ABSTRACT

Title of Document: QUESTIONING THE VALIDITY OF THE IAT: KNOWLEDGE OR ATTITUDE?

.

Eric F. Siegel, M.A., 2006

Directed By: Professor, Harold Sigall, Department of Psychology

A series of experiments was conducted to determine if the Implicit Association Test is a valid measure of attitude, or if it instead reflects knowledge of the groups being tested. Participants were given positive or negative descriptions groups of which they had no previous knowledge. They then completed an IAT testing their attitudes towards these groups. The results revealed that the IAT was sensitive to whether the participants believed these descriptions were accurate. This indicates that the IAT measures more than simple knowledge, it is sensitive to whether knowledge is endorsed. These findings support the validity of the IAT as an attitude measure.

# QUESTIONING THE VALIDITY OF THE IAT: IS IT KNOWLEDGE OR ATTITUDE?

By

Eric F. Siegel

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Arts
2006

Advisory Committee:
Professor Harold Sigall, Chair
Professor Charles Stangor
Associate Professor David Huber

# Acknowledgements

I'd like to acknowledge the help of Dr. Sigall and Dr. Huber, whose guidance and support made this research possible.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Literature Review

## *1.1 Introduction*

The study of attitudes is one of the largest areas of research in social psychology. Although *attitude* has many definitions, we will define it as a favorable or unfavorable evaluation of someone (or something) that is exhibited through beliefs, feelings, or behaviors (Greenwald and Banaji, 1995). One branch of attitude research focuses on the relationship between attitude and behavior. Unfortunately, the association between attitude and behavior has proven tenuous at best; people will often express attitudes that have little connection to their subsequent behaviors (Festinger, 1964; Ajzen & Fishbein, 1977; Fazio & Zanna, 1981; Kelman, 1974; Wicker, 1969). There are many possible reasons for the separation between attitudes and behavior. One reason that is often an issue in research is fear that an attitude is socially unpopular and will lead to criticism from others; researchers call this concern social desirability (Crosby, Bromley, & Saxe, 1980; Fazio, Jackson, Dunton, & Williams, 1995).

Participants are often concerned about social desirability in an experimental setting because they know their actions are being scrutinized and recorded. Social desirability can lead participants to hide their true attitudes, which is a serious threat to an experiment's validity (Clark & Tifft, 1966; Knudson, Pope, & Irish, 1967; Paulhaus, 1984). This problem is particularly acute for research on attitudes towards a racial or ethnic group, because modern standards of political correctness have made it socially unacceptable to express racist attitudes (Sigall & Page, 1971). It is difficult

to get accurate measures of prejudice, because people are unwilling to express attitudes which could be considered "racist."

Recently, researchers have developed a series of implicit measurers which they argue can bypass social desirability effects (Greenwald, McGhee, & Schwartz, 1995; Hetts, Sakuma, & Pelham, 1999; Jones, Pelham, Mirenberg, & Hetts, 2002; von Hippel, Sekaquaptewa, & Vargas, Wittenbrink, Judd, & Park, 1997). Implicit tests are designed so that the test taker cannot deliberately control his responses. Proponents of these measures argue this renders them immune to social desirability. One of the most popular implicit measures is the Implicit Association Test (IAT) (Greenwald, et al., 1998). Every IAT features two groups, and the IAT measures the strength of the association between each group and both a positive and a negative valence. The relative strength of the associations are used to indicate the test taker's attitude towards the two groups.

The IAT has enjoyed explosive popularity within the past few years. It has been used in hundreds of psychological studies and has begun to see use in mainstream applications, such as political polling and as an educational tool in the classroom. (Greenwald & Nosek, 2001; Kester, 2001). Despite the measures' popularity, numerous critics have questioned the IAT's validity as an attitude measure (Nosek, Greenwald, & Banaji, in press). This research will focus in on one specific criticism of the IAT raised by Karpinski and Hilton (2001). They argue the IAT does not measure attitude towards the groups; instead it measures knowledge of how they are commonly portrayed in the test taker's culture. The goal of this research is to determine whether this criticism of the IAT is valid.

The following paper will first review social desirability and the development of implicit measures designed to overcome its influence on attitude measurement. We then turn to the IAT itself; the processes that underlie its ability to measure attitude, how it is administered and scored, and the evidence that it measures knowledge rather than attitude. The paper describes a series of experiments conducted to test the core validity of the IAT. Results and implications of the experiment and plans for future research are then reviewed.

## *1.2 Social-Desirability and Implicit Measurement*

Designing valid methods of attitude measurement has been a persistent issue in psychological research[MSOffice2]. One challenge has been controlling for *social-desirability:* deception by participants designed to convey attitudes that should lead to praise or hide attitudes that could lead to criticism. (Crosby et al., 1980; Crowne & Marlowe, 1960; Dovidio et al., 1997; Fazio et al., 1995). Concern about social desirability can cause participants to hide their true attitudes when participating in an experiment, which can seriously undermine the validity of an attitude measure. Social-desirability has a particularly strong effect on measures of socially sensitive topics, such as measures of sexual habits (Knudson, Pope, & Irish, 1967), deviant behaviors (Clark & Tifft, 1966), and stereotypes (Sigall & Page, 1971). Because of the threat social desirability poses to test validity, it is critical to both understand the nature of social desirability and develop effective methods for its control.

Researchers can use measures of social desirability to control for its effects on an attitude measure. They can administer one of these measures along with the primary measure of their research. The results of the social desirability measure are

used to control for social desirability in the main measure. This can be done by covarying out the effect of social desirability, or dropping individual questions from the main measure that are strongly affected by social desirability (Norman, 1967; Karson & O'Dell, 1976).

Psychologists have been designing measures of social desirability for almost half a century. The first test developed to measure social desirability was the L-Scale, it was designed as an addition to the Minnesota Muliphasic Personality Inventory (MMPI[MSOffice4]) (Edwards, 1957). This scale was later modified for use on a non-clinical population by Crown[BU5]e and Marlowe (1960) and renamed the Marlowe-Crowne Scale. This scale is based on the model that social-desirability is a stable, single factor personality construct. Responses to the scale indicate the degree to which a person will respond according to what he believes is socially acceptable in any given situation. The Marlowe-Crowne scale remains the most popular measure of social desirability, but other independent measures of social-desirability exist [see Paulhaus (1991) for a review of these measures].

A dual factor model of social desirability was developed more recently by Paulhus (1984). This model divides social desirability is into two distinct factors:

- *impression management*-- deliberate, conscious deception on an attitude measure designed to garner praise and avoid criticism
- *self-deception* -- unconscious deception on an attitude measure designed to garner praise and avoid criticism. Participant believe their answers are accurate, but they do not predict their future behavior.

Paulhaus (1984) developed a measure based on this two factor model called the Balanced Inventory of Desirable Responding (BIDR) that has also been widely used in psychological research to measure social-desirability.

## *1.3 Methods of Implicit Measurement*

Researchers have attempted to sidestep the difficulties posed by social-desirability altogether by developing *implicit measures* of attitude (Greenwald & Banaji, 1995). An implicit measure is a test in which the construct being measured is not directly related to the content of the test itself; this makes it difficult to consciously control responses. (Spence, 2006; De Houwer, 2006) In theory, the oblique nature of implicit measures neutralizes any attempt by subjects to deliberately modify, alter or revise their responses (Banse, Seis, & Zerbes, 2001; Egloff & Schmukle, 2002; Kim, 2003). Therefor, social desirability should have no effect on an implicit measure. Over the years, a number of different types of implicit measures have been developed.

### 1.3.1 Projective Measures

The earliest forms of implicit measures were projective measures such as the Rorschach inkblot test and the Thematic Apperception Test (TAT). Projective measures are used to infer attitudes or desires from reactions to neutral stimuli. The type of neutral stimuli varies by test; the Rorschach test asks participants to describe their immediate impressions of an abstract picture (Grønnerød, 2003; Weiner, 2006), while the TAT requires participants to create a story to explain an ambiguous scene. For example, a participant might be shown a picture of a man and a woman waiting at

a bus stop and asked to describe their relationship. The researcher interprets the participants' reactions to the pictures and makes an assessment of their personality traits and mental state. While these measures are still used in clinical settings, research has demonstrated that these measures lack validity (Garb, Wood, Nezworski, Grove, & Stejskal, 2001; Lilienfeld, Wood, & Garb, 2000).

### 1.3.2 Biological Measures

Biological measures such as Galvanic Skin Response (GSR) and ElectroEncephograph (EEG) can also be used as implicit measures of attitude (Kozel, Revell, Lorberbaum, 2004, Kaiser, 2005). These measures detect arousal levels through subtle physiological reactions. These arousal levels can be used to indirectly infer attitude. The GSR detects arousal by measuring the electrical resistance of the skin. The test is administered by attaching two electrodes to the skin. A base measure is acquired, which is then compared to a measure taken after the experimental manipulation. An increase in the participant's level of skin conductivity indicates an increase in arousal. The GSR can be used to study attitude through these arousal responses. For example, negative attitudes towards a group can be measured through fear responses to members of that group (Lissek, Baas, & Pine, 2005), or it can be used to detect sexual arousal to asses sexual preference (Sarlo, Palomba, & Buodo, 2005). The EEG is a diagnostic test that measures electrical activity in the brain using recording equipment attached to the scalp by delicate electrodes. The resulting EEG data is a record of brain activity, and can identify which areas react to specific stimuli. The EEG can be used to study a variety of mental processes which are influenced by attitudes, including emotional responses (Tomarken, Davidson,

Henriques, 1990), memory (Buckner, Kelley, & Petersen, 1999), and attention (Müller & Gruber, 2001). For example, recently it has been demonstrated exposure to disliked objects can evoke activity in the amygdale, which controls fear responses (Phelps, O'Connor, Cunningham, Funayama, Gatenby, Gore, & Banaji, 2000).

The usefulness of biological measures is hampered by the difficulty of administering them properly. Scores can vary due to differences in the model and age of the equipment (Nuwer 1988, 1997). Results from biological measures can also be confounded by participant's age, gender, and state of alertness. These issues make the widespread use of these measures problematic

### 1.3.3 Reaction Time Measures

Reaction time measures are a relatively new form of implicit measurement. They gauge how strongly the participant associates different concepts and uses these associations to infer his attitudes. A *concept* is a person's mental representation of a person, group, or attribute. Reaction time measures can detect associations between three different types of concepts:

- An *object concept* is a mental representation of a group, individual, inanimate object, or institution.
- An *attribute concept* is a mental representation of a trait or characteristic.
- A *valence concept* is mental representation of positivity or negativity.

