the "thin ideal" for body-size.

IRAP has great potential, which could be applied in many areas in psychology and education. I have envisioned the possibility in the future of extending IRAP research to various clinical and possibly criminal cases to identify a bias or a strong adherence to a particular thought pattern at an individual level. In such circumstances, it would provide vital information to predict classes of behaviour. I used the external self-report measurement (BIQ) for the prediction of anti-fat behaviour; however, I would consider that it would be significant for future IRAP research to obtain data from actual behaviour to evaluate whether implicit attitudes (BIRRs) could be significant indicators that signal the probability of a particular class of behaviour.

References

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). Exemplar training and a derived transformation of function in accordance with symmetry. *The Psychological Record, 51*, 287-308.

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). Exemplar training and a derived transformation of function in accordance with symmetry: II. *The Psychological Record, 51*, 589-603.

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.

Barnes-Holmes, D., Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The Implicit Relational Assessment Procedure as a response-time and event-related-potentials methodology for testing natural verbal relations: a preliminary study. *The Psychological Record, 58*, 497-516.

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*, 169-177.

Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The implicit relational assessment procedure: exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record, 60*, 57-80.

Barnes-Holmes, D., Waldron, D., Barnes-Holmes, Y., & Stewart, I. (2009). Testing the validity of the Implicit Relational Assessment Procedure and the Implicit Association Test: measuring attitudes toward Dublin and country life in Ireland. *The Psychological Record, 59*, 389-406.

Bassili, J. N., & Smith, M. C. (1986). On the spontaneity of trait attribution: covering evidence for the role of cognitive strategy. *Journal of Personality and Social Psychology, 50*, 239-245.

Baum, W. M. (2005). *Understanding behaviorism: Behavior, culture, and evolution.* Malden, MA: Wiley-Blackwell.

Biglan, A., & Hayes, S. C. (1997). Should the behavioral sciences become more pragmatic? The case for functional contextualism in research on human behavior. *Applied and Preventive Psychology, 5,* 47-57.

Bessenoff, G. R., & Sherman, J. W. (2000). Automatic and controlled components of prejudice toward fat people: evaluation versus stereotype activation. *Social Cognition, 18,* 329-353.

Brady, N. C., & McLean, L. K. S. (2000). Emergent symbolic relations in speakers and nonspeakers. *Research in Developmental Disabilities, 21*, 197-214.

Brochu, P. M., & Morrison, M. A. (2007). Implicit and explicit prejudice toward overweight and average-weight men and women: testing their correspondence and relation to behavioral intentions. *The Journal of Social Psychology, 147,* 681-706.

Burkholder, E. O., & Peláez, M. (2000). A behavioral interpretation of Vygotsky's theory of thought, language, and culture. *Behavioral Development Bulletin 9*, 7-9.

Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality & Social Psychology, 66*, 882-894.

De Houwer, J. (2002). The Implicit Association Test as a tool for studying dysfunctional associations in psychopathology: strengths and limitations. *Journal of Behavior Therapy and Experimental Psychiatry, 33*, 115-133.

Drake, C. E., Seymour, K. H., & Habib, R. (2016). Testing the IRAP: Exploring the reliability and fakability of an idiographic approach to interpersonal attitudes. *The Psychological Record, 66*, 153-163.

Expósito, P. M., López, M. H., & Valverde, M. R. (2015). Assessment of implicit anti-fat and pro-slim attitudes in young women using the Implicit Relational Assessment Procedure. *International Journal of Psychology and Psychological Therapy, 15*, 17-32.

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

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013-1027.

Ferguson, M. J., & Bargh, J. A. (2007). Beyond the attitude object: Automatic attitudes spring from object-context. In B. Wittenbrink & N. Schwartz (Eds.), *Implicit measures of attitudes.* (pp. 216-246). New York, NY: Guilford.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics: (and sex and drugs and rock 'n' roll)* (4th ed.). Los Angeles, CA: Sage.

Fletcher, L., & Hayes, S. C. (2005). Relational frame theory, acceptance and commitment therapy, and a functional analytic definition of mindfulness. *Journal of Relational-Emotive & Cognitive-Behaviour Therapy, 23*, 315-336.