An *association* is a cognitive link between two concepts. If two concepts are associated, activation of one of the concepts speeds up the activation of the associated concept from memory (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). For

example, since people are likely to see nurses in a hospital, when someone goes to a hospital, it will be easier for him to remember what he knows about nurses. (Neely, 1976, 1977). *Accessibility* is the strength of an association --- the greater the accessibility of an association, the more one of the associated concepts will aid in the activation of the other. Accessibility is increased every time an association is utilized (Higgins, 1996).

When taking a reaction time measure, participants are asked to identify or categorize sets of words or pictures which represent groups the experimenter is researching. If two of the groups are associated, the participants are able to identify a word from one of the sets faster when it is paired with a word from an associated set (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The method used to pair the different sets varies depending on the measure. The decrease in response time is used to infer the associations between the groups tested by the measure. When a reaction time measure is used to measure attitude, a set positive words and a set of negative words are included in the measure. If words from a group set can be identified faster when they are paired with positive words, it indicates the group has an association with positivity, which indicates a positive attitude towards that group. If the words can be identified faster when they are paired with negative words, it indicates the participant has a negative attitude towards that group (Fazio, Sanbonmatsu, Powell, and Kardes, 1986, Fazio, 2001).

*1.4 Association Formation*

Associations are formed when people are exposed to information about a particular object. If the information describes the object as having a certain attribute, a simple association is formed that links the corresponding object concept to the corresponding attribute concept. For example, if you are told the Yankees are an arrogant baseball team, you would create an association between your conception of the Yankees with your conception of the attribute arrogance (Greenwald et al, 2002). This association would cause you to remember the Yankees are arrogant whenever you thought about them later. As shown in figure one, an object concept can be associated with multiple attribute concepts. The links between an object concept and multiple attribute concepts is called a *cognitive network*.

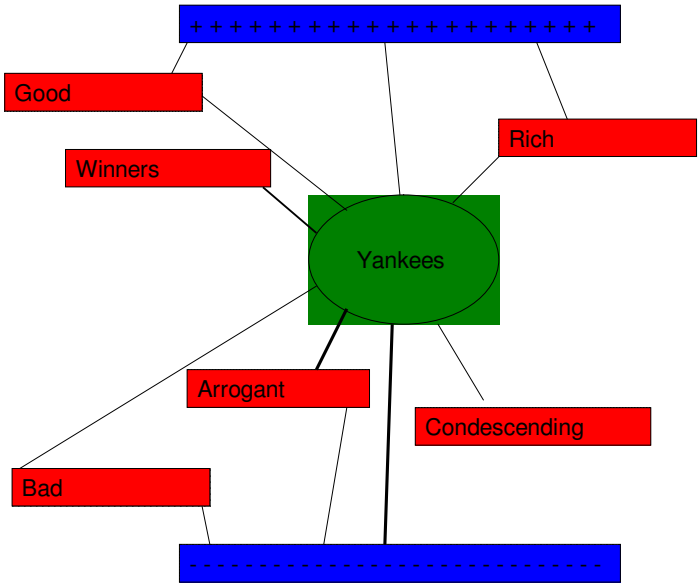


**Figure 2:** This figure demonstrates a cognitive association network. The object concept (Yankees) has cognitive associations with the attribute concepts used to describe the Yankees (Arrogant) and an affective association with the valence concepts based on their evaluation (+ + + + and - - - -). The thickness of the connecting lines demonstrates the associations can very in strength.

An unconscious association can also form between an object or attribute concept and a valence concept. If a person perceives a certain trait as good or bad, an association forms between the corresponding attribute concept and the appropriate valence concept (Greenwald et al., 2002). If that attribute concept has an association with an object concept, an indirect association forms between the object concept and the valence concept. If an object concept has a strong indirect association with a valence association, a direct association can form between the object concept and the valence concept. An indirect association between an object and valence concept is functionally identical to a direct association (Greenwald et al., 2002). Both of types of associations are demonstrated in figure one. A direct association exists between the

Yankees and negativity (----), and an indirect association exists with the attribute concept "arrogant" acting as an intermediary. The formation of a valence association can be described using the Yankees example. If you saw arrogance as a negative trait, you would form an indirect association between your concept of the Yankees and your concept of negativity. This association will cause a negative emotional reaction whenever you encounter the Yankee's logo, and would predispose you to make negative judgments about the Yankees.

Associations form whenever people receive information about a specific object. One important issue is whether associations form when people don't believe the information they are based on is true. The answer to this question depends on what process people use to judge the accuracy of information. Gilbert, Krull, and Malone (1990) argue that when information is first learned, it is assumed to be accurate. Later, the new information is reviewed and judged for accuracy. They supported this model with a series of studies. In their research, participants were told a series of facts about an imaginary animal. They were told some of these facts were true and others were false. They found that if the participants read a fact and were distracted with a short tone before they were told the fact was false, they were more likely to accidentally say it was true when they were questioned later. However, they were not likely to say a true fact was false if they were distracted by the tone. Gilbert et al. (1990) argued this was because the distraction prevented them from labeling the fact as false. The implication of this finding was that the fact was initially encoded as "true" by the participants.

The Gilbert (1990) model can be applied to the associations that form from new information. When people initially encounter a fact about an object, they will assume it is true and form the appropriate association. Later, the information's veracity will be judged. If the information is judged to be false, it will still be stored in memory but will "tagged" as untrue. The association that formed based on this information should also continue to exist, but will also be "tagged" as an inaccurate association. If two concepts are associated, exposure to one concept should still speed up the activation of an associated concept from memory, even if the association itself is perceived as inaccurate. To put it another way, associations should cause both endorsed and unendorsed information to be activated, but the endorsed information should determine unconscious attitudes, while the unendorsed information will be disregarded.

### *1.5 The Relationship Between Conscious and Unconscious Attitudes*

Since reaction time measures detect attitudes indirectly through associations, it is important to define how these associations are related to attitudes. Reaction time measures are implicit, which means participants' responses are largely automatic and outside of conscious control. The associations detected by reaction time measures are also unconscious and cannot be deliberately changed. [MSOffice12]Because of this, implicit measures exclusively detect unconscious attitudes, which are distinct from conscious attitudes. Greenwald defines *unconscious attitudes* as "introspectively unidentified traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects" (Greenwald & Banaji, 1995, pg 5). In most cases, the associations between object concepts and valence concepts should mirror

unconscious attitudes. (Fazio, 2001). If a person has an unconscious negative attitude towards a group, there should be a stronger association between that group's object concept and the negative valence concept. Conversely, if a person has a positive unconscious attitude towards a group, there should be a stronger association between that group's object concept and the positive valence concept. Reaction time measures are used to infer unconscious attitudes from these associations.

The relationship between unconscious and conscious attitudes is still a matter of debate (Fazio, 2001). One theory holds that conscious and unconscious attitudes are separate constructs. (Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooner, 2000) In contrast to unconscious attitudes, conscious attitudes are the result of rational, conscious processing of relevant information. Conscious attitudes are also influenced by *situational motivations*; motivations specific to the situation that affect the formation of the attitude. Social desirability is one type of situational motivation, but they can be any drive or goal that alters the attitude, such as making a correct decision, impressing others, or avoiding a conflict. Because conscious attitudes are the result of a deliberate process, they are predictive of controlled, deliberate behavior. In contrast, unconscious attitudes have been shown to predict automatic behavior which is outside of conscious control, such as body language (Fazio, 1990; Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooner, 2000). It is also possible that unconscious and conscious attitudes can influence each other; however the exact relationship between them is still unclear (Nozek, 2005).

Another view is that the unconscious attitudes are the only "true" attitudes. According to this view, situational motivations do not change or create attitudes; they

only influence the expression of the previously formed, unconscious attitudes. This means any recorded differences between unconscious and professed attitudes are due to the intervention of situational motivations (Fazio, 1990; Fazio & Olson, 2003). When situational motivations are in direct opposition to unconscious attitudes, the unconscious attitudes will be actively suppressed. However, processing situational motivations and suppressing unconscious attitudes requires cognitive resources. Therefore, the effect of situational motivations on attitude expression depends on the availability of cognitive resources (Bassili, 1995, 1996; Fazio, Powell, & Williams, 1989; Fazio & Williams, 1986). This means people will express previously held, unconscious attitudes if they are tired or distracted, but if there is time to consider all information carefully, situational motivations will have a comparably greater influence. The accessibility of the associations that underlie the attitude also determines whether the attitude can be suppressed. The more accessible the association, the more effort it takes to suppress. (Logan, 1988; Fazio, 1986).

The Yankees example can be used to demonstrate how situational motivations can influence unconscious attitudes. If a person who had a negative attitude towards the Yankees moved to New York, he would be motivated by social desirability concerns to say he liked the Yankees. He would therefore suppress any negative attitude they had about the Yankees. However, suppressing implicit attitudes requires mental effort, so if his resources were low, he would be unable to prevent his negative attitude from being expressed.

*1.6 Reaction Time Measures and Attitude*

Reaction time measures test for attitude by detecting for the association between an object concept and a valence concept. In order to test for this association, stimuli representing a valence concept are presented in conjunction with stimuli representing an object concept. If the two concepts are associated, stimuli representing one of the concepts will be recognized faster if it is displayed in conjunction with a stimulus representing the associated concept. (De Houwer, 2001; Fazio, 2001; Fazio et al, 1986 Greenwal et al., 1998). An association with a positive valence indicates a positive attitude, and an association with a negative valence indicates a negative attitude. For example[MSOffice15], a reaction time measure testing attitudes towards insects would display words or pictures related to insects (e.g. wasp, fly) in conjunction with both positive and negative words (e.g. wonderful or awful). If the test taker was able to identify or categorize negative words faster when they were preceded by insect words, it would imply the test taker has a negative attitude towards insects. Reaction time measures can also be used to test for a cognitive association between an object concept and an attribute concept. A reaction time measure of gender stereotypes would test whether exposure to words representing women (skirt, makeup) speeds up the identification of words representing weakness (passive, helpless). Supporters of reaction time tests argue they can be used to infer whether test taker believes the group posses a certain trait (Greenwald, et al., 2002; Poehlman, Uhlmann, Greenwald, & Banaji, in press).