Gifford, E. V., & Hayes, S. C. (1999). Functional contexualism: A pragmatic philosophy for behavioural science. In W. O'Donohue & R. Kitchener (Eds.), *Handbook of behaviourism* (pp.285-327). San Diego, CA: Academic Press.

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics, 115*, 791-810.

Golijani-Moghaddam, N., Hart, A., & Dawson, D. L. (2013). The implicit relational assessment procedure: Emerging reliability and validity data. *Journal of Contextual Behavioral Science, 2*, 105-119.

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

Greenwald, T., Banaji, M., & Nosek, B. (1998). *Project Implicit: Take a test*. Retrieved from https://implicit.harvard.edu/implicit/selectatest.html

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*, 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

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

Guerin, B. (1994). Special section on attitudes and behavior in social psychology: attitudes and beliefs as verbal behavior. *The Behavior Analyst, 17*, 155-163.

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fisher, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup and ingroup face stimuli. *Neuroreport, 11*, 2351-2354.

Hayes, S. C. (1993). Analytic goals and the varieties of scientific contexualism. In S. C. Hayes, L. J. Hayes, H. W. Resse, &T. R. Sarbin (Eds.) *Varieties of scientific contexualism* (pp.11-27). Reno, NV: Context Press.

Hayes, S. C. (1994). Relational frame theory as a behavioral approach to verbal events. Chapter in S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.), *Behavior analysis of language and cognition* (pp. 9-30). Reno, NV: Context Press.

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

Hayes, S. C., Blackledge, J. T., & Barnes-Holmes, D. (2001). Language and cognition: Constructing an alternative approach within the behavioral tradition. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.) *Relational frame theory: A post-Skinnerian account of human language and*

cognition. (pp.3-20). New York, NY: Kluwer Academic/Plenum Publishers.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65, 185-241.

Horne, P. J., & Lowe, C. F. (1997). Toward a theory of verbal behavior. *Journal of the Experimental Analysis of Behavior, 68,* 271-296.

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 S. Dymond & B. Roche (Eds.), *Advances in relational frame theory and contextual behavioural science*. (pp.97-125) Oakland, CA: New Harbinger.

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

Hussey, I., Thompson, M., McEnteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science, 4*, 157-162.

Lipkens, R., Hayes, L. J., & Hayes, S. C. (1993). Longitudinal Study of the Development of Derived Relations in an Infant. *Journal of Experimental Child Psychology, 56*, 201-239.

Lloyd, K. E. (1994). Do as I say, not as I do. *The Behavior Analyst / MABA*, 17, 131-139.

Lowe, C. F., Horne, P. J., Harris, F. D., & Randle, V. R. (2002). Naming and categorization in young children: Vocal tact training. *Journal of the Experimental Analysis of Behavior, 78*, 527-549.

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*, 253-268.

Moore, J. (2015). *From a behavioral point of view: A psychological primer*. Cornwall-on-Hudson, NY: Sloan Publishing.

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

Nuttin J. M. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15*, 353-361.

O'Brien, K. S., Hunter, J. A., & Banks, M. (2007). Implicit anti-fat bias in physical educators: physical attributes, ideology and socialization. *International Journal of Obesity, 31*, 308-331.

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

Roddy, S., Stewart, I., & Barnes-Holmes, D. (2010). Anti-fat, pro-slim, or both? Using two reaction-time based measures to assess implicit attitudes to the slim and overweight. *Journal of Health Psychology*, *15*, 416-425.

Roche, B., & Barnes, D. (1997). A transformation of despondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior, 67*, 275-301.

Rudy, E. B., Estok, P. J., Kerr, M. E., & Menzel, L. (1994). Research incentives: money versus gifts. *Nursing Research, 43*, 253-255.

Setchell, J., Watson, B., Jones, L., Gard, M., & Briffa, K. (2014). Physiotherapists demonstrated weight stigma: A cross-sectional survey of Australian physiotherapists. *Journal of Physiotherapy, 60*, 157-162.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior*, 37, 23-44.

Sidman, M., Willson-Morris, M., & Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: The role of naming. *Analysis and Intervention in Developmental Disabilities, 6*, 1-19.

Skinner, B. F. (1938). *The behaviour of organisms: An experimental analysis.* New York, NY: Appleton-Century.

Skinner, B. F. (1957). *Verbal behaviour.* New York, NY: Appleton-Century-Crofts.

Teachman, B. A. (2007). Evaluating implicit spider fear associations using the go/no-go association task. *Journal of Behaviour Therapy of Experimental Psychiatry, 38*, 156-167.

Teachman, B. A., & Brownell, K. D. (2001). Implicit anti-fat bias among health professionals: is anyone immune? *International Journal of Obesity, 25*, 1525-1531.

Vanman, E. J., Paul, B. Y., Ito, T. A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of personality and social psychology, 73*, 941-960.

Vygotsky, L. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature*, 217, 972-974.

Watt, A., Keenan, M., Barnes, D., & Cairns, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record, 41*, 33-50.

Whigham, L. D., Schoeller, D. A., Johnson, L. K., & Atkinson, R. L. (2013). Effect of clothing weight on body weight. *International Journal of Obesity, 37*, 160-161.

World Health Organization. (2006). BMI Classification. Retrieved from http://www.who.int/mediacentre/factsheets/fs311/en/

Appendices

Appendix A: Information Sheet (Experiment Brief)

A Research Project: Investigating influences of incentives on implicit attitudes toward body size

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 Rebecca Sargisson, phone 07 557 8673, email: rebeccas@waikato.ac.nz

What is it about?

This experiment is designed to examine how language and cognition works. Using a computer program known as the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008), we will investigate if the IRAP will reveal any particular implicit bias against "fat" and "slim" people which is a common problem in some cultures. You will be asked questions relating to silhouettes of slim and overweight people. In addition, at the end of the experiment, you will be asked to report your height and weight, which will be recorded anonymously.

This is a Master's thesis research project supervised by Dr Timothy Edwards, phone 07 557 9409 and Dr Rebecca Sargisson, phone 07 557 8673.

How much of your time will participation involve?

Around 40 to 60 minutes.

Will your participation in the project remain confidential?

If you agree to take part, your name will not be recorded either on the questionnaires (including on the data sheet for height and weight) or the IRAP program. The information will also not be disclosed to other parties. Your responses to the questions and data from the IRAP program will be used for the purpose of this project only. You can be assured that if you take part in this project you will remain anonymous.

What kind of materials and tasks involved in this experiment?

- The IRAP computer program
- A keyboard – either pressing 'd' or 'k' keys
- Self-report questionnaires – using a paper and pencil
  - Explicit Anti-Fat Attitudes Questionnaire (AFA; Crandall, 1994)
  - Behavioural Intention Questionnaire (BIQ; Brochu and Morrison, 2007; Roddy et al., 2009)
  - General demographic information questionnaire (GIQ)
- A measuring tape and electric scale for height and weight measurements (to calculate Body Mass Index)

What are the advantages of taking part?

- You may find the project interesting because the IRAP is relatively a new measurement tool for assessing implicit attitudes based on behavioural analysis.
- If you complete the experiment, you will receive either course credit (only applicable for undergraduate psychology papers) or enter a draw to win one of five $20 iTunes

vouchers. Even if you decide to quit in the middle of the experiment, you are still entitled to enter the draw to win the voucher.

- If you are interested to know more about the studies of implicit attitude concerning overweight prejudice, a list of references will be provided for your information.

What are the disadvantages of taking part?

- You may feel a minor eyestrain from staring at the monitor during the IRAP testing. You can take a short break between the IRAP sessions. The length of each block trial is about 1 minute.
- If you feel any mental distress after participating in this research, please contact the university's Student Counselling Services for support. To make a confidential appointment, contact Student Health or phone 07 838 4037. www.waikato.ac.nz/go/counselling

What happens now?

If you agree to participate, please sign the attached consent form and the researcher will provide further instructions.

References

Barnes-Holmes, D., Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The Implicit Relational Assessment Procedure as a response-time and event-related-potentials methodology for testing natural verbal relations: a preliminary study. *The Psychological Record, 58*(4), 497-516.

Brochu, P. M., & Morrison, M. A. (2007). Implicit and explicit prejudice toward overweight and average-weight men and women: testing their correspondence and relation to behavioral intentions. *The Journal of Social Psychology, 147*(6), 681-706.

Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality & Social Psychology, 66*, 882-894.