*1.7 The Implicit Association Test*

One of the most popular reaction time measures is the Implicit Association Test (IAT), which was created by Greenwald, McGhee, and Schwartz (1998). Since its development, the IAT has in research of many topics, including self-esteem (Robinson, Meier, Zetocha, & McCaul, 2005), smoking attitudes, (Huijding, De Jong, Wiers, & Verkooijen 2005); and brand preferences (Maison, Greenwald, & Bruin, 2004). However, the original, and most controversial, use of the IAT is as a test of racism (Greenwald, McGhee, & Schwartz, 1998).

The IAT measures the relative preference for one group compared to another. Flowers and insects, Democrats and Republicans, and Blacks and Whites are examples of groups that have been tested for a relative preference (Greenwald, et al., 1998). One assumption built into the format of the IAT is the test taker has a positive attitude towards one of the groups and a negative attitude towards the other (Blanton & Jaccard, 2006). The two groups the researcher is interested in testing for preference are called the *target sets*. A list of positive words and a list of negative words are also included in the IAT. These two additional lists are called *valence sets.* The valence sets are often adjectives, such as "wonderful" or "awful." but can also be nouns, such as "cancer" or "joy." The IAT consists of a series of trials divided into blocks. In each trial, a word or picture from one of the sets is displayed in the center of a computer screen. The test taker must categorize these stimuli based on their set by pressing a key on a keyboard or response box. Generally, there is a key on the left and a key on the right side of the answer box, and each set is assigned to one of the answer keys.

The IAT consists of five blocks of 20-60 trials. Each block has a different combination of sets assigned to the answer keys:

I. In the first block, participants must categorize words or pictures belonging to the target sets. One set is assigned to the right answer key, and the other to the left answer key.

II. In the second block, they must categorize the words belonging to the valence sets. Once again, each set is assigned to a specific answer key.

III. In the third block, words or pictures from all four sets must be categorized. However, there are still only two answer keys. One target set and one valence set is assigned to each answer key.

IV. In the fourth block, participants must once again categorize words or pictures belonging to the target sets. However, the target set assigned to the right answer key in block one is now assigned to the left answer key, and the target set assigned to the left answer key in block one is now assigned to the right answer key.

V. In the fifth block, the test taker must categorize stimuli from all four sets. The two target sets retain the button assignments from block four. This means the target sets now share button assignments with the opposite valence set they did in block three. [BU18]

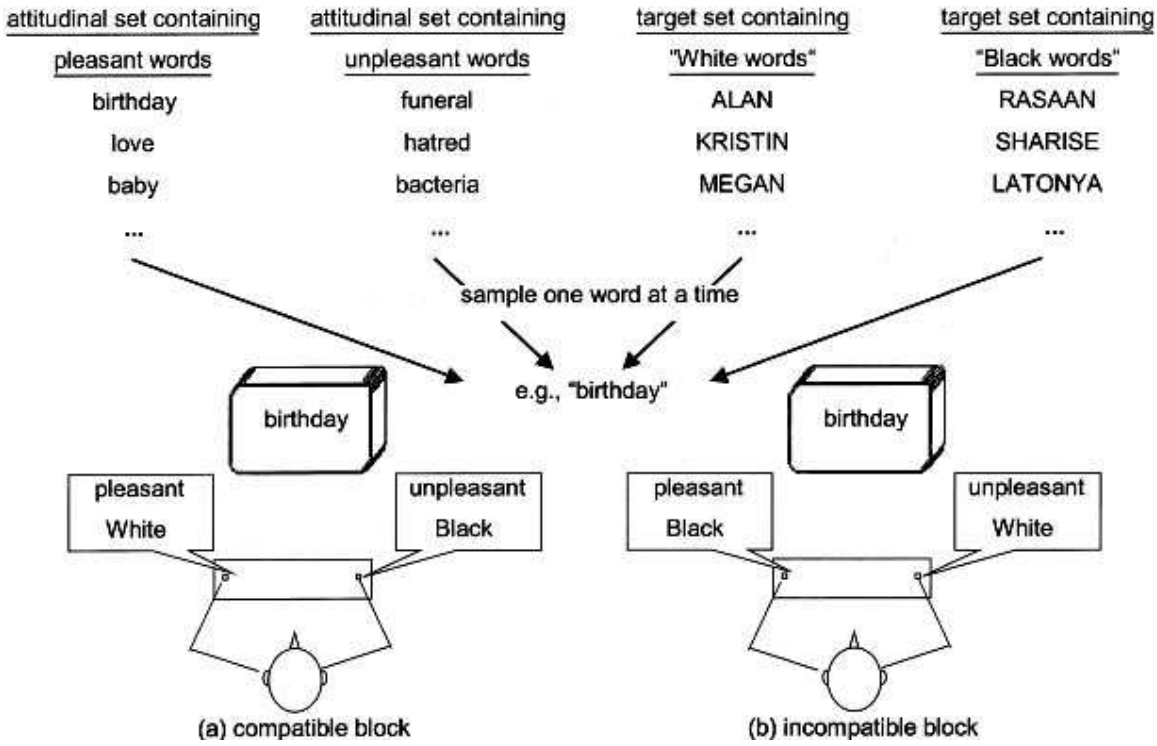


Figure 2 - Illustration of the IAT's main blocks (i.e. , Blocks 3 and 5). One word at a time is sampled randomly from the four word sets and presented on the computer screen (e. g. *birthday*). The participant has to categorize the word as belonging to one out of four categories using two response keys. The response key assignments differ in the (a) compatible block and (b) incompatible block. (Figure from Brendle, Markman, & Messner, 2001)

If the test taker has a positive attitude towards one of the target sets, that target set should have a stronger association with the positive valence set. The test taker should also have a negative attitude toward the target set that has a stronger association with the negative valence set. These associations are measured by how fast the participants can categorize the members of the different sets. This speed is measured by *Reaction time*; the amount of time it takes the test taker to identify a word and push the correct answer key for a single trial of the IAT. Reaction time in the IAT tends to be between 200 and 800 milliseconds (ms). The associations between the target sets and valence sets are measured by the difference in average reaction time between the two IAT blocks in which all four sets are presented to the participants (block three and block five). The block where each target set is assigned

to the same answer key as the valence set it is associated with is called the *compatible block*. The test taker is able to categorize the all of the words quickly in this block. The block in which each target set is assigned to the opposite answer key as the valence set it is associated with is called the *incompatible block*. In this block, the test taker will hesitate and answer more slowly compared to the compatible block. The researcher normally designates which block is the compatible block and which block is the incompatible block based on his prediction of which group will be liked and which will be disliked by the participants. Figure two shows an example of a Black/White IAT. The figure demonstrates the how the target sets share an answer key with a different valence set in the two critical blocks of the IAT. To obtain an IAT score, an average the reaction time is calculated for all the trials in the compatible block, and a second average is calculated for the incompatible block. Then, the average reaction time in the compatible block is subtracted from the average reaction time of the incompatible block to obtain a difference score, the *IAT effect*. The larger the IAT effect, the greater the test taker's preference the group paired with the positive valence set in the compatible block (Greenwald et al., 1998).

The process underlying the IAT can be explained using a random walk model (Nosofsky & Palmeri, 1997, Brendl, Markman, & Messner, 2001). In a random walk decision process, information is received over time that eventually leads to a specific decision or response. When the amount of information in favor of the response reaches a decision threshold, the response will be activated. In the case of the IAT, the participants are exposed to a word from one of the sets, and they process the word's identity over time (in this case a few hundred milliseconds), and eventually

reach a decision threshold which causes them to hit the appropriate response key. The amount of time it takes a participant to reach this decision threshold is his reaction time for that trial. In the IAT, a target set's association with one of the valence sets influences reaction time in the compatible and incompatible blocks. When the two sets are associated, words from the target set are perceived as having a secondary identity as a word from the valence set (Brendl, Markmen, & Messner, 2001). For example, in a Flower/Insect IAT "wasp" would be part of the target set "insect" but it would also have a secondary identity as a "bad" word.

In the incompatible block, the target sets are assigned to opposite answer key from the valence they are associated with. The participants therefore receive information that pushing both keys is the correct response. The participants know from the IAT's instructions and from the blocks of the IAT preceding the incompatible block that they are supposed to categorize target sets based on their group identity, so they will almost always answer correctly. However, the participant must suppress the target words' secondary identity as valence words, so it takes longer to reach a decision threshold for the target sets (Nosofsky & Palmeri, 1997, Brendl, Markman, & Messner, 2001). This effect is called response competition. In the compatible block, the target set's dual identity as a valence set increases response speeds, because both the association and the target set's identity leads the participant to push the correct answer key. This effect is called response facilitation. The more strongly the target set is associated with the valence set, the greater the influence on the participants' response speeds.

*1.8 The Validity of the IAT*

A number of research efforts support the validity of the IAT as an attitude measure. IAT scores have been correlated to a variety of behaviors, such as ratings of attractiveness (McConnell & Leibold, 2001), roommate preferences (Lambert, Payne, Ramsey, & Shaffer, 2005), and even discriminatory hiring practices (Ziegert, & Hanges, 2004). The IAT is also sensitive to stimuli that should change the test takers' attitudes, such as diversity seminars (Rudman, Ashmore, & Gary, 2001) and positive Black role models (Dasgupta, & Greenwald, 2001).

Even though the performance on the IAT has been linked to a variety of behaviors, research has found a weak correlation between the IAT and explicit measures of racism (Wilson, Lindsey, & Schooler, 2000; Karpinski & Hilton, 2004). Greenwald et al. (2002) argue this low correlation doesn't invalidate the IAT, but instead brings the validity of the explicit measures into question. They argue explicit measures are susceptible to social-desirability, while the IAT is immune. If there is a low correlation between the IAT and an explicit measure, the IAT is the more valid attitude measure, and the explicit measure is being influenced by the participants' social desirability concerns. (Greenwald, et al., 2002, Poehlman & Uhlmann, 2006).

However, the critics of the IAT argue it does not measure the participants' attitudes, but a separate construct that in some cases appears to be attitude. The criticism we focus on in our research originated from research by Karpinski and Hilton (2001). They argued the IAT is a measure of environmental association rather than attitude:

"The IAT may tap the associations a person has been exposed to in his or her environment. According to the environmental association model of the IAT, a high

> score on a White/Black IAT, for example, should not be seen as indicating that the individual has more favorable evaluations of Whites compared with Blacks. Instead, the score may simply indicate that the individual has been exposed to a larger number of positive-White and negative-Black associations than negative-White and positive-Black associations." (Karpinski & Hilton, 2001, pg 776)

If the IAT is a measure of the associations participants have been exposed to, it would not account in any way for whether the participants believe those associations are true. If Karpinski and Hilton (2001) are correct, it would mean the IAT effect measures the participants' knowledge of how the groups in the IAT are perceived by others, rather than their attitudes towards the groups. This criticism can be tied into Gilbert's (1990) theory of fact labeling. According to this model, people should have associations based on both endorsed and unendorsed information, but their attitudes should only be based on the endorsed associations. Since unendorsed associations still speed up the activation of the associated concepts, the IAT may detect both endorsed and unendorsed associations. Unendorsed association would not indicate attitude; instead they are an indication of the test taker's knowledge about the group. If this is the case, the IAT could not be considered an attitude measure, but rather a simple memory exercise.

Karpinski and Hilton (2001) supported their criticism of the IAT with a series of studies. Their first study compared the IAT's ability to predict behavior against other explicit measures. In one study, they administered an IAT and explicit measures assessing participants' attitudes towards apples and candy bars. They found that there was no correlation between the IAT and the explicit measures. However, the explicit measures predicted whether the participants would prefer to receive a free apple or candy bar. In contrast, the IAT did not predict the participants' choice of food. This

experiment demonstrated that explicit measures are more effective at predicting behavior than the IAT for a behavior unaffected by social desirability concerns.

In a second experiment, Karpinski and Hilton (2001) demonstrated that IAT results could be manipulated through a task that created associations between the target sets and the valence sets of the IAT. Participants performed a memory task in which they were asked to memorize a series of word pairs. The words belonged to either a "youth" target set, or an "elderly" target set. In one condition, the youth set was paired with positive words while the elderly set was paired with negative words; in the other condition the sets were given the reverse associations. Participants then completed an elderly/youth IAT and a series of explicit measures that tested for attitudes towards the elderly and youths. Karpinski and Hilton (2001) found that the IAT recorded a preference for whichever group was paired with the positive words in the memory exercise. In contrast, the explicit measures recorded a bias for youth and against the elderly, regardless of the associations that were learned in the memory task. These results demonstrated IAT performance can be based on simple exposure to information, without taking into account whether the test taker believes that information is accurate. The results were particularly effective at demonstrating this point, because the participants were told that the memory task was unrelated to the rest of the experiment. This means that IAT performance can be influenced by associations which aren't even related to the groups being tested. Although there was no manipulation check testing whether the participants suspected the memory task was related to the IAT, it seems unlikely that they would make the connection. Additionally, the participants' knowledge of the memory task's purpose should not

affect the experiment, because IAT performance is resistant to conscious manipulation.

### *1.9 The Present Research*

Karpinski and Hilton's (2001) study supports their assertion that the IAT may measure knowledge rather than attitude. However, there were several shortcomings in their research which the present research will address. First, by using preexisting groups in their study, it is unclear how the participants' preexisting attitudes towards the groups affected the results. For example, in the first experiment, both the IAT and explicit measures detected a preference for apples. It could be argued apples and candy bars are portrayed both positively and negatively in American culture. Karpinski and Hilton (2001) argue apples have a more positive image than candy bars in American culture. However, it is possible that the larger number of commercials and advertisements for candy bars create positive associations which balance out the negative. Therefore, it is uncertain what IAT results would indicate a true attitude versus an environmental association in the case of apples and candy bars.

Karpinski and Hilton's (2001) second experiment was also confounded by the possibility the participants had preexisting associations. It is possible that the participants had both positive and negative associations about the elderly and youth. Instead of creating associations, the learning exercise may have made one set of associations temporarily more accessible. For example, in our society, the elderly are often stereotyped as being set in their ways and a burden on younger generation, but are also viewed as experienced and patient. It is possible that the participants in Karpinski's study knew, or even believed, both of these stereotypes equally, but that

the priming procedure strengthened one stereotype temporarily. If this effect was only temporary, it would mean the IAT could normally be an effective measure of attitude, and the learning task merely disrupted its effectiveness for a short time.

An effective method of controlling the effect of pre-existing knowledge on the IAT results would be to administer an IAT featuring groups the participants have never heard of before. To accomplish this, we utilized fictional groups and an experimental methodology created by Gregg, Seibt, & Banaji (2005). Their research focused on the difficulty of changing pre-existing attitudes held about a group. To test this, Gregg, et al. (2005) created two fictional groups. The participants were taught to associate one group with a positive valence and the other with a negative valence. After learning these associations, the participants were told the associations they learned were incorrect and were taught the opposite associations. Gregg, et al. (2005) found that the participants' IAT scores did not respond to the retraining, but they did change their responses on explicit measures. This experiment was meant to demonstrate unconscious associations are difficult to change after their initial formation. However, Gregg et al. (2005) were assuming that the IAT is an accurate measure of affective association. An alternative explanation is that the explicit measures reflected the participants' updated attitudes, but the IAT reflected their initially learned associations, which they no longer endorsed.

In the current experiment, descriptive information about these fictional groups will be presented in such a way that the participants will learn associations based on the information, but will not endorse what they learn. This will allow us to determine whether an IAT score reflects endorsement, or simple knowledge.

# Chapter 2: Experiment 1

## *2.1 Purpose and Overview*

### 2.1.1 Purpose

The purpose of experiment 1 was to provide an initial test of whether the IAT measures simple associations between words, or if it can capture attitudes. To eliminate the influence of prior knowledge, the two groups used in the experiment were fictitious. The groups and the names of the group members were taken from a previous experiment (Gregg et al., 2005) and had been tested to make sure initial attitudes towards them would be neutral. Participants were then taught to associate each group with a specific set of adjectives. In one condition, the adjectives were presented with a negative modifier which nullified the meaning of each adjective. In a second condition, the adjectives were unmodified. If IAT scores are not sensitive reversed meanings of the negated words, it would mean the IAT measures simple associations between the words. These simple associations have little to do with unconscious attitude. If the IAT is able to account for the reversed meanings of the negated words, it would suggest the IAT reflects the attitudes which form in response to those words.

### 2.1.2 Overview

In the context of a memory experiment, participants were taught about two fictitious groups. Through a learning task, the participants were taught to associate one of the groups with a set of positive adjectives and the other group with a set of negative adjectives. In the "not present" condition, all of the adjectives were preceded

by the word "not" when the participants were learning the associations. In the "not absent" condition, the adjectives were unmodified. The participants were then given an explicit measure which tested their explicit group preference. After completing the explicit measure, the participants were asked to take an IAT which featured the two fictitious groups they had learned about. Our main hypothesis was that the explicit measure would reflect the reversed meaning of the adjectives in the "not present" condition, but the IAT would not be sensitive to the adjectives reversed meanings. In the "not absent condition, the two measures would record a similar attitude in the participants.

The experiment also counterbalanced for the order of the compatible and incompatible blocks in the IAT, and the adjective list with which each group was associated. Therefore, the experiment had a split plot design, with type of measure (IAT vs. explicit) as a within subject variable and the "not condition" (not present vs. not absent), IAT block order, and the pairing of the adjective/group associations (Luupite positive vs. Niffite positive) as between subject variables.

*2.2 Participants*

Participants were drawn from the introductory psychology research pool at a large state university in the Mid-Atlantic. All of the participants were over the age of 18. Participants were given an hour of research credit for their participation. Data was recorded for 29 participants in total.

*2.3 Procedure*

After reading and signing the consent form, participants were seated in a small room containing a computer and given oral instructions complemented by on-screen text. The participants were told that the purpose of the experiment was to test how knowledge about a group affects one's ability to identify and categorize its members. The researcher explained that the experiment used groups the participant had never heard of before in order to prevent the influence of prior knowledge and/or bias.

The two groups were called the Luupites and the Niffites. There were eight members in each group, who could be identified by the structure of their names: Luupite names all contained double vowels and ended in –lup, Niffite names all contained double consonants and ended in –nif. The groups were adopted from a previous experiment (Gregg et al., 2005) which demonstrated the initial attitudes towards them would be neutral Participants were told in reality they were ancient historical groups, but that the group's names and the names of the group members had been changed so that they wouldn't be recognized.

The participants were told that they would learn about the groups through a learning exercise in which each group would be paired with a set of adjectives. The format of the learning exercise was outlined, and a brief explanation of the IAT's format was also presented. The IAT was described as the "testing phase" of the experiment, in which the participants would be tested to see if they had learned what the groups were like. They were told that the better they had learned about the groups in the learning task, the faster they would be able to perform on the IAT.

After receiving these instructions, the participants began the learning task. Upon completion, participants continued on to the explicit measure, and finally the IAT. The participants did not receive further oral instructions from the experimenter before taking the explicit measure or the IAT. Instead, on screen instructions were presented that explained the format of each new section of the experiment. After completing the IAT, the participants were debriefed about the purpose of the experiment.

### 2.3.1 The Learning Task

The learning task consisted of a series of trials that were similar in format to the IAT. In each trial, a name from one of the two groups appeared in the center of the computer screen. The name presented in each trial was randomly selected from a list of all the names. The participants had to categorize the names into Luupites and Niffites by pressing the correct key on an answer box. There was a right answer key and a left answer key. In each trial, the Luupites were assigned to one answer key, and the Niffites were assigned to the other. The names of the two groups were displayed in the upper corners of the screen, and their location (right or left corner) corresponded to the answer key assignments. The answer key assignments were randomly changed in each trial to prevent the participants from associating the groups with a specific answer key. If the participants correctly categorized a name, a blue "C" appeared on the screen; an incorrect response resulted in a red "X"

Twenty practice trials were conducted to familiarize the participants with the format of the learning task. After this practice phase, the main learning task began. During the learning task, each name was preceded by an adjective. There was a list of

positive adjectives and a list of negative adjectives. The adjectives were flashed on the screen for 200 ms., sufficient time for the participants to consciously detect and recognize the adjectives. The adjectives were immediately followed by a name from the appropriate group. Throughout the entire learning phase, one set of adjectives (positive or negative) preceded the names of one group, and the other set of adjectives preceded the names of the other group. The participants were told by the experimenter that these adjectives were accurate descriptions of the names and the groups to which the names belonged. Both groups were presented in an equal number of trials in the learning phase.