Roddy, S., Stewart, I., & Barnes-Holmes, D. (2010). Anti-fat, pro-slim, or both? Using two reaction-time based measures to assess implicit attitudes to the slim and overweight. *Journal of Health Psychology, 15*(3), 416-425.

Appendix B: Consent Form

School of Psychology

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

CONSENT FORM

A completed copy of this form should be retained by both the researcher and the participant.

Research Project: Investigating influences of incentives on implicit attitudes toward body size

Please complete the following checklist. Tick (✓) the appropriate box for each point.	YES	NO
1. I have read the Participant Information Sheet (or it has been read to me) and I understand it.		
2. I have been given sufficient time to consider whether or not to participate in this study		
3. I am satisfied with the answers I have been given regarding the study and I have a copy of this consent form and information sheet		
4. I understand that taking part in this study is voluntary (my choice) and that I may withdraw from the study at any time without penalty		
5. I have the right to decline to participate in any part of the research activity		
6. I know who to contact if I have any questions about the study in general.		
7. I understand that my participation in this study is confidential and that no material, which could identify me personally, will be used in any reports on this study.		
8. I wish to receive a copy of the findings. If yes, provide email address:_____		

Declaration by participant:

I agree to participate in this research project and I understand that I may withdraw at any time. If I have any concerns about this project, I may contact the convenor of the Psychology Research and Ethics Committee (Dr Rebecca Sargisson, phone 07 557 8673, email: rebeccas@waikato.ac.nz)

Participant's name (Please print):

Signature: Date:

Declaration by member of research team:

I have given a verbal explanation of the research project to the participant, and have answered the participant's questions about it. I believe that the participant understands the study and has given informed consent to participate.

Researcher's name (Please print): Tokiko Taylor

Signature: Date:

Appendix C: IRAP rules and stimuli

Table 1. Rules and stimuli for an IRAP investigating body-size bias

Rule for a consistent block: Answer as if you are *anti-fat* and *pro-slim*
Rule for an inconsistent block: Answer as if you are *pro-fat* and *anti-slim*

***Label 1*: Silhouette of overweight individual**	***Label 2:* Silhouette of slim individual**
***Target 1*: Positive**	***Target 2:* Negative**
Desirable Active Disciplined Attractive Healthy Good	Undesirable Lazy Sloppy Ugly ill Bad

***Label 1:* Silhouettes of overweight individuals**

***Label 2:* Silhouettes of slim individuals**

***Figure 1*.** IRAP picture stimuli – 12 silhouettes of female and male overweight vs. slim individuals. These images are from the stimulus set used in Weight IAT tests for public viewers in the Project Implicit web site founded by Greenwald, Banaji, and Nosek (1998) for their social implicit studies.

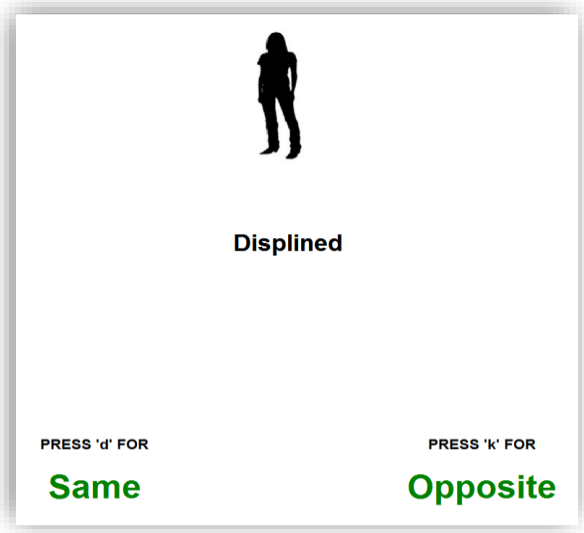
Appendix D: IRAP Experimenter's Script

Before commencing the experiment, make sure that IRAP is up and running. The monitor screen should be showing the following message "*During the next phase, answer as if you are anti-fat and pro-slim. This is just a practice. Please avoid the red X.*"

Experimenter: "*The IRAP task will ask you to follow a rule and test how easy or difficult you find it to follow that rule. As you can see on the screen, you need to answer as if you are anti-fat and pro-slim. You might not personally agree with this, but that is fine*."