The form in which the adjectives were presented to the participants was varied by condition:

1. In the *not present* condition, all the adjectives in the learning task were displayed with the word "not" preceding them. For example, if the Luupites were paired with the list of positive adjectives, the Luupite names would preceded by the adjectives "not wonderful" and " not fantastic," while the Niffite names would be preceded by the words " not horrible" or "not awful."
2. In the *not absent* condition, the adjectives were presented unmodified. For example, if the Luupites were paired with the positive adjectives, the Luupite names would preceded by the adjectives "wonderful" and "fantastic," while the Niffite names would be preceded by words "horrible" or "awful."

The adjective list associated with each group was counterbalanced:

1. For half the participants, the Luupites were associated with the positive word list and the Niffites with the negative word list.

2. For the other participants, the Niffites were associated with the positive word list and the Luupites with the negative word list.

### 2.3.2 Explicit Measure

Participants' attitudes towards the Luupites and the Niffites were also assessed with an explicit measure. The participants were instructed that the measure used the number pad on the keyboard rather than the response box, and to respond based on how they felt about the two groups. The participants were first asked to complete the statement "I think the [Luupites/Niffites] are…" Participants responded on a seven point scale anchored with the phrases "very bad" (1) and "very good" (7). Then they were asked to complete the statement "I like the [Luupites/Niffites]…" The response scale was anchored with the phrases "not at all" (1) and "very much." (7). Both questions were asked once about each group, so there were a total of four questions.

### 2.3.3 The IAT

The participants' implicit associations were assessed using a Luupite/Niffite, good/bad IAT. Participants were given instructions on the computer screen before each block of the IAT. The instructions explained the new button assignments featured in each block, and instructed the participants to be as quick and accurate as possible. The IAT consisted of five blocks:

- Block 1, participants classified the adjectives into categories based on their valence (Bad and Good).
- Block 2, they classified the names of group members into categories based on their group membership (Niffite and Luupite).

- Block 3 was a combination of the first two blocks in which all four groups were presented.
- Blocks 4 and 5 corresponded to Blocks 2 and 3, except that responses for the Niffites and Luupites now have the button assignments from block 3.

As in the learning task, the groups featured in each block appeared in the upper corners of the screen, and their location corresponded to the button assignments for that block.

The Luupite and Niffite names and the adjective lists utilized in this IAT were the same as the ones featured in the learning task. It is important to note that in the "not present" condition, the "nots" were absent during the IAT. Furthermore, the designation of blocks as compatible and incompatible was based on which adjective list each group was associated with in the learning task, regardless of the presence of the "nots." For example, if the Luupite names were associated with the adjective "not wonderful" in the learning task, the IAT block where the Luupites were assigned to the same answer key as the positive word list was considered the compatible block.

The order of the IAT blocks was also counterbalanced across participants:

1. For half the participants, the compatible block was presented in block three of the IAT and the incompatible block was presented in block five.
2. For the rest of the participants, the incompatible block was presented in block three, and the compatible block was presented in block five.

## *2.4 Results*

### 2.4.1 Explicit Measure

A Cronbach's Alpha was calculated for responses to the explicit questions to see if the scores could be combined. We predicted that participants in the not present and not absent conditions would have opposite scores on the measure, so a separate Alpha was calculated for each condition. We also predicted the participants would give opposite responses for the questions referring to the Luupites and the Niffites, so we calculated separate Alphas for each group. Table 1 lists the results for the alpha tests, which indicated it would be acceptable to combine the responses for the two questions concerning each group. This created one attitude score for the group associated with the positive adjectives, and a second score for the group associated with the negative adjectives. These designations did not take into account the presence of the "nots," so in the not present condition, the group paired with the "not negative" adjectives was considered the negative group. Each participant's average rating for the negative group was subtracted from their average rating of the positive group. This created a relative preference score which could be compared to the relative preference score of the IAT.

Table 1
Cronbach's Alpha- Experiment 1

| | Not Absent Condition | | Not Present Condition | |
|---|---|---|---|---|
| Condition | Positive Group | Negative Group | Positive Group | Negative Group |
| α | .922 | .976 | .89 | .991 |
| N | 16 | 16 | 13 | 13 |

After creating the explicit preference scores, a 2 (not present vs. not absent) by 2 (IAT block order) by 2 (adjective/group association) univariate ANOVA was performed to determine if the method factors affected responses. The only significant finding was a main effect for the "not" variable, $F(29) = 21.296$, $p \leq .05$. Because the group association and block order had no effect on the results, they were not included as variables in further statistical tests.

Overall, the participants had a greater preference for the positive group in the not absent condition ($M=2.096$, $SD= 4.317$), but the not present condition, the participants showed a preference for the group associated with the negative words ($M= -4.516$, $SD= 2.141$). As we predicted, the participants rated the group associated with the positive adjectives positively in the "not absent" condition, but in the "not present" condition they reversed the valence of the adjectives and rated the groups accordingly.

### 2.4.2 The IAT

As with the explicit measure, we first performed a data reduction in order to calculate a single score for IAT performance. Since IAT scores are based on reaction times, calculating a single IAT score can be a complex process. Reaction time data can be difficult to analyze due to the high variability in response times, both between participants and within each participant's set of trials. In addition, IAT trial response time distributions tend to have a positive skew. The method used to convert IAT data into a normal distribution is a subject of debate among IAT researchers. The algorithm we utilized was based on the one recommended by Greenwald and Banaji (2003). We did not use the exact formula recommended by Greenwald and Banaji

(2003), because it was intended for performing correlational analyses. In order to perform an analysis of variance it was necessary to perform additional data transformations to standardize the data across participants. Our final scoring algorithm was as follows:

1. Eliminate trials greater than 10000 ms
2. Log transform trial scores
3. Transform each subject's log transformed trial scores into z-scores based on that subject's (and only that subject's) scores.
4. Calculate an average of the participants' performance for the compatible block and an average for the incompatible block (blocks three and five)
5. Calculate a difference score between the two z-scores (from step 4).
6. Transform the difference scores into z-scores based on the mean and SD of all the participants' difference scores.

This created a single score for each participant. A higher score indicated a greater preference for the positive group over the negative group.

As with the explicit measure, a 2 (not present vs. not absent) by 2 (IAT block order) by 2 (adjective/group association) univariate ANOVA was performed on the participants' IAT scores to determine if the method factors effected the results. The results revealed that the "not" manipulation was the only variable to have a significant effect on the IAT, $F(29) = 4.189$, $p = .05$. Because the group association and the IAT block order had no effect on the results, they were not included as variables in further statistical tests of the IAT.

Overall, the participants showed a greater preference for the group associated with the positive word list in the not absent condition (M= 58 ms., SD= 167 ms.) but in the not present condition, the preference was for the group associated with the negative adjectives in the associative learning exercise (M= -78 ms., SD= 87 ms.). Contrary to our predictions, IAT scores reflected the reversed meanings of the adjectives in the not present condition.

Table 2
Means of Participant Response on The IAT & Explicit Measure; Experiment 1

| | IAT | | Questionnaire | |
|---|---|---|---|---|
| Condition | Not Absent | Not Present | Not Absent | Not Present |
| Mean | 58 | -78 | 2.096 | -4.516 |
| Standard Deviation | 167 | 87 | 4.317 | 2.141 |
| N | 16 | 13 | 16 | 13 |

### 2.4.3 Implicit and Explicit Comparison

A final set of analyses was performed to test if the participants demonstrated a different degree of preference on the two types of measures. A significant correlation was found between participants' performance on the implicit and explicit measures r (29) = .431, p=.02. In order to assess whether the relationship between the two measures interacted with the "not" manipulation, separate correlations were calculated for the not present and not absent conditions. Neither correlation proved to be significant (not present condition: r (13) = .167, p=.586, not absent condition: r (16) = .049, p=.858). Finally, a 2 (not present vs not absent) by 2 (implicit measure vs explicit) repeated measures ANOVA was performed. The results revealed there was

no significant interaction between measure and reversal, $F(29) = 2.977$, $p = .096$. Both measures were sensitive to the reversed meaning of the adjectives in the not present condition, and responses on the two measures appeared to be similar.

*2.5 Discussion*

The purpose of this experiment was to demonstrate that IAT scores are based on simple semantic associations between words. If the IAT is based on such simple associations, it can not reflect participants' attitudes. To test this, an experimental condition was created in which the participants learned associations between two fictional groups and positive or negative words. The meaning of the adjectives was manipulated by the presence or absence of a "not" modifier. The "not" manipulation should reverse the associations formed between the groups and a valence concept. These associations were recorded by the explicit measure. However, the associations between the group names and adjectives should not reflect the reversed meanings of the adjectives. We hypothesized these associations would be reflected in the IAT scores.

Contrary to our predictions, our results showed that IAT scores reflected the reversed meanings of the adjectives. This finding demonstrates the IAT is more than a measure of simple associations between words; it reflects associations with a valence based on the meanings of those words. These findings demonstrate that the IAT can reflect associations which unconscious attitudes are based on.

# Chapter 3: Experiment 2

The first experiment demonstrated that the IAT reflects evaluations of a group rather than associations between specific words and a group. These are the types of associations which create unconscious attitudes. However, not all evaluations of a group are endorsed, and it is possible that the IAT reflects these unendorsed evaluations. We therefore conducted a second experiment in which endorsement of the learned associations was manipulated so that the overall association between group and valence would not be endorsed by the participants.

## *3.1 Overview*

The procedure for the second experiment was similar to the first. In the context of a memory experiment, participants were taught to associate adjectives with the Luupites and Niffites. Through the learning task, the participants learned to associate one of the groups with a set of positive adjectives and the other group with a set of negative adjectives. One difference from the previous experiment was the "not present" variable was eliminated and replaced by an endorsement variable: we manipulated whether the participants' endorsed the associations between the groups and the adjectives.