Experimenter: "*During each trial you will see a silhouette image of either an overweight or a slim person at the top of the screen, and either a positive or negative word, such as 'desirable' or 'undesirable' in the middle of the screen.*"

Experimenter: "*Your two response options are at the bottom of the screen. Press the 'd' key for the left option and 'k' key for the right option. These options can swap sides. You are about to begin the practice phase and after the practice, you need to pass certain performance criteria that I will explain to you bit by bit.*"



An image of an IRAP trial – this visual aid will not be shown to the participant. The experimenter will explain it verbally without using any visual aids.

****** Only for the incentive group********

Experimenter: "*If you meet the criteria, which require you to respond accurately and quickly during the practice phase and maintain a good performance until the end of the IRAP test, you will earn a $20 voucher in addition to the participation reward of either course credits or entry into a draw to win a $20 iTunes voucher. The extra $20 voucher is a part of this experimental design, so, you were not informed about it at the beginning of the experiment briefing. Even if you do not pass the criteria, you can still obtain the participation rewards. Would you still like to continue?*"

If the participant declines the offer, he or she will be thanked and dismissed.

****

Experimenter: "*Ok, let's start practicing. Go as slowly you need to get them all right according to the rule*."

When the participant gets a trial wrong for the first time, the participant may ask about the red "X" sign in the middle.

Experimenter: "*You will receive the red 'X' if you do not follow the rule, which is to answer as if you are anti-fat and pro-slim in this block. You need to give it the 'right' answer according to the rule. The 'X' sign will not disappear until you give the correct answer. You can then proceed to the next trial*."

When the participant completes the first block, accuracy feedback (% correct) will be shown on the monitor. This continues until the end of the second block. Again accuracy feedback will be presented. Before proceeding to the next practice phase, the experimenter will introduce a speed criterion, which requires participants to answer within two seconds in each trial.

Experimenter: "*From the next practice phase, you will be asked to be both accurate and quick. From now on the task will show a red exclamation mark at the bottom of the screen if you take more than the two seconds to answer in any given trial. When you answer accurately you will naturally start to go very quickly*."

If a participant is responding very quickly but sacrificing accuracy, the experimenter needs to stop this quickly and effectively. The experimenter may stop them mid-block and emphasize that they need to concentrate on being accurate first.

Experimenter: "*If you see a few consecutive red 'X's, it means that you are going too fast. Try slowing down a little bit, as you need to focus on answering correctly. Then you will speed up naturally.*"

When the participant meets the criteria

Experimenter: "*Very good. You have shown that you learned how to complete the task. Keep going. The program will remind you what the new rule is before each block, and you will do six more block trials. Keep being as accurate as possible, and when you are responding accurately you will naturally go quickly. You will still be given feedback after each block so that you can see how you are doing. If you find your scores have dropped, you can take a break between blocks*."

******* For the incentive group**

Experimenter: "*Very good. You've shown that you have learned how to do the task.* ***Now keep going. If you maintain your good performance till the end of the test period, you can earn the $20 voucher.*** *The program will remind you what the new rule is before each block, and you will do six more block trials. Keep being as accurate as possible, and when you are responding accurately you will naturally go quickly. You will still be given feedback after each block so that you can see how you are doing. If you find your scores have dropped, you can take a break between blocks*."

****

If the participant does not meet the criteria, they will go through the test blocks as well for a post-hoc data analysis. Press **Ctrl**, **Alt,** and **Delete** keys to force-quit the practice program and begin the test blocks of the IRAP program.

Experimenter: "*Ok, you finished the practice phase. You did not pass the criteria, but it is ok. Let's try the test blocks. The program will remind you what the new rule is before each block, and you will do six more block trials. Keep being as accurate as you can, and when you are responding accurately you will naturally go quickly. You will still be given feedback after each block so that you can see how you are doing. If you find your scores have dropped, you can take a break between blocks*."

Appendix E: General Information Questionnaire

Please answer the following questions:

	General Information Questionnaire			
1	I identify my gender as ….	Female	Male	Transgender
2	What is your age?			
3	Have you ever tried to lose weight?	Yes		No
4	Have you taken the IRAP test before?	Yes		No
5	IF you select "Yes," how many times have you taken the IRAP experiments so far?			