The participants were given the same explicit measure and IAT as in the previous experiment. We predicted that the explicit measure would reflect a relatively neutral preference for the groups when the participants were taught not to endorse the associations. In contrast, IAT scores would record a preference for the positive group. Also, a second independent variable was added: the length of the learning phase was

varied between participants. We predicted increased training of the associations would lead to a greater preference on the IAT. We hypothesized that increasing the length of the training would increase the strength of the association between the groups and the adjectives, but would have no effect on how much those associations are endorsed. Therefor, if IAT scores reflected the length of the learning phase, it would indicate the IAT measures strength of association rather than endorsement. The experiment also counterbalanced for the order of the compatible and incompatible blocks in the IAT, and the adjective list with which each group was associated. Therefore, the experiment had a split plot design, with type of measure (IAT vs. explicit) as a within subject variable and the endorsement condition (endorsed vs. unendorsed), IAT block order, and the pairing of the adjective/group associations (Luupite positive vs. Niffite positive) as between subject variables.

### *3.2 Participants*

The participants for the experiment were from the introductory psychology research pool from a large state university in the mid-Atlantic. Participants were given an hour of research credit for their participation. Data was recorded for a total of 45 participants.

### *3.3 Procedure*

The basic procedure was identical to that of the first experiment: Participants completed the learning task, then the explicit measure, and then the IAT. The primary variation from the previous procedure occurred during the learning task.

### 3.3.1 The Learning Task

The procedure for the learning task was similar to that of the previous experiment. However, the learning task featured a new experimental manipulation:

1. In the endorsement condition, the participants were told that the adjective sets accurately described the groups they were paired with in the learning task, and that the adjectives could be used to form an accurate impression of the two groups. Since they had no knowledge to the contrary, the participants should have believed the adjectives accurately described the groups.
2. In the non-endorsement condition, the participants were told even though each adjective set was paired with a specific group, they did not accurately describe the group. In order to convince the participants that the adjective sets did not describe the group they were paired with, the participants were asked to flip a coin to determine which adjective set each group was paired with in the learning task. Therefore, the participants should not have believed the adjectives accurately described the groups.

The length of the associative learning task was also manipulated:

1. In the short condition, the learning task consisted of 120 trials
2. In the long condition, it consisted of 240 trials.

The two method factors that were counterbalanced in the first experiment were also counterbalanced in the second experiment. The first was the assignment of adjective list to group in the learning phase:

1. In the first condition, the Luupites were associated with the positive word list and the Niffites with the negative word list.

2. In the second condition, the Niffites were associated with the positive word list and the Luupites with the negative word list.

In the non-endorsement condition, the assignment of the adjective lists was based on the participant's coin flip. In the endorsement condition, the adjective lists were randomly assigned to a group. The participants were not aware the adjective lists were randomly assigned in the endorsement condition.

### 3.2.2 Explicit Measure

In addition to the four questions featured in Experiment one, a new question was added to the explicit measure: "How much do you really know about the [Luupites/Niffites]?" [1]

### 3.3.3 The IAT

After completing the explicit measure, the participants took the Luupite/Niffite, good/bad IAT. The instructions and procedure for the IAT were

[1] This question was added because during pilot testing, some participants demonstrated a preference for one group over the other on the explicit measure in the non-endorsement condition. When questioned during debriefing, the participants said even though they didn't really believe the groups were good or bad, they felt compelled, when asked, to have some form of group preference. These responses indicated that because of their prior participation in experiments, some of the participants' believed it was unusual or unacceptable to give a neutral answer on an attitude questionnaire.

identical to the previous experiment. The order of the IAT blocks were counterbalanced to control for order effects:

1. In one condition, the compatible block was presented in block three of the IAT and the incompatible block was presented in block five.
2. In the second condition, the incompatible block was presented in block three, and the compatible block was presented in block five.

After completing the IAT, the participants were debriefed about the purpose of the study.

## *3.4 Results*

### 3.4.1 Explicit Measure

The method used to calculate the explicit scores was identical to the method used in the previous experiment. The alphas indicated it would be acceptable to combine responses to the two questions asked about each group. They were combined using the same method as in the previous experiment.

Table 3
*Cronbach's Alpha- Experiment 2*

| | Endorsement Condition | | Non-Endorsement Condition | |
|---|---|---|---|---|
| Condition | Positive Group | Negative Group | Positive Group | Negative Group |
| α | .751 | .817 | .603 | .63 |
| N | 19 | 19 | 26 | 26 |

After calculating the explicit preference scores, A 2 (endorsement vs. non-endorsement) by 2 (implicit measure vs. explicit measure) by 2 (short learning phase vs. long learning phase) by 2 (IAT block order) by 2 (adjective/group association)

repeated measures ANOVA was run, and the results revealed that the endorsement manipulation was the only variable to have a significant effect on the explicit measure, $F(29)= 52.443$, $p \leq .001$. Participants demonstrated a greater preference for the positive group in the endorsement condition ($M=4.5$, $SD= 1.87$) compared to the non-endorsement condition ($M= .31$, $SD= .99$). These results supported our prediction that the participants endorsed the information they learned about the groups.

### 3.4.2 The IAT

The method used to calculate IAT scores was identical to the method used in the previous experiment. After calculating the IAT scores, a 2 (endorsement vs. non-endorsement) by 2 (implicit measure vs. explicit measure) by 2 (short learning phase vs. long learning phase) by 2 (IAT block order) by 2 (adjective/group association) univariate ANOVA was performed. The only significant finding was a main effect of the endorsement variable, $F (1)= 43.676$, $p\leq .001$. The IAT recorded a much greater preference for the positive group in the endorsement condition ($M= 196$, $SD= 240$) than in the non-endorsement condition ($M=52$, $SD=199$). An independent samples t-test revealed this difference was significant, $t (43)= 2.209$, $p=.033$. Contrary to our predictions, the IAT was sensitive to the participants' endorsement of the group/valence associations.

Table 4
*Means of participant Response on The IAT & Explicit Measure; Experiment 2*

| IAT | Questionnaire |
|---|---|

| Condition | Endorsement | Non-Endorsement | Endorsement | Non-Endorsement |
|---|---|---|---|---|
| Mean | 196.085 | 52 | 4.5 | .31 |
| Standard Deviation | 240 | 199 | 1.87 | .99 |
| N | 26 | 19 | 26 | 19 |

3.4.3 Implicit and Explicit Comparison

A final set of analyses was performed to test if the participants demonstrated a different degree of preference on the two types of measures. There was a weak correlation between participants' performance on the implicit and explicit measures, r (45) = .241, p=.111. In order to assess whether the relationship between the two measures interacted with the endorsement manipulation, separate correlations were calculated for the endorsement and non-endorsement conditions. Neither correlation proved to be significant (endorsement: r (19) = .046 p= .851; non-endorsement: r (26) = -.127, p= .537). A 2 (endorsement vs. non-endorsement) by 2 (implicit measure vs. explicit measure) repeated measures ANOVA revealed there was no significant interaction between the type of measure and the endorsement variable, F (44)= 1.277, p = .265. Although the similarity in preference recorded on the two measures was weaker than in the previous experiment, there was still no significant difference in the participants' responses on the two measures.

*3.5 Discussion*

In this experiment, the IAT demonstrated sensitivity to the participants' endorsement. Specifically, when participants believed the adjectives associated with a

group did not describe them, the IAT evaluations were neutral. In both conditions, the participants' preference on the IAT corresponded to the preference recorded by the explicit measure. These results imply the IAT is sensitive to whether participants endorsed the associations they learned. Furthermore, IAT scores were not affected by the length of the learning phase. This implies the strength of the learned associations do not affect IAT scores. However, it is also possible that the variation in length was not sufficient to create a variation in strength.

However, it is possible that the participants in the non-endorsement condition put less effort into learning the associations, because they were told the associations were meaningless. If they didn't learn the associations, then the neutral IAT response would reflect a lack of knowledge rather than a lack of endorsement. This possibility could not be evaluated based on the findings in experiment two, so a third study was designed to control for participant learning.

# Chapter 4: Experiment 3

## *4.1 Overview*

A new type of question was added to the learning phase in order to clarify whether the participants learned the associations. These new questions tested the participants to see if they knew which adjective set was associated with the Luupites and which was associated with the Niffites. Other than this new question type, the procedure for experiment three was similar to the previous experiment. We predicted that participant's IAT scores would demonstrate a preference for the positive group in the non-endorsement condition, if the participants had learned the associations. In

contrast, the explicit measure would show no preference between the groups when the associations were not endorsed.

The length of the learning task was not varied in this experiment because length had no effect in the previous experiment. Additionally, the IAT block order and the adjective/group associations had no significant effect in the previous two experiments, so they were also not included. This created a 2 (endorsement vs. non-endorsement) by 2 (implicit measure vs. explicit measure) split plot design.

### *4.2 Participants*

The participants for the experiment were drawn from the introductory psychology research pool of a large state university in the mid-Atlantic. Each participant was given an hour of research credit for their participation. Data was recorded for 49 participants in total.

### *4.3 Procedure*

The procedure of the third experiment was similar to that of the second experiment, with a few modifications. During the learning task, each Luupite and Niffite name was paired with a specific adjective from the appropriate list. For example, if the positive words were associated with the Luupites as a whole, the adjective "wonderful" might be specifically paired with the name "Neenolup." In the learning phase, each group name would be preceded only by the adjective with which it was paired.

Participants were presented with a test question every tenth trial of the learning task. The test questions had a similar format to the regular trials of the learning task;

one of the adjectives from the sets was presented in the middle of the screen. Two names from the group the adjective's set was associated with appeared in the upper corners of the screen. One of the names presented was paired with the adjective on the screen. Just as with a regular trial from the learning phase, the location of the names corresponded to the buttons they were assigned to on the answer box. The participant had to identify which specific name the adjective always preceded by pushing the appropriate button.

After completing the learning task, the participants completed the explicit measure, and then the Luupite/Niffite IAT. The procedure and instructions for these measures were identical to those of the previous experiment.

## *3.4 Results*

### 3.4.1 Test Questions

The participants' responses to the test questions were analyzed to determine if they had successfully learned the associations. The participants' accuracy score was calculated based on the percentage of the test questions they had answered correctly. Participants were coded as having learned or not learned the associations based on whether their accuracy was above 50%, which indicated better than chance performance on the test questions. Only four participants performed below 50% accuracy. There was no difference in the results if these participants were excluded, so all of the participants' data were utilized in the analysis.

3.4.2 Explicit Measure

The method used to calculate the participants' explicit scores was identical to the method used in the previous experiment. The alpha scores indicated it would be acceptable to combine responses to the two questions asked about each group, so they were combined using the same method as in the previous experiment.

Table 5
*Cronbach's Alpha- Experiment 3*

| | Endorsement Condition | | Non-Endorsement Condition | |
|---|---|---|---|---|
| Condition | Positive Group | Negative Group | Positive Group | Negative Group |
| α | .723 | .69 | .814 | .765 |
| N | 23 | 23 | 26 | 26 |

There was a greater preference for the positive group by the participants in the endorsement condition (M= 3.09 SD= 1.64) than in the non-endorsement condition (M=1.2857 SD=1.75), an independent samples t-test revealed this difference was significant: t (47)= -3.678, p= .001. As we predicted, the explicit measure was sensitive to the participants' endorsement of the information they learned about the groups.

4.3.3 The IAT

The method used to calculate IAT scores was identical to the method used in the previous experiment. After calculating each participant's IAT score, descriptive statistics revealed that a much greater preference was shown for the positive group

(M=167.085, SD= 208.829) in the endorsement condition than in the non-endorsement condition (M=17.8200 SD=98.140). An independent samples t-test revealed this difference was significant, t (47) = -3.331, p = .002. Contrary to our predictions, the IAT was sensitive to whether the participants endorsed the associations they learned about the groups.

Table 6
*Means of participant Response on The IAT & Explicit Measure; Experiment 3*

| | IAT | | Questionnaire | |
|---|---|---|---|---|
| Condition | Endorsement | Non-Endorsement | Endorsement | Non-Endorsement |
| Mean | 167 | 17 | 3.09 | 1.2857 |
| Standard Deviation | 208 | 98.140 | 1.64 | 1.75 |
| N | 21 | 28 | 21 | 28 |

*3.4.5 Implicit and Explicit Comparison*

A bivariate correlation was used to compare participants' performance on the implicit and explicit measures. There was a weak but non-significant correlation between participants' performance on the implicit and explicit measures, r (49) = .199, p= .126. In order to assess whether the relationship between the two measures was affected by the independent variable, separate correlations were calculated for the endorsement and non-endorsement conditions; neither correlation proved to be significant (endorsement: r (21) = .192, p= .404; non-endorsement: r (28) = -.142, p= .472). A 2 (endorsement vs. non-endorsement) by 2 (implicit measure vs. explicit measure) repeated measures ANOVA results revealed there was no significant interaction between the type of measure and the endorsement variable F (48)= .169,

p= .651. These results offer no evidence that the endorsement manipulation affected the two measures differently. However, individual participants' did not have similar responses on the two measures.

*4.6 Discussion*

The results for this experiment were almost identical to the previous experiment. The IAT still recorded a preference for the positive group in the endorsement condition, which disappeared in the non-endorsement condition. There was still a moderate, though weaker, correlation between the IAT and explicit measure. However, the inclusion of the test questions eliminated the possibility that the participants did not learn the associations in the non-endorsement condition. This strengthens the conclusion that IAT performance reflects endorsed associations rather than simple knowledge.

# Chapter 6: General Discussion

In a series of experiments, the IAT's ability to distinguish between knowledge about groups and the attitudes towards those groups was tested. In Experiment 1, we tested whether the IAT detects simple associations between words, rather than associations between group concepts and valence concepts. This is important because associations with valence concepts are the primary determinant of attitude (Greenwald et al, 2002). To test this, we had participants learn to associate groups with sets of positive or negative adjectives. In one condition, all of the adjectives were preceded by a negative modifier which reversed their meaning. If scores IAT did not reflect the negation of the words, it would mean it detected simple associations rather than

associations to an overall valence. We found that the IAT detected the reversed valence of the negated adjectives. This indicated that the IAT measures the associations between a group and a valence, which are the primary source of unconscious attitudes.

After establishing that the IAT detects associations with a valence, we tested whether it only detected associations which were endorsed. In Experiment 2, we once again had participants learn to associate groups with sets of positive or negative adjectives. The participants were told that the adjectives either did or did not describe the groups. When the participants were told the adjectives did not describe the groups, they should have still formed the valence associations through the learning exercise, but they should not endorse those associations. The IAT only detected a preference in the participants when they were told the adjectives accurately described the groups. These results indicate that the IAT is sensitive to the participants' endorsement of the learned associations. In Experiment 3, we replicated the results of Experiment 2 while controlling for the participants' success at learning the associations. We again demonstrated that the IAT measures more than simple knowledge; we showed that it is sensitive to participants' endorsement of the information they have about a group. Since only information people believe is true should determine attitudes, these results support the IAT's use as an attitude measure.

As a group, these experiments address a critical issue concerning the validity of the IAT. Our research can be tied to Gilbert's (1990) theory of how information is falsified. Gilbert argues information that is unendorsed, or believed to be false, is still stored in memory, but is "tagged" as false. When this information is recalled later, it

is discounted and has little influence on attitudes. It is possible that unconscious that are unendorsed are also tagged as false. Critics such as Karpinski et al. (2001) suggest that, rather than measuring attitudes, the IAT reflects the participants' knowledge of the groups being tested. In Gilbert's framework, this would mean IAT scores are based on both associations tagged "accurate," and associations tagged "false." The IAT could not be an effective attitude measure if it is influenced by them associations tagged false, because such associations have little effect on attitude.

Karpinski et al. (2001) supported this criticism by demonstrating that IAT scores could be influenced by unendorsed information learned through a task similar to the one we used in our experiments. However, they used groups which the participants already had knowledge of, so it is possible that their learning task merely primed pre-existing endorsed associations. Consequently, the effect of the associative learning task on the IAT recorded by Karpinski et al. (2001) might not have demonstrated that the IAT measures knowledge rather than association. Our research was an attempt to support Karpinski's criticism while controlling for the participants' prior knowledge of the groups. This was accomplished by utilizing groups previously unknown to participants. We then taught the participants associations which we told them were not accurate, so they should have been labeled "false." If the IAT measures knowledge rather than association, these "false" associations should have had a strong influence on IAT scores. We found the "false" associations appeared to have very little influence on IAT scores. This implies that the IAT is only sensitive to knowledge that the test taker believes is true. Rather than supporting

Karpinski et al's (2001) findings, our results imply that the IAT is sensitive to the participants' endorsement.

While the current research supports the validity of the IAT as a measure of attitude, there was only a low to moderate correlation between IAT scores and the explicit measure in our experiments. This leaves open questions about the IAT's validity. Supporters of the IAT attribute low correlations to the effect of social desirability on explicit attitude measures. A recent meta-analysis by Poehlman and Uhleman (2006) found that the correlation between the IAT and explicit measures is lowest in studies of racism, where social desirability should have a strong effect. However, the present study utilized groups which were unfamiliar to the participants, so social desirability should not have been a factor. Since social desirability is not a plausible explanation for the low correlation, another factor must affect IAT performance, and is responsible for the low correlation.

One explanation for the low correlation is that the IAT measures preference for one group over another; it measures positive attitudes towards one group and negative attitude towards the other. The attitude towards just one of the groups can not be determined from IAT scores. In contrast, explicit measures tend to detect attitudes towards a single group. This makes it hard to compare IAT scores to most explicit measures because they are measuring different constructs. (Ashburn-Nardo, Voils, & Montieth; Blanton & Jaccard, 2006). This concern could be addressed by modifying the IAT so that it includes a third, neutral group. Researchers could compare the increase or decrease in reaction time for trials featuring the two groups being tested to the third, neutral group. This would allow IAT scores to be split into

two separate, one for each group. These group specific measures of bias may have higher correlations to explicit measures of bias.

The test taker's cognitive ability is another factor which may interfere with the correlation between the IAT and explicit measures. The influence of individual cognitive factors on the IAT is an area that needs further study, and is a potential explanation for the low correlation between the IAT and explicit measures. There is empirical evidence that cognitive skill can influence IAT performance (McFarland & Crouch, 2002; Mierke, & Klauer, 2001, 2003, 2005). Mierke & Klauer (2001) conducted research on the influence of task switching on the IAT. They found that on individual trials in the incompatible block of the IAT, participants performed more slowly on trials featuring a word from a target set which were preceded by a trial featuring an adjective. This was because the adjective primed the participants to focus their attention on the word's valence information, and they have to refocus attention to the next word's group identity; which is called "task-switching." Participants were able to perform faster on trials featuring group members that are preceded by trials also featuring group members. These trials are called "task congruent" because attention is already focused on group identity. Cai, Sriram, Greenwald and McFarland (2004) claim this effect can be neutralized using their recommended scoring algorithm. The effect of cognitive skill on IAT performance, and the ability of the scoring algorithm to counter it, needs to be tested experimentally.

The IAT has tremendous potential as a research tool for psychologists. If it is an effective attitude measure, it has numerous applications both in academic research and in the public sphere. However, before the IAT can be confidently used, it is

critical that its validity as an attitude measure be confirmed. The present research tested one such challenge to the IAT, and the results supported its use as an attitude measure. If research continues to support the IAT's validity, it will make a welcome addition to the tools of psychological assessment.

# Bibliography

Amodio, D. M., Harmon-Jones, E., & Devine, P. G., (2001). *Journal of Personality and Social Psychology,* 738-753.

Ashburn-Nardo, L., Voils, C.I., Monteith, M.J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology*, *81*, 789-799.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84,* 888-918.

Banse, R., Seis, J. & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity and controllability of the IAT. *Zeitschrift fur Experimentelle Psychologie, 48,* 145-160.

Bassili, J.N. (1995). Response latency and the accessibility of voting intentions: What contributes to

accessibility and how it affects vote choice. *Personality and Social Psychology Bulletin*, *21*, 686–695.

Bassili, J.N. (1996). Meta-judgmental versus operative indexes of psychological attributes: The case of measure of attitude strength. *Journal of Personality and Social Psychology*, *71*, 637–653.

Blanton, H., Jaccard, J., Gonzales, P.M. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology, 42*, 192-212.

Brendl, M.C.. Markman, A.B., Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 760-773.

Crowne, D. P., Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Dasgupta, N., Greenwald, A.G. (2001). On the malleability of automatic attitudes: combating prejudice with images of liked and disliked individuals. *Journal of Personality and Social Psychology*. *81,* 800-814.

De Houwer, J. (2006) What are implicit measures and indirect measures of attitudes? *Social Psychology Review*.

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled processes, *Journal of Personality and Social Psychology,* 56, 5-18.