Thank you. Please return this form to the experimenter.

Appendix F: Anti-Fat Attitudes (AFA) questionnaire to test explicit weight stigma

NB: the language in this questionnaire has been adapted slightly to be suitable for professional Australian participants, for original wording please see the reference at the bottom of the page.

Please answer as honestly as possible. Your answers are anonymous.

Dislike

	Very strongly disagree	Strongly disagree	Disagree	Disagree somewhat	Unsure	Agree somewhat	Agree	Strongly agree	Very strongly agree
1. Few of my friends are overweight or obese.	○	○	○	○	○	○	○	○	○
2. I tend to think that people who are overweight are a little untrustworthy.	○	○	○	○	○	○	○	○	○
3. Although some overweight people must be intelligent, generally I think they tend not to be.	○	○	○	○	○	○	○	○	○
4. I have a hard time taking overweight people too seriously.	○	○	○	○	○	○	○	○	○
5. Fat people make me somewhat uncomfortable.	○	○	○	○	○	○	○	○	○
6. If I were an employer, I might avoid hiring an overweight person.	○	○	○	○	○	○	○	○	○
7. I dislike people who are overweight or obese.	○	○	○	○	○	○	○	○	○

Fear

	Very strongly disagree	Strongly disagree	Disagree	Disagree somewhat	Unsure	Agree somewhat	Agree	Strongly agree	Very strongly agree
8. I feel disgusted with myself when I gain weight.	○	○	○	○	○	○	○	○	○
9. One of the worst things that could happen to me would be if I gained 10kgs.	○	○	○	○	○	○	○	○	○
10. I worry about becoming fat.	○	○	○	○	○	○	○	○	○

Willpower

	Very strongly disagree	Strongly disagree	Disagree	Disagree somewhat	Unsure	Agree somewhat	Agree	Strongly agree	Very strongly agree
11. People who weigh too much could lose at least some part of their weight through a little exercise.	○	○	○	○	○	○	○	○	○
12. Some people are overweight because they have no willpower.	○	○	○	○	○	○	○	○	○
13. It is people's own fault if they are overweight.	○	○	○	○	○	○	○	○	○

1. Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. The Journal of Personality and Social Psychology, 66(5), 882-894.

2. Setchell, J., Watson, B., Jones, L., Gard, M., & Briffa, K. (2014). Physiotherapists demonstrate weight stigma: A cross-sectional survey of Australian physiotherapists. Journal of Physiotherapy, 60(3), 157-162.

Appendix G: Behaviour Intentions Questionnaire (BIQ)

Seth

Daniel

Lisa

Joy

Behaviour Intentions Index

1. How likely is it that you would want to get to know **Daniel** better?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

2. How likely is it that you would ask **Daniel** if you could copy his notes from a class you missed?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

3. How likely is it that you would want to work on a class project with **Daniel**?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

4. How likely is it that you invite **Daniel** to study group for an exam?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

5. How likely is it that you want to become friends with **Daniel**?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

Behaviour Intentions Index

1. How likely is it that you would want to get to know **Seth** better?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

2. How likely is it that you would ask **Seth** if you could copy his notes from a class you missed?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

3. How likely is it that you would want to work on a class project with **Seth**?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

4. How likely is it that you invite **Seth** to study group for an exam?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

5. How likely is it that you want to become friends with **Seth**?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

Behaviour Intentions Index

1. How likely is it that you would want to get to know **Joy** better?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

2. How likely is it that you would ask **Joy** if you could copy her notes from a class you missed?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

3. How likely is it that you would want to work on a class project with **Joy**?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

4. How likely is it that you invite **Joy** to study group for an exam?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

5. How likely is it that you want to become friends with **Joy**?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

Behaviour Intentions Index

1. How likely is it that you would want to get to know **Lisa** better?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

2. How likely is it that you would ask **Lisa** if you could copy her notes from a class you missed?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

3. How likely is it that you would want to work on a class project with **Lisa**?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

4. How likely is it that you invite **Lisa** to study group for an exam?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	

5. How likely is it that you want to become friends with **Lisa**?

Very Unlikely	1	2	3	4	5	6	7	Very Likely
	O	O	O	O	O	O	O	