Dovidio, J. F., & Gaertner, S. L. (1986). The aversive form of racism, Dovidio, J. F., & Gaertner, S. L. (Eds.), *Prejudice, discrimination, and racism,* 61-89.

Eagly, A.H., Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt Brace Jovanovich.

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

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.

Edwards, K. (1990). The interplay of affect and cognition in attitude formation and change. *Journal of Personality and Social Psychology, 59,* 202-216.

Edwards, AL., Diers, C. J., Walker, J. N., (1962) Response sets and factor loadings on sixty-one personality scales. *Journal of Applied Psychology, 46,* 220-225.

Egloff, B. & Schmukle, S.C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology,* 83, 1441-1455.

Farley, S.D., Sasson, M.F. (2003). Relative influences of affect and cognition on behavior: are feelings or beliefs more related to blood donation intentions? *Experimental Psychology*; *50*, 55-62

Fazio, R. H. (1986). How do attitudes guide behaviour? In R. M. Sorrentino, & E. T. Higgins (Eds.), The handbook of motivation and cognition: Foundations of social behavior (pp. 204–243). New York: Guilford.

Fazio, R. H. (1990a). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75-109). New York: Academic Press.

Fazio, R.H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R.E. Petty & J.A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Erlbaum.

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

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

Fazio, R. H., & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude-perception and attitude-behaviour relations: An investigation of the 1984 presidential election. *Journal of Personality and Social Psychology, 51*, 505–514.

Fazio, R. H. & Zanna, M. P. (1981). Direct experience and attitude—behavior consistency. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 161—202). San Diego, CA: Academic Press.

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility.

Fazio, R.H. (2001) On the automatic activation of associated evaluations: An overview, *cognition and emotion*, *15*, 115–141.

Fazio, R.H., Powell, M.C., & Williams, C.J. (1989). The role of attitude accessibility in the attitude to behaviour process. *Journal of Consumer Research*, *16*, 280–288.

Frey, D., & Gaertner, S. L. (1986). Helping and the avoidance of inappropriate interracial behavior: a strategdy that can perpetuate a non-prejudiced self-image. *Journal of Personality and Social Psychology, 50,* 1083-1090.

Gaertner, S.L. (1973). Helping behavior and discrimination among liberals and conservatives. *Journal of Personality and Social Psychology,* 23, 335-341.

Gaertner, S. L, Bickman, L. (1971). Effects of race on the elicitation of helping behavior: the wrong number technique. *Journal of Personality and Social Psychology,* 20, 218-222.

Garb, H. N., Wood, J. M., Nezworski, M., Grove, W. M., & Stejskal, W. J. (2001). Towards a resolution of the Rorschach controversy. *Psychological Assessment, 13*, 433-448.

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

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

Greenwald, A.G., & Nosek, B.A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48,* 85-93.

Greenwald, A.G., Banaji, M.R., Rudman, L.A., Farnham, S.D., Nosek, B.A., Mellott, D.S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept, *Psychological Review. 109*, 3-25.

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.

Grønnerød, C. (2003). Temporal stability in the Rorschach method: A meta-analytic review. *Journal of Personality Assessment*, *80*, 272-293

Hetts, J.J., Sakuma, M., & Pelham, B.W. (1999). Two roads to positive regard: Implicit and explicit self-evaluation and culture. *Journal of Experimental Social Psychology, 35,* 512-559.

Houston, D. A., & Fazio, R. H. (1989). Biased processing as a function of attitude accessibility: Making objective judgments subjectively. *Social Cognition, 7,* 51–66.

Huijding, J., de Jong, P.J., Wiers, R.W., & Verkooijen, K. (2005). Implicit and explicit attitudes toward smoking in a smoking and a nonsmoking setting. *Addictive Behaviors, 30*, 949-961.

Jackson, D.N., Messick, S. (1962). Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal & Social Psychology, 65*. 285-299

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced-choice offer a solution? *Human Performance, 13,* 371–388.

Jones, J.T., Pelham, B.W., Mirenberg, M.C., & Hetts, J.J. (2002). Name letter preferences are not merely mere exposure: Implicit egotism as self regulation. *Journal of Experimental Social Psychology,* 38, 170-177.

Kaiser, D.A. (2005) Basic principles of quantitative EEG. *Journal of Adult Development, 12,* 99-104.

Katz I. R. Hass, G., (1988). Racial ambivalence and American value conflict: correlational and priming studies of dual cognitive structures *Journal of Personality and Social Psychology*, 893-899.

Karson, S., O'Dell, J.W. (1976). A guide to the clinical use of the 16 PF. Oxford, England: Inst for Personality & Ability Test,. xiii, 160.

Kelman, H.C. (1974). Attitudes are alive and well and gainfully employed in the sphere

of action. American Psychologist, 29, 310-324.

Kester, J. D. (2001). A revolution in Social Psychology. *The American Psychological Society Observer 14*, 9-26.

Kim, D-Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, *66*, 83–96.

Kinder, D.R. Sears, D.O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life, *Journal of Personality & Social Psychology.* 40(3), *414-431. American Psychological Assn, US*

Knudson, D. D., Pope, H., & Irish, D. P. (1967). Response differences to questions on sexual standards. *Public Opinion Quarterly, 31*, 290-297.

Kovel, J. (1970). *White Racism: A Psychohistory*. New York: Pantheon

Kozel, A.F., Revell, L.J., Lorberbaum, J. P. (2004). A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men; *Journal of Neuropsychiatry & Clinical Neurosciences*, *16*, 295-305.

Lambert, A.J., Payne, K.B., Ramsey, S., Shaffer, L.M. (2005). On the predictive validity of implicit attitude measures: The moderating effect of perceived group variability. *Journal of Experimental Social Psychology, 41*, 114-128.

Leite, W.L., Beretvas, N.S. (2005). Validation of scores on the Marlowe-Crowne social desirability scale and the balanced inventory of desirable responding. *Educational and Psychological Measurement, 65*, 140-154.

Lilienfield, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*, 27-66.

Lissek, S, Baas, J, Pine, D. (2005) Sensation Seeking and the Aversive Motivational System. *Emotion*, *5*, 396-407

Logan, G. D. (1988). Toward an instance theory of automization. *Psychological Review, 95,* 492–527.

Macrae, C. N., Stangor, C., & Alan, A.B. (1994). Activating social stereotypes: a functional analysis. *Journal of Experimental Social Psychology,* 30, 370-389.

Maison, D., Greenwald, A.G., & Bruin, R.H., (2004). Predictive validity of the implicit association test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology. 14*, 405-415.

McConahay, J. B. *(1986)* Modern racism, ambivalence, and the Modern Racism Scale. Dovidio, J. F. Gaertner, S L. (Eds). Prejudice, discrimination, and racism. 91-125.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37,* 435–442.

Millar, M. G., & Millar, K. U. (1990). Attitude change as a function of attitude type and argument type. *Journal of Personality and Social Psychology, 59,* 217-228.

Heggestad, E., Morrison, M., Reeve, C.L., McCloy, R., (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance, *Journal of Applied Psychology, 91*, 9–24

Müller, M.M., Gruber, T., Keil, A. (2000). Modulation of induced gamma band activity in the human EEG by attention and visual information processing. *International Journal of Psychophysiology, 38,* 283-299.

Neely, J.H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory and Cognition*, *4*, 648–654.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading spreading activation and limited-capacity attention. *Journal of Experimental Psycyology: General*, *106*, 225–254.

Norman, W. T. (1967). Personality measurement, faking and detection: An assessment method for use in personnel selection. In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 374–391). New York: McGraw-Hill.

Nosofsky, R. M. & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266-300.

Nozek, B. (2005) Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology,* 134, 565–584.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), Automatic Processes in Social Thinking and Behavior. Psychology Press.

Nuwer, M. R. (1988). Quantitative EEG: I. Techniques and problems of frequency analysis and topographic mapping. *Journal of Clinical Neurophysiology, 5,* 1-43.

Nuwer, M. (1997). Assessment of digital EEG, quantitative EEG, and EEG brain mapping: Report of the American Academy of Neurology and the American Clinical Neurophysiology Society. *Neurology, 49*, 277-292

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), Measures of personality and social psychological attitudes (pp. 17-59). San Diego: Academic Press.

Petty & J. A Krosnick (Eds.), *Attitude strength: Antecedents and consequences. Ohio State University series on attitudes and persuasion: Vol. 4* (pp. 247–282). Hillsdale, NJ: Erlbaum.

Plant, A.E. Devine, P.G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75,* 811-832.

Poehlman A.T. Uhlmann E. Greenwald A. G. Banaji M. R. (in press) Understanding and using the implicit association test: III meta-analysis of predictive validity

Robinson, M.D., Meier, B.P., Zetocha, K.J., & McCaul, K.D. (2005). Smoking and the implicit association test: when the contrast category determines the theoretical conclusions. *Basic and Applied Social Psychology*. *27*, 201-212.

Rudman, L.A., Ashmore, R.D., & Gary M.L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Experimental Social Psychology, 81,* 856-868

Sarlo, M; Palomba, D; Buodo, G. (2005); Blood pressure changes highlight gender differences in emotional reactivity to arousing pictures. *Biological Psychology*, *70*. 188-196

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. P*sychological Review*, *84*, 127–190.

Sigall, H., & Page, R. (1971): Current Stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology, 18*, 247-255

Strach, F., Deutsch, R (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247.

Spence, A. (2006). Using implicit tasks in attitude research: A review and a guide. *Social Psychology Review*.

Tomarken, A.J., Davidson, R.J., Henriques, J.B. (1990)Resting frontal brain asymmetry predicts affective responses to films. *Journal* of *Personality* and *Social Psychology*, *Vol 59*, 791-801.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The linguistic intergroup bias

as an implicit indicator of prejudice. *Journal of Experimental Social Psychology, 33,* 490-509.

Weiner, I B. (2006). The Rorschach inkblot method. In Archer, R.P. (Eds) Publishers. *Forensic uses of clinical assessment instruments*. pp. 181-207 Lawrence Erlbaum Associates

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

Wittenbrink, B., Judd, C. M. & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262-274.

Wittenbrink, B., Judd, C. M, & Park, B (2001). Affective versus conceptual judgments in automatic attitude activation. *Journal of Experimental Social Psychology, 37, 244-252*.

Ziegert, J.C., Hanges, P.J., (2005). Employment discrimination: the role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, *90*, 553-562